Modern views of machine learning for precision psychiatry

Zhe Sage Chen, Prathamesh (Param) Kulkarni, Isaac R. Galatzer-Levy, Benedetta Bigio, Carla Nasca and Yu Zhang

1Department of Psychiatry, New York University Grossman School of Medicine, New York, NY 10016, USA
2Department of Neuroscience and Physiology, New York University Grossman School of Medicine, New York, NY 10016, USA
3The Neuroscience Institute, New York University Grossman School of Medicine, New York, NY 10016, USA
4Department of Biomedical Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201, USA
5Headspace Health, San Francisco, CA 94102, USA
6Meta Reality Lab, New York, NY, USA
7Department of Bioengineering, Lehigh University, Bethlehem, PA 18015, USA
8Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015, USA

*Correspondence: zhe.chen@nyulangone.org (Z.S.C.), yuzi20@lehigh.edu (Y.Z.)
https://doi.org/10.1016/j.patter.2022.100602

SUMMARY

In light of the National Institute of Mental Health (NIMH)’s Research Domain Criteria (RDoC), the advent of functional neuroimaging, novel technologies and methods provide new opportunities to develop precise and personalized prognosis and diagnosis of mental disorders. Machine learning (ML) and artificial intelligence (AI) technologies are playing an increasingly critical role in the new era of precision psychiatry. Combining ML/AI with neuromodulation technologies can potentially provide explainable solutions in clinical practice and effective therapeutic treatment. Advanced wearable and mobile technologies also call for the new role of ML/AI for digital phenotyping in mobile mental health. In this review, we provide a comprehensive review of ML methodologies and applications in precision psychiatry. We argue that advances in ML-powered modern technologies will create a paradigm shift in the current practice in diagnosis, prognosis, monitoring, and treatment of mental illnesses. We discuss conceptual and practical challenges in precision psychiatry and highlight future research opportunities in ML.

INTRODUCTION

Mental health issues are an epidemic in the United States and the world. According to the National Institute of Mental Health (NIMH), nearly one in five American adults suffer from a form of mental illness or psychiatric disorder (www.nimh.nih.gov/health/statistics/). According to the Centers for Disease Control and Prevention (CDC), the COVID-19 pandemic has witnessed a significant impact on our lifestyle and considerably elevated adverse mental health conditions caused by fear, worry, and uncertainty. Increased suicide rates, opioid abuse, and antidepressant usage have been observed in both adults and teenagers. The diagnosis and treatment of mental health has imposed a burden to the healthcare system and society. In the
of dysfunction in psychological or neurobiological systems; it attempts to bridge the power of multi-disciplinary (such as the genetics, neuroscience, and behavioral science) research approaches. The current gold standard for diagnosis and treatment outcome in mental disorders—the Diagnostic and Statistical Manual of Mental Disorders (DSM), maintained by the American Psychiatric Association (APA)—is often based on the clinician’s observations, behavioral symptoms, and patient reporting, which are all susceptible to a high degree of variability. Therefore, it is imperative to develop quantitative neurobiological markers for mental disorders while accounting for their heterogeneity and comorbidity.

One important goal in neuropsychiatry research is to identify the relationship between neurobiological/neuropsychological findings and clinical behavioral/self-report observations. Machine learning (ML) and artificial intelligence (AI) have generated growing interests in psychiatry because of their strong predictive power and generalization ability for prognosis and diagnosis applications. The interest of applying ML/AI in psychiatry has grown steadily in the past two decades, as reflected in the number of PubMed publications (Figure 1A). To improve mental health outcomes with digital technologies, the so-called “digital psychiatry” focuses on developing ML/AI methods for assessing, diagnosing, and treating mental health issues. A recent global survey has indicated that psychiatrists were somewhat skeptical that AI could replace human empathy, but many predicted that “man and machine” would increasingly collaborate in undertaking clinical decisions, and psychiatrists were optimistic that AI might improve efficiencies and access to mental care and reduce costs.

The past two decades have witnessed substantial growth of ML applications for psychiatry in the literature, reflected in many applications and reviews. Although multiple reviews of ML for psychiatry are available, the majority of reviews are restricted to relatively narrow scopes. In this paper, we try to provide a comprehensive review of ML and ML-powered technologies in mental health applications. Our view is “modern” in the sense that the development of new technologies, consumer market demand, and public health crises (such as COVID-19) have constantly redefined the role of ML and reshaped our thinking in precision psychiatry. Specifically, we will cover state-of-the-art methodological developments in ML, multimodal neuroimaging, large-scale circuit modeling, neuromodulation, and human-machine interface. Due to space limitations, our reviewed literature is by no means exhaustive. To distinguish our review from others, we will focus on several issues central to the ML applications for psychiatry: generalizability, interpretability, causality, and clinical and behavioral integration.

Our view about this emerging field is cautiously optimistic for several reasons. First, with an increasing amount of data and computational power, there is a growing demand for psychiatrists to use ML to reevaluate clinical, behavioral, and neuroimaging data. The interests in mental health funding from the industry have also grown substantially (Figure 1B). Second, it is becoming increasingly important to leverage the power of ML and develop explainable AI (XAI) tools for unbiased risk diagnosis, personalized medicine recommendation, and precise neurostimulation. The integration of ML with neuroimaging can potentially help us identify and validate biomarkers in diagnosis...
and treatment of mental illnesses. Third, there is a growing demand for psychiatrists in the United States, and the shortage is even more acute in poorer countries. ML/AI technologies may change the practice of psychiatry for both clinicians and patients. Finally, advanced technologies such as social media, multi-media, and mobile and wearable devices also call for the development of ML/AI tools to assist the assessment, diagnosis, or treatment of individuals who are mentally ill or at risk. The meanings of ML and AI are relatively broad in our current review and generally cover a wide range of analytic or predictive tools that are designed for finding structures or regularity of data; therefore, ML under our discussion also includes data mining and knowledge discovery. From now on, we will use ML and AI interchangeably throughout the paper.

BACKGROUND OF NEUROIMAGING

Advances in neuroimaging
Neuroimaging provides a window to probe human brains in terms of both structural and functional forms and offers various resolutions to examine brain activity at macroscopic, mesoscopic, and microscopic scales across spatial and temporal domains (Figure 1C).

Our understanding of brain and behavior relationships has expanded exponentially over the last few decades. While this improvement may be attributed to a multitude of factors, advancement in neuroimaging has played a prominent role. Ranging from increased utilization of structural neuroimaging techniques to the significant scientific advancements brought about by the increased availability of functional neuroimaging, these technologies have provided significant benefits to improved understanding of neural correlates and discovery of biomarkers in psychiatric disorders. Some of the most common neuroimaging methods for probing brain function include the utilization of magnetic resonance imaging (MRI), diffusion MRI (dMRI), functional MRI (fMRI), electroencephalography (EEG), magnetoencephalography (MEG), electrocorticography (ECoG), functional near-infrared spectroscopy (fNIRS), and positron emission tomography (PET). To date, EEG and fMRI are two most commonly used imaging modalities for precision psychiatry. Specifically, EEG is low cost and easy to operate, making it more appealing for clinical practice or home use.

Neuroimaging analysis
These rich neuroimaging modalities allow us to comprehensively probe brain functions. Numerous research efforts have been devoted to revealing the neurobiological basis of various psychiatric disorders using advanced neuroimaging analyses. Under specifically designed cognitive paradigms, task-related neuroimaging allows us to examine the relationship between brain activities (e.g., event-related potential and spectral perturbation and reward or emotional processing-related functional activation) and cognitive dysfunctions. A promising direction for probing brain function using neuroimaging is to investigate brain connectivity (or connectome). Studying the resting-state brain connectome provides an elegant way to characterize the complex brain architecture and uncover brain dysfunctions in intrinsic brain networks. Increasing neuroimaging studies suggest that functional connectivity may fluctuate rather than being stationary during an entire session of data collection. Studies examining spatiotemporal dynamics of brain networks have recently received growing attention and may reveal meaningful brain states associated with different psychiatric conditions.

Another promising approach to establish robust biomarkers for psychiatry is to combine multiple neuroimaging modalities in a data-driven manner, which offers opportunities to exploit cross-modality complementary information that a single modality approach may not capture.

Feature engineering
The input data comprise features that are fed to ML algorithms. All ML methods will benefit from proper feature engineering (including but not limited to imputation, scaling, standardization, normalization, transformation, and one-hot encoding). Incorporating knowledge-driven feature engineering into the analysis of specific neuroimaging modalities has shown significant promise in enhancing the model performance and its physiological interpretability. For instance, spatial filtering (or source localization) followed by orthogonalizing the resulting time series and calculating their power envelope correlations can mitigate the effects of volume conduction and source leakage, which may lead to more accurate quantification of functional connectivity in EEG or MEG.

To fully understand the brain structural and functional organization, we argue that neuroimaging, when combined with modern ML and other ML–powered technologies, can provide powerful tools in advancing diagnosis, prognosis, and intervention of psychiatric disorders.

HOW CAN ML HELP PSYCHIATRY?

Psychiatry versus other medicine disciplines
The nature and etiology of mental illnesses remain unclear and challenging to study. Traditional studies for the neurobiology of psychiatric disorders have followed a categorical classification framework using a case-control design whereby all patients with a given diagnosis are compared with healthy individuals. The symptom-based diagnosis covered hundreds of thousands of different symptom combinations, which has caused extensive clinical heterogeneity. It is increasingly recognized that existing clinical diagnostic categories could misrepresent the causes underlying mental disturbance. The conventional case-control design often fails to match a clinically useful decision process in the absence of differential diagnostic specificity, which is due to its limited strengths in delineating the significant clinical and neurological heterogeneity of psychiatric disorders. On the other hand, previous studies have broadly explored the group effects of neurobiology to explain its connection to behavior and disease. However, such group-level analyses cannot fully capture individual-level brain abnormality that is crucial for developing personalized medicine.

In addition, many psychiatric disorders may be considered as falling along multiple dimensions. Cooccurrence of multiple psychiatric disorders might reflect different patterns of symptoms resulting from shared risk factors and perhaps the same underlying disease processes. The high comorbidity in these disorders
Table 1. Categories of ML, concepts, typical methods, and their representative applications

| Learning category | Concepts | Representative methods | Applications |
|-------------------|----------|------------------------|--------------|
| Supervised        | learning from labeled data to predict class/clinical measures | SVM, random forest, sparse learning, ensemble learning | Disease diagnosis, prognosis, treatment outcome prediction |
| Unsupervised      | learning from unlabeled data to uncover structure and identify subgroups | Hierarchical clustering, K-means, PCA, CCA | Disease subtyping, normative modeling, identify behavioral and neurobiological dimension |
| Semi-supervised   | learning from both labeled and unlabeled data to perform supervised or unsupervised tasks | multi-view learning, Laplacian regularization, semi-supervised clustering | multi-modal analysis, joint disease subtyping and diagnosis, prediction with incomplete data |
| Deep              | learning hierarchies and non-linear mappings of features for higher-level representations, can be either supervised or unsupervised | CNN, deep autoencoder, GCN, RNN, LSTM, GAN | a large class of generic learning problems |
| Reinforcement     | solving temporal credit assignment problems, optimal control, trial-and-error learning | temporal difference learning, Q-learning, actor-critic model, dynamic programming | online control, modeling of decision-making and choiced behaviors |

significantly affects the characterization of psychopathology according to the traditional diagnostic categories. Conventional studies focusing on a single diagnostic domain are therefore insufficient in uncovering the neural correlates of comorbidity among multiple disorders or identifying the dimensions of neural circuits and behavioral phenotypes.

Clinical need driving ML applications in mental health

Despite the rapid progress in psychiatric studies, several areas appear highly underexplored but may carry substantial potentials for achieving major breakthroughs toward precision psychiatry. First, the capacity to dissect inter- and intra-individual variability is crucial for understanding the neural basis of variation in human cognition and behavior. Studies focusing on the level of individuals may find greater success over conventional group-level analyses. Translational study-orientated approaches for psychiatric neuroimaging may further enhance the ability to find statistically significant effect sizes that can be used in individuals.40

Second, identifying subgroups (i.e., subtypes) in psychiatric disorders may delineate disease heterogeneity. Increasing evidence suggests that data-driven subtyping may drive novel neurobiological phenotypes associated with distinct behavior and cognitive functioning.15 These stratified phenotypes may help improve the predictability of clinical outcomes and serve as potential biomarkers for treatment selection.52 However, subtyping analysis is widely viewed as hypothesis generating and poses significant challenges related to reproducibility and physiological interpretability.43,47 Linking subtype identification to a particular outcome or question using well-designed ML approaches is vital to address these challenges.48

Third, another promising area focuses on transdiagnostic approaches to uncover neural correlates of specific domains (such as cognition, arousal, and emotion regulation), which have been implicated in psychopathology across the diagnostic spectrum.49 Recent ML efforts have been dedicated to identifying transdiagnostic brain dysfunctions and dimensions of psychopathology to improve understanding of comorbidity among psychiatric disorders.50–53 Importantly, leveraging “big data” from a longitudinal perspective offers a promising way to track the neurobiological and phenotypic trajectories that have been rarely examined in previous cross-sectional psychiatric studies.54–56 Ultimately, such longitudinal studies may help reveal the neural mechanism underlying the disease progression and provide new insights for the development of timely interventions.

It should be noted that the presence of confounding effects is one of the most critical challenges in psychiatry studies.57,58 For example, the site effects or unmatched phenotypic information (e.g., demographics and clinical measures) may distort the apparent relationship between input features and output. Medications were also found to strongly alter brain activity and connectivity,59,60 Therefore, inappropriately modeling those confounders can lead to erroneous findings. To date, mental health studies have been done to control the impact of confounders on biomarker quantification.61,62

These new frontiers in studying psychiatric disorders can be substantially empowered by ML methodologies summarized in Table 1. The applications include stratifying patients into clinically meaningful subtypes, discovering novel transdiagnostic disease dimensions, and tailoring treatment decisions to individual patients. Together, these research outcomes can deliver a significant promise in promoting the development of objective biomarker-based precision psychiatry.

The applications of ML in psychiatry can be mainly categorized according to their clinical purposes: diagnosis, prognosis, treatment, and readmission. In contrast to most medical disciplines, traditional diagnoses in psychiatry remain restricted to subjective symptoms and observable signs and therefore call for a paradigm shift. ML offers a new paradigm to achieve automated and more objective assessments for various psychiatric conditions. For disease diagnosis, supervised classification can be used to identify discriminative biomarkers that distinguish a specific disorder from healthy condition or other mental illnesses.63 Unsupervised clustering is useful in identifying disease subtypes for dissect clinical and biological heterogeneity, thus offering new ways of defining psychiatric conditions.52,64 For disease prognosis, classification models can be built to distinguish different course trajectories (e.g., progressor versus non-progressor), whereas regression models are useful for predicting symptom development during the course of the disease.53
treatment studies, individual responses to the treatment can be predicted using classification methods that distinguish responder from non-responder. Regression-based approaches can also be utilized to predict changes in post-treatment symptoms.66 For readmission evaluation, supervised classification can be used to predict whether an individual would be rehospitalized or to detect the relapse trajectory.67,68 With a properly designed strategy, feature reduction/mapping approaches and knowledge-driven feature engineering can be integrated into the ML model training to identify more informative and interpretable biomarker patterns.46

It should be noted that ML is an ever-growing big discipline and covers many categories and emergent topics, each with a different technical focus. A standard taxonomy of ML typically includes supervised, unsupervised, and reinforcement learning paradigms.69,70 There are also various extensions or special treatment of each category or combined (e.g., semi-supervised learning, kernel learning, ensemble learning, deep learning; see Figure 2). Here, our rationale for reviewing specific ML methods is simply based on their applicability in existing mental health applications.

In the following subsections, we will review several key ML paradigms in mental health applications based on neuroimaging, behavioral, and clinical measurements. A tabular review of representative applications is shown in Table 2. Specifically, we focus on the review of neuroimaging-based psychiatric studies, and detailed reviews of the other data domains (such as genetic, clinical, behavioral, and social media data) will be presented in later sections. For the sake of space limitation, we will not include exhaustive reviews of all ML paradigms (such as reinforcement learning, active learning, and transfer learning) in this section but rather provide some reference pointers whenever necessary.

**Supervised and unsupervised learning**

ML holds substantial promise in promoting research from small case-control studies to those with large transdiagnostic samples and from prior specified brain regions to whole-brain circuit dysfunction for individual-level precision medicine.29,105,106 In a new era of evidence-based psychiatry tailored to individual patients, objectively measurable endophenotypes could allow for early disease detection, personalized treatment selection, and dosage adjustment to reduce the burden of disease.18,107,108 These promising applications in psychiatric disorders have been enabled by leveraging the powerful strength of modern ML techniques.21,63,106,116

---

**Figure 2. Various ML models for mental health applications**

(A) Left: multi-modal supervised classification scheme. Three modality-specific factors are optimized on the training data: classifier types, parameters, and weights. The final diagnostic classification is based on a weighted sum of decision values, where weights correspond to those estimated during training. Right: feature importance maps of functional neuroimaging modalities.

(B) Unsupervised learning. Left: whole-brain functional-connectivity matrix averaged across all subjects. $z =$ Fisher-transformed correlation coefficient. Right: hierarchical clustering analysis (copyright Springer Nature; figure reproduced with permission).

(C) Semi-supervised learning pipeline for phenotype stratification based on EHRs (Beaulieu-Jones and Greene; figure reproduced with permission).

(D) Deep neural networks (DNNs) for group-level and individualized treatment predictions. Future data points could then be used to forecast symptom onset, treatment response, or other mental health-related variables (Koppe et al.; Creative Commons licenses 4.0).
Table 2. Representative ML applications in psychiatry based on neuroimaging and clinical data

| Application       | Learning category | ML method                  | Mental disorder       | Data type                                | Reference         |
|-------------------|-------------------|----------------------------|-----------------------|------------------------------------------|-------------------|
| Diagnosis         | supervised classification, deep | dynamic GCN | ADHD                  | rs-fMRI + phenotypic data                | Zhao et al. 71    |
|                    | supervised classification | ensemble learning | ADHD                  | multi-modal task fMRI                   | Luo et al. 72      |
|                    | supervised classification, deep | GCN             | ASD                   | task fMRI                               | Li et al. 73       |
|                    | supervised classification, deep | ensemble learning + GCN | ASD                   | rs-fMRI                                 | Khosla et al. 74   |
|                    | supervised classification | PCA + LASSO   | bipolar               | dMRI + cognitive data                   | Wu et al. 76       |
|                    | supervised classification | RVM            | PTSD                  | rs-fMRI                                 | Zhu et al. 76      |
|                    | supervised classification, deep | ICA + LSTM | schizophrenia         | fMRI                                    | Yan et al. 77      |
|                    | supervised classification, deep | SVM             | schizophrenia         | sMRI                                    | Mikolas et al. 78  |
|                    | supervised classification, deep | CNN             | depression             | rs-EEG                                  | Uyulan et al. 79   |
|                    | supervised classification, deep | autoencoder + DNN, SVM, random forest | ASD                   | rs-fMRI                                 | Heinsfeld et al. 80|
|                    | semi-supervised classification | GNN            | ASD                   | rs-fMRI + phenotypic data                | Parisot et al. 81  |
|                    | unsupervised, subtyping | normative modeling + clustering | PTSD                  | rs-fMRI                                 | Maron-Katz et al. 82|
|                    | unsupervised, subtyping | CCA + hierarchical clustering | depression           | rs-fMRI                                 | Drysdale et al. 15 |
|                    | unsupervised, subtyping | sparse K-means       | PTSD, depression       | rs-EEG                                  | Zhang et al. 42    |
|                    | supervised, transdiagnostic | latent class analysis | ADHD                  | task fMRI                               | Lecei et al. 83    |
|                    | supervised, transdiagnostic | normative modeling + GP regression | multiple disorders | rs-fMRI                                 | Parkes et al. 84   |
|                    | supervised, transdiagnostic | sparse CCA          | multiple disorders     | rs-fMRI                                 | Xia et al. 81      |
|                    | supervised, transdiagnostic | PLS             | multiple disorders     | rs-fMRI                                 | Kebets et al. 82   |
| Prognosis          | supervised classification | GP classifier      | depression             | task fMRI                               | Schmaal et al. 85  |
|                    | supervised classification | LASSO            | psychosis             | rs-EEG                                  | Ramyead et al. 86  |
|                    | supervised classification | SVM              | psychosis, depression  | multi-modal task fMRI                   | Koutsouleris et al.87 |
|                    | supervised classification, deep | DNN             | PTSD                  | rs-fMRI/task fMRI                       | Sheynin et al. 88  |
|                    | supervised classification, deep | SVM             | schizophrenia         | sMRI                                    | Nieuwenhuis et al. 89 |
|                    | supervised classification, deep | SVM, random forest, DNN | schizophrenia         | task fMRI                               | Smucny et al. 90   |
|                    | supervised regression     | LASSO            | substance use          | MRI/task fMRI                           | Bertocci et al. 91 |
|                    | supervised regression, deep | SVR + LSTM      | PTSD                  | MEG                                     | Zhang et al. 92    |

(Continued on next page)
Supervised learning

Supervised learning, being the most popularly used ML category, has been widely applied to neuroimaging-based predictive modeling tasks for psychiatric disorders. Classic supervised-learning algorithms include logistic regression, support vector machine (SVM), and random forest. Given the high-dimensional nature of neuroimaging data, these approaches are commonly accompanied by a feature selection step to obtain low-dimensional representations. Connectome-based predictive modeling is one of such approaches that combine simple linear regression and feature selection to predict individual differences in traits and behavior from connectivity data. Least absolute shrinkage and selection operator (LASSO) provides an alternative approach that performs simultaneous feature selection and prediction to learn a compact feature pattern for accurate prediction of a specific disorder or clinical outcome. Relevance vector machine (RVM) builds upon a probabilistic framework by leveraging automatic relevance determination to learn a sparse solution and penalize unnecessary model complexity. RVM has recently demonstrated its strength in quantifying neuroimaging biomarkers for post-traumatic stress disorder (PTSD) diagnosis as well as for treatment outcome prediction in depression. As an extension of the conventional single-task methods, multi-task learning (MTL) approaches have been increasingly employed to exploit complementary features jointly from multiple views of neuroimaging data.

Due to the complex nature of the brain’s function, informative features may not be observable in the raw high-dimensional feature space. To address this challenge, latent space-based supervised learning has been developed to uncover latent dimensions of neural circuits in psychiatric disorders. For example, a sparse latent space regression algorithm tailored for EEG data was developed to identify anti-depressant-responsive brain signatures in depression. By jointly estimating the spatial filters and regression weights under a convex optimization framework, the ML model was able to successfully reveal treatment-predictive signatures in a low-dimensional latent space (see case study 1 below). To address comorbidities among psychiatric disorders, dimensional approaches have been developed using statistical models capable of discovering the complex linear relationship between high-dimensional datasets. For instance, low-dimensional representations of depression-related connectivity features have been successfully identified by applying canonical correlation analysis (CCA) to resting-state fMRI (rs-fMRI) connectivity and clinical symptoms. The discovered representations defined two disease dimensions corresponding to an anhedonia-related component and an anxiety-related component, respectively. A similar dimensional analysis was also utilized to examine the neural correlates of neuropsychiatric symptoms in dementia. Using CCA, two latent modes were identified with distinct neuroanatomical bases of common and mood-specific factors of the symptoms. A sparse CCA approach has been applied to reveal linked dimensions of psychopathology and functional connectivity in brain networks for psychiatric disorders.

Ensemble learning

Although ML approaches have been extensively designed for supervised learning, using a single model may not produce the optimal generalization performance for a complex prediction task. By combining multiple ML models to reduce variance, ensemble learning outperforms a single model in prediction and has proven successful in discovering robust biomarkers for psychiatric disorders. For instance, multi-atlas ensemble-learning algorithms have been proposed for improved schizophrenia detection and autism spectrum disorder (ASD) diagnosis. By
utilizing multi-modal neuroimaging including sMRI, fMRI, and DTI, a bagging-based SVM produced significant improvement in prediction of adult outcomes in childhood-onset attention-deficit/hyperactivity disorder (ADHD). Based on the selective ensemble algorithm, a sparse multi-view prediction model has been designed with rs-fMRI connectivity for ASD diagnosis; this model combined multiple classifiers under a bootstrap framework and significantly outperformed other single-model approaches.

Although sophisticated models of supervised learning often produce better classification or prediction performance, their interpretability decreases at the cost of increasing model complexity. We will discuss the interpretable ML methods in more detail later (XAI in psychiatry). Additionally, labeled data require ground-truth knowledge, which is not always accurate or reliable in the case of mental disorders. For instance, a skin cancer diagnosis may rely on training samples that have been biopsied and cataloged, leaving no doubt as to whether they are malignant or not; however, there is no equivalent of the biopsy in mental disorder.

**Unsupervised learning**

Unsupervised learning relaxes the assumption of labeled samples and can be useful, e.g., for exploratory data analysis, feature engineering, or cluster analysis. Unsupervised learning aims to uncover the intrinsic data structure by either identifying potential clusters (e.g., using latent class analysis or K-means clustering) or learning a feature mapping that satisfies certain criteria (e.g., using principal-component analysis [PCA]). Identifying patient subtypes offers a promising strategy to delineate neurobiological heterogeneity in psychiatric disorders. With rs-fMRI, hierarchically clustering was applied to successfully identify four subtypes of functional connectivity in depression. These subtypes were found to correlate with differing clinical-symptom profiles and predict responsiveness to brain stimulation therapy. From rs-EEG, two transdiagnostic subtypes were identified using sparse K-means clustering with distinct power envelope connectivity patterns and found to respond differentially to anti-depressant medication and psychotherapy. As a non-distance probability-based clustering approach, latent class analysis has also been applied to discover subgroups in psychiatric disorders. A proof-of-concept study was conducted using latent class analysis to identify ADHD subtypes from fMRI activation profiles and revealed that the subtype with attenuated brain activity showed fewer behavior problems in daily life. By leveraging data resources from multiple time points, psychiatric studies have been shifting from cross-sectional analysis to longitudinal modeling. Finite mixture modeling became increasingly popular for the analysis of longitudinally repeated-measure data, which can identify latent classes following similar paths of temporal development. Typical finite mixture models include growth mixture modeling, group-based trajectory modeling, and latent transition analysis. The use of latent growth mixture modeling (LGMM) and group-based trajectory modeling has become increasingly popular in studying psychiatric disorders, such as depression, anxiety, and ASD. They offer flexible ways to identify latent subpopulations that manifest heterogeneous symptom trajectories. LGMM approaches have also been successfully used to predict the PTSD course among the population at risk. As an extension of latent class analysis (LCA) to longitudinal data, latent transition analysis (LTA) may predict the longitudinal service use for individuals with substance use disorder (SUD). Together, these approaches provide powerful tools to delineate longitudinal heterogeneity and the corresponding distinctive phenotypes during the course of psychiatric disorders.

**Semi-supervised learning**

Semi-supervised learning is an ML approach that combines supervised learning and unsupervised learning. Popular semi-supervised-learning techniques include self-training, mixture models, co-training and multi-view learning, graph-based methods, and semi-supervised clustering. These methods have been increasingly applied to psychiatric studies. By unifying autoencoder and classification, a semi-supervised model was developed for ASD diagnosis. A semi-supervised classification has been devised using graph convolutional networks and applied to the population graph-based diagnosis of ASD. A semi-supervised clustering has also been designed by extending SVM with implicit clustering driven by a convex polytope to form a method called heterogeneity through discriminative analysis, which can achieve joint disease subtyping and diagnosis. This approach has shown strong in delineating neurostructural heterogeneity in bipolar and major depressive disorders (MDDs) and schizophrenia, as well as in youth with internalizing symptoms. Additionally, semi-supervised learning has gained increasing mental health applications in digital data from electronic health records (EHRs), social media, and mobile phones. See ML-powered technologies for psychiatry for a detailed discussion.

**Normative modeling**

Normative modeling is an emerging and innovative framework for mapping individual differences at the level of a single subject with respect to a reference model. It has been increasingly used in mental health to parse the substantial neurobiological heterogeneity by quantifying individual deviations. By building a normative model of neuroimaging data on a large-scale healthy population, brain abnormalities of individual patients can be quantified by examining their statistical differences from the distribution of the norm. Gaussian process (GP) regression-based normative modeling has been applied to quantify individual deviations and dissect neurobiological heterogeneity in various psychiatric disorders. With this tool, an association was successfully discovered between transdiagnostic dimensions of psychopathology and an individual’s unique deviations from normative neurodevelopment in brain structure. By combining tolerance interval-based normative modeling and clustering analysis, individual abnormalities in rs-fMRI were accurately quantified to define two stable subtypes in patients with PTSD. The two subtypes showed distinct patterns of functional connectivity with respect to the healthy population and differed clinically on levels of reexperiencing symptoms. These novel data-driven approaches provide useful techniques to identify “abnormal” subtypes in patients, thereby advancing clinical and mechanistic investigations in psychiatric disorders. More recently, an autoencoder model has been utilized to realize normative modeling for detecting microstructural deviations in Chamberland et al.

**Deep learning**

Deep learning consists of a collection of methods that use multi-layered-architecture (≥2 hidden layers) artificial neural networks
for ML tasks. Through a specifically designed deep neural network structure, high-level feature representations can be learned from raw features. Deep learning thus holds promise in offering an end-to-end analytic framework for disease diagnosis and prediction. With the advancement in neuroimaging technologies, an increasing number of large-scale multi-center datasets have been established for building powerful ML models to fully explore the informative feature representations from complex brain and genomic data. By training on these large-scale datasets, deep learning can learn robust neuroimaging representations and outperform standard ML methods in a variety of application scenarios in mental health.17,140–142

**Deep autoencoder**

The deep autoencoder, also known as stacked autoencoder (SAE), aims to learn latent representations of input data through an encoder and uses these representations to reconstruct output data through a decoder. By stacking multiple layers of autoencoders, a deep autoencoder is formed to discover more complicated and potentially non-linear feature patterns. Deep autoencoder has been applied to extract low-dimensional features from the amplitude of low-frequency fluctuations in fMRI.143 Clustering analysis with the latent features uncovered by deep autoencoder further identified two subtypes within major psychiatric disorders including schizophrenia, bipolar disorder, and MDD. A deep-learning model was also designed based on a sparse SAE and applied to lower the dimensionality of fMRI connectivity. The sparsity constraint used in this model yielded interpretable neural patterns for improved ASD diagnosis.144 Deep autoencoder has also been applied to implement normative modeling with structural MRI for the quantification of individual abnormalities in neuropsychiatric disorders, including schizophrenia and ASD.145 The abnormal features extracted using the normative model led to improved diagnosis performance compared with the traditional case-control analysis. Recently, a deep contrast variational autoencoder was used to extract brain and genomic data. By training on these large-scale datasets, deep learning can learn robust neuroimaging representations and outperform standard ML methods in a variety of application scenarios in mental health.17,140–142

**Convolutional neural networks (CNNs)**

Different from conventional multi-layer perceptron or autoencoder assigning a different weight to each input feature, CNNs were designed to better capture the spatial and local structure information from pixels or voxels.110,147 Due to its strength in utilizing neighborhood information to learn hierarchies of features,148 CNNs have been one of the most successful deep-learning models applied in various medical applications. A diagnosis model was established through EEG-based image construction coupled with the CNN for accurate detection of MDD.149 This model provided an end-to-end framework to successfully identify translational biomarkers from rs-EEG in distinguishing depressive patients from healthy people. With whole-brain structure MRI, a three-dimensional (3D) CNN model has also been designed to automatically extract multi-layer high-dimensional features for the diagnosis of conduct disorder.149

**Graph neural networks (GNNs)**

Though deep-learning models have shown strengths in capturing complex neuroimaging patterns, they may not generalize well to non-Euclidean data types (e.g., brain networks). In contrast, GNNs provide a clever way of learning the deep graph structure of non-Euclidean data, leading to enhanced performance in various network neuroscience tasks.150 For instance, a framework based on graph convolutional networks (GCNs) has been designed for the diagnosis of ASD.151 By building a population graph that integrates rs-fMRI data as node features and phenotypic measures as edges, the designed model outperformed other state-of-the-art methods. An inductive GNN model was also devised to embed the graphs containing different properties of task fMRI and drive interpretable connectome biomarkers for ASD detection.152 More recently, a novel GNN model was developed to incorporate dynamic graph computation and feature aggregation of 2-hop neighbor nodes into graph convolution for brain network modeling.71 This dynamic GNN significantly improved the performance in ADHD diagnosis and revealed the circuit-level association between connectomic abnormalities and symptom severity.

**Recurrent neural networks (RNNs)**

As a specific extension of the feedforward neural network, RNNs have the ability to learn features and long-term dependencies from sequential and time-series data. Long-short-term memory (LSTM) models are the most popular RNN and have shown advantages in capturing temporal dynamic information of neuroimaging data for various psychiatric disorder studies.153 An LSTM-based RNN architecture was built with the time course of fMRI-independent components to exploit the temporal information, which yielded an improved diagnosis of schizophrenia.77 By combining RNNs with other deep neural networks, novel ML models have also been proposed to model the spatiotemporal dynamics in neuroimaging data. A spatiotemporal CNN model was proposed for 4D modeling of fMRI, with confirmed robustness in identifying key features in the default mode network.154 LSTMs have also been applied to incorporate multi-stage neuroimaging data into longitudinal analytic frameworks for modeling the trajectories of psychopathology development in various psychiatric disorders. A recent LSTM-based model was built with MEG data to achieve accurate longitudinal tracking of pathological brain states and prediction of clinical outcomes in PTSD.153

**Generative adversarial networks (GANs)**

As one type of generative model, GANs have gained considerable attention in computer vision and natural language processing and also have become increasingly popular in neuroimaging analysis.116 GANs consist of two competing neural networks (one as generator and the other as discriminator) and can learn deep feature representations without extensive labeled data. Due to this unique advantage, GANs have been increasingly applied in data augmentation to enhance the sample size for model training.153 Moreover, GANs have been used to impute missing values in multi-modal datasets, a common problem in psychiatric studies, rather than discarding an entire multi-variate data point.154 The adversarial model has also been incorporated into other ML models for specific applications in psychiatric studies. For instance, the discriminative and generative components were incorporated in LSTM to form a MTL approach for fMRI-based classification, which resulted in an improved diagnosis of ASD compared with the standard LSTM.156 By integrating GANs with group independent component analysis (ICA), a functional connectivity-based deep-learning model was developed for the diagnosis of MDD and schizophrenia.156 Specifically, the generator with fake connectivity was trained to...
match the discriminator with real connectivity in the intermediate layers, whereas a new objective loss was determined for the generator to improve the diagnosis accuracy. More recently, a confounder-free deep-learning framework was designed by incorporating the concept of GAN into the model training.62 This end-to-end approach is capable of simultaneously learning informative features and controlling for confounder effects to improve model performance.

The strength of deep-learning algorithms is that they can learn complex predictor-response mappings, but the power also comes at the cost of requiring a very large sample size for model optimization. This poses potential overfitting and interpretability challenges in psychiatric applications.17

Key ML concepts for precision psychiatry
Regardless of the ML paradigms in psychiatric applications, there are some common themes that distinguish human intelligence from automated or human-in-the-loop machine intelligence. In a recently published white paper, “Machine intelligence for healthcare,” four important features are emphasized for ML systems.157 These concepts are broadly applicable to precision psychiatry.19

- Trustworthiness: the ability to access the validity and reliability of an ML-derived output across varying inputs and environments. In other words, psychiatrists need to be able to evaluate the limitations of an ML system and confidently apply system-derived information for psychiatric evaluation.
- Explainability: the ability to understand and evaluate the internal mechanism of a machine. The development of ML systems will need to account for data quality, quality metrics for the system’s functioning and impact, standards for applications in the environment, and future updates to the system.
- Usability: the extent to which an ML system can be used to achieve specified goals with effectiveness, efficiency, and patient satisfaction in multiple environments. These applications need to be scalable across multiple settings while preventing additional burdens on providers and patients.
- Transparency and fairness: the right to know and understand the aspects of an input that could influence outputs (clinical decision support) from the system. Such factors should be available to the people who use, regulate, and are affected by any type of care decision that employs the ML system. The potential bias in the data or the system needs to be identified and informed prior to decision-making.

The first two features are related to interpretability, which we will discuss in more detail in XAI in psychiatry. The other two features will be discussed in discussion and conclusion.

Case studies
To help the reader get a concrete idea of the reviewed ML techniques in psychiatric applications, here we present several case studies to illustrate the strengths in prediction/classification diagnosis analytics. These representative case studies employ different ML strategies and cover different data modalities, including rs-EEG, task fMRI, and ECoG.

Case study 1: Sparse latent space learning for EEG-based treatment prediction in depression
Anti-depressants have shown only modest superiority over placebo, which is partly because the clinical diagnosis of MDD encompasses biologically heterogeneous conditions that relate differentially to treatment outcomes. It is important to develop a robust neurobiological signature for an anti-depressant-responsive phenotype that determines which patients will benefit from medications. To address the challenge, Wu et al.99 developed a sparse EEG latent space regression (SELSER) model to predict the treatment outcome. Specifically, SELSER optimizes the spatial filters and regression weights in conjunction under a convex optimization framework and identifies an anti-depressant-responsive EEG signature for MDD (Figure 3A). The identified signature accurately predicts anti-depressant outcomes (n = 228). A neurophysiologically interpretable cortical pattern was further observed through a source mapping from the scalp spatial pattern, mainly contributed by the right parietal-occipital regions and the lateral prefrontal regions (Figure 3B). The validation on an independent cohort showed that the treatment outcomes predicted by the brain signature are significantly higher in a partial responder group versus a treatment-resistant group, demonstrating its further clinical utility in the broader construct of treatment resistance in depression.

Case study 2: Unsupervised learning-based identification of neurophysiological subtypes in psychiatric disorders
Neurobiological heterogeneity has a substantial impact on treatment outcome independent of pre-treatment clinical symptoms. For example, although psychotherapy is currently the most effective treatment for PTSD, many patients are nonetheless non-responsive and display differences in brain function relative to responsive patients. Using sparse K-means clustering, Zhang et al.18 developed a data-driven framework to achieve simultaneous feature selection and subtyping on the high-dimensional power envelope connectivity of rs-EEG source-reconstructed signals. This approach successfully identified two transdiagnostic subtypes with distinct functional connectivity patterns in PTSD and MDD (n = 648), which were prominently within the frontoparietal control network and default mode network (Figure 3C). Importantly, linear mixed models in an intent-to-treat analysis on symptom severity revealed that the two subtypes differentially responded to psychotherapy and anti-depressant versus placebo. An RVM-based classification analysis further confirmed that the EEG connectivity-driven subtypes were distinguishable using rs-fMRI connectivity. The discriminative pattern identified from fMRI was also consistent with the EEG connectivity pattern (Figure 3D).

Case study 3: Classification of anxious versus non-anxious brains from fear extinction learning task-based fMRI
Using a neuroimaging cohort study (n = 304 adults, 92 patients with anxiety, 74 trauma-exposed individuals, 138 matched controls), Wen et al.158 examined how the fMRI activations of 10 brain regions that were commonly activated during fear conditioning and extinction (Figure 4A) might distinguish anxious or trauma-exposed brains from controls. They proposed a CNN classifier (Figure 4B) to map fear-induced fMRI activities in space and time to a prediction probability score indicating that the
subject belongs to the anxious group. The CNN achieved an area under the receiver operating characteristic curve (AUC) of 0.84 ± 0.01, 0.75 ± 0.03 sensitivity, and 0.77 ± 0.02 specificity in 5-fold cross-validation (Figure 4C), outperforming other ML methods (e.g., SVM and random forest). The prediction score was also found to correlate with the anxiety sensitivity index (ASI) in the control group (Figure 4D). Furthermore, control analyses were performed to demonstrate the specificity of the fear network in discrimination (Figure 4E).

Case study 4: Decoding mood state from multi-site intracranial brain activity
From intracranial ECoG signals and simultaneously collected self-reported mood-state measurements over multiple days in seven patients with epilepsy, Sani et al. developed a dynamic state-space model (SSM) framework to track the patients’ mood-state variations over time (Figure 5A). The modeling framework consists of unsupervised- and supervised-learning components (Figure 5B). The spectro-spatial features were extracted from the mood-predictive network within the limbic brain region. The neural decoders were also highly predictive of the immediate mood scaler (IMS) points at the population level. Furthermore, the same trained decoder could be used for mood-state prediction across hours and days and generalized across a wide range of IMS. In cross-validation, the decoders could predict IMS variations that covered 73% and 33% ± 7.2% of the total possible IMS range across all seven subjects and within individuals, respectively (Figure 5C). These results suggest that ML-based decoders can predict mood-state variations from brain activity across multiple days of recordings in patients.

ML-POWERED TECHNOLOGIES FOR PSYCHIATRY
ML can be applied to a wide range of digital platforms, including software (e.g., mobile apps), hardware (e.g., wearable devices, robots), social services (e.g., online chatbots), and clinical practice (e.g., EHRs). In this section, we will review various ML-powered technologies in the non-neuroimaging domains and highlight the emerging digital platforms for precision psychiatry.
A recent McKinsey study showed that use of telehealth has increased by 38-fold compared with the pre-COVID baseline.\textsuperscript{160} With a steep increase in teletherapy demand and consumption, many companies (such as Talkspace and Headspace Health) provide services that include chat-based conversations with licensed mental health professionals. The definition of teletherapy today has expanded to include these newer modalities of care delivery. These advances in care delivery have enabled collecting massive amounts of text, audio, and video data on a regular basis, which was previously only available in controlled research settings. Furthermore, the recent advancements of natural language processing, speech, and video analysis technologies, combined with the ML tools, have generated numerous innovations in this emerging field. The global psychiatrist community is increasingly aware of these developments. For example, a recent survey among more than 700 psychiatrists showed that 49% believed that in the next 5–10 years, ML technology will help analyze patient information to establish prognosis, and 54% believed that this technology can help synthesize patient information to reach a diagnosis.\textsuperscript{12}

ML can be applied across all stages of a patient’s journey:\textsuperscript{161,162} risk assessment, diagnosis, prognosis, treatment, and relapse in a variety of disorders.\textsuperscript{163} where the analysis can be applied to natural language, speech, facial expressions, body language, and social media, as well as traditional clinical surveys and neuroimaging data.\textsuperscript{22,164} Table 3 summarizes recent representative studies that use ML to support various stages of patient journey. Applying ML can build personalized models that are optimized for each patient,\textsuperscript{8} as opposed to traditional models that are only optimized for group effects. Furthermore, given the inter- and intra-disorder variability between clinical diagnosis and symptoms, ML methods such as MTL can be used to model differential diagnoses between disease categories. All of these mentioned ML applications can be considered to be the first level of precision psychiatry.

However, the amount of precision that can be modeled using ML is far beyond the first level.\textsuperscript{181,182} During psychiatric evaluation, psychiatrists may try to build a mental model of what is going on in the patient’s life in about 30 min. They aim to understand as much as possible about the patient’s history within a short time, define what “normal” looks like for the patient, and identify deviations from the normal. This is often done by asking questions to the patient and examining their speech, body language, and behavioral responses. It is very challenging and nearly unrealistic to expect psychiatrists to build an accurate baseline model of the patient’s entire life within such a short time span while interacting with the patient in a compromised psychological state. ML can help by building baseline models specific to each patient before their visit and present the bounds for various observations as a reference to psychiatrists during the exam.\textsuperscript{183} This can be viewed as the second level of precision psychiatry (Figure 6A). Take MDD as an example: Figure 6B shows how ML can be applied at different stages of a patient’s journey.

In the following subsections, we describe how ML can be applied to clinically relevant data and to support one or more stages of the patient’s journey.

**Mobile and sensing technologies**

The development of smartphones, smart watches, and other wearable sensing devices have enabled us to access more
information about our physical and mental health than ever.\textsuperscript{184} Specifically, several types of signals are relevant for mental health monitoring and assessment (Figure 7A):

- Behavioral and physical signals: location (e.g., GPS coordinates), mobility (e.g., accelerometer).
- Multi-media signals: face expression, speech patterns.
- Social signals: social interactions (e.g., call and text message logs), communication patterns, engagement, online gaming.
- Physiological signals: skin conductance, heart rate variability (HRV), eye movement, electrodermal activity (EDA).
- Sleep activity: sleep duration, sleep staging, phone on/off status.

These signals have different implications and relevance to mental illnesses. Although a single signal may not be indicative of a mental disorder, combination of these physical/physiological/social cues may reveal important clues of individual mental health. In what follows, we will focus on the analysis of multi-media, language, and social media data for development of their mental health applications.

### Speech and video analyses

To date, voice and visual (video of facial expressions and body language behaviors) data have gained increasing attention in the studies of mental disorders. ML technologies using speech samples obtained from the clinic or accessed remotely may help identify biomarkers to improve diagnosis and treatment. Since the early days of practice, psychologists have already used auditory and visual cues to assist with the diagnosis of mental illnesses.\textsuperscript{185} Furthermore, speech and video are not only readily available in traditional teletherapy settings but are also easily interpretable as the most natural form of human communication.

#### Audio and speech features

Acoustic features derived from audio data have been found to be relevant in many mental disorders,\textsuperscript{175,186} including speech analysis for patients with depression, bipolar, and schizophrenia. Table 4 lists some commonly used acoustic features in the analysis of mental illnesses.\textsuperscript{187} These categories have enabled standardization and interpretation of ML-analyzed speech data in clinical applications.

Models built from speech-based features may be effective in predicting the diagnosis of depression and suicidality.\textsuperscript{188} Applications for depression include predicting the presence, severity, and score.\textsuperscript{165,173} These models use prosodic, spectral, or other features computed from raw speech data to quantify flattered speech, slow speech, and other relevant markers. The target outcome variable is derived from a clinically valid scale such as a patient health questionnaire (PHQ-9). Furthermore, models for suicidality that explore similar features have been used in multi-class settings to differentiate among healthy, depressed, and suicidal speech.

One key challenge in applying speech-based models in clinical practice is the lack of longitudinal data validation in real-world settings. However, this issue is starting to get addressed in recent studies,\textsuperscript{197} which detect manic and depressive speech from recordings of outgoing calls from phone conversations of consenting participants. Another challenge is the lack of large labeled datasets for evaluating performance across various methods. To this end, it is noted that companies like Ellipsis Health\textsuperscript{166,198} have used deep learning and transfer learning to predict depression and anxiety scores with high accuracy based on a large labeled dataset of over 10,000 unique speakers. Human-level accuracy using only 20–30 s of an audio clip has been reported in some commercial applications to detect depression.\textsuperscript{199,200}

#### Visual features

Although body language and facial expressions have always formed a key part of a psychiatric exam, ML has only recently been applied to analyze such data objectively. To date, most work has focused on suicidal ideation,\textsuperscript{189} depression,\textsuperscript{168,169,190} schizophrenia,\textsuperscript{178} and ASDs.\textsuperscript{192} Features derived from overall facial expression, eyes, gait, and posture (Table 4) have been shown to be relevant across many mental disorders.

---

**Figure 5. Illustrations of concepts and major findings in case study 4**

(A) Schematic of cross-validation. An IMS point (e.g., SN) is left out as the test IMS to be predicted. The other IMS points (i.e., training IMS, using S1 to SN–1) and the associated neural activity are used within the modeling framework to train a neural encoding model.

(B) Main components of the modeling framework based on both unsupervised and supervised learning.

(C) Cross-validated prediction of the mood state is shown against the true measured mood state (copyright Springer Nature; figures were modified from Sani et al.\textsuperscript{159} with permission).
Studies in suicidal ideation have mainly focused on using interpretable ML for characterizing the disorder. This makes the ML models more applicable in augmenting human caregivers by bringing up a specific insight that they would like to measure. In depression studies, some approaches have also involved fusion of video features derived from each frame that are used to train a sequential DNN,\textsuperscript{201} and some have used pre-training to compensate a relatively small sample size of depression datasets.\textsuperscript{202} While these models perform very well on the same held-out test set, their clinical applications remain limited due to a lack of interpretability. To improve interpretability, depression activation maps were developed to highlight the facial areas corresponding to the depression severity as learned by the ML model.\textsuperscript{203} Meanwhile, utilization of pre-defined features has been most effective in providing interpretable results.\textsuperscript{202,204}

**Natural language processing (NLP)**

NLP techniques enable computers to analyze, understand, and derive meaning from text and speech in a similar manner to humans. NLP techniques can enable mental health professionals to evaluate language patterns to help identify and predict psychiatric illness (Table 4). Language is not only one of the primary expressions of human behavior that carries a variety of implicit and explicit markers relevant to mental health\textsuperscript{205,206} but is also more abundantly available compared with speech data. For example, social media platforms contain a large quantity of real-world language data, whereas speech data are rarely available at that scale. There are two types of NLP applications for detecting specific mental health symptoms. The first type of applications is directly applied to the patient data, varying from predicting the risk of suicide and early psychiatric readmission to identifying phenotypes and comorbidities. The second type of applications is applied to EHRs and clinical records (tests, transcripts), which can be used for automating chart reviews, clustering patients into phenotype subtypes, and predicting patient-specific outcomes. The EHRs (including pathology reports, lab results, clinical tests, and clinical session transcripts) are systematic collections of longitudinal, patient-centered clinical records. Patients’ EHRs consist of both structured and unstructured data. The structured data include information about a patient’s diagnosis, medications, and laboratory test results, and the unstructured data include information in the form of clinical notes.

### Table 3. Representative ML applications of multi-media data in mental disorders

| Study | Data source(s) | Patient journey stage | ML approach | Test sample size |
|-------|----------------|-----------------------|-------------|-----------------|
| **Depression spectrum** | | | | |
| Vázquez-Romero et al.\textsuperscript{165} | audio – clinical interviews | diagnosis | CNN ensemble | 47 speakers |
| Harati et al.\textsuperscript{166} | audio – answers to personal questions | diagnosis | transfer learning | 3,078 speakers |
| Huang et al.\textsuperscript{167} | audio – clinical interviews | diagnosis | SVM with speech landmark features | 47 speakers |
| Zhu et al.\textsuperscript{168} | face video – reading and personal questions | diagnosis | CNN | 50 videos |
| Shao et al.\textsuperscript{169} | gait-only video – casual walking in a corridor | diagnosis | LSTM + CNN weighted fusion | 40 videos |
| Lu et al.\textsuperscript{170} | language – answers to personal questions | diagnosis | LSTM fine-tuned with health forum data | 2,425 subjects |
| Eichstaedt et al.\textsuperscript{171} | language – Facebook posts | risk assessment | logistic regression | 68 patients |
| Sun et al.\textsuperscript{172} | audio, video – clinical interviews | diagnosis | transformer + multi-modal fusion | 56 subjects |
| **Bipolar spectrum** | | | | |
| Weiner et al.\textsuperscript{173} | audio – verbal fluency tasks | relapse | SVM | 56 subjects |
| Palmius et al.\textsuperscript{174} | sensor – GPS | diagnosis | linear regression | 36 subjects |
| **PTSD** | | | | |
| Marmar et al.\textsuperscript{175} | audio – clinical interviews | diagnosis | random forest | 43 veterans |
| Mallol-Ragolta et al.\textsuperscript{176} | audio, video, skin conductance | relapse | SVM | 110 subjects |
| **Schizophrenia spectrum** | | | | |
| Tahir et al.\textsuperscript{177} | audio – clinical interviews | diagnosis | SVM | 70 subjects |
| Abbas et al.\textsuperscript{178} | video – neutral open-ended questions | diagnosis | logistic regression | 16 subjects |
| Birnbaum et al.\textsuperscript{179} | language – internet search queries | relapse | random forest | 23 subjects |
| Birnbaum et al.\textsuperscript{180} | audio, video – clinical interviews | diagnosis | gradient boosting | 17 subjects |
socioeconomic features can be combined with the language data. Symptoms derived from free texts in EHRs have been used for prediction of bipolar disorder, sitational aggression, and suicidal ideation with achieved comparable performance to clinicians. Furthermore, discharge summaries from EHRs have also been used to predict relapse. Aside from symptoms, a variety of relevant mental health data (such as the intervention status and physical health comorbidities) can be routinely extracted from EHRs using NLP methods. Privacy concerns around EHR data sharing remain one of the key challenges in validating generalization of NLP methods. Encouragingly, there has been a growing interest in using transformers for generating artificial mental health clinical notes to mitigate this issue.

The advances in text-based mental health interventions (e.g., Talkspace and CrisisTextLine) have made transcripts of clinical sessions easily amenable to NLP. Aside from developing models for detecting suicidal ideation, NLP can also be applied to these datasets to identify the population-level trend, such as the increase in anxiety and decrease in quality of personal relationships during the COVID-19 pandemic. Since language data are ubiquitous, one of the NLP challenges in mental health applications is data standardization. Depending on the task, different types of data may yield different levels of “signal.” For example, to predict first-episode psychosis, language data from clinical tests have higher performance compared with transcripts of free speech. On the other hand, data collected from free-speech samples for diagnostic purposes can be highly effective for developing a language-based depression screening that generalizes well across various age groups.

Social media
To date, social media companies have collected a large amount of language data that may contain clinically relevant information. This information can not only be extracted on a population level, such as the notable rise in cognitive distortions over time, but can also be attributed on an individual level, all of which have made social media a powerful tool to support mental health risk assessment and diagnosis. Language from Facebook posts, for example, has been shown to contain markers for depression. Ruminations and sadness can be detected in such data up to 6 months prior to a clinical diagnosis. Models applied to Facebook and other platforms have successfully predicted the diagnosis of psychosis, anorexia, anxiety, and stress levels. In addition to the language analysis for the user posts and comments, ML models may also process media data such as Instagram images or integrate images and text to infer the user’s mental state. Entries of online search also form a complementary and equally compelling dataset alongside social media activity.

Recent developments of transformer models, including those learning multilingual language representations, have enabled researchers to apply powerful NLP models to detect depression or self-harm from social media data. Furthermore, specialized language representations that were trained on mental health-specific conversations and that became publicly available have been shown to improve performance compared with non-specific representations. Finally, an ML technique known as transformers can assess text responses via NLP and predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy.

While social media solves the scale issue with millions of samples available, most social media data lack clinically valid labels. Most reported studies have relied on using labels from self-disclosure of mental illness, which are not only inaccurate but also bring additional issues of defining a healthy control. Despite the challenges, the validity of social media data has been repeatedly proven to support mental health diagnosis and risk assessment.

Sensing technologies and mobile mental health
Smartphones, wearables, and other devices equipped with ambient sensors (Figure 7B) are increasingly capable of recording physiological measurements that are known to affect mental health. In addition, some of the less obvious measurements (such as keystroke usage patterns) have been shown to...
be implicated in mental health.\(^{227,228}\) Additionally, online gaming behaviors, such as interaction patterns with non-player characters (NPCs) and other game behavior patterns, can be used to measure cognitive performance and their relationship with mental illness.\(^{229,230}\)

Measurements from mobile sensors may constitute valuable sources of mental health data (Table 5) and can be useful at various levels of granularity: from raw sensor data (e.g., the accelerometer) to derived high-level features (such as psychomotor activity). This has inspired many corporations to invent technologies for detecting depression and cognitive decline based on data collected from their wearable devices.\(^{231}\) Sensor-based measurements are found to be correlated with high-stress levels and a variety of ailments including depression, anxiety, psychosis, and bipolar disorder.\(^{232,233}\) Since sensor-based data are widespread and readily available, they offer an opportunity to build baseline models for individual users, which can be then used to identify significant physiological changes in users and further inform clinical interventions. Devices that collect various data streams from patients, such as surveys, cognitive tests, social medial interactions, GPS coordinates, and behavioral patterns (e.g., keyboard typing), have a great potential for monitoring, managing, and predicting the individual’s mental health.\(^{234,235}\) Overall, longitudinal quantification of these data streams may result in clinically meaningful markers that can be used to refine diagnostic processes, tailor treatment choices, improve condition monitoring for actionable outcomes (such as early signs of relapse), and develop new intervention models.\(^{236–239}\)

**Commercial and research platforms and services**

While studies have demonstrated promising results in ML applications for the individual’s journey in mental health, their broad applicability in clinical practice remains limited. Table 6 lists examples of platforms and services that use ML for mental health applications. While most platforms focus on risk assessment based on single modality, the initial commercial viability of these platforms is still promising for the success of using ML in mental health because they enable a collection of large amounts of data that can be used to further validate biomarkers.

There are between 10,000 and 20,000 smartphone apps that digitize mindfulness or cognitive behavioral therapy (CBT) techniques,\(^{244}\) allowing the user to engage in psychotherapy at a greatly reduced price compared with in-person therapy. However, the quality is highly variable, and the mechanisms used to validate them are often dubious. Moreover, since this area is relatively new, the industry and governmental standards to validate such a technology are still in the early phase. We will briefly outline two interrelated areas of development: digital measurements and digital interventions.

**Digital measurements**

We are entering a new era of digital psychiatry.\(^{11,245}\) In 2016, the Harvard professor Jukka-Pekka Onnela coined the term “digital phenotype,”\(^{246}\) which refers to the use of mobile devices and other digital data sources to measure behavior and physiology for understanding brain activity that is relevant to pathological states. These techniques utilize measurement paradigms from translational neuroscience, which were developed in laboratory settings such as direct quantification of motor (i.e., movement, muscle activation) and physiological activity (i.e., heart rate, electro-dermal response) more than traditional clinical scales or self-report scales. The advantage of this approach is that it better aligns with emerging knowledge of rapid-acting biological processes and provides high measurement accuracy through direct rapid sampling, which is in contrast to traditional clinical measures that are taken sporadically over a long period.\(^{164,247,248}\) These measurement approaches have relevance in multiple areas including treatment development, treatment selection, and ongoing monitoring.

Medications that target mental health conditions have a significant history of failure. Most psychiatric medications were discovered capriciously rather than being developed based on

---

**Table 5. Mobile Sensors and Mental Health Applications**

| Sensor Type | Application |
|-------------|-------------|
| Accelerometer | Measuring psychomotor activity |
| Heart rate | Monitoring physiological activity |
| GPS coordinates | Tracking location and movement |
| Social medial interactions | Assessing interaction patterns |
| Cognitive tests | Evaluating cognitive performance |

---

**Diagram A: ML applications in mobile health**

- Microphone
- FaceTime
- GPS
- Mobility
- Call Logs
- SMS / Email
- Screen Time
- App Usage
- Light Sensor
- Browser History

**Diagram B: Different types of data collection strategies for digital measurement tools**

- Wearable e.g., smart watches, podometers, restrictively monitors
- Electronic behavior e.g., social media interactions and key frame activity
- In-home sensors e.g., camera, motion sensors, sleep monitors
- Handheld devices e.g., smartphones and tablets with GPS
- Biological devices e.g., MRI scanners, genetic sequencing

**Diagram C: A technological infrastructure for the integration of digital measurement tools**

- Sensor data transfers
- Platform data transfers
- EHR

---

**Figure 7. Illustrations of ML-powered technologies for mental health**

(A) ML applications in mobile health.
(B) Different types of data collection strategies for digital measurement tools.
(C) A technological infrastructure for the integration of digital measurement tools. Independent platforms for measurement of health will have their own data repositories, depicted as clouds. These data could be safely transferred across platforms using transfer tools such as secure application program interfaces (APIs), depicted using dashed arrows. Such tools could allow for both unidirectional and bidirectional movement of data. ML can be applied to integrate all measures for clinical decision-making (B and C are reproduced from Vázquez-Romero and Gallardo-Antolín\(^{164}\) with permission).
knowledge of the underlying biological mechanisms. As new medications emerge from basic and translational neuroscience research, both drug developers and clinicians struggle with how to measure the effects of new treatments and how to properly target old treatments. For example, traditional anti-depressant medications are designed to slowly titrate serotonin levels, resulting in slow global effects over a 2- to 4-week period. Correspondingly, measures of depression based on the DSM, query about the presence of depressive states over a 2-week period. Similarly, medications are designed to slowly titrate serotonin levels, resulting in slow global effects over a 2- to 4-week period. Correspondingly, measures of depression based on the DSM, query about the presence of depressive states over a 2-week period. For example, traditional anti-depressant medications such as ketamine and psilocybin affect specific depressive symptoms in minutes. Further, the mechanistic effects, and thus the need for measurement, is much more specific and granular. In fact, most classes of anti-depressants including serotonin reuptake inhibitors (SSRIs) and psilocybin, act on serotonin receptors that ultimately impact peripheral motor and physiological activity. Serotonin regulation will likely have a higher likelihood of capturing both pathology and treatment effects.

As an example, research effort has been dedicated to using computer vision and voice to directly quantify motor activity. Some recent work has demonstrated that digital phenotyping parameters that reflect gross motor activity including speech characteristics (rate of speech, tone) and facial/head movements are associated with suicidal risk. SSRI response in MDD, negative symptomatology in schizophrenia, and Parkinsonian tremor. Such approaches are now being commercialized for all phases of drug development from proof of concept to direct measurement in order to make decisions about ongoing treatment needs. Such measures solve many of the current problems in clinical measurement since they can be captured remotely in an automated way. These measures can also be captured at a much higher frequency and provide a sensitive numeric value.

Meanwhile, these new approaches to measurement have significant challenges. First, methods that are adapted from the laboratory often lack the tight experimental control necessary to interpret the data correctly. For example, a rapid change in physiological responses (such as adrenaline and cortisol, or HRV) can indicate stress but also exercise or other forms of exertion. Second, while the scientific basis of these measurement paradigms may be sound, commercial approaches are rarely validated to the extent required to be of clinical utility, or they are rarely sufficiently transparent in their approach to be used for regulatory approval.

**Digital interventions**

The other rapidly emerging area of mental health technology are digital approaches to clinical care. We will briefly outline some of the leading approaches. Importantly, digital approaches to clinical care are often aligned with a digital measurement approach as these approaches are “blind” without some sort of remote data. A number of companies such as Mindstrong Health, IesoTrigger Health, and Headspace Health have attempted to integrate digital phenotyping to identify when patients are in acute clinical need. However, it is unclear how accurate these methods are, as they are typically unpublished. This has led to the development of models that can identify patterns in patient’s and clinician’s language that are markers of improved outcomes; these can be further used to measure success of various therapy modalities and treatment design, as well as to improve care quality.

Digital therapeutics proposes the use of mobile devices to offer CBT, mindfulness, or other validated psychotherapy in an
A central goal of precision psychiatry is to integrate all clinical, physiological, neuroimaging, and behavioral data to derive reliable individualized diagnosis and therapeutics. Importantly, health-related data are produced daily, especially from personal devices. The most essential effort in multi-modal data analysis tasks is to explore the relationship between modalities, complementarity, shared versus modality-specific information, and other mutual properties. Multi-modal data fusion techniques present a framework to infer information on how different data modalities interact and can be integrated for improved disease prediction. In this section, we will review several data fusion methods in diagnostic analytics (popular ML methods for multi-modal fusion). We will focus on multi-modal neuroimaging data (multi-modal neuroimaging studies), and then extend the discussion to other modalities including vocal and visual expression data (multi-modal fusion of non-imaging data).

### Popular ML methods for multi-modal fusion

In the past decades, numerous research efforts have been dedicated to developing powerful ML methods for multi-modal data fusion. Some commonly used approaches are summarized below (Figure 8).

#### Multi-variate correlation analysis

CCA is a standard statistical method based on second-order statistics for data fusion. It aims at finding a pair of linear transformations to drive latent variables (also known as [aka] canonical variates) that have maximized correlation between two different data modalities. For a more general setting, multi-set/multi-way CCA (mCCA) has been developed as an extension of the standard CCA to multi-modal fusion by maximizing the overall correlation among latent variables from more than two sets of modalities. Similar to CCA, PLS and its extensions, i.e., multiway PLS (N-PLS), provide alternative approaches to integrating multi-modal data by maximizing the covariance between latent variables from different modalities.

#### Matrix and tensor factorization

Based on matrix and tensor factorization techniques, joint blind source separation (BSS) approaches have been developed and successfully applied to multi-modal fusion of biomedical data. As a typical example, joint ICA (JICA) aims to maximize the independence among jointly estimated components from multiple modalities that are assumed to share the same statistics for data fusion.

---

**Table 5. Mobile sensor measurements and potential applications for mental health monitoring**

| Measurement | Feature | Effect in mental health |
|-------------|---------|-------------------------|
| Movement    | psychomotor agitation | increased in anxiety |
| Location    | social avoidance       | increased in MDD     |
| Social activity | call/text volume      | reduced in MDD      |
| Keystroke   | keystroke latency      | impaired in ADHD    |
| Heart rate  | heart rate variability | impaired in stress  |
| Gaming      | NPC interactions       | impaired in social anxiety |

---

**Table 6. Commercial and research platforms and services for mental health applications**

| Platform     | Primary data source | Mental health appl.     |
|--------------|---------------------|------------------------|
| Woebot       | language            | depression, anxiety    |
| Mindstrong   | keystrokes          | serious mental illness |
| Sonde Health | voice               | mental fitness         |
| Ellipsis Health | voice           | stress                  |
| Amazon Halo  | voice               | emotion detection      |
| Apple Watch  | mobility, sleep     | depression              |
| Alphabet Fitbit | skin conductance  | stress                  |
| Kintegi      | voice               | depression, anxiety    |
| Bewie        | raw data from smartphone | multiple             |
| MindDoc      | language (ask daily question) | depression |
| Clarigent Health | voice              | suicide risk           |

---

**MULTI-MODAL DATA FUSION IN DIAGNOSTIC ANALYTICS**

A central goal of precision psychiatry is to integrate all clinical, physiological, neuroimaging, and behavioral data to derive reliable individualized diagnosis and therapeutics. Importantly, health-related data are produced daily, especially from personal devices. The most essential effort in multi-modal data analysis tasks is to explore the relationship between modalities, complementarity, shared versus modality-specific information, and other mutual properties. Multi-modal data fusion techniques present a framework to infer information on how different data modalities interact and can be integrated for improved disease prediction. In this section, we will review several data fusion methods in diagnostic analytics (popular ML methods for multi-modal fusion). We will focus on multi-modal neuroimaging data (multi-modal neuroimaging studies), and then extend the discussion to other modalities including vocal and visual expression data (multi-modal fusion of non-imaging data).
Multivariate Correlation Analysis
- mCCA
- N-PLS

Matrix and Tensor Factorization
- JICA
- IVA
- JICA + mCCA
- CMFT

Deep Learning-based Fusion
- Multitask GNN
- SAE
- Multimodal GNN
- Combined Models

Multi-modal Fusion Approaches
- CIFA
- JIVE

Mixing matrix. The JICA approach involves concatenating modality features alongside each other and then performing ICA on the composite feature matrix. Independent vector analysis (IVA) is another extension of ICA to multiple datasets. IVA makes use of dependence across datasets by defining source component vectors concatenating a specific source estimated from multiple modalities. Coupled matrix and tensor factorization (CMFT) was also developed to simultaneously factorize multiple datasets in the form of matrices and high-order tensors using tensor decomposition, showing strength in capturing the potential multi-linear structure for multi-modal fusion. Besides extracting shared common components, some multi-modal fusion tasks are also interested in deriving individual components that are modality specific. Common and individual feature analysis (CIFA) and joint and individual variation explained (JIVE) models have been proposed to achieve this goal. By jointly decomposing multiple feature matrices, CIFA and JIVE are able to simultaneously estimate common and individual feature subspaces. A further extension of CIFA has been achieved by leveraging high-order tensor factorization, which provides an efficient way to perform a multi-dimensional fusion of multiple data modalities.

Multi-kernel learning
Multi-kernel learning (MKL) has won many successful applications in multi-modal data fusion due to the full utilization of multiple kernels that enable simultaneous learning from various modalities with heterogeneous data. Different kernels naturally correspond to different modalities, such as neuroimaging, clinical behavior, speech features, etc., which may provide complementary information to drive improved modal learning performance. The MKL problem can be set as a linear combination of kernel matrices or a non-linear function with specified forms of regularization. MKL may be designed under different ML models including SVM, GP, and clustering. Among them, MKL-SVM has been most popularly applied to integrate heterogeneous data modalities in studies of mental health.

Deep learning-based fusion
Cutting-edge deep-learning techniques have become increasingly popular for deep multi-modal fusion. Data fusion through deep learning allows integration of multiple modalities based on learned high-level feature representations that are theoretically more comparable with each other and more informative for predicting the target. By exploiting cross-modal manifolds as a feature graph, a deep manifold-regularized learning model was recently designed to integrate transcriptions and electrophysiology data from neuronal cells, and yielded promising performance for phenotype prediction.

GNNs show capability in information fusion for multi-modal causability by defining casual links between features with graph structures, thereby enhancing the explainability of the derived multi-modal feature representation. By extending GNNs to multi-modal structures, deep-representation approaches have also been designed for integrating brain networks constructed from diverse modalities.

Multi-modal neuroimaging studies
Neuroimaging data types are intrinsically dissimilar in nature, having different spatial and temporal resolutions. Instead of feeding the entire dataset into a combined analysis, an alternate approach is to reduce each modality to low-dimensional (latent) features of selected brain activity or structure and then explore associations across these feature sets through variations across individuals. Exploiting such latent feature representations from multiple neuroimaging modalities for diagnosis has generally been shown to improve performance compared with using a single modality alone. Multi-modal fusion allows for integration of neuroimaging data modalities from different scales of spatial and temporal resolutions. Combining multi-modal neuroimaging offers an elegant way to exploit complementary information for more accurate and robust characterization of brain dysfunctions and hence is instrumental in optimal decisions for diagnosis, prognosis, and treatment in psychiatry.

A combination of mCCA and JICA was successfully applied to fMRI and DTI fusion in the diagnosis of schizophrenia and bipolar disorder. CMTF has been applied to identify diagnostic biomarkers of schizophrenia by integrating sMRI, fMRI, and EEG. MKL-SVM has been successfully applied to integrate multi-modal structural neuroimaging for predicting differential diagnosis between bipolar and unipolar depression and to combine sMRI and fMRI for improved classification of trauma survivors with and without PTSD. More recently, it also showed efficacy in the diagnosis of early adolescent ADHD by integrating sMRI, fMRI, and DTI. In learning low-dimensional representations of fMRI and sMRI, the fMRI can be split into several independent component networks, each treated as a separate modality along with the structural scan for learning using autoencoders. Furthermore, MKL methods have been used for diagnosing schizophrenia by combining markers from MRI and DTI. A multi-modal GCN was designed to integrate functional and structural connectomics data for an improved prediction of phenotypic characterizations in ASD. By combining multiple typical neural network structures, multi-modal deep-learning models have also been developed to effectively integrate fMRI connectivity and sMRI features and also genomic data for discovering schizophrenia-associated brain dysfunction. Methods for learning joint representations from neuroimaging and non-neuroimaging data are still in early development, and there is an opportunity for ML methods to evolve for this task. For example, transformer networks with late fusion can be used to learn joint representations from various modalities such as EEG and eye movement signals.

Multi-modal fusion of non-imaging data
Multi-modal approaches consist of combining data from various sources to jointly arrive at an answer. Given how little is conclusively known about which type of data, neuroimaging, social

Figure 8. Summary of typical approaches for multi-modal data fusion in psychiatry studies
ML FOR MOLECULAR PHENOTYPING IN PSYCHIATRY

Molecular phenotyping is referred to as the technique of quantifying pathway reporter genes (i.e., pre-selected genes that are modulated specifically by metabolic and signaling pathways) in order to infer activity of these pathways. Mapping genes and genomics to behaviors can identify risk factors and biomarkers in mental disorders. The brain is the central organ exposed to stressors and external behavioral interventions and is therefore vulnerable to changes in multiple interacting biological networks at the systems level. ML methods may play an active role in capturing the complexities of interacting variables within and across multiple levels (Figure 9A). For instance, at the molecular level, ML may help identify mechanistic-based phenotyping models as new targets for prevention and treatment of mood and cognitive disorders. The advent of unbiased next-generation sequencing (NGS) has prompted the development of bioinformatics and ML tools to profile and decode large molecular datasets (e.g., transcriptomics, epigenomics, metabolomics) at the genome-wide level in health and disease states (Figure 9B). To date, increasing applications of ML methods have integrated these multi-level molecular datasets with clinical characteristics to map specific neurobiological substrates into the complexity of symptom clusters, which may further help the classification of diseases, prediction of treatment outcomes, and selection of personalized treatment.

Animal models have been playing a vital role in precision psychiatry for understanding disease mechanisms and predicting treatment responses. Gene expression studies that integrate neuroscience, ML, and bioinformatics approaches can contribute to advancing understanding of the molecular basis of MDD and bridge the knowledge gap between animals and humans. Using RNA sequencing assays and gene coexpression network analyses (based on hierarchical clustering to identify gene modules), differential gene expression profiles have been shown across six key brain regions in post-mortem tissues of patients with MDD compared with age- and sex-matched controls, along with remarkable sex differences in these molecular pathways. Recent work using RNA sequencing assays at single-nucleus resolution (snRNA-seq) and t-distributed stochastic neighbor embedding (t-SNE) analyses showed cell-type-specific transcriptomic profiles in the post-mortem dorsolateral prefrontal cortex (PFC) that are differentially regulated in MDD cases. Importantly, these gene expression studies in humans were supported by findings in rodents showing a brain that continually changes with experience. Several studies based on RNA-seq assays and bioinformatic analyses have showed striking transcriptomic differences in the ventral and dorsal hippocampus in the responses to stress—a primary risk factor for multiple psychiatric diseases—with the ventral hippocampus being sensitive not only to the effects of stress but also a target for the responses to next-generation anti-depressants.

The expansion of NGS to single-cell resolution assays provides opportunities for advanced bioinformatics and ML approaches to analyze large datasets, which include denoising and dimensionality reduction, cell-type classification, gene regulatory network inference, and multi-modal data integration. For instance, the software toolkit for single-cell genomics, Seurat (https://satijalab.org/seurat/), combines unsupervised non-linear dimensionality reduction, K-nearest neighbor graph analysis for cell-type clustering, and weighted nearest-neighbor analysis for multi-modal data integration. Deep-learning approaches, such as the deep autoencoder, provide analytic tools for denoising and dimensionality reduction. Autoencoders can also be used in a supervised manner for transfer learning across datasets, such as to learn the embedding from a larger previously annotated dataset and to transfer this knowledge to cluster new datasets. The combination of multi-modal data generated from the simultaneous assessment of transcriptomic profiles...
Integration of multidimensional factors for new mechanistic treatment models

It has been increasingly recognized that mood and cognitive disorders are unlikely to be only contributed by the brain. Instead, growing evidence has suggested that they are system-level disorders affecting multiple interacting biological pathways. Involving dynamic cross-talk between the brain and the body, using hierarchical clustering to integrate in vivo molecular measures of brain metabolism with clinical symptoms in patients with MDD, recent work has shown that the specific neurobiological substrates map into discrete clinical symptoms, including anhedonia. Furthermore, the integration of multi-dimensional factors spanning mitochondrial metabolism, cellular aging, metabolic function, and childhood trauma may provide more detailed signatures than individual factors to predict longitudinal changes in depression severity in response to the metabolic agents used as anti-depressant treatment. Furthermore, deployment of multi-omics approaches and a random forest classifier has achieved 85% sensitivity and 77% specificity in prediction of the PTSD status. This system-level diagnostic panel of multiple molecular and physiological measures outperformed separate panels composed of each individual data type, showing certain mitochondrial metabolites as the most important predictors.

Another example of ML applications includes the integration of multi-dimensional phenotypic measures to identify those mechanisms that pre-dispose apparently healthy individuals to develop maladaptive coping strategies from those that confer resilience. A recent study has used a high-throughput unbiased automated phenotyping platform to collects more than 2,000 behavioral features and applied supervised ML to minimize Bayesian misclassification probability. The results have demonstrated that such a rich set of behavioral alterations can distinguish susceptible versus resilient phenotypes after exposure to social defeat stress (SDS) in rodents. Furthermore, the ML classifier may integrate a priori constructs (such as the measures of anxiety and immune system function) and predict whether a given animal developed SDS-induced social withdrawal or remained resilient. Additionally, integration of features can improve the classification sensitivity (~80%), which is better than the performance derived from either individual measure alone.

The development of personalized psychiatry strategies for diagnosis and treatment will benefit from meeting the demand enforced by the recent advent of molecular biology protocols, which may provide opportunities to capture CNS nanovesicles (known as exosomes) and examine specific neurobiological substrates (e.g., transcriptomic profiles). ML-based dynamic network analyses will also enable us to link brain molecular targets and signaling pathways with other levels of analyses and to incorporate the brain-body relationship to redefine thinking about the mechanisms throughout the complex disease course.

**XAI IN PSYCHIATRY**

XAI aims to provide strong predictive values along with a mechanistic understanding of AI by combining ML techniques with effective explanatory techniques. XAI has found emergent applications in medicine, finance, economy, security, and defense. In psychiatry, XAI can help clarify the link between neural circuits and behavior and improve our understanding of therapeutic strategies to enhance cognitive, affective, and social functions. XAI distinguishes the standard AI in two important ways: (1) it promotes transparency, interpretability, and generalizability and (2) transforms classical “black box” ML models into “glass box” models while achieving comparable or improved performance. From the diagnosis or prognosis perspective, it is crucial to know whether the ML solutions are explainable to the point of providing mechanistic insights into the way brains execute a particular function or complex behaviors. For instance, an ML-produced classification function to predict a disease outcome would need to not only report a probability outcome but also address additional questions for the end user: why is it this outcome instead of the alternative? How reliable is the outcome? When will it fail if something is missing or misrepresented? When and why is the prediction wrong? Accordingly, a model with high interpretability is often accompanied with parameter/structure/connectivity constraints or some prior domain knowledge. These explainable models can be continuously adapted such that an iterative process may be required to force ML models to fit models with specific interpretations.

**Interpretability and interpretable ML models**

A model is interpretable if its outcome and operation that leads to the outcome can be understood by human users. In terms of taxonomy, intrinsic interpretability of ML models is attributed to their simple structures, such as short decision trees or sparse linear models. Post-hoc interpretability is referred to the application of interpretation methods after model training (online resource: https://christophm.github.io/interpretable-ml-book/). Interpretation may appear in different forms: (1) finite...
feature summary statistics, (2) meaningful model parameters, or (3) easy visualization of the model outcome (e.g., feature summary or decision boundary). Interpretability and explainability are two similar concepts and are sometimes used interchangeably. Interpretability often, but not always, implies explainability. A model is highly explainable if it shares at least one of the following properties of explainability:335 high portability (regarding the range where the model can be applied), high expressive power (regarding the model strength in explaining the outcome), low translucency (regarding the model dependency on specific conditions), low algorithmic complexity, and informative constraints. Generally, there is a trade-off between model explainability and performance. For instance, a constrained linear or bilinear model may fit many of these criteria, but the linear model does not warrant a good performance. Additionally, a model that is potentially explainable does not guarantee explainability. For example, codependence of input variables may make explanations ambiguous; latent variables of probabilistic generative models may face the problem of “explaining away.”334 Here, we briefly mention several classes of interpretable ML models.

Hybrid rule-based ML models
This type of ML models can be used for generating rules, such as a decision rule set: IF (condition) THEN (outcome 1) ELSE (outcome 2) statement, where the conditional clause will be learned from data.341 This type of model has more expressive power but less portability.

Constrained ML models
This type of ML models imposes parameter constraints to avoid overfitting and enhance interpretability. Examples of such include the constrained convolutional filters in the CNN model342 or constrained mixture models used for clustering.343 As a result, these constrained models have low translucency.

Feedback ML models
ML models can be provided with user feedback in the human-in-the-loop system, where the user feedback is treated as a constraint in the optimization problem.344,345 The feedback may appear as a form of rule sets that are either known or unknown in advance. Feedback can also help interpolate missing data and constrain the solution. Iterating feedback-rule optimization steps can generate more accurate rule sets. This type of model has good expressive power and high portability.

Circuit-level modeling for computational psychiatry
Rooted in ML, computational psychiatry shares a similar goal with XAI and tries to combine multiple levels and types of computation with behavioral and neuroimaging data in an effort to compute understanding, prediction, and treatment of mental illness.346 The levels of computation range from short to long timescales (min/h/days/weeks) and focus on the changes in brain activity and behavior. The types of computation include circuit-level modeling, data-driven analytics, and theory-driven algorithmic development. Two complementary approaches have been proposed in computational psychiatry: (1) data-driven approaches, which apply ML methods to high-dimensional multi-modal data to tackle classification and prediction problems (how can ML help psychiatry?), and (2) theory-driven approaches (such as reinforcement learning), which develop algorithmic or mechanistic models to test hypotheses.

In the second approach, an important research topic is circuit-level computational modeling of macroscopic or mesoscopic brain dynamics for mentally ill brains in task or resting-state conditions.347–349 A common strategy is to first use a biologically inspired model to simulate neural activity based on a network of interacting neural masses and next, within each brain area, to model the neuronal population activity as the Wilson-Cowan neural mass model, with each consisting of excitatory and inhibitory populations.350 Furthermore, individual brain nodes are coupled together according to the empirically derived anatomical network.351 The computational model can be driven by an empirical hypothesis or EEG/fMRI data.

One data-driven macroscopic-level modeling approach is dynamic causal modeling (DCM). DCM has been widely used in characterizing the effective connectivity of a functional network based on task or rs-fMRI,352,353 where the model parameters are inferred from unsupervised learning. By incorporating prior knowledge or hypotheses of network connections, DCM may reveal important brain mechanisms and offer experimental predictions. One potential application of DCM is to characterize the neural plasticity in human brains, especially the change in functional connectivity informed by neuroimaging studies. The functional connectivity can either change gradually during the course of tasks or be induced by neurostimulation. These changes are often, but not always, associated with changes in functional activation of specific brain regions.

CLOSING THE LOOP FOR TESTING CAUSALITY THROUGH NEUROSTIMULATION

From the treatment perspective, it is critical to obtain an improved understanding of brain dynamics that are responsible for dysfunctional cognitive functions and maladaptive behaviors in mental illnesses. To find the hidden cause, the concept of causality requires special attention in perturbing the brain activity. Because of the complexity of the human brain and brain-behavior relationship, adaptive closed-loop neurostimulation provides a critical step to understand healthy and diseased brains.354,355 Neuroimaging provides a passive sensing approach to observe the (correlational) brain-behavior relationship. However, correlation is different from causation. Correlational dependencies describe associations of measurements that experiments do not control, whereas causal dependencies link a dependent variable to an experimentally controlled variable. The key concept in causal inference is to introduce randomization to perturb the mapping. The relationship between every dependable variable and the randomized variable is causal, whereas the relationship between non-randomized variables and behavior remains correlational.356 A closed-loop experimental design would help to test the potential causality.354 In human experiments, we classify closed-loop testing into two categories: one being fully automated, and the other being closed human in the loop.

One big challenge in human psychiatric neuroscience is the causality gap.357 Statistical causality or Granger causality between two variables is not equivalent to brain-behavioral causality. To identify an effective treatment strategy for mental illnesses, it is critical to causally modulate neural circuitry that is
responsible for maladaptive behaviors. Human neuroimaging alone only demonstrates correlations but not causation. To understand the causal mechanisms, it is imperative to close the loop in experiments by perturbing the brain circuits and measuring its outcome, as commonly done in animal experiments. Unfortunately, a rigorous and causal grounding of clinical symptoms and behavior is still missing. Since the clinical symptoms are diverse, how to define the dimension of brain function that defines one or few clinical symptoms and how to effectively manipulate them remains unknown. Furthermore, closed-loop neurostimulation has a conceptual link to XAI for mental health studies and can be seen as an extension in the design of human brain-machine interface (BMI) to causally test brain-behavior mapping.

Neuromodulation therapies have become increasingly popular in treating various neuropsychiatric and neurological disorders. Temporally precise neuromodulation tools provide a plausible means to perturb or stimulate the brain. Clinically used neuromodulation methods include invasive deep brain stimulation (DBS), non-invasive transcranial magnetic stimulation (TMS), non-invasive transcranial direct/alternating current stimulation (tDCS/tACS), and transcranial focused ultrasound stimulation (tFUS). A review of advances in neuromodulation technologies for treating mental disorders can be found in the literature. To date, repetitive TMS (rTMS) has been cleared by the FDA for the treatment of depression and has recently been used in studies of neural functioning and behavior. Along this research line, ML may potentially help addressing three important questions (where/when/how) to achieve precision neuromodulation in psychiatry.

For the where question, depending on the neuromodulation techniques, delivery of target-specific stimulation requires active and scheduled stimulation strategies to identify behaviorally activated targets. In the case of depression, neurostimulation can have multiple potential targets or modes of action, but how to identify the optimal target to achieve effective treatment outcomes remains unexplored. For the when and how questions, compared with open-loop stimulation, closed-loop stimulation can deliver temporally precise stimulation triggered by detected features, symptoms, or user demand. Traditional neurostimulation strategies are designed in an on/off stimulation fashion, where the stimulation is determined by pre-selected parameters. However, these stimulation parameters may not be optimal. To accommodate an adaptive subject-specific stimulation strategy, adaptive stimulation uses neurofeedback to adjust the stimulation parameters or control policy to achieve various optimality criteria. Therefore, ML can play a guiding role in online adaptive stimulation.

For instance, the feedback loop can analyze the neural signal’s oscillatory patterns or other reliably detectable biosignals (e.g., biochemical, electromyographic, and mechanical signals) to classify or detect the critical brain state for delivery of closed-loop neurostimulation. Additionally, reinforcement learning can be applied to learn a state-action value function to identify the best excitability brain state, where the state corresponds to the neural activity (e.g., the amplitude of evoked potentials, characteristics of brain connectivity) and the action corresponds to on/off stimulation mode.

Neurostimulation can not only induce changes in behavior but can also induce plasticity in brain connectivity. Simultaneous or post-neurostimulation neuroimaging provides a window of examining the change in brain network connectivity patterns. Brain connectivity and dynamics can be studied from a network communication and control perspective. The distinction between a healthy and a pathological brain can be characterized by their different efficiency to route information between distributed brain nodes, to control or modulate the target node under specific constraints, or to influence its behavior for performing specific tasks (“cognitive control”). Therefore, the well-established network and graph theories can be used to study the change in brain connectivity by a controller (neurostimulator). Specifically, the control-theoretic models have also been applied to quantify the response of brain networks to exogenous and endogenous perturbations. Several important research questions can be studied along this line: (1) can a target node stimulation rewire brain connectivity in evoked and steady-state conditions? (2) Can the neurostimulation-induced change of evoked or resting-state brain connectivity distinguish a pathological from a healthy brain? (3) Given a controller’s energy constraint, what is the optimal neurostimulation policy? Will alternate or simultaneous neurostimulations at multiple sites influence the network connectivity more effectively or bring additional benefit in treatment? (4) Can the induced brain patterns or changes in network connectivity predict the treatment outcome? ML may address these questions by providing individualized treatment-response likelihood in precision psychiatry.

Finally, we suggest that combining multiple efforts of XAI and neurostimulation in the loop (“neuroimaging → circuit modeling → neurostimulation → observing behaviors → revising models”; Figure 10B) will provide an effective pathway toward a better understanding of brain-behavior causation as well as individualized precision treatment in psychiatry. For instance, looping neurostimulation and DCM may provide a way to test the impact of neurostimulation on neural plasticity that underlies the change in adaptive or maladaptive behaviors.

DISCUSSION AND CONCLUSION

Challenge and opportunities
The past few decades have witnessed growing interests and rapid developments in ML methods for precision psychiatry. However, caution was also raised regarding the unrealistic hope for ML applications in clinical practice, and the field is still facing both conceptual and practical challenges.

At the conceptual level, first, the term disorder was used to specifically avoid the term disease, implying that precise mechanistic understanding is still missing in psychiatry; this further makes it very difficult to build clinical inference models for mental disorders. As a result, it is still infeasible to develop treatments that target underlying physiological risk factors in a similar manner as other medical disciplines (e.g., treating hypertension in heart diseases). Furthermore, each mental disorder has overlapping symptoms with varying degrees, bringing difficulties to uniquely define the psychiatric disorder. Second, many disorders are presented as a spectrum (e.g., ASD, generalized anxiety spectrum, and schizophrenia spectrum) and vary across different patients, creating a wide range of subtypes and subject variability within the same type of mental disorder. Third, due to various genetic, biochemical, and neuropathological factors, the
same mental disorder may have different causes and symptoms in different age/gender/race populations. Fourth, overlapping symptoms can be found in many mental disorders, making the diagnosis less precise or more error prone. For example, changes in sleep and energy level, often found in depression and generally measured using the PHQ-9 questionnaire, are very common across many other disorders. One goal in precision psychiatry is to fully dissect the mechanisms and causally reveal the many-to-one relationship. This can be catalyzed by rigorous measurements and quantification of neural and behavioral data relevant to mental health.

At the practical level, many challenges also remain in effective applications of ML for mental health.

**Sample size**

Datasets used in many ML applications have a small sample size, especially by the standard of ML-based speech/image/video applications. Neuroimaging data collection from mental health patients is limited to one-shot examples, which creates large data variability in addition to the intrinsic heterogeneity and disease comorbidity. Recent developments in foundation models and their mental health applications may help overcome this challenge—for example, by sharing the pre-trained language model. However, further caution is needed to ensure appropriate validation methods on the problem-specific data. Reproducibility is the main bottleneck to biomarker discovery for any mental disorder.

**Data quality**

The lack of standardization in data acquisition and varying degrees of data quality present a challenge in rigor and reproducibility. For example, studies using social media data often rely on the mental disorder labels based on users’ self-identification instead of rigorous clinical diagnosis. This can lead to a post containing “I am depressed” being labeled as a depression sample regardless of the underlying clinical symptoms. Terms such as depressed or anxious have colloquial uses that can differ from clinical criteria, leading to inaccurately labeled samples. Furthermore, collection of properly matched control samples remains difficult given the diversity of age, gender, race, education, family history, and lifestyle among users. Furthermore, there are also intrinsic dataset biases during data collection.

**Data privacy and security**

Advances in sensing technologies enable us to collect a large amount of personal data, including location, face images, speech conversations, and social interactions. However, how to store and process these data without the leakage risk of privacy information remains an important challenge. While research studies have rules and regulations (e.g., internal review board) to ensure ethical use, social media data are collected on a massive scale by companies. Due to the lack of sufficient regulation, such data have not been treated as personally identifiable information (PII) that can be used to inform the user’s health, creating a major obstacle in securing identifiable user data. To help alleviate this issue, regulations in the United States such as the Health Insurance Portability and Accountability Act (HIPAA) can be used to govern PII acquired by all social media companies and commercial entities.

**Social implications and environmental factors**

Gender and race are critical factors in mental health. According to WHO, mental disorders have a long history of gender bias. In terms of the gender risk factor, females are more likely to suffer from depression and anxiety, whereas there is more prevalence of autism in males. In terms of gender treatment bias, women are more likely to be diagnosed with depression compared with men, and women are more likely to be prescribed with mood-altering psychotropic drugs. ML may play a role in uncovering the gender or race risk factor and minimize the diagnosis or treatment bias related to these social factors.

**Generalizability**

The standard ML generalization issue becomes even more pronounced in mental health applications, especially due to the poor data quality and small sample size. Most ML studies use cross-validation to report the performance but lack independent validation datasets to assess generalizability. Furthermore, very few studies test generalization across data sources and experimental conditions. For example, it is important to test how well ML models trained from speech data from clinical interviews will perform on non-clinical speech data.

**Algorithmic bias**

Digital mental health inherits a long history of bias in psychiatry, which can be found at all stages of a patient journey. In addition to biological underpinnings, the domains of data (such as language) also represent social underpinnings, and therefore it is important to consider how socioeconomic factors may influence measurements. Using training and validation sets that are representative across all demographics can not only help address some of these issues but also uncover new symptom expressions in various groups. This is even more important for ML approaches that inherit biases from other ML models.

**Interpretability**

The ability to understand which latent factors contribute most to the outcome is the key for advancing clinical understanding of mental disorders for mental health professionals as well as for establishing the trust for the users of mental health technology. This is also an important dimension to improve precision in mental health. The choice of the interpretation method, either model specific (such as analyzing attention weights of a transformer) or model agnostic (such as local interpretable model-agnostic explanations [LIME]), is very specific to the nature of the problem. While various interpretation methods can be used to identify model functioning, it is important to note that the interpretation results can only be trusted as long as the issues of generalizability and data quality are addressed. In other words, model interpretation methods may produce erratic results with insufficient or poor-quality data. Furthermore, there is always an “explainability-performance” trade-off. For instance, deep learning may outperform traditional methods at the cost of reduced interpretability. Despite the rapid progress in the development of XAI and interpretable ML techniques, the challenge of ML explainability still remains in psychiatric applications, especially when dealing with data of mixed modality, data of high dimensionality, and data measured at varying timescales.

**Causal inference**

Most ML applications in mental health have focused on integrating information from multiple data sources and reaching a diagnosis decision faster. However, diagnosing a mental disorder, even with a highly interpretable model, neither speaks to the underlying causes nor has limited implications on treating the causes. ML-based causal inference methods may help with precision treatment design. Recent developments in...
ultra-high-field neuroimaging with sufficient temporal and spatial resolution may provide a means for developing inference models for mental disorders.

**Clinical integration**

It is important to consider the clinical need from a user experience perspective, varying from the mental health professionals using the application to patients who are mentally ill. Part of this work, such as conducting user research in various demographics, lies outside of the ML domain; however, such cross-functional research can inform the best practice in ML model development. This type of thinking with the end goal in mind is important for successful translation of precision psychiatry research to widespread clinical practice. Additionally, considering a possible disparity of interests between the clinical and ML communities, cross-disciplinary dialogues and collaborations between two communities would help the deployment of ML solutions in clinical practice.

**Ethical considerations**

ML applications in mental health also raise important ethical considerations. For example, ML models for risk assessment can lead to early screening that may help with early treatment. However, when screening techniques are available outside clinical settings, it can create the risk of misinterpretation by patients, which may negatively affect treatment-seeking behavior or trigger self-harming thoughts in patients. Other ethical questions related to increasing the risk of self-harm arise inherently from ML that use foundation models like Generative Pre-trained Transformer 3 (GPT-3), which should be fully considered before deployment in clinical settings.

**Applications of new ML technologies**

In addition to the opportunities arising from addressing the above-mentioned challenges, precision psychiatry is accompanied by plenty of opportunities in future ML applications.

**Data-centric approach**

In the data-driven ML view (“ML system = model/algorithm + data”), data are powerful. However, medical data are costly to collect and noisy. Currently, there is an ML paradigm shift from model-centric to data-centric ([https://datacentricai.org/](https://datacentricai.org/)), which advocates using good “small” data instead of simply collecting from big, but possibly noisy, data. The good-quality criteria include (1) consistency, (2) coverage of important cases, and (3) inclusion of timely feedback from user or production data. Unlike the model-centric ML approach that focuses on modifying the model/algorithm (while fixing the data) to improve the performance, a data-centric ML approach involves building ML systems with quality data, with a goal to systematically process the data (while fixing the model) to improve the ML performance. The modification of the available data may include data regeneration, data augmentation, and label refinement strategies to improve data consistency. For instance, confident learning has been proposed to estimate label uncertainty and identify label errors, based on the principles of pruning noisy data, counting with probabilistic thresholds to estimate noise, and ranking examples to train with confidence. The iterative process of two approaches can bootstrap the system performance.

**Data augmentation approach**

To deal with the small sample size issue in patient data collection, one ML approach is to create synthetic data (as a data augmentation strategy) or increase the number of training instances. Deep-learning methods such as GAN and its variants have served as a powerful tool to generate synthetic brain scan images, speech, video, physiological data, and EHRs. However, unlike traditional ML/computer vision studies, the value of synthetic training samples remains unclear in psychiatric applications. Whether ML-augmented samples can generate clinically meaningful and diversified samples that match psychiatric heterogeneity would require future in-depth investigations.

**Automated-learning approach**

In contrast to the human-in-the-loop solutions, automated ML (autoML) and automated deep learning (autoDL) represent a new paradigm that aims to automate the data analysis pipeline while minimizing the need of human intervention during the course of modeling and training. This has become increasingly important since the volume of social media and multi-media data streams is so overwhelming that even a small effort of human involvement would make the task prohibitive.

**Data-integration approach**

Integration of multi-modal data is critical for psychiatric diagnostics and monitoring. Therefore, it is urgently needed to develop weakly supervised, interpretable, multi-modal deep-learning pipelines to fuse histopathology, genomics, neuroimaging, and behavioral data, as well as to develop multi-modal fusion algorithms for speech, video, and EHRs, to assist both psychiatrists and patients. Because of the nature of multi-modality, not all data can be quantified in the Euclidean space. Graph and geometric deep learning may play a role in this research direction. Finally, expert-augmented ML (EAML) methods that automatically extract problem-specific human expert knowledge and integrate it with ML to build robust, dependable, and data-efficient predictive models will also have great potentials for psychiatric applications.

**Conclusion**

To date, there is still a lack of biomarkers and individualized treatment guidelines for mental illnesses. In this review, we have shown that ML technologies and data analytics can be used for various stages of a patient journey: detection/diagnosis, prognosis, treatment selection/optimization, outcome monitoring/tracking, and relapse prevention. We predict that the multi-modal integration of neuroimaging, ML, genetics, behavioral neuroscience, and mobile health will open doors for new method developments and technology inventions. First, making brain scans more accessible will be the key to clinical applications of neuroimaging techniques. Using real-time fMRI, ML can guide neurofeedback-based intervention and provide closed-loop treatment or rehabilitation. As a “real-time mirror” of psychiatry, mind-control intervention can improve behavioral outcomes. Second, data-driven ML methods can identify sub-types of symptoms and cognitive deficits and develop model-based phenotyping. Third, combination of ML methods with large EHR databases may accommodate a personalized psychiatry. Fourth, when developing ML-powered technologies for psychiatry, it is imperative to consider concerns and feedback from various stakeholders, including knowledgeable experts (clinical and ML experts, technology or engineer experts), decision-makers (hospital administrators, institutional leaders, state and federal governments), and end users (physicians, nurses,
patients, friends, and family. Finally, an integration of medications, wearable devices, mobile health apps, social support, and online education will be essential to improve mental health and assist therapeutic outcomes in the new era of digital psychiatry. Future precision psychiatry will leverage ML and all technologies to provide individualized custom packages that are built upon the patient’s needs and specific pathology.

**BIOGRAPHY**

Zhe Sage Chen is an associate professor at the New York University School of Medicine, with joint appointment at the Departments of Psychiatry, Neuroscience, and Physiology and the Department of Biomedical Engineering at the NYU Tandon School of Engineering. Prof. Chen is the codirector of Computational Psychiatry Program at NYU and the director of the Computational Neuroscience, Neuroengineering, and Neuropsychiatry Laboratory. His research interests include computational neuroscience, computational psychiatry, neural engineering, brain-machine interface, and machine learning, with more than 120 peer-reviewed articles in these areas. More details about his lab can be found on his website: http://cn3laboratory.org/.

Prathamesh (Param) Kulkarni received a PhD degree in electrical engineering from the University of Houston. He is currently leading a machine-learning team at Headspace Health, Inc. Dr. Kulkarni is an expert in the digital health space with a specialization in AI and behavioral health. He has successfully built digital health companies from the ground up. Previously, as the cofounder of an innovative behavioral health company AwareHealth, he led the company from inception to product-market fit leading product, sales, and tech.

Isaac R. Galatzer-Levy received a PhD degree in Clinical Psychology from Columbia University. He completed post-doctoral training in neurobiology of stress and bioinformatics at NYU School of Medicine. He served on the research faculty at NYU School of Medicine Department of Psychiatry from 2015 to 2019. Subsequently, Dr. Galatzer-Levy led clinical and algorithm development across multiple successful startups in the AI-driven digital health space. He is currently working at the Meta Reality Lab. Dr. Galatzer-Levy has published more than 100 articles in the fields of neuroscience, psychiatry, psychology, medicine, and bioinformatics.

Benedetta Bigio is a Computational Engineer and Research Assistant Professor at NYU Grossman School of Medicine. Her work utilizes integrative computational and bioinformatic approaches applied to rodent and human models of main diseases, including mood and cognitive disorders. Using genome-wide assays and bioinformatic and machine-learning algorithms, her work identified novel molecular networks of mitochondrial metabolism in resilience versus susceptibility to stress.

Carla Nasca is an assistant professor at NYU Grossman School of Medicine. Her research focuses on developing an innovative model of epigenetic mechanisms of neuroplasticity to stress with a new angle on the communication between mitochondria and cell nuclei. Her laboratory also uses computational approaches together with emerging technologies to isolate patient-derived exosomes to identify in vivo neural mechanisms of main CNS diseases (mood and cognitive disorders) and potential biomarkers of responses to novel anti-depressant treatment and transmagnetic brain stimulation.

Yu Zhang is currently an assistant professor of bioengineering at Lehigh University. He was a post-doctoral research fellow at the Wu Tsai Neurosciences Institute and the Department of Psychiatry and Behavioral Sciences at Stanford University Medical School and in the Biomedical Research Imaging Center at the University of North Carolina, Chapel Hill. He also worked as an assistant professor at the School of Information Science and Engineering at East China University of Science and Technology, China. His research interests include computational neuroscience, pattern recognition, machine learning, signal processing, brain-computer interface, and medical imaging computing.

**ACKNOWLEDGMENTS**

The research was partially supported from the US National Science Foundation (CBET-1835000 to Z.S.C.), the National Institutes of Health (R01NS121776 and R01-MH118928 to Z.S.C.), and an Alzheimer’s Association Grant (AARG-22-972541 to Y.Z.). We thank W. Wu for valuable comments and R. MacKay for help in English proofreading.

**DECLARATION OF INTERESTS**

The authors declare no competing financial interests.

**INCLUSION AND DIVERSITY**

We support inclusive, diverse, and equitable conduct of research. One or more of the authors of this paper self-identifies as a gender minority in their field of research. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

**REFERENCES**

1. Czeisler, M.E., Lane, R.I., Petrosky, E., Wiley, J.F., Christensen, A., Njai, R., Weaver, M.D., Robbins, R., Facer-Childs, E.R., Barger, L.K., et al. (2020). Mental health, substance use, and suicidal ideation during the COVID-19 pandemic United States, June 24–30, 2020. MMWR Morb. Mortal. Wkly. Rep. 69, 1049–1057.

2. Greenberg, P.E., Fournier, A.A., Sisitsky, T., Simes, M., Berman, R., Koenigsberg, S.H., and Kessler, R.C. (2021). The economic burden of adults with major depressive disorder in the United States (2010 and 2016). PharmacoEconomics 39, 653–665.

3. The Lancet. 20 years of precision medicine in oncology The Lancet 397:1781, https://doi.org/10.1016/S0140-6736(21)01099-0.

4. Insel, T.R., and Cuthbert, B.N. (2015). Brain disorders? precisely. Science 348, 499–500.

5. Fernandes, B.S., Williams, L.M., Steiner, J., Leboyer, M., Carvalho, A.F., and Berk, M. (2017). The new field of ‘precision psychiatry. BMC Med. 15, 80–87.

6. Insel, T., Cuthbert, B., Garvey, M., Heinsen, R., Pine, D.S., Quinn, K., Samslov, C., and Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am. J. Psychiatr. 167, 748–751.

7. Insel, T.R. (2014). The nimh research domain criteria (RDoC) project: precision medicine for psychiatry. Am. J. Psychiatr. 171, 395–397.

8. Bzdok, D., and Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: opportunities and challenges. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 3, 223–230.

9. Zhou, Z., Wu, T.C., Wang, B., Wang, H., Tu, X.M., and Feng, C. (2020). Machine learning methods in psychiatry: a brief introduction. Gen. Psychiatr. 33, e100171.
10. Allen, M., and Salmon, A. (2020). Synthesizing artificial patient-level data for open science—an evaluation of five methods. Preprint at medRxiv. https://doi.org/10.1101/2020.10.09.20210138.

11. Burr, C., Morley, J., Taddeo, M., and Floridi, L. (2020). Digital psychiatry: risks and opportunities for public health and wellbeing. IEEE Trans. Technol. Soc. 1, 21–33.

12. Doraiswamy, P.M., Blease, C., and Bodner, K. (2020). Artificial intelligence and the future of psychiatry: insights from a global physician survey. Artif. Intell. Intell. Med. 102, 101753.

13. Thukral, A., Ershad, F., Enan, N., Rao, Z., and Yu, C. (2018). Soft ultrathin silicon electronics for soft neural interfaces: a review of recent advances of soft neural interfaces based on ultrathin silicon. IEEE Nanotechnol. Mag. 12, 21–34.

14. Guggenmos, M., Schmack, K., Veer, I.M., Lett, T., Sekutowicz, M., Sebold, M., Garbusow, M., Sommer, C., Wittchen, H.U., Zimmermann, U.S., et al. (2020). A multimodal neuroimaging classifier for alcohol dependence. Sci. Rep. 10, 2986.

15. Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetchko, R.N., Zebley, B., Oathes, D.J., Etkin, A., et al. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat. Med. 23, 28–38.

16. Beaulieu-Jones, B.K., and Greene, C.S.; Pooled Resource Open-Access ALS Clinical Trials Consortium (2016). Semi-supervised learning of the electron health record for phenotype stratification. J. Biomed. Inf. 64, 168–178.

17. Koppe, G., Meyer-Lindenberg, A., and Durstewitz, D. (2021). Deep learning for small and big data in psychiatry. Neuropsychopharmacology 46, 176–190.

18. Rutledge, R.B., Chekroun, A.M., and Huys, Q.J. (2019). Machine learning and big data in psychiatry: toward clinical applications. Curr. Opin. Neurol. 55, 152–159.

19. Chandler, C., Foltz, P.W., and Elvevåg, B. (2020). Using machine learning in psychiatry: the need to establish a framework that nurtures trustworthiness. Schizophr. Bull. 46, 11–14.

20. Galatzer-Levy, I.R., Ruggles, K.V., and Chen, Z. (2018). Data science in the research domain criteria era: relevance of machine learning to the study of stress pathology, recovery, and resilience. Chronic Stress 2, 1–14.

21. Shatte, A.B.R., Hutchinson, D.M., and Teague, S.J. (2019). Machine learning in mental health: a scooping review of methods and applications. Psychol. Med. 49, 1426–1448.

22. Su, C., Xu, Z., Pathak, J., and Wang, F. (2020). Deep learning in mental health outcome research: a scooping review. Transl. Psychiatry. 10, 116–126.

23. Liu, G.D., Li, Y.C., Zhang, W., and Zhang, L. (2020). A brief review of artificial intelligence applications and algorithms for psychiatric disorders. Engineering 6, 462–467.

24. Thieme, A., Belgraves, D., and Doherty, G. (2020). Machine learning in mental health: a systematic review of the hci literature to support the development of effective and implementable ml systems. ACM Trans. Comput. Hum. Interact. 27, 1–53.

25. Durstewitz, D., Kope, G., and Meyer-Lindenberg, A. (2019). Deep neural networks in psychiatry. Mol. Psychiatr. 24, 1583–1598.

26. Hedderich, D.M., and Eckhoff, S.B. (2021). Machine learning for psychiatry: getting doctors at the black box? Mol. Psychiatr. 26, 23–25.

27. Bracher-Smith, M., Crawford, K., and Escott-Price, V. (2021). Machine learning for genetic prediction of psychiatric disorders: a systematic review. Mol. Psychiatr. 26, 70–79.

28. Allen, S. (2020). Artificial intelligence and the future of psychiatry. IEEE Pulse 11, 2–5.

29. Etkin, A. (2019). A reckoning and research agenda for neuroimaging in psychiatry. Am. J. Psychiatr. 176, 507–511.
49. Sargent, K., Chavez-Baldini, U., Master, S.L., Verweij, K.J.H., Lok, A., Sutherland, A.L., Vulin, N.C., Denys, D., Smit, D.J.A., and Nieman, D.H. (2021). Resting-state brain oscillations predict cognitive function in psychiatric disorders: a transdiagnostic machine learning approach. Neuroimage. Clin. 30, 102617.

50. Barch, D.M. (2017). The neural correlates of transdiagnostic dimensions of psychopathology. Am. J. Psychiatr. 174, 613–615.

51. Xia, C.H., Ma, Z., Cincir, R., Gu, S., Betzel, R.F., Kaczekurkin, A.N., Calkins, M.E., Cook, P.A., García de la Garza, A., Vandekar, S.N., et al. (2018). Linked dimensions of psychopathology and connectivity in functional brain networks. Nat. Commun. 9, 1–14.

52. Kebets, V., Holmes, A.J., Orban, C., Tang, S., Li, J., Sun, N., Kong, R., Poldrack, R.A., and Yeo, B.T.T. (2019). Somatosensory-motor disconnection spans multiple transdiagnostic dimensions of psychopathology. Biol. Psychiatr. 86, 779–791.

53. McTeague, L.M., Rosenberg, B.M., Lopez, J.W., Carreon, D.M., Huemer, J., Jiang, Y., Chick, C.F., Eickhoff, S.B., and Etkin, A. (2020). Identification of common neural circuit disruptions in emotional processing across psychiatric disorders. Am. J. Psychiatr. 177, 411–421.

54. Wachinger, C., Nho, K., Saykin, A.J., Reuter, M., and Rieckmann, A.; Alzheimer’s Disease Neuroimaging Initiative (2018). A longitudinal imaging genetics study of neuroanatomical asymmetry in Alzheimer’s disease. Biol. Psychiatr. 84, 522–530.

55. Vidal-Ribas, P., Benson, B., Vitale, A.D., Keren, H., Harrewijn, A., Fox, N.A., Pine, D.S., and Stringaris, A. (2019). Bidirectional associations between stress and reward processing in children and adolescents: a longitudinal neuroimaging study. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 4, 893–901.

56. Roeckner, A.R., Oliver, K.I., Lebois, L.A.M., van Rooij, S.J.H., and Stevens, J.S. (2021). Neural contributors to trauma resilience: a review of longitudinal neuroimaging studies. Transl. Psychiatr. 11, 1–17.

57. Smith, S.M., and Nichols, T.E. (2018). Statistical challenges in “big data” human neuroimaging. Neuroim 97, 263–268.

58. Duncan, N.W., and Northoff, G. (2013). Overview of potential procedural and participant-related confounds for neuroimaging of the resting state. J. Psychiatr. Neurosci. 38, 84–96.

59. Blume, W.T. (2006). Drug effects on eeg. J. Clin. Neurophysiol. 23, 306–311.

60. Linke, A.G., Olson, L., Gao, Y., Fishman, I., and Müller, R.A. (2017). Psychotropic medication use in autism spectrum disorders may affect functional brain connectivity. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 2, 518–527.

61. Yu, M., Lin, K.A., Cook, P.A., Phillips, M.L., McInnis, M., Fava, M., Trivedi, M.H., Weissman, M.M., Shihorna, R.T., and Sheline, Y.I. (2018). Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. Hum. Brain Mapp. 39, 4213–4227.

62. Zhao, Q., Aided, E., and Pohl, K.M. (2020). Training confounder-free deep learning models for medical applications. Nat. Commun. 17, 1–9.

63. Nielsen, A.N., Barch, D.M., Petersen, S.E., Schlagger, B.L., and Greene, D.J. (2020). Machine learning with neuroimaging: evaluating its applications in psychiatry. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 5, 791–798.

64. Pelin, H., Ising, M., Stein, F., Meintz, S., Meller, T., Brosch, K., Winter, N.R., Krug, A., Leenings, R., Lemke, H., et al. (2021). Identification of transdiagnostic psychiatric disorder subtypes using unsupervised learning. Neuropsychopharmacology 46, 1985–1905.

65. Janssen, R.J., Mourão-Miranda, J., and Schnack, H.G. (2018). Making individual prognoses in psychiatry using neuroimaging and machine learning. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 3, 798–808.

66. Gao, S., Calhoun, V.D., and Sui, J. (2018). Machine learning in major depression: from classification to treatment outcome prediction. CNS Neurosci. Ther. 24, 1037–1052.

67. Cearns, M., Opel, N., Clark, S., Kaehler, C., Thalamanuthu, A., Heindel, W., Winter, T., Teissmann, H., Minnerup, H., Dannowski, U., et al. (2019). Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. Transl. Psychiatr. 9, 1–9.

68. Gueorguieva, R., Chekroud, A.M., and Krystal, J.H. (2017). Trajectories of relapse in randomised, placebo-controlled trials of treatment discontinuation in major depressive disorder: an individual patient-level data meta-analysis. Lancet Psychiatr. 4, 230–237.

69. Murphy, K.P. (2012). Machine Learning: A Probabilistic Perspective (MIT Press).

70. Emmert-Streib, F., and Dehmer, M. (2022). Taxonomy of machine learning paradigms: a data-centric perspective. WIREs Data Min. Knowl. Discov. 2022, e1470.

71. Zhao, K., Duka, B., Xie, H., Oathes, D.J., Calhoun, V., and Zhang, Y. (2022). A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD. Neuroimage 246, 118774.

72. Luo, Y., Alvarez, T.L., Halperin, J.M., and Li, X. (2020a). Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. Neuroimage. Clin. 26, 102238.

73. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Ribas, L.M., Ventola, P., and Duncan, J.S. (2021). Braingraph: interpretable brain graph neural network for fMRI analysis. Med. Image Anal. 74, 102333.

74. Khosla, M., Jamison, K., Kuceyeski, A., and Sabuncu, M.R. (2019). Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction. Neuroimage 199, 651–662.

75. Wu, M.J., Mwangi, B., Bauer, I.E., Passos, I.C., Sanches, M., Zunta-Soares, G.B., Meyer, T.D., Hasan, K.M., and Soares, J.C. (2017). Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. Neuroimage 145, 254–264.

76. Zhu, H., Yuan, M., Qiu, C., Ren, Z., Li, Y., Wang, J., Huang, X., Liu, S., Gong, Q., Zhang, W., and Zhang, Y. (2020). Multivariate classification of earthquake survivors with post-traumatic stress disorder based on large-scale brain networks. Acta Psychiatr. Scand. 147, 285–298.

77. Yan, W., Calhoun, V., Song, M., Cui, Y., Yan, H., Liu, S., Fan, L., Zuo, N., Yang, Z., Xu, K., et al. (2019). Discriminating schizophrenia using recurrent neural network applied on times of multi-site fMRI data. EBioMedicine 47, 543–552.

78. Mikolas, P., Hlinka, J., Skoch, A., Pitra, Z., Frodl, T., Spaniel, F., and Hasin, J. (2018). Disease prediction using graph convolutional networks. Transl. Psychiatr. 8, 1–17.

79. Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., and Rueckert, D. (2018). Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer’s disease. Med. Image Anal. 48, 117–130.

80. Maron-Katz, A., Zhang, Y., Narayan, M., Wu, W., Toll, R.T., Naparstek, S., De Los Angeles, C., Longwell, P., Shipgel, E., Newman, J., et al. (2020). Individual patterns of abnormality in resting-state functional connectivity reveal two data-driven PTSD subgroups. Am. J. Psychiatr. 177, 244–253.

81. Lecei, A., van Hulst, B.M., de Zeeuw, P., van der Pluijm, M., Rijks, Y., and Durston, S. (2019). Can we use neuroimaging data to differentiate between subgroups of children with ADHD symptoms: a proof of concept study using latent class analysis of brain activity. Neuroimage. Clin. 21, 101601.

82. Maron-Katz, A., Zhang, Y., Narayan, M., Wu, W., Toll, R.T., Naparstek, S., De Los Angeles, C., Longwell, P., Shipgel, E., Newman, J., et al. (2020). Individual patterns of abnormality in resting-state functional connectivity reveal two data-driven PTSD subgroups. Am. J. Psychiatr. 177, 244–253.

83. Lecei, A., van Hulst, B.M., de Zeeuw, P., van der Pluijm, M., Rijks, Y., and Durston, S. (2019). Can we use neuroimaging data to differentiate between subgroups of children with ADHD symptoms: a proof of concept study using latent class analysis of brain activity. Neuroimage. Clin. 21, 101601.

84. Parkes, L., Moore, T.M., Calkins, M.E., Cook, P.A., Cieslak, M., Roaf, D.R., Wolf, D.H., Gur, R.C., Gur, R.E., Satterthwaite, T.D., and Bassett, D.S. (2021). Transdiagnostic dimensions of psychopathology explain...
individuals’ unique deviations from normative neurodevelopment in brain structure. Transl. Psychiatr. 11, 1–13.

85. Schmaal, L., Marquand, A.F., Rhebergen, D., van Tol, M.J., Ruhé, H.G., van der Wee, N.J.A., Veitman, D.J., and Penninx, B.W.J.H. (2019). Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: a multivariate pattern recognition study. Biol. Psychiatry. 78, 278–286.

86. Ramyead, A., Studerus, E., Komter, M., Uttinger, M., Gschwandtner, U., Fuhr, P., and Riecher-Rössler, A. (2016). Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naive at-risk patients. World J. Biol. Psychiatr. 17, 285–295.

87. Koutsouleris, N., Dwyer, D.B., Degenhardt, F., Maj, C., Urquijo-Castro, S., Wolf, L., Ben-Zion, Z., Sheynin, J., Reznik, S., Keynan, J.N., Chang, J.C., Lin, H.Y., Lv, J., Tseng, W.Y.I., and Gau, S.S.F. (2021). Zhang, J., Wong, S.M., Richardson, J.D., Jetly, R., and Dunkley, B.T. (2020). Use of machine learning for predicting escitalopram treatment outcome in young children with autism. Transl. Psychiatr. 6, e948.

88. Koutsouleris, N., Dwyer, D.B., Degenhardt, F., Maj, C., Urquijo-Castro, S., Wolf, L., Ben-Zion, Z., Sheynin, J., Reznik, S., Keynan, J.N., Chang, J.C., Lin, H.Y., Lv, J., Tseng, W.Y.I., and Gau, S.S.F. (2021). Zhang, J., Wong, S.M., Richardson, J.D., Jetly, R., and Dunkley, B.T. (2020). Use of machine learning for predicting escitalopram treatment outcome in young children with autism. Transl. Psychiatr. 6, e948.
canonical correlation analysis for brain imaging genetics. Med. Image Anal. 76, 102297.

119. Krak, S., Park, S., Kim, J., Park, S., and Lee, J.-Y. (2020). Multivariate neuroanatomical correlates of behavioral and psychological symptoms in dementia and the moderating role of education. Neuroimage. Clin. 28, 102452.

120. Kalmady, S.V., Greiner, R., Agrawal, R., Shivakumar, V., Narayanasamy, J.C., Brown, M.R.G., Greenshaw, A.J., Dursun, S.M., and Venkat-subramanian, G. (2019). Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. NPJ Schizophr. 5, 2–11.

121. Wang, J., Zhang, L., Wang, Q., Chen, L., Shi, J., Chen, X., Li, Z., and Shen, D. (2020). Multi-class and classification based on functional connectivity and functional correlation tensor via multi-source domain adaptation and multi-view sparse representation. IEEE Trans. Med. Imag. 39, 3137–3147.

122. Elmer, J., Jones, B.L., and Nagin, D.S. (2018). Using the beta distribution in group-based trajectory models. BMC Med. Res. Methodol. 18, 1–5.

123. van der Nest, G., Lima Passos, V., Candel, M.J., and van Breukelen, G.J. (2021). Towards a neurodevelopmental framework for understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 6, 102297.

124. Quaaq, M., van de Mortel, L., Thomas, R.M., and van Wingen, G. (2021). Deep learning applications for the classification of psychiatric disorders using neuroimaging data: systematic review and meta-analysis. Neuroimage. Clin. 30, 102594.

125. Chang, M., Wolfer, F.Y., Gong, X., Chen, X., Tang, L., Feng, R., Dong, S., Duan, J., Chen, Y., Zhang, R., et al. (2021). Identifying and validating sub-types within major psychiatric disorders based on frontal–posterior functional imbalance via deep learning. Mol. Psychiat. 26, 2991–3002.

126. Almuqhim, F., and Saeed, F. (2021). A sparse autoencoder and deep-neural network model for detecting autism spectrum disorder (ASD) using fMRI data. Front. Comput. Neurosci. 15, 654315.

127. Aglinias, K., Hartshorne, J.K., and Anzellotti, S. (2022). Contrastive machine learning reveals the structure of neuroanatomical variation within autism. Science 376, 1070–1074.

128. Anwar, S.M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., and Khan, M.K. (2018). Medical image analysis using convolutional neural networks: a review. J. Med. Syst. 42, 1–13.

129. Zhang, J., Li, X., Li, Y., Wang, M., Huang, B., Yao, S., and Shen, L. (2020). Three dimensional convolutional neural network-based classification of conduct disorder with structural mri. Brain Imaging Behav. 14, 2333–2340.

130. Varol, E., Sotiras, A., and Davatzikos, C. (2017). Alzheimer’s Disease Neuroimaging Initiative (2017). Hydro: revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. Neuroimage 145, 346–364.

131. Varol, E., Sotiras, A., and Davatzikos, C. (2019). Neuroanatomical heterogeneity of schizophrenia revealed by semi-supervised machine learning methods. Schizophr. Res. 214, 43–50.

132. Varol, E., Sotiras, A., and Davatzikos, C. (2019). Neuroanatomical heterogeneity of schizophrenia revealed by semi-supervised machine learning methods. Schizophr. Res. 214, 43–50.

133. Varol, E., Sotiras, A., Baller, E.B., Barzilay, R., Calkins, M.E., Chand, G.B., Cui, Z., Erus, G., Fan, Y., Gur, R.E., et al. (2020). Neurostructural heterogeneity in youths with internalizing symptoms. Biol. Psychiatry. 87, 473–482.

134. Varol, E., Sotiras, A., Baller, E.B., Barzilay, R., Calkins, M.E., Chand, G.B., Cui, Z., Erus, G., Fan, Y., Gur, R.E., et al. (2020). Neurostructural heterogeneity in youths with internalizing symptoms. Biol. Psychiatry. 87, 473–482.

135. Yazdavar, A.H., Al-Omait, H.S., Ebrahim, M., Bajaj, G., Banerjee, T., Thrunarayan, K., Pathak, J., and Sheht, A. (2017). Semi-supervised approach to monitoring clinical depressive symptoms in social media. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp. 1191–1198.

136. Dong, G., Tang, M., Cai, L., Barnes, L.E., and Boukhecheba, M. (2021). Semi-supervised graph instance transformer for mental health inference. In Proc. 20th IEEE International Conference on Machine Learning and Applications (ICMLA) (IEEE), pp. 1221–1228.

137. Rutherford, S., Kia, S.M., Wolfers, T., Fraza, C., Zabihi, M., Dinga, R., Berthet, P., Worker, A., Verdi, S., Ruhe, H.G., et al. (2022). The normative modeling framework for computational psychiatry. Nat. Protoc. 17, 1711–1734.

138. Marquand, A.F., Rezek, I., Buitelaar, J., and Beckmann, C.F. (2016). Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. Biol. Psychiatr. 80, 552–561.

139. Chamberland, M., Genc, S., Tax, C.M.W., Shastin, D., Koller, K., Raven, E.P., Cunningham, A., Doherty, J., van den Bree, M.B.M., Parker, G.D., et al. (2021). Detecting microstructural deviations in individuals with deep diffusion MRI tractometry. Nat. Comput. Sci. 1, 598–606.

140. Shen, D., Wu, G., and Suk, H.J. (2017). Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19, 221–248.

141. Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., and Calhoun, V. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. Nat. Commun. 12, 1–17.

142. Quaak, M., van de Mortel, L., Thomas, R.M., and van Wingen, G. (2021). Deep learning applications for the classification of psychiatric disorders using neuroimaging data: systematic review and meta-analysis. Neuroimage. Clin. 30, 102594.

143. Anwar, S.M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., and Khan, M.K. (2018). Medical image analysis using convolutional neural networks: a review. J. Med. Syst. 42, 1–13.

144. Amla, M., and Saeed, F. (2019). Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: a large-scale multi-sample study. Hum. Brain Mapp. 40, 944–954.

145. Aglinias, K., Hartshorne, J.K., and Anzellotti, S. (2022). Contrastive machine learning reveals the structure of neuroanatomical variation within autism. Science 376, 1070–1074.

146. Chand, G.B., Cui, Z., Erus, G., Fan, Y., Gur, R.E., et al. (2020). Neuroanatomical heterogeneity of schizophrenia as disorders of network dynamics. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 6, 102297.
Dvornek, N.C., Li, X., Zhuang, J., and Duncan, J.S. (2019). Jointly discriminative and generative recurrent neural networks for learning from fMRI. In International Workshop on Machine Learning in Medical Imaging (Springer), pp. 382–390.

Zhao, J., Huang, J., Zhi, D., Yan, W., Ma, X., Yang, X., Li, X., Ke, Q., Jiang, T., Calhoun, V.D., and Sui, J. (2020). Functional network connectivity (FNC)-based generative adversarial network (GAN) and its applications in classification of mental disorders. J. Neurosci. Methods 347, 108756.

Cutillo, C.M., Sharma, K.R., Foschini, L., Kundu, S., Mackintosh, M., and Mandli, K.D.; Multi Healthcare Workshop Working Group (2020). Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit. Med. 3, 1–5.

Wen, Z., Marin, M.F., Blackford, J.J., Chen, Z.S., and Milad, M.R. (2021). Fear-induced brain activations distinguish anxious and trauma-exposed brains. Transl. Psychiatr. 11, 1–10.

Sani, O.G., Yang, Y., Lee, M.B., Dawes, H.E., Chang, E.F., and Shanechi, M.M. (2018). Mood variations decoded from multi-site intracranial human brain activity. Nat. Biotechnol. 36, 954–961.

Besteussen, O., Gilbert, G., Harris, A., and Rost, J. (2021). Telehealth: a quarter-trillion-dollar post-covid-19 reality?. https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/telehealth-a-quarter-trillion-dollar-post-covid-19-reality.

Lee, E.E., Torous, J., De Choudhury, M., Depp, C.A., Graham, S.A., Kim, H.C., Paulus, M.P., Krystal, J.H., and Jeste, D.V. (2021). Artifical intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 6, 856–864.

Moustafa, A.A. (2021). Big Data in Psychiatry and Neurology (Academic Press).

Graham, S., Depp, C., Lee, E.E., Nebecker, C., Tu, X., Kim, H.C., and Jeste, D.V. (2019). Artificial intelligence for mental health and mental illnesses: an overview. Curr. Psychiatr. Res. 21, 11–18.

Abbas, A., Schultebraucks, K., and Galatzew-Lewr, I.R. (2021). Digital democratization of mental health: challenges, promises, and future directions. Psychiatr. Ann. 51, 14–20.

Vázquez-Romero, A. and Gallardo-Antolin, A. (2020). Automatic detection of depression in speech using ensemble convolutional neural networks. Entropy 22, 688.

Harati, A., Shiri-Lberg, E., Rutowski, T., Chiebek, P., Lu, Y., and Oliveira, R. (2021). Speech-based depression prediction using encoder-weight-only transfer learning and a large corpus. In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'21) (IEEE), pp. 7273–7277.

Huang, Z., Epps, J., and Joachim, D. (2019). Investigation of Speech Landmark Patterns for Depression Detection (IEEE Transactions on Affective Computing).

Zhu, Y., Shang, Y., Shao, Z., and Guo, G. (2018). Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. IEEE Trans. Affect. Comput. 9, 578–584.

Shao, W., You, Z., Liang, L., Hu, X., Li, C., Wang, W., and Hu, B. (2021). A multi-modal gait analysis-based depression detection system. IEEE J. Biomed. Health Inform. 1.

Lu, Y., Harati, A., Rutowski, T., Oliveira, R., Chiebek, P., and Shiri-Lberg, E. (2020). Robust speech and natural language processing models for depression screening. In Proc. IEEE Signal Processing in Medicine and Biology Symposium (SPMB) (IEEE), pp. 1–5.

Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Unger, L.H., Crutchley, P., Preolit-Pietro, D., Asch, D.A., and Schwartz, H.A. (2018). Facebook language predicts depression in medical records. Proc. Natl. Acad. Sci. USA 115, 11203–11208.

Sun, H., Liu, J., Chai, S., Oiu, Z., Lin, L., Huang, X., and Chen, Y. (2021). Multi-modal adaptive fusion transformer network for the estimation of depression level. Sensors 21, 4764.

Weiner, L., Guidi, A., Doignon-Camus, N., Giersch, A., Bertschy, G., and Vanello, N. (2021). Vocal features obtained through automated methods in verbal fluency tasks can aid the identification of mixed episodes in bipolar disorder. Transl. Psychiatr. 11, 1–9.

Palmius, N., Tsanas, A., Saunders, K.E.A., Bilderbeck, A.C., Geddes, J.R., Goodwin, G.M., and De Vos, M. (2017). Detecting bipolar depression from geographic location data. IEEE Trans. Biomed. Eng. 64, 1761–1771.

Mamar, C.R., Brown, A.D., Qian, M., Laska, E., Siegel, C., Li, M., Abu-Amara, D., Tsiartas, A., Richey, C., Smith, J., et al. (2019). Speech-based markers for posttraumatic stress disorder in us veterans. Depress. Anxiety 36, 607–616.

Mallol-Ragolta, A., Dharmia, S., and Boutt, T.E. (2018). A multimodal approach for predicting changes in ptsd symptom severity. In Proceedings of the 20th ACM International Conference on Multimodal Interaction, pp. 324–333.

Tahir, Y., Yang, Z., Chakraborty, D., Thalmann, N., Thalmann, D., Maniam, Y., Binte Abdul Rashid, N.A., Tan, B.L., Lee Chee Keong, J., and Dauwels, J. (2019). Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. PLoS One 14, e0214314.

Abbas, A., Yadav, V., Smith, E., Ramjas, E., Rutter, S.B., Benavidez, C., Koemshahargy, V., Zhang, L., Guan, L., Rosenfield, P., et al. (2021). Computer vision-based assessment of motor functioning in schizophrenia: use of smartphone apps for remote measurement of schizophrenia symptomatology. Digit. Biomark. 5, 29–36.

Birnbaum, M.L., Kulkarni, P.P., Van Meter, A., Chen, V., Rizvi, A.F., Are- nare, E., De Choudhury, M., and Kane, J.M. (2020). Utilizing machine learning on internet search activity to support the diagnostic process and relapse detection in young individuals with early psychosis: feasibility study. JMIR Ment. Health 7, e19348.

Birnbaum, M.L., Abrami, A., Heisng, S., Ali, A., Arenare, E., Agurto, C., Lu, N., Kane, J.M., and Cecchi, G. (2022). Acoustic and facial features from clinical interviews for machine learning-based psychiatric diagnosis: an algorithm development. JMIR Ment. Health 9, e26499.

Bickman, L. (2020). Improving mental health services: a 50-year journey from randomized experiments to artificial intelligence and precision mental health. Adm. Pol. Ment. Health 47, 795–843.

Wilkinson, J., Arnold, K.F., Murray, E.J., van Smeden, M., Carr, K., Sippy, R., de Kamps, M., Beam, A., Konigorski, S., Lippert, C., et al. (2020). Time to reality check the promises of machine learning-powered precision medicine. Lancet. Digit. Health 2, e67–e680.

Barron, D. (2021). Reading Our Minds: The Rise of Big Data Psychiatry (Columbia Global Reports).

Abdullah, S., and Choudhury, T. (2018). Sensing technologies for monitoring serious mental illnesses. IEEE Multimedia 25, 61–75.

Kraepelin, E. (1921). Manic depressive insanity and paranoia. J. Nerv. Ment. Dis. 53, 350.

Low, D.M., Bentley, K.H., and Ghosh, S.S. (2020). Automated assessment of psychiatric disorders using speech: a systematic review. Laryngoscope Investig. Otolaryngol. 5, 96–116.

Eyben, F., Scherer, K.R., Schuller, B.W., Sundberg, J., Andre´, E., Busso, N., Quatieri, T.F. (2015). A review of depression and suicide risk assessment methods of the 20th ACM International Conference on Multimodal Interaction, pp. 36–43.

Beyen, F., Scherer, K.R., Schluteller, B.W., Sundberg, J., Andra, E., Busso, N., Quatieri, T.F. (2016). The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Trans. Affect. Comput. 7, 190–202.

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., and Quatieri, T.F. (2015). A review of depression and suicide risk assessment using speech analysis. Speech Commun. 71, 10–49.

Galatzew-Leyr, I., Abbas, A., Ries, A., Homan, S., Seis, L., Koemshahargy, V., Yadav, V., Colla, M., Scheerer, H., Vetter, S., et al. (2021). Validation of visual and auditory digital markers of suicidality in acute suicidal psychiatric inpatients: proof-of-concept study. J. Med. Internet Res. 23, e25199.

Song, S., Shen, L., and Valstar, M. (2018). Human behaviour-based automatic depression analysis using hand-crafted statistics and deep learned spectral features. In Proc. 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) (IEEE), pp. 158–165.
207. (2015). National electronic health records survey: 2015 specialty and
191. Abbas, A., Hansen, B.J., Koesmahargyo, V., Yadav, V., Rosenfield, P.J.,
192. de Belen, R.A.J., Bednarz, T., Soumya, A., and Del Favero, D. (2020).
195. Rutowski, T., Shriberg, E., Harati, A., Lu, Y., Chlebek, P., and Oliveira, R.
202. He, L., Niu, M., Tiwari, P., Marttinen, P., Su, R., Jiang, J., Guo, C., Wang,
209. Walsh, C.G., Ribeiro, J.D., and Franklin, J.C. (2018). Predicting suicide
210. Rumshisky, A., Ghassemi, M., Naumann, T., Szolovits, P., Castro, V.M.,
213. Bantilan, N., Malgaroli, M., Ray, B., and Hull, T.D. (2021). Just in time
214. Raveau, M.P., Gon˜i, J., Rodriguez, J., Paiva, I., Barriga, F., Hermosilla,
215. Kesari, G. (2021). AI can now detect depression from your voice, and it's
216. Bathina, K.C., Ten Thij, M., Lorenzo-Luaces, L., and Bollen, J. (2021).
217. Guntuku, S.C., Buffone, A., Jaidka, K., Eichstaedt, J.C., and Ungar, L.H.
218. Rı´ssola, E.A., Losada, D.E., and Crestani, F. (2021). A survey of compu-
219. Karam, Z.N., Provost, E.M., Singh, S., Montgomery, J., Archer, C., Har-
220. Martı´ nez-Castan˜o, R., Htait, A., Azzopardi, L., and Moshfeghi, Y. (2021).
221. Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., and Cambria, E. (2021). Men-
222. Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard,
226. Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard,
230. Le GlaZ, A., Haralambus, Y., Kim-Dufor, D.H., Lenca, P., Billot, R., Ryan,
231. Le D.V., Montgomery, J., Kirkby, K.C., and Scanlan, J. (2018). Risk pre-
diction using natural language processing of electronic mental health re-
cords in an inpatient forensic psychiatry setting. J. Biomed. Inf. 86, 49–58.
232. Walsh, C.G., Ribeiro, J.D., and Franklin, J.C. (2018). Predicting suicide
233. Castro, V.M., Minnier, J., Murphy, S.N., Kohane, I., Churchill, S.E.,
234. Morgan, S.E., Diederken, K., Vérites, P.E., Io, S.H.Y., Wang, B., Thompson,
235. de Belen, R.A.J., Bednarz, T., Soumya, A., and Del Favero, D. (2020).
236. Abbas, A., Hansen, B.J., Koesmahargyo, V., Yadav, V., Rosenfield, P.J.,
237. de Belen, R.A.J., Bednarz, T., Soumya, A., and Del Favero, D. (2020).
238. Castro, V.M., Minnier, J., Murphy, S.N., Kohane, I., Churchill, S.E.,
239. Morgan, S.E., Diederken, K., Vérites, P.E., Io, S.H.Y., Wang, B., Thompson,
240. Abbas, A., Hansen, B.J., Koesmahargyo, V., Yadav, V., Rosenfield, P.J.,
241. de Belen, R.A.J., Bednarz, T., Soumya, A., and Del Favero, D. (2020).
242. Castro, V.M., Minnier, J., Murphy, S.N., Kohane, I., Churchill, S.E.,
243. Morgan, S.E., Diederken, K., Vérites, P.E., Io, S.H.Y., Wang, B., Thompson,
244. Abbas, A., Hansen, B.J., Koesmahargyo, V., Yadav, V., Rosenfield, P.J.,
245. de Belen, R.A.J., Bednarz, T., Soumya, A., and Del Favero, D. (2020).
246. Castro, V.M., Minnier, J., Murphy, S.N., Kohane, I., Churchill, S.E.,
247. Morgan, S.E., Diederken, K., Vérites, P.E., Io, S.H.Y., Wang, B., Thompson,
248. Abbas, A., Hansen, B.J., Koesmahargyo, V., Yadav, V., Rosenfield, P.J.,
249. de Belen, R.A.J., Bednarz, T., Soumya, A., and Del Favero, D. (2020).
250. Castro, V.M., Minnier, J., Murphy, S.N., Kohane, I., Churchill, S.E.,
251. Morgan, S.E., Diederken, K., Vérites, P.E., Io, S.H.Y., Wang, B., Thompson,
sensing and machine learning: a survey. Pervasive Mob. Comput. 51, 1–26.

227. Dagum, P. (2018). Digital biomarkers of cognitive function. NPJ Digit. Med. 1, 1–3.

228. Zulueta, J., Piscitello, A., Rasic, M., Easter, R., Babu, P., Langenecker, S.A., McNinnis, M., Ajlone, O., Nelson, P.C., Ryan, K., and Leow, A. (2018). Predicting mood disturbance severity with mobile phone keystroke metadata: a biometric digital phenotyping study. J. Med. Internet Res. 20, e241.

229. Mandryk, R.L., and Birk, M.V. (2019). The potential of game-based digital biomarkers for modeling mental health. JMIR Ment. Health 6, e13485.

230. Dechant, M., Frommel, J., and Mandry, R. (2021). Assessing social anxiety through digital biomarkers embedded in a gaming task. In Proc. 2021 CHI Conference on Human Factors in Computing Systems.

231. Winkler, R. (2021). Apple is working on iPhone features to help detect disorders and symptoms: systematic review. JMIR Ment. Health 8, e13485.

232. Dagum, P. (2018). Digital biomarkers of cognitive function. NPJ Digit. Med. 1, 1–3.

233. Chikersal, P., Doryab, A., Tumminia, M., Villalba, D.K., Dutcher, J.M., Liu, Bohn, D. (2020). Amazon announces halo, a fitness band and app that scans your body and voice. https://www.theverge.com/2020/8/27/21402493/amazon-halo-band-health-fitness-body-scan-tone-emotion-activity-sleep.

234. fitbit (2022). Understand your stress so you can manage it. https://www.fitbit.com/global/us/technology/stress.

235. Mohr, D.C., Zhang, M., and Schueller, S.M. (2017). Personal sensing: unstructured and telehealth approaches for psychiatric disorders and symptoms: systematic review. JMIR Ment. Health 6, e9819.

236. Chikersal, P., Doryab, A., Tumminia, M., Villalba, D.K., Dutcher, J.M., Liu, X., Cohen, S., Creswell, K.G., Mankoff, J., Creswell, J.D., et al. (2021). Detecting depression and predicting its onset using longitudinal symptom captures by passive sensing: a machine learning approach with robust feature selection. ACM Trans. Comput. Hum. Interact. 28, 1–41.

237. Torous, J., Kiang, M.V., Lorme, J., and Onnela, J.P. (2016). New tools for research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Health 2, e16.

238. Mohr, D.C., Zhang, M., and Schueller, S.M. (2017). Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu. Rev. Clin. Psychol. 13, 23–47.

239. Buckvale, K., Venkatesh, S., and Christensen, H. (2019). Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit. Med. 2, 1–11.

240. Insel, T.R., and Cuthbert, B.N. (2018). Digital phenotyping: a global tool for psychiatry. World Psychiatr. 77, 276–277.

241. Benoit, J., Onyeaka, H., Keshavan, M., and Torous, J. (2020). Systematic review of digital phenotyping and machine learning in psychosis spectrum illnesses. Harv. Rev. Psychiatr. 28, 296–304.

242. Mendes, J.P.M., Moura, I.R., Van de Ven, P., Viana, D., Silva, F.J.S., Coutinho, L.R., Teixeira, S., Rodrigues, J.J.P.C., and Teles, A.S. (2022). Sensing apps and public data sets for digital phenotyping of mental health: systematic review. J. Med. Internet Res. 24, e28735.

243. Prochaska, J.J., Vogel, E.A., Chiang, A., Kendra, M., Baioocchi, M., Pajarito, S., and Robinson, A. (2021). A therapeutic relational agent for reducing problematic substance use (woebot): development and usability study. J. Med. Internet Res. 23, e24850.

244. Bohn, D. (2020). Amazon announces halo, a fitness band and app that scans your body and voice. https://www.theverge.com/2020/8/27/21402493/amazon-halo-band-health-fitness-body-scan-tone-emotion-activity-sleep.

245. fitbit (2022). Understand your stress so you can manage it. https://www.fitbit.com/global/us/technology/stress.

246. Galatzor-Levy, I.R., and Bryant, R.A. (2013). Estimation of clinical tremor using spatio-temporal adversarial autoencoder. In Proc. 25th International Conference on Pattern Recognition (ICPR’20) (IEEE), pp. 8259–8266.

247. Stein, D.J., Fineberg, N.A., and Chamberlain, S.R. (2021). Mental Health in a Digital World (Elsevier).

248. Jacobs, B.L. (1994). Serotonin, motor activity and depression-related disorders. Am. Sci. 82, 456–463.

249. Gigliucci, V., O’Dowd, G., Casey, S., Egan, D., Gibney, S., and Harkin, A. (2013). Ketamine elicits sustained antidepressant-like activity via a serotonin-dependent mechanism. Psychopharmacology 228, 157–166.

250. Abbas, A., Saeder, C., Yadav, V., Koesmahargyo, V., Aghjayan, A., Marecki, S., Evans, M., and Galatzor-Levy, I.R. (2021). Remote digital measurement of facial and vocal markers of major depressive disorder severity and treatment response: a pilot study. Front. Digit. Health 3, 610006.

251. Zhang, L., Koesmahargyo, V., and Galatzor-Levy, I. (2021). Estimation of clinical tremor using spatio-temporal adversarial autoencoder. In Proc. 25th International Conference on Pattern Recognition (ICPR’20) (IEEE), pp. 8259–8266.

252. Ewbank, M.P., Cummings, R., Tablan, V., Batteup, S., Catarino, A., Martin, A.J., and Blackwell, A.D. (2020). Quantifying the association between psychotherapy content and clinical outcomes using deep learning. JAMA Psychiatr. 77, 35–43.

253. Economides, M., Martman, J., Bell, M.J., and Sanderson, B. (2018). Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. Mindfulness 9, 1584–1593.

254. Kunkle, S., Yip, M., Hunt, J., Watson, X., Uddai, D., Arean, P., Nierenberg, A., and Naslund, J.A. (2021). Association between care utilization and anxiety outcomes in an on-demand mental health system: retrospective observational study. JMIR Form. Res. 5, e24682.

255. Patel, N.A., and Butte, A.J. (2020). Characteristics and challenges of the clinical pipeline of digital therapeutics. NPJ Digit. Med. 3, 1–5.

256. Insel, T.R. (2019). Bending the curve for mental health: technology for a public health approach. Am. J. Publ. Health 109, S168–S170.

257. Folk, J.B., Schiel, M.A., Oblath, R., Feuer, V., Sharma, A., Khan, S., Doan, B., Kulkinda, C., Ramtekkar, U., Hawks, J., et al. (2022). The transition of academic mental health clinics to telehealth during the covid-19 pandemic. J. Am. Acad. Child Adolesc. Psychiatr. 61, 277–290.e2.

258. Wagner, B., Horn, A.B., and Maercker, A. (2014). Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. J. Affect. Disord. 152, 113–121.

259. Lahat, D., Adali, T., and Jutten, C. (2015). Multimodal data fusion: an overview of methods, challenges, and prospects. Proc. IEEE 103, 1449–1477.

260. Croitor-Sava, A.R., Martinez-Bisbal, M.C., Laudadio, T., Piquer, J., Celda, B., Heerschap, A., Sima, D.M., and Van Huffel, S. (2011). Fusing in vivo nd ex vivo nmr sources of information for brain tumor classification. Meas. Sci. Technol. 22, 114012.
264. Adali, T., Levin-Schwartz, Y., and Calhoun, V.D. (2015). Multimodal data fusion using source separation: two effective models based on ICA and IVA and their properties. Proc. IEEE. Inst. Electr. Electron. Eng. 103, 1478–1493.

265. Adali, T., Levin-Schwartz, Y., and Calhoun, V.D. (2015). Multimodal data fusion using source separation: application to medical imaging. Proc. IEEE 103, 1494–1506.

266. Calhoun, V., Sui, J., and Qi, S. (2020). Multimodal fusion signature as transdiagnostic psychiatric biomarker. Biol. Psychiatry 87, S37.

267. Zhang, Y.D., Dong, Z., Wang, S.H., Yu, X., Yao, X., Zhou, Q., Hu, H., Li, M., Jiménez-Mesa, C., Ramirez, J., et al. (2020). Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation. Inf. Fusion 64, 149–187.

268. Correa, N.M., Adali, T., Li, Y.O., and Calhoun, V.D. (2010). Canonical correlation analysis for data fusion and group inferences. IEEE Signal Process. Mag. 27, 39–50.

269. de Cheveigne, A., Di Liberto, G.M., Arzouanian, D., Wong, D.D.E., Hjortkjar, J., Fuglaang, S., and Parra, L.C. (2019). Multiway canonical correlation analysis of brain data. Neuroimage 186, 728–740.

270. Chen, X., Wang, Z.J., and McKeown, M. (2016). Joint blind source separation: application to medical imaging. Proc. IEEE 103, S163–S172.

271. Silva, R.F., and Plis, S.M. (2019). How to integrate data from multiple biological layers in mental health? In Personalized Psychiatry (Springer), pp. 135–159.

272. Zhou, G., Zhao, Q., Zhang, Y., Adali, T., Xie, S., and Cichocki, A. (2016). Linked component analysis from matrices to high-order tensors: applications to biomedical data. Proc. IEEE 104, 310–331.

273. Calhoun, V.D., and Adali, T. (2009). Feature-based fusion of medical imaging data. IEEE Trans. Inf. Technol. Biomed. 13, 711–720.

274. Calhoun, V.D., Liu, J., and Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. Neuroimage 45, S163–S172.

275. Acar, E., Bro, R., and Smilde, A.K. (2015). Data fusion in metabolomics using coupled matrix and tensor factorizations. Proc. IEEE 103, 1602–1620.

276. Zhou, G., Cichocki, A., Zhang, Y., and Mandic, D.P. (2016). Group component analysis for multiblock data: common and individual feature extraction. IEEE Transact. Neural Networks Learn. Syst. 27, 2426–2439.

277. Lock, E.F., Hoadley, K.A., Marron, J.S., and Nobel, A.B. (2013). Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. Ann. Appl. Stat. 7, 523–542.

278. Rakotomamonjy, Á., Bach, F., Canu, S., and Grandvalet, Y. (2008). Simplemik. J. Mach. Learn. Res. 9, 2491–2521.

279. Mariette, J., and Villa-Vialaneix, N. (2018). Unsupervised multiple kernel learning for heterogeneous data integration. Bioinformatics 34, 1009–1015.

280. Squarzina, L., Castellani, U., Bellani, M., Perlini, C., Lasalvia, A., Dusi, N., Bonetto, C., Cristofalo, D., Tosato, S., Rambaldelli, G., et al. (2017). Classification of first-episode psychosis in a large cohort of patients using support vector machine and multiple kernel learning techniques. Neuroimage 15, 238–245.

281. Dyrbø, M., Grothe, M., Kirse, T., and Teipel, S.J. (2015). Multimodal analysis of functional and structural disconnection in a Alzheimer’s disease using multiple kernel svm. Hum. Brain Mapp. 36, 2118–2131.

282. Zhang, D., Wang, Y., Zhou, J., Yuan, H., and Shen, D. (2011). Alzheimer’s Disease Neuroimaging Initiative (2011). Multimodal classification of Alzheimer’s disease and mild cognitive impairment. Neuroimage 55, 856–867.

283. Ramachandram, D., and Taylor, G.W. (2017). Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process. Mag. 34, 96–108.

284. Zhou, T., Thung, K.H., Zhu, X., and Shen, D. (2019). Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Hum. Brain Mapp. 40, 1001–1016.

285. Nguyen, N.D., Huang, J., and Wang, D. (2022). A deep manifold-regularized learning model for improving phenotype prediction from multi-modal data. Nat. Comput. Sci. 2, 38–46.

286. Holzinger, A., Maile, B., Saranti, A., and Pfeifer, B. (2021). Towards multimodal causality with graph neural networks enabling information fusion for explainable ai. Inf. Fusion 71, 28–37.

287. Dsouza, N.S., Neelb, M.B., Crocetti, D., Robinson, J., Mostofsky, S., and Venkataraman, A. (2021). M-GCN: A multimodal graph convolutional network to integrate functional and structural connectomics data to predict multi-dimensional phenotypic characterizations. In Medical Imaging with Deep Learning (PMLR), pp. 119–130.

288. Zhang, W., Zhan, L., Thompson, P., and Wang, Y. (2020). Deep representation learning for multimodal brain networks. In International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer), pp. 613–624.

289. Kong, Z., Sun, L., Peng, H., Zhan, L., Chen, Y., and He, L. (2021). Multiplex graph networks for multimodal brain network analysis. Preprint at arXiv. https://doi.org/10.48550/arXiv.2108.00158.

290. Tulası, E.E., Metlin, B., Tarhan, N., and Arkan, M.K. (2019). Multimodal neuroimaging: basic concepts and classification of neuropsychiatric diseases. Clin. EEG Neurosci. 50, 20–33.

291. Sui, J., Pearlson, G., Caprihan, A., Adali, T., Klehi, K.A., Liu, J., Yamamoto, J., and Calhoun, V.D. (2011). Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model. Neuroimage 57, 839–855.

292. Acor, E., Schenker, C., Levin-Schwartz, Y., Calhoun, V.D., and Adali, T. (2019). Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data. Front. Neurosci. 13, 416.

293. Vai, B., Parenti, L., Bolletini, I., Cara, C., Verga, C., Meloni, E., Mazza, E., Poletti, S., Colombo, C., and Benedetti, F. (2020). Predicting differential diagnosis between bipolar and unipolar depression with multiple kernel learning on multimodal structural neuroimaging. Eur. Neuropsychopharmacol. 34, 28–38.

294. Zhang, G., Wu, Q., Zhu, H., He, L., Huang, H., Zhang, J., and Zhang, W. (2018). Multimodal mri-based classification of trauma survivors with and without post-traumatic stress disorder. Front. Neurosci. 10, 292.

295. Zhou, X., Lin, Q., Gui, Y., Wang, Z., Liu, M., and Lu, H. (2021). Multimodal MRI Images-based diagnosis of early adolescent attention-deficit/hyperactivity disorder using multiple kernel learning. Front. Neurosci. 15, 710133.

296. Geenjaa, E., Lewis, N., Fu, Z., Venkatdas, R., Plis, S., and Calhoun, V. (2021). Fusing multimodal neuroimaging data with a variational autoencoder. In Proc. 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (IEEE), pp. 3630–3633.

297. Liu, J., Wang, X., Zhang, X., Pan, Y., Wang, X., and Wang, J. (2018). Mm: classification of schizophrenia using multi-modality multi-atlas feature representation and multi-kernel learning. Multimed. Tools Appl. 77, 29651–29667.

298. Plis, S.M., Amin, M.F., Chekroud, A., Hjelm, D., Damara, S., Lee, H.J., Bustillo, J.R., Chou, K., Pearson, G.D., and Calhoun, V.D. (2018). Reading the (functional) writing on the (structural) wall: multimodal fusion of brain structure and function via a deep neural network based translation approach reveals novel impairments in schizophrenia. Neuroimage 181, 734–747.

299. Rahaman, M.A., Chen, J., Fu, Z., Lewis, N., Iraji, A., and Calhoun, V.D. (2021). Multi-modal deep learning of functional and structural neuroimaging and genomic data to predict mental illness. In Proc. 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (IEEE), pp. 3267–3272.

300. Akhonda, M.A.B.S., Levin-Schwartz, Y., Calhoun, V.D., and Adali, T. (2022). Association of neuroimaging data with behavioral variables: a
class of multivariate methods and their comparison using multi-task fMRI data. Sensors 22, 1224.

301. Wang, Y., Jiang, W.B., Li, R., and Lu, B.L. (2021). Emotion transformer fusion: complementary representation properties of EEG and eye movements on recognizing anger and surprise. In Proc. IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (IEEE), pp. 1575–1578.

302. Jaques, N., Taylor, S., Sano, A., and Picard, R. (2015). Multi-task, multi-kernel learning for estimating individual wellbeing. Proc. NIPS Workshop on Multimodal Machine Learning 896, 3.

303. Lam, G., Dongyan, H., and Lin, W. (2019). Context-aware deep learning for multi-modal depression detection. In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE), pp. 3946–3950.

304. Parra, F., Benezeth, Y., and Yang, F. (2022). Automatic assessment of emotion dysregulation in american, french, and tusnian adults and new developments in deep multimodal fusion: cross-sectional study. JMIR Ment. Health 9, e43333.

305. Doryab, A., Villalba, D.K., Chikersal, P., Dutcher, J.M., Tumminia, M., Liu, X., Cohen, S., Creswell, K., Mankoff, J., Creswell, J.D., and Dey, A.K. (2019). Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data. JMIR Mhealth Uhealth 7, e13209.

306. He, J., Mai, S., and Hu, H. (2021). A unimodal reinforced transformer with time squeeze fusion for multimodal sentiment analysis. IEEE Signal Process. Lett. 28, 992–996.

307. Dibeklio€glu, H., Hammal, Z., and Cohn, J.F. (2018). Dynamic multimodal measurement of depression severity using deep autoencoding. IEEE J. Biomed. Health Inform. 22, 525–536.

308. Herzog, D.P., Beckmann, H., Lieb, K., Ryu, S., and Müller, M.B. (2018). Understanding and predicting antidepressant response: using animal models to move toward precision psychiatry. Front. Psychiatr. 9, 512.

309. Bale, T.L., Abel, T., Akil, H., Carlezon, W.A., Jr., Moghaddam, B., Nestler, E.J., Nestler, E.J., Pessler, K.J., and Thompson, S.M. (2019). The critical importance of basic animal research for neuropsychiatric disorders. Neuropsychopharmacology 44, 1349–1353.

310. Labonté, B., Engmann, O., Purushothaman, I., Menard, C., Wang, J., Tan, C., Scarpa, J.R., Moy, G., Loh, Y.H.E., Cahill, M., et al. (2017). Sex-specific transcriptional signatures in human depression. Nat. Med. 23, 1102–1111.

311. Nagy, C., Mastra, M., Tanti, A., Sumerdan, M., Théroux, J.F., Davoli, M.A., Perlmutter, K., Yerko, V., Wang, Y.C., Tripathy, S.J., et al. (2020). Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. Nat. Neurosci. 23, 771–781.

312. McEwen, B.S., Bowles, N.P., Gray, J.D., Hill, M.N., Hunter, R.G., Karrat, E.J., Nasca, C., and McEwen, B. (2016). Epigenetics and energetics in ventral hippocampus mediate rapid antidepressant action: implications for treatment resistance. Proc. Natl. Acad. Sci. USA 113, 7906–7911.

313. Flight, M.H. (2013). Antidepressant epigenetic action. Nat. Rev. Neurosci. 14, 226.

314. Stuart, T., and Satija, R. (2018). Integrative single-cell analysis. Nat. Rev. Genet. 20, 257–272.

315. Petegrosso, R., Li, Z., and Kuang, R. (2020). Machine learning and statistical methods for clustering single-cell rna-sequencing data. Brief. Bioinform. 21, 1209–1223.

316. Amodio, M., van Dijk, D., Srinivasan, K., Chen, W.S., Mohnen, H., Moon, K.R., Campbell, A., Zhao, Y., Wang, X., Venkataswamy, M., et al. (2019). Integrating analysis of single-cell data with deep multitasking neural networks. Nat. Methods 16, 1139–1145.

317. Schultebraucks, K., Qian, M., Abu-Amara, D., Dean, K., Laska, E., Siegel, C., Guffanti, A., Wolff, G., Kaidar-Person, A., et al. (2021). Multidimensional predictors of antidepressant responses: integrating mitochondrial, genetic, metabolic and environmental factors with clinical outcomes. Neurobiol. Stress 15, 100407.

318. Dean, K.R., Hamamieh, R., Meillon, S.H., Abu-Amara, D., Flory, J.D., Guffanti, G., Wang, K., Daigle, B.J., Jr., Gautham, A., Lee, I., et al. (2020). Multi-omic biomarker identification and validation for diagnosing warzone-related post-traumatic stress disorder. Mol. Psychiatr. 25, 3337–3349.

319. Schultebraucks, K., Qian, M., Abu-Amara, D., Dean, K., Laska, E., Siegel, C., Guffanti, A., Hamamieh, R., Misjanaw, B., et al. (2021). Pre-deployment risk factors for ptsd in active-duty personnel deployed to Afghanistan: a machine-learning approach for analyzing multivariate predictors. Mol. Psychiatr. 26, 5011–5022.

320. Lorsch, Z.S., Ambesi-Impiombato, A., Zenowich, R., Morganstern, I., Leahy, E., Bansal, M., Nestler, E.J., and Hanania, T. (2021). Computational analysis of multidimensional behavioral alterations after chronic social defeat stress. Biol. Psychiatr. 89, 920–928.

321. Alexsandrov, V., Brunner, D., Hanania, T., and Leahy, E. (2015). High-throughput analysis of behavior for drug discovery. Eur. J. Pharmacol. 750, 82–89.

322. Nasca, C., Menard, C., Hodes, G., Bigio, B., Pena, C., Lorsch, Z., Zelli, D., Ferris, A., Kana, V., Purushothaman, L., et al. (2019). Multidimensional predictors of susceptibility and resilience to social defeat stress. Biol. Psychiatr. 86, 483–491.
335. Gunning, D., and Aha, D. (2019). DARPA’s explainable artificial intelligence (XAI) program. AI Mag. 40, 44–58.
336. Roessner, V., Rothe, J., Kohls, G., Schomerus, G., Ehrlich, S., and Beste, C. (2021). Taming the chaos? Using eXplainable Artificial Intelligence (XAI) to tackle the complexity in mental health research. Eur. Child Adolesc. Psychiatry. 30, 1143–1146.
337. Thorsen-Meyer, H.C., Nielsen, A.B., Nielsen, A.P., Kaas-Hansen, B.S., Toft, P., Schierbeck, J., Strom, T., Chmura, P.J.M., Heimann, M., Dybdahl, L., et al. (2020). Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digit. Health 2, e179–e191.
338. Sheu, Y.H. (2020). Illuminating the black box: interpreting deep neural network models for psychiatric research. Front. Psychiatr. 11, 551299.
339. Molnar, C. (2020). Interpretable machine learning: a guide for making black box models explainable. https://www.amazon.com/Interpretable-Machine-Learning-Making-Explainable/dp/B09T7MVHW8.
340. Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (Elsevier).
341. Letham, B., Rudin, C., McCormick, T.H., and Madigan, D. (2015). Interpretable classifiers using rules and bayesian analysis: building a better stroke prediction model. Ann. Appl. Stat. 9, 1350–1371.
342. Li, Y., Dzirasa, K., Carin, L., and Carlson, D.E. (2017). Targeting EEG/LFP synchrony with neural nets. In Advances in Neural Information Processing Systems 30, I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, and S. Vishwanathan, et al., eds. (Curran Associates, Inc.).
343. Zou, J.Y., and Adams, R.P. (2012). Priors for diversity in generative latent variable models. . Advances in Neural Information Processing Systems (NIPS), 25.
344. Nair, R., Mattetti, M., Daly, E., Wei, D., Aikan, O., and Zhang, Y. (2021). What changed? interpretable model comparison. In Proc. Int. Joint Conf. Artificial Intelligence (IJCAI’21).
345. Daly, E., Mattetti, M., Aikan, O., and Nair, R. (2021). User driven model adjustment via boolean rule explanation. In Proc. 35th Conf. Artificial Intelligence (AAAI’21), pp. 5896–5904.
346. Huys, Q.J.M., Maia, T.V., and Frank, M.J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. Nat. Neurosci. 19, 404–413.
347. Breakspear, M. (2017). Dynamic models of large-scale brain activity. Nat. Neurosci. 20, 340–352.
348. Murray, J.D., Demirtas, M., and Anticevic, A. (2018). Biophysical modeling of large-scale brain dynamics and applications for computational psychiatry. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 3, 777–787.
349. Papadopoulos, L., Lynn, C.W., Battaglia, D., and Bassett, D.S. (2020). Relations between large-scale brain connectivity and effects of regional stimulation depend on collective dynamical state. PLoS Comput. Biol. 16, e1008144.
350. Wilson, H.R., and Cowan, J.D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. 12, 1–24.
351. Chaudhuri, R., Knoblauch, K., Gariel, M.A., Kennedy, H., and Wang, X.J. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. Neuron 88, 419–431.
352. Friston, K.J., Harrison, L., and Penny, W. (2003). Dynamical causal modeling. Neuroimage 19, 1273–1302.
353. Friston, K.J., Kahan, J., Biswal, B., and Razi, A. (2014). A DCM for resting state fMRI. Neuroimage 94, 396–407.
354. Chen, Z.S., and Pesaran, B. (2021). Improving scalability in systems neuroscience. Neuron 109, 1776–1790.
355. Nair, K., Haslacher, D., Dayan, E., Censor, N., Cohen, L.G., and Soekadar, S.R. (2022). Breaking the boundaries of interacting with the human brain using adaptive closed-loop stimulation. Prog. Neurobiol. 216, 102311.
356. Zajzer, M., and Afraz, A. (2017). Navigating the neural space in search of the neural code. Neuron 93, 1003–1014.
357. Etkin, A. (2018). Addressing the causality gap in human psychiatric neuroscience. JAMA Psychiatr. 75, 3–4.
358. Fellous, J.M., Sapiro, G., Rossi, A., Mayberg, H., and Ferrante, M. (2019). Explainable artificial intelligence for neuroscience: behavioral neurostimulation. Front. Neurosci. 13, 1346.
359. Lewis, P.M., Thomson, R.H., Rosenfeld, J.V., and Fitzgerald, P.B. (2016). Brain neuromodulation techniques: a review. Neuroscientist 22, 406–421.
360. Romei, V., Thut, G., and Silvanto, J. (2016). Information-based approaches of noninvasive transcranial brain stimulation. Trends Neurosci. 39, 782–795.
361. Lo, M.C., and Widge, A.S. (2017). Closed-loop neuromodulation systems: next generation treatments for psychiatric illness. Int. Rev. Psychiatr. 29, 191–204.
362. Chen, A.C., Osth, D.J., Chang, C., Bradley, T., Zhou, Z.W., Williams, L.M., Glover, G.H., Deisseroth, K., and Etkin, A. (2013). Causal interactions between fronto-parietal central executive and default-mode networks in humans. Proc. Natl. Acad. Sci. USA 170, 19944–19949.
363. Hobot, J., Klineciewicz, M., Sandberg, K., and Wierczon, M. (2020). Causal inferences in repetitive transcranial magnetic stimulation research: challenges and perspectives. Front. Hum. Neurosci. 14, 586448.
364. Akhtar, H., Bilkhari, F., Nazir, M., Anwar, M.N., and Shahzad, A. (2016). Therapeutic efficacy of neurostimulation for depression: techniques, current modalities, and future challenges. Neurosurg. Bull. 29, 115–126.
365. Pineau, J., Guez, A., Vincent, R., Panuccio, G., and Avoli, M. (2009). Treating epilepsy via adaptive neurostimulation: a reinforcement learning approach. Int. J. Neural Syst. 19, 227–240.
366. Tafazoli, S., MacDowell, C.J., Che, Z., Letai, K.C., Steinhardt, C.R., and Buschman, T.J. (2020). Learning to control the brain through adaptive closed-loop patterned stimulation. J. Neural Eng. 17, 056007.
367. Gao, O., Neumann, M., Jovanov, I., Lesi, V., Kamaravelu, K., Grill, W.M., and Pajic, M. (2020). Model-based design of closed loop deep brain stimulation controller using reinforcement learning. In Proc. 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPSS) (IEEE). https://doi.org/10.1109/ICCPSS48487.2020.00018.
368. Hebb, A.O., Zhang, J.J., Mahoor, M.H., Taiakos, C., Mattlack, C., Chizeck, H.J., and Pouratian, N. (2014). Creating the feedback loop: closed-loop neurostimulation. Neurosurg. Clin. 25, 187–204.
369. Bauer, R., and Bharagabhi, A. (2015). Reinforcement learning for adaptive threshold control of restorative brain-computer interfaces: a Bayesian simulation. Front. Neurosci. 9, 36.
370. Tang, E., and Bassett, D.S. (2018). Control of dynamics in brain networks. Rev. Mod. Phys. 90, 031003.
371. Srivastava, P., Nozari, E., Kim, J.Z., Ju, H., Zhou, D., Becker, C., Pasqualett, F., Pappas, G.J., and Bassett, D.S. (2020). Models of communication and control for brain networks: distinctions, convergence, and future outlook. Natw. Neurosci. 4, 1122–1159.
372. Zhang, X., Braun, U., Tost, H., and Bassett, D.S. (2020). Data-driven approaches to neuroimaging analysis to enhance psychiatric diagnosis and therapy. Biol. Psychiatr. Cogn. Neurosci. Neuroimaging 5, 780–790.
373. Fox, M.D., Buckner, R.L., Liu, H., Chakravarty, M.M., Lozano, A.M., and Pascual-Leone, A. (2014). Resting-state networks link invasive and noninvasive brain stimulation across diverse psychiatric and neurological diseases. Proc. Natl. Acad. Sci. USA 117, E4367–E4375.
374. Zandvakili, A., Philip, N.S., Jones, S.R., Tyrka, A.R., Greenberg, B.D., and Carpenter, L.L. (2019). Use of machine learning in predicting clinical response to transcranial magnetic stimulation in comorbid posttraumatic stress disorder and major depression: a resting state electroencephalography study. J. Affect. Disord. 252, 47–54.
375. Vollmer, S., Mateen, B.A., Bohnen, G., Kiraly, F.J., Giani, R., Jonsson, P., Cumbers, S., Jonas, A., McAllister, K.S.L., Myles, P., et al. (2020).
Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 368, l6927.

376. Kessler, R.C., Chiu, W.T., Demler, O., Merikangas, K.R., and Walters, E.E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. Arch. Gen. Psychiatry. 62, 617–627.

377. Born, J., Beymer, D., Rajan, D., Coy, A., Mukherjee, V.V., Manica, M., Prasanna, P., Ballah, D., Guindy, M., Shaham, D., et al. (2021). On the role of artificial intelligence in medical imaging of COVID-19. Patterns 2, 100269.

378. Pendse, S.R., Nkemelu, D., Bidwell, N.J., Jadhav, S., Pathare, S., De, M., and Kumar, N. (2022). From treatment to healing: envisioning a decolonial digital mental health. In Proc. CHI Conference on Human Factors in Computing Systems (CHI’22).

379. Palaniyappan, L. (2021). More than a biomarker: could language be a biosocial marker of psychosis? npj Schizophr. 7, 1–5.

380. Luo, Y., Peng, J., and Ma, J. (2020b). When causal inference meets deep learning. Nat. Mach. Intel. 2, 426–427.

381. Prosperi, M., Guo, Y., Sperrin, M., Koopman, J.S., Min, J.S., He, X., Rich, S., Wang, M., Buchan, I.E., and Bial, J. (2020). Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nat. Mach. Intel. 2, 369–375.

382. Neuner, I., Veselinovic, T., Ramkran, S., Rajkumar, R., Schnellbaecher, G.J., and Shah, N.J. (2022). 7T ultra-high-field neuroimaging for mental health: an emerging tool for precision psychiatry? Transl. Psychiatr. 12, 1–10.

383. Rajpurkar, P., Chen, E., Banerjee, O., and Topol, E.J. (2022). AI in health and medicine. Nat. Med. 28, 31–38.

384. Grande, D., Mitra, N., Iyer, P., Merchant, R.M., Asch, D.A., Sharma, M., and Cannuscio, C.C. (2022). Consumer willingness to share personal digital information for health-related uses. JAMA Netw. Open 5, e2144787.

385. Davidson, B.I. (2022). The crossroads of digital phenotyping. Gen. Hosp. Psychiatri. 74, 126–132.

386. Korngiebel, D.M., and Mooney, S.D. (2021). Considering the possibilities and pitfalls of generative pre-trained transformer 3 (gpt-3) in healthcare delivery. NPJ Digit. Med. 4, 1–3.

387. Polyzotis, N., and Zaharia, M. (2021). What can data-centric AI learn from data and ML engineering? Preprint at arXiv. https://arxiv.org/pdf/2112.06439.pdf.

388. Northcutt, C., Jiang, L., and Chuang, I. (2021). Confident learning: estimating uncertainty in dataset labels. J. Artif. Intell. Res. 70, 1373–1411.

389. Chen, R.J., Lu, M.Y., Chen, T.Y., Williamson, D.F.K., and Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. Nat. Biomed. Eng. 5, 493–497.

390. Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., Chen, Y., and Zhou, X. (2020). Generative adversarial networks and its applications in biomedical informatics. Front. Public Health 8, 164.

391. Geng, D., Alkachroum, A., Melo Cicchi, M.A., Jagid, J.R., Caijgas, I., and Chen, Z.S. (2021). Deep learning for robust detection of interictal epileptic discharges. J. Neural. Eng. 18, 056015.

392. Weldon, J., Ward, T., and Brophy, E. (2021). Generation of synthetic electronic health records using a federated gan. Preprint at arXiv. https://doi.org/10.48550/arXiv.2109.02543.

393. Chen, L., Xia, C., and Sun, H. (2020). Recent advances of deep learning in psychiatric disorders. Precis. Clin. Med. 3, 202–213.

394. Zhang, Z., Cui, P., and Zhu, W. (2022). Deep learning on graphs: a survey. IEEE Trans. Knowl. Data Eng. 34, 249–270.

395. Bronstein, M.M., Bruna, J., Cohen, T., and Velickovic, P. (2021). Geometric deep learning: grids, groups, graphs, geodesics, and gauges. Preprint at arXiv. https://doi.org/10.48550/arXiv.2104.13478.

396. Gennatas, E.D., Friedman, J.H., Ungar, L.H., Pirracchio, R., Eaton, E., Reichmann, L.G., Interian, Y., Luna, J.M., Simone, C.B., 2nd, Auerbach, A., et al. (2020). Expert-augmented machine learning. Proc. Natl. Acad. Sci. USA 117, 4571–4577.

397. Habtewold, T.D., Rodijk, L.H., Limburg, E.J., Sidorenkov, G., Boezen, H.M., Bruggeman, R., and Alizadeh, B.Z. (2020). A systematic review and narrative synthesis of data-driven studies in schizophrenia symptoms and cognitive deficits. Transl. Psychiatr. 10, 1–24.

398. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V.X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., et al. (2019). Do no harm: a roadmap for responsible machine learning for health care. Nat. Med. 25, 1337–1340.