

Modern Views of Machine Learning for Precision Psychiatry

Zhe Sage Chen^{a,b,c,d,*}, Prathamesh (Param) Kulkarni^e, Isaac R. Galatzer-Levy^{f,a}, Benedetta Bigio^a,
Carla Nasca^{a,c}, Yu Zhang^{g,h,*}

^aDepartment of Psychiatry, New York University Grossman School of Medicine, New York, NY 10016, USA

^bDepartment of Neuroscience and Physiology, New York University Grossman School of Medicine, New York, NY 10016, USA

^cThe Neuroscience Institute, New York University Grossman School of Medicine, New York, NY 10016, USA

^dDepartment of Biomedical Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201, USA

^eHeadspace Health, San Francisco, CA 94102, USA

^fMeta Reality Lab, New York, NY, USA

^gDepartment of Bioengineering, Lehigh University, PA 18015, USA

^hDepartment of Electrical and Computer Engineering, Lehigh University, PA 18015, USA

Abstract

Summary: In light of the 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 paper, we provide a comprehensive review of the ML methodologies and applications by combining neuroimaging, neuromodulation, and advanced mobile technologies in psychiatry practice. We further review the role of ML in molecular phenotyping and cross-species biomarker identification in precision psychiatry. We also discuss explainable AI (XAI) and neuromodulation in a closed-human-in-the-loop manner, and highlight the ML potential in multimedia information extraction and multimodal data fusion. Finally, we discuss conceptual and practical challenges in precision psychiatry and highlight ML opportunities in future research.

Keywords: Machine learning (ML), artificial intelligence (AI), deep learning, precision psychiatry, digital psychiatry, computational psychiatry, neuroimaging, neurobiomarker, molecular biomarker, digital phenotyping, multimodal data fusion, neuromodulation, causality, explainable AI (XAI), teletherapy

1. Introduction

Mental health issues are 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 uncertainty¹. Increased suicide rates, opioid abuse, 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 the society. In the United States alone, the economic burden of depression alone is estimated to be at least \$210 billion annually². Precision medicine (or personalized medicine) is an innovative approach to tailoring disease prevention, diagnosis, and treatment that accounts for the differences in subjects' genes, environments, and

lifestyles. The goal of precision medicine is to target timely and accurate diagnosis/prognosis/therapeutics for the individualized patient's health problem, and further provide feedback information to patients and surrogate decision-makers. Recent decades have witnessed various degrees of successes in precision medicine, especially in oncology (Lancet, vol. 397, 2021, p. 1781). Traditional diagnoses of mental illnesses rely on physical exams, lab tests, psychological and behavioral evaluation. Meanwhile, precision psychiatry has increasingly received its deserved attention^{3,4}. Although psychiatry has not yet benefited fully from the advanced diagnostic and therapeutic technologies that have an impact on other clinical specialties, these technologies have the potential to transform the future psychiatric landscape.

The NIMH's RDoC (Research Domain Criteria) initiative aims to address the heterogeneity of mental illness and provide a biology-based (as opposed to symptom-based) framework for understanding these mental illnesses in terms of varying degrees of dysfunction in psychological or neurobiological systems; it attempts to bridge the power of multidisciplinary (such as the genetics, neuroscience, and behavioral science) research approaches^{5,6}. The current gold standard for diagnosis

*Corresponding authors:

zhe.chen@nyulangone.org (Z.S. Chen, ORCID: 0000-0002-6483-6056)
yuzi20@lehigh.edu (Y. Zhang, ORCID: 0000-0003-4087-6544)

and treatment outcome in mental disorders—the DSM (Diagnostic and Statistical Manual of Mental Disorders) 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/neurophysiological 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^{7,8,9}. 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 outcome with digital technologies, the so-called “digital psychiatry” focuses on developing ML/AI methods for assessing, diagnosing, and treating mental health issues¹⁰. 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 costs¹¹.

The past two decades have witnessed substantial growth of ML applications for psychiatry in the literature, reflected in many applications and reviews^{12,13,14,15,16,17,18,19,20,21,22}. 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 a 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 limitation, 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, clinical and behavioral integration.

Our view about this emerging field is cautiously optimistic for several reasons. First, with 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 artificial intelligence (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.

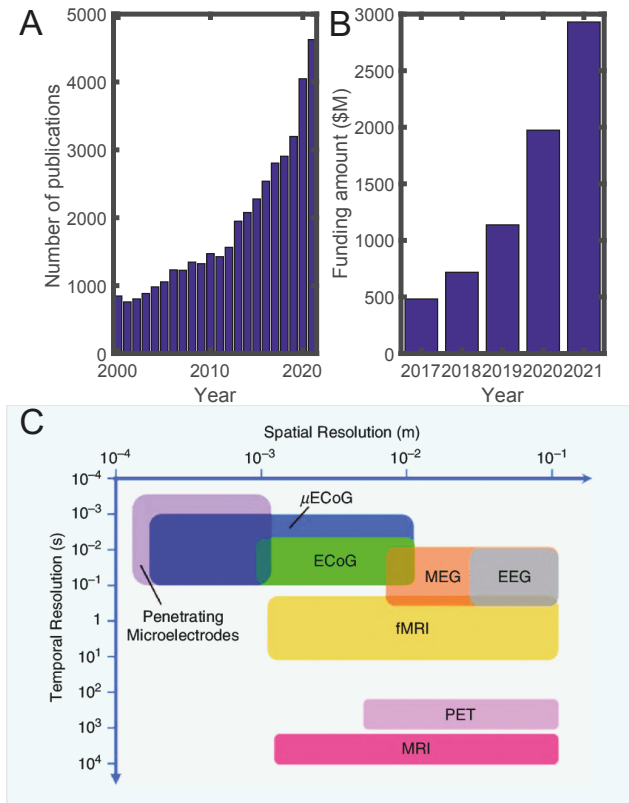


Figure 1: (A) The number of PubMed publications with keywords “machine Learning or AI” and “psychiatry or mental health” in the title or abstract (Year 2000-2021). (B) Growth of mental health tech funding in the US market (Year 2017-2021, data source: cbinsights.com). (C) Human neuroimaging at various spatial and temporal resolution (© IEEE, figure reproduced from²⁴ with permission).

Third, there is a growing demand of psychiatrists in the US, and the shortage is even more acute in poorer countries²³. ML/AI technologies may change the practice of psychiatry for both clinicians and patients. Finally, advanced technologies such as social media, multimedia, 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 paper, 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.

2. Background of Neuroimaging

2.1. 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 role²⁵. 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 disorders²⁶. Some of the most common neuroimaging methods²⁷ 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.

2.2. 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 allow us to examine the relationship between brain activities (e.g., event-related potential and spectral perturbation, and reward or emotional processing-related functional activation^{28,29}) and cognitive dysfunctions. A promising direction for probing brain function using neuroimaging is to investigate brain connectivity (or connectome)³⁰. Studying the resting-state brain connectome provides an elegant way to characterize the complex brain architecture and uncover brain dysfunctions in intrinsic brain networks³¹. Increasing neuroimaging studies suggest that functional connectivity may fluctuate, rather than being stationary during an entire session of data collection³². Studies examining spatiotemporal dynamics of brain networks have recently received growing attention and may reveal meaningful brain states associated with different psychiatric conditions³³. 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 capture³⁴.

2.3. 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)³⁵. 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^{36,37,38}.

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.

3. How Can ML Help Psychiatry?

3.1. Psychiatry vs. 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^{39,40}. 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 neurobiological 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. Co-occurrence 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 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.

3.2. 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 behavior⁴¹. 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

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	Multimodal analysis, Joint disease subtyping and diagnosis, Prediction with incomplete data
Deep	Learning hierarchies and nonlinear 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

enhance the ability to find statistically significant effect sizes that can be used in individuals⁴².

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 distinctive behavior and cognitive functioning⁴³. These stratified phenotypes may help improve the predictability of clinical outcomes and serve as potential biomarkers for treatment selection³⁸. However, subtyping analysis is widely viewed as hypothesis-generating and poses significant challenges related to reproducibility and physiological interpretability^{44,39}. Linking subtype identification to a particular outcome or question using well-designed ML approaches is vital to address these challenges⁴⁵.

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 spectrum⁴⁶. Recent ML efforts have been dedicated to identifying transdiagnostic brain dysfunctions and dimensions of psychopathology to improve understanding of comorbidity among psychiatric disorders^{47,48,49,50}. 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^{51,52,53}. 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^{54,55}. 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^{56,57}. 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^{58,59}.

These new frontiers in studying psychiatric disorders can be substantially empowered by ML methodologies summarized in Table 1. The applications include stratifying patients into clin-

ically 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 illnesses⁶⁰. Unsupervised clustering is useful in identifying disease subtypes for dissect clinical and biological heterogeneity, thus offering new ways of defining psychiatric conditions^{38,61}. 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 disease⁶². For 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 symptoms⁶³. For readmission evaluation, supervised classification can be used to predict whether an individual would be rehospitalized or to detect the relapse trajectory^{64,65}. 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 patterns⁴².

It should be noted that ML is an ever-growing big discipline, and cover many categories and emergent topics, each with a different technical focus. A standard taxonomy of ML typically include supervised, unsupervised and reinforcement learning paradigms^{66,67}. There are also various extensions or special treatment of each category or combined (e.g., semi-supervised learning, kernel learning, ensemble learning, deep learning). Here, our rationale of reviewing specific ML methods is simply based on their applicability in existing mental health applica-

tions.

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.

3.3. 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^{102,25,103}. 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^{12,104,105}. These promising applications in psychiatric disorders have been enabled by leveraging the powerful strength of modern ML techniques^{15,60,106,107}.

Supervised Learning. Supervised learning, being the most popularly used ML category, has been widely applied to neuroimaging-based predictive modeling tasks for psychiatric disorders¹⁰⁸. 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^{109,110} is one of such approaches that combine simple linear regression and feature selection to predict individual differences in traits and behavior from connectivity data. 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 outcome⁹⁹. Relevance vector machine (RVM) builds upon a probabilistic framework by leveraging automatic relevance determination to learn a sparse solution and penalize unnecessary model complexity^{111,112}. RVM has recently demonstrated its strength in quantifying neuroimaging biomarkers for PTSD diagnosis⁷³ as well as for treatment outcome prediction in depression⁹⁷. 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^{113,114,115}.

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 antidepressant-responsive brain signatures in depression⁹⁶. 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 symptoms⁴³. 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 symptoms¹¹⁶. A sparse CCA approach has been applied to reveal linked dimensions of psychopathology and functional connectivity in brain networks for psychiatric disorders⁴⁸. This approach successfully identified interpretable dimensions, involving mood, psychosis, fear, and externalizing behavior, guided neural circuit patterns across the clinical diagnostic spectrum. The partial least squares (PLS) approach was also applied to identify latent components linking a broad set of behavioral measures to functional connectivity⁴⁹. The latent components defined distinct dimensions with dissociable brain functional signatures, and provided potential intermediate phenotypes spanning diagnostic categories. These dimensional analytics hold great promise in uncovering novel transdiagnostic phenotypes for developing targeted interventions.

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 detection¹¹⁷ and ASD diagnosis⁷¹. By utilizing multimodal neuroimaging including sMRI, fMRI, and DTI, a bagging-based SVM produced significant improvement in prediction of adult outcomes in childhood-onset ADHD⁶⁹. Based on the selective ensemble algorithm, a sparse multi-view prediction model has been designed with rs-fMRI connectivity for ASD diagnosis¹¹⁸; 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 (Section 7). Additionally, labeled data require the ground truth knowledge, which is not always accurate or re-

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	68
	Supervised Classification	Ensemble learning	ADHD	Multimodal	69
	Supervised Classification, Deep	GCN	ASD	Task fMRI	70
	Supervised Classification, Deep	Ensemble learning + GCN	ASD	rs-fMRI	71
	Supervised Classification	PCA + LASSO	Bipolar	dMRI + Cognitive data	72
	Supervised Classification	RVM	PTSD	rs-fMRI	73
	Supervised Classification, Deep	ICA + LSTM	Schizophrenia	fMRI	74
	Supervised Classification	SVM	Schizophrenia	sMRI	75
	Supervised Classification, Deep	CNN	Depression	rs-EEG	76
	Supervised Classification, Deep	Autoencoder + DNN, SVM, Random forest	ASD	rs-fMRI	77
	Semi-supervised Classification	GNN	ASD	rs-fMRI + Phenotypic data	78
	Unsupervised, Subtyping	Normative modeling + clustering	PTSD	rs-fMRI	79
	Unsupervised, Subtyping	CCA + Hierarchical clustering	Depression	rs-fMRI	43
	Unsupervised, Subtyping	Sparse K-means	PTSD, Depression	rs-EEG	38
	Unsupervised, Subtyping	Latent class analysis	ADHD	Task fMRI	80
	Supervised, Transdiagnostic	Normative modeling + GP regression	Multiple disorders	rs-fMRI	81
	Unsupervised, Transdiagnostic	Sparse CCA	Multiple disorders	rs-fMRI	48
	Supervised, Transdiagnostic	PLS	Multiple disorders	rs-fMRI	49
Prognosis	Supervised Classification	GP classifier	Depression	Task fMRI	82
	Supervised, Classification	LASSO	Psychosis	rs-EEG	83
	Supervised Classification	SVM	Psychosis, Depression	multimodal	84
	Supervised Classification, Deep	DNN	PTSD	rs-fMRI / task fMRI	85
	Supervised Classification	SVM	Schizophrenia	sMRI	86
	Supervised Classification, Deep	SVM, random forest, DNN	Schizophrenia	Task fMRI	87
	Supervised Regression	LASSO	Substance use	MRI/task fMRI	88
	Supervised Regression, Deep	SVR+LSTM	PTSD	MEG	89
	Supervised Classification	SVM	ADHD	sMRI	90
	Supervised Classification	SVM	Psychosis	sMRI	91
Treatment prediction	Supervised Classification	GP classifier	PTSD	MRI/rs-fMRI	92
	Supervised Classification	SVM	Schizophrenia	rs-fMRI	93
	Supervised Classification	SVM	Depression	rs-EEG	94
	Supervised Classification	SVM + GP classifier	Depression	sMRI	95
	Supervised Regression	Latent space learning	Depression	rs-EEG	96
	Supervised Regression	RVM	Depression	Task fMRI	97
	Supervised Regression	MVPA	ASD	Task fMRI	98
	Supervised Regression	LASSO	Anxiety	rs-fMRI	99
	Supervised Classification	SVM	Depression	Multimodal	64
	Supervised Classification	Growth mixture modeling	Depression	Clinical data	65
Readmission	Supervised Classification	Decision tree	Bipolar	EHR	100
	Supervised Classification	Ensemble learning	Substance use	Phenotypic data	101

liable in the case of mental disorders. For instance, the 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 PCA). Identifying patient subtypes offers a promising strategy to delineate neurobiological heterogeneity in psychiatric disorders³⁹. With rs-fMRI, hierarchical clustering was applied to successfully identify four subtypes of functional connectivity in depression⁴³. 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 antidepressant medication and psychotherapy³⁸. 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 profiles⁸⁰ and reveal 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 modeling²⁵. 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^{119,120}. 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^{121,122,123}. LGMM approaches have also been successfully used to predict the PTSD course among the population at risk¹²⁴. 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)¹²⁵. 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 cluster-

ing¹²⁶. These methods have been increasingly applied to psychiatric studies. By unifying autoencoder and classification, a semi-supervised model was developed for ASD diagnosis¹²⁷. A semi-supervised classification has been devised using graph convolutional networks and applied to the population graph-based diagnosis of ASD⁷⁸. 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 diagnosis¹²⁸. This approach has shown strength in delineating neurostructural heterogeneity in bipolar and major depressive disorders (MDDs)¹²⁹, schizophrenia¹³⁰, as well as in youth with internalizing symptoms¹³¹. Additionally, semi-supervised learning has gained increasing mental health applications in digital data from electronic health records (EHRs), social media and mobile phones^{132,133,134}. See Section 4 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 model¹³⁵. It has been increasingly used in mental health¹³⁶ 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 individual's unique deviations from normative neurodevelopment in brain structure⁸¹. 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 PTSD⁷⁹. 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, autoencoder model has been utilized to realize normative modeling for detecting microstructural deviations in¹³⁷.

3.4. 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

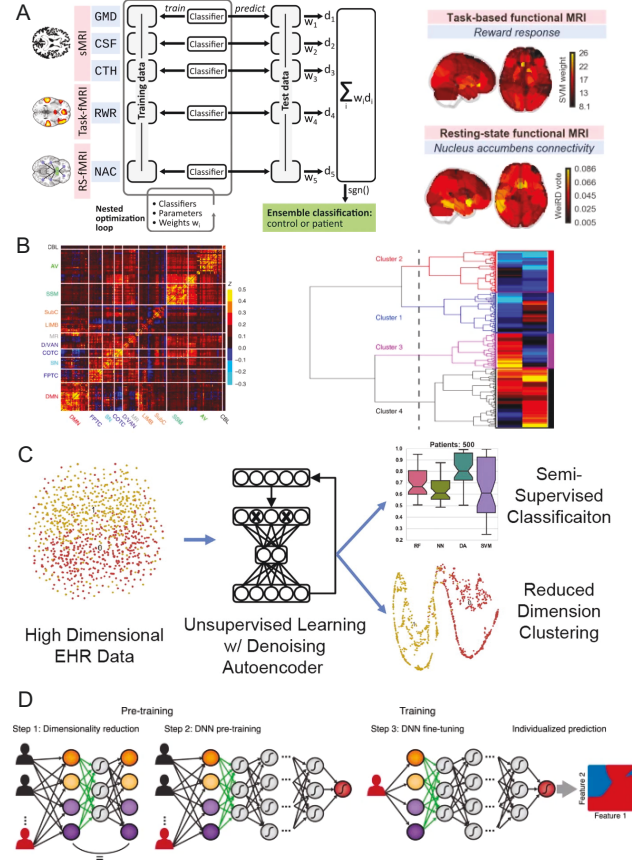


Figure 2: Various ML models for mental health applications. (A) Left: Multi-modal supervised classification scheme. Three modality-specific factors are optimized on the training data: classifier types, parameters and weights. The final diagnostic classification is based on a weighted sum of decision values, where weights correspond to those estimated during training. Right: Feature importance maps of functional neuroimaging modalities (¹³⁸). (B) Unsupervised learning. Left: whole-brain functional-connectivity matrix averaged across all subjects. z = Fischer transformed correlation coefficient. Right: Hierarchical clustering analysis. (© Springer Nature.⁴³ Figure reproduced with permission). (C) Semi-supervised learning pipeline for phenotype stratification based on EHRs (¹³⁹, 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 (²⁰, Creative Commons licenses 4.0).

the 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^{140,141,20,142}.

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, deep autoencoder is formed to discover more complicated and potentially nonlinear feature patterns. Deep autoencoder has been applied to extract low-dimensional features from the amplitude of low-frequency fluctuations in fMRI¹⁴³. Clustering analysis with the latent features uncovered by deep autoencoder further identified two sub-

types within major psychiatric disorders including schizophrenia, bipolar disorder, and MDD. A deep learning model was also designed based on a sparse stacked autoencoder and applied to lower the dimensionality of fMRI connectivity. The sparsity constraint used in this model yielded interpretable neural patterns for improved ASD diagnosis¹⁴⁴. 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 ASD¹⁴⁵. 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 neuroanatomical features from MRI data to identify brain dysfunction that can be attributed to ASD and not to other causes of individual variation¹⁴⁶.

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^{147,107}. Due to its strength in utilizing neighborhood information to learn hierarchies of features¹⁴⁸, 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 MDD⁷⁶. This model provided an end-to-end framework to successfully identify translational biomarkers from resting-state EEG in distinguishing depressive patients from healthy people. With whole-brain structure MRI, a 3D CNN model has also been designed to automatically extract multilayer high-dimensional features for the diagnosis of conduct disorder¹⁴⁹.

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 tasks¹⁵⁰. For instance, a framework based on graph convolutional networks (GCNs) has been designed for the diagnosis of ASD⁷⁸. 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 detection⁷⁰. 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 modeling⁶⁸. 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 feed-forward neural network, RNNs have the ability to learn features and long-term dependencies from sequential and time-series data. Long-short-term memory (LSTM) mod-

els are the most popular RNN and have shown advantages in capturing temporal dynamic information of neuroimaging data for various psychiatric disorder studies¹⁵¹. 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 schizophrenia⁷⁴. By combining RNNs with other deep neural networks, novel machine learning models have also been proposed to model the spatio-temporal dynamics in neuroimaging data. A spatio-temporal CNN model was proposed for 4D modeling of fMRI, with confirmed robustness in identifying key features in the default mode network¹⁵². 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 PTSD⁸⁹.

Generative Adversarial Networks (GANs). As one type of generative model, GANs have gained considerable attention in computer vision and natural language processing and also become increasingly popular in neuroimaging analysis¹⁰⁷. 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 training¹⁵³. Moreover, GANs have been used to impute missing values in multimodal datasets, a common problem in psychiatric studies, rather than discarding an entire multivariate data point¹⁵⁴. 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 multitask learning approach for fMRI-based classification, which resulted in an improved diagnosis of ASD compared with the standard LSTM¹⁵⁵. By integrating GANs with group ICA, a functional connectivity-based deep learning model was developed for the diagnosis of MDD and schizophrenia¹⁵⁶. 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 training⁵⁹. 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 applications²⁰.

3.5. 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 intelli-

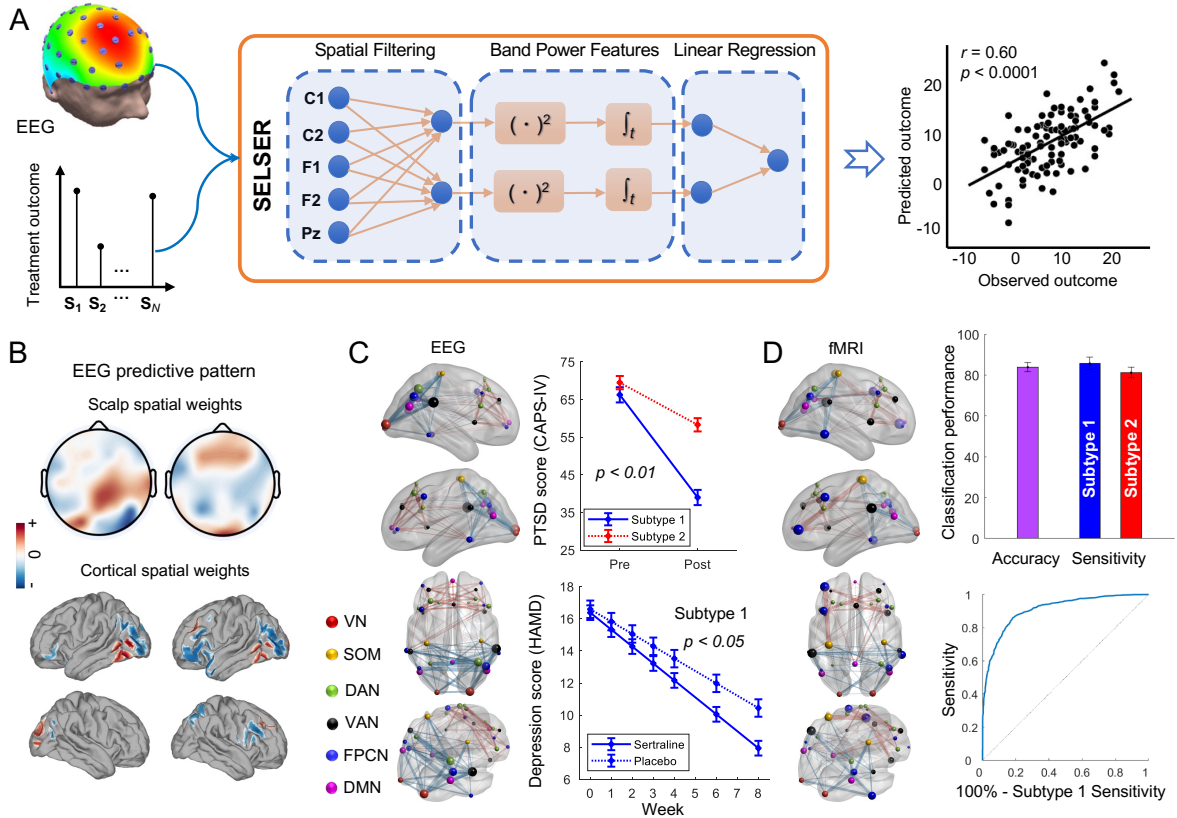


Figure 3: Concepts and major findings in case studies 1 and 2. (A) Illustration of the sparse EEG latent space regression (SELSER) framework in Case Study 1 for treatment outcome prediction. (B) Interpretable cortical pattern derived from the scalp pattern (© Springer Nature, figures are modified from⁹⁶ with permission). (C) Distinctive EEG connectivity profiles were identified by sparse K-means for defining psychiatric subtypes in Case Study 2 on PTSD and MDD. The two identified subtypes were further found to predict treatment responsiveness to psychotherapy and antidepressant medication. (D) The EEG connectivity-defined subtypes are distinguishable by rs-fMRI connectivity patterns derived from an RVM-based classifier (© Springer Nature, figures are modified from³⁸ with permission).

gence. In a recently published white paper “*Machine intelligence for healthcare*”, four important features are emphasized for ML systems¹⁵⁸. These concepts are broadly applicable to precision psychiatry¹³.

- **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 Section 7. The other two features will be discussed in Section 9.

3.6. 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. Antidepressants 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

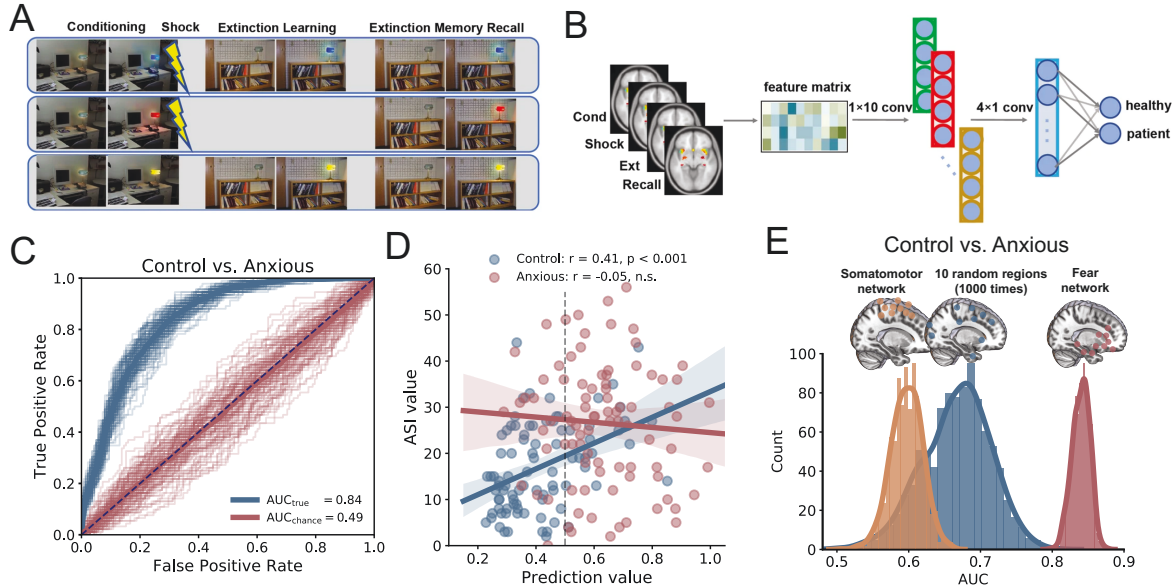


Figure 4: Illustrations of concepts and major findings in case study 3. (A) Experimental paradigm. (B) Schematic of the CNN. (C) AUC curves produced by CNN vs. chance level. (D) The prediction score positively correlated with the anxiety sensitivity index (ASI) for the control group ($r = 0.41$, $p < 0.001$), but at the chance level for anxious brains ($r = -0.05$, $p = 0.65$). (E) Distribution of AUCs based on brain activations within the 10-node fear randomly selected brain regions (Figures were adapted from ¹⁵⁷ with permission).

signature for an antidepressant-responsive phenotype that determines which patients will benefit from medications. To address the challenge, Wu et al.⁹⁶ 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 antidepressant-responsive EEG signature for MDD (Figure 3A). The identified signature accurately predicts antidepressant 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.³⁸ 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 ($n = 648$), 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 antidepressant 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 vs. non-anxious brains from fear extinction learning task-based fMRI. Using a neuroimaging cohort study ($n = 304$ adults, 92 anxiety patients, 74 trauma-exposed individuals, 138 matched controls), Wen et al.¹⁵⁷ 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 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

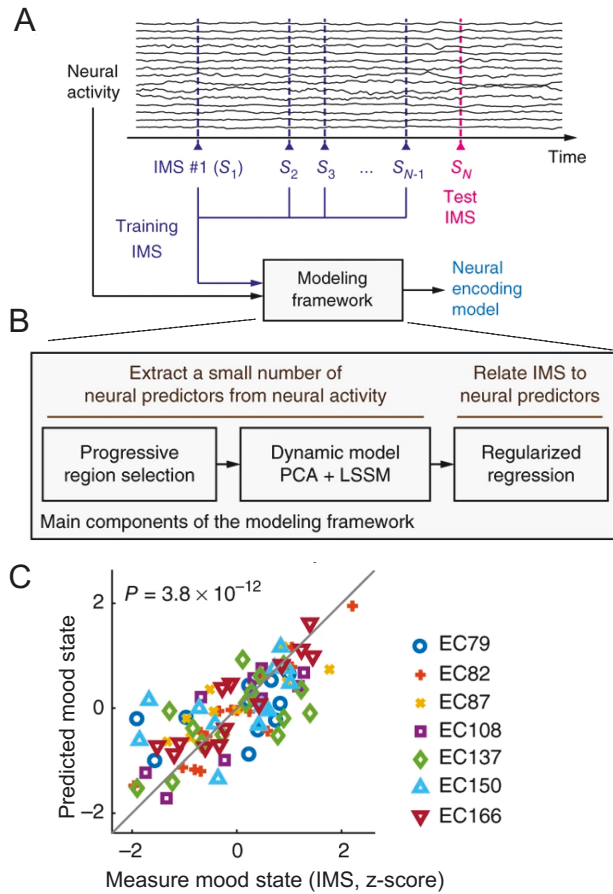


Figure 5: Illustrations of concepts and major findings in case study 4. (A) Schematic of cross-validation. An IMS point (e.g., S_N) is left out as the test IMS to be predicted. The other IMS points (i.e., training IMS, using S_1 to S_{N-1}) and the associated neural activity are used within the modeling 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 (© Springer Nature; Figures were modified from ¹⁵⁹ with permission).

multiple days in seven epilepsy patients, Sani et al. ¹⁵⁹ 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 $33 \pm 7.2\%$ 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.

4. 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 as compared to the pre-COVID baseline 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 ¹¹.

ML can be applied across all stages of a patient's journey ^{161,162}: risk assessment, diagnosis, prognosis, treatment, and relapse in a variety of disorders ¹⁶³, where the analysis can be applied to natural language, speech, facial expressions, body language, social media, as well as traditional clinical surveys and neuroimaging data ^{16,164}. 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 ⁷, 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 multi-task learning 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 ^{165,166}. During psychiatric evaluation, psychiatrists may try to build a mental model of what is going on in the patient's life in about 30 minutes. 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 build-

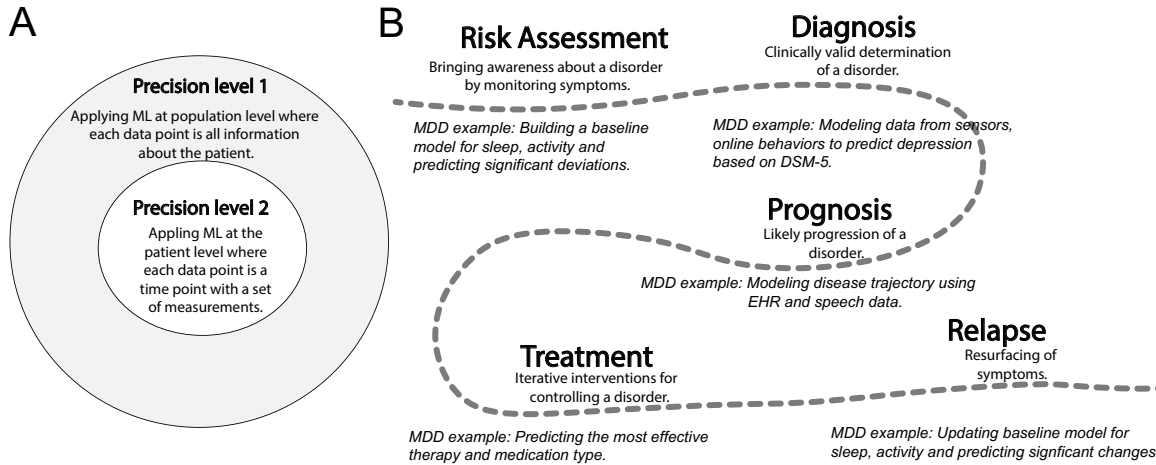


Figure 6: (A) Two levels of precision in applying ML for mental health. (B) Examples of ML applications at various stages of a patient's journey in case of major depressive disorder (MDD).

ing baseline models specific to each patient before their visit and present the bounds for various observations as a reference to psychiatrists during the exam¹⁶⁷. 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.

4.1. 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 ever¹⁸⁴. 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)
- Multimedia 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 multimedia, language, and social media data for development of their mental health applications.

4.2. 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 the diagnosis of mental illnesses¹⁸⁵. 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^{178,186}, including speech analysis for patients with depression, bipolar, and schizophrenia. Table 4 lists some commonly used acoustic features in the analysis of mental illnesses¹⁸⁷. 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 suicidality¹⁸⁸. Applications for depression include predicting the presence, severity, and score^{168,176}. These models use prosodic, spectral or other features computed from raw speech data to quantify flattened 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 studies¹⁸⁹, which detect manic and depressive speech from recordings of outgoing calls from phone conversations of consenting participants. Another challenge is the lack of

Table 3: Representative ML applications of multimedia data in mental disorders.

Study	Data source(s)	Patient journey stage	ML approach	Test sample size
Depression spectrum				
¹⁶⁸	Audio - clinical interviews	Diagnosis	CNN ensemble	47 speakers
¹⁶⁹	Audio - answers to personal questions	Diagnosis	Transfer learning	3078 speakers
¹⁷⁰	Audio - clinical interviews	Diagnosis	SVM w/ speech landmark features	47 speakers
¹⁷¹	Face video -reading & personal questions	Diagnosis	CNN	50 videos
¹⁷²	Gait-only video - casual walking in a corridor	Diagnosis	LSTM+CNN weighted fusion	40 videos
¹⁷³	Language - answers to personal questions	Diagnosis	LSTM fine-tuned w/ health forum data	2425 subjects
¹⁷⁴	Language - Facebook posts	Risk assessment	Logistic regression	68 patients
¹⁷⁵	Audio, video - clinical interviews	Diagnosis	Transformer + multimodal fusion	56 subjects
Bipolar spectrum				
¹⁷⁶	Audio - verbal fluency tasks	Relapse	SVM	56 subjects
¹⁷⁷	Sensor - GPS	Diagnosis	Linear regression	36 subjects
PTSD				
¹⁷⁸	Audio - clinical interviews	Diagnosis	Random forest	43 veterans
¹⁷⁹	Audio, video, skin conductance	Relapse	SVM	110 subjects
Schizophrenia spectrum				
¹⁸⁰	Audio - clinical interviews	Diagnosis	SVM	70 subjects
¹⁸¹	Video - neutral open-ended questions	Diagnosis	Logistic regression	16 subjects
¹⁸²	Language - internet search queries	Relapse	Random forest	23 subjects
¹⁸³	Audio, video - clinical interviews	Diagnosis	Gradient boosting	17 subjects

large labeled datasets for evaluating performance across various methods. To this end, it is noted that companies like Elipsis Health^{190,169}, 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 seconds of audio clip has been reported in some commercial applications to detect depression^{191,192}.

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 ideation¹⁹³, depression^{171,194,172}, schizophrenia¹⁸¹ and autism spectrum disorders¹⁹⁵. Features derived from overall facial expression, eyes, gait, and posture (Table 4) have been shown to be relevant across many mental disorders.

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 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 DNN¹⁹⁶, and some have used pre-training to compensate a relatively small sample size of depression datasets¹⁹⁷. 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 model¹⁹⁸. Meanwhile, utilization of predefined features has been most effective in providing interpretable results^{199,197}.

4.3. 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 profession-

als 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^{205,206}, but is also more abundantly available compared to 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 EHR 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.

Massive EHR datasets have provided opportunities to adapt ML approaches to track and identify target areas for quality improvement in mental health care. According to a 2015 national survey, 61.3% of US psychiatrists use EHRs²⁰⁷. The EHR language is at least one level abstracted from the patient's symptoms, consisting of clinical notes. However, the unique advantage of EHR data is the ease with which demographic and socioeconomic features can be combined with the language data. Symptoms derived from free texts in EHRs have been used for prediction of bipolar disorder²⁰¹, situational aggression²⁰⁸, and suicidal ideation²⁰⁹, with achieved comparable performance to clinicians. Furthermore, discharge summaries from EHRs have also been used to predict relapse²¹⁰. Aside from symptoms, a

Table 4: Multimodal data features and their uses in mental health.

Features	Example(s)	Relevant in mental disorder(s)	Study
Acoustic			
Source of sound features	Jitter	Increase with depression severity	188
Filtering features by vocal and nasal tracks	First resonant peak in the spectrum	increase with bipolar severity	176
Spectral features of speech	Mel frequency cepstral coefficients	A variety of disorders	186
Prosodic features of speech	Pause duration	Higher in schizophrenia	186
Video			
Facial	Smile duration, eyebrow movement, disgust expression	Increased disgust expression in suicidal ideation	193
Eyes	Gaze angle	More non-mutual gazes in MDD	194
Gait	Arm swing and stride	Reduced arm swing in MDD	172
Posture	Head pitch variance, upper body movements	Reduced head movement in schizophrenia	200
		Higher head movement in ASD	195
Language			
Grandiosity	Unrealistic sense of superiority	Increased in bipolar	201
Semantic coherence	Flow of meaning	Decreased in psychosis	202
Rumination	Repetitive thought patterns	Increased in MDD	203
Self-focus	Self-referent information	Increased in stress	204

variety of relevant mental health data (such as the intervention status and physical health comorbidities) can be routinely extracted from EHRs using NLP methods²¹¹. 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^{211,212}.

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 ideation²¹³, 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 pandemic²¹⁴. 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 the first episode psychosis, language data from clinical tests has higher performance compared to transcripts of free speech²⁰². 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^{173,203,190}.

4.4. Social Media

To date, social media companies have collected a large amount of language data which 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 time²¹⁵, but also be attributed on an individual level²¹⁶, 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. Rumination and sadness can be detected in such data up to 6 months prior to a clinical diagnosis¹⁷⁴. Models applied to Facebook and other social platforms (e.g., Twitter and Reddit) have been successful in predicting the diagnosis of psychosis, anorexia, anxiety, and stress levels^{204,217,218}. In addition to the language analysis for the user posts and comments,

ML models may also process media data such as Instagram images²¹⁹, or integrate images and text to infer the user’s mental state^{220,219}. 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^{221,222}. Furthermore, specialized language representations that were trained on mental health specific conversations and became publicly available²²³, have been shown to improve performance compared to 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 accuracy²²⁴.

While social media solves the scale issue with millions of samples available, most social media data lack clinically-valid labels²²⁵. 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.

4.5. 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 health²²⁶. 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

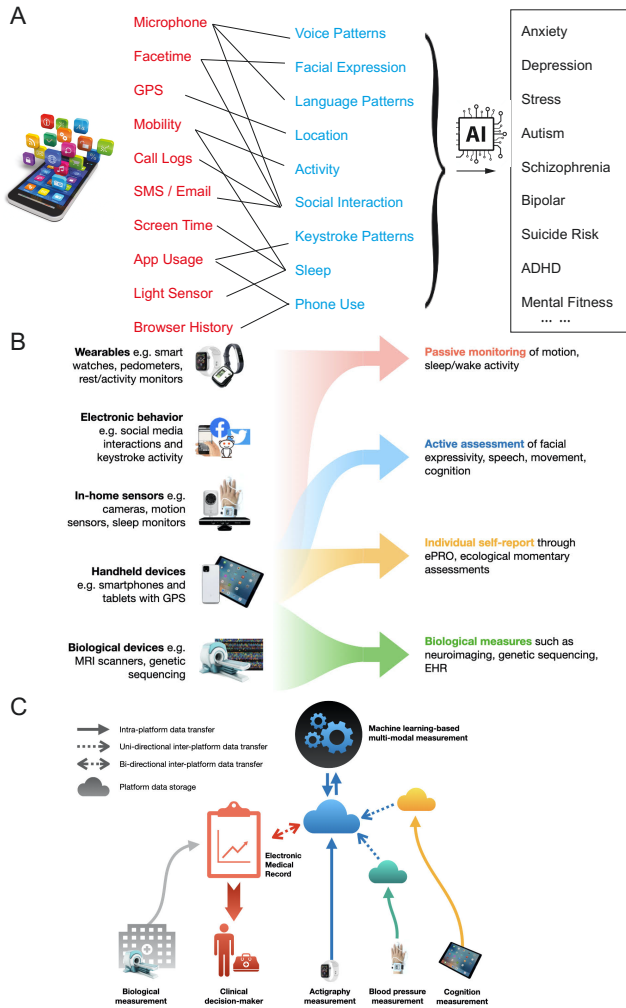


Figure 7: Illustrations of ML-powered technologies for mental health. (A) ML applications in mobile health. (B) Different types of data collection strategies for digital measurement tools. (C) A technological infrastructure for the integration of digital measurement tools. Independent platforms for measurement of health will have their own data repositories, depicted as clouds. This data could be safely transferred across platforms using transfer tools such as secure APIs (application program interfaces), 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 (panels B and C are reproduced from¹⁶⁴ with permission).

based on data collected from their wearable devices²³¹. 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 media interactions, GPS coordinates, and behavioral patterns (e.g., keyboard typing), have a great potential for monitoring, managing and predicting the individual's mental

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 ²⁴⁰	Language	Depression, Anxiety
Mindstrong ²²⁷	Keystrokes	Serious mental illness
Sonde Health ¹⁷⁰	Voice	Mental fitness
Ellipsis Health ¹⁶⁹	Voice	Stress
Amazon Halo ²⁴¹	Voice	Emotion detection
Apple Watch ¹⁹²	Mobility, Sleep	Depression
Alphabet Fitbit ²⁴²	Skin conductance	Stress
Kintsugi ²⁴³	Voice	Depression, Anxiety
Bewie ²³⁴	Raw data from smartphone	Multiple
MindDoc	Language (ask daily question)	Depression
Clarigent Health	Voice	Suicide risk

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^{236,237,238,239}.

4.6. Commercial and Research Platforms and Services

While studies have demonstrated promising results in ML applications for the individual's journey in mental health, the 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 to 20,000 smartphone apps that digitize mindfulness or cognitive behavioral therapy (CBT) techniques²⁴⁴, allowing the user to engage in psychotherapy at a greatly reduced price compared to in-person therapy. However, the quality is highly variable and the mechanisms used to validate them is 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^{10,245}. 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^{247,248,249}. 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 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 antidepressant medications are designed to slowly titrate serotonin levels, resulting in slow global effects over a 2-4 week period. Correspondingly, measures of depression based on the DSM, query about the presence of depressive states over a 2-week period. New classes of anti-depressant such as ketamine and psilocybin/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 (SSRI) and psilocybin/psilocybin, and ketamine, act on serotonin receptors that ultimately impact peripheral motor and physiological activity^{250,251}. Serotonin regulation will likely have a direct effect on depression symptoms such as psychomotor retardation, but the direct effect on feelings of guilt is minimal. As such, methods used to directly measure motor output 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 risk²⁵², SSRI response in MDD²⁵³, negative symptomatology in schizophrenia²⁰⁰, and Parkinsonian tremor²⁵⁴. 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

heart rate variability) 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 Health²²⁷, IesoTrigger Health²⁵⁵, and Headspace Health^{256,257}, 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 outcomes²⁵⁸; these can be further used to measure success of various therapy modalities, treatment design, as well as to improve care quality²⁵⁹.

Digital therapeutics proposes the use of mobile devices to offer CBT, mindfulness, or other validated psychotherapy in an automated fashion. These app-based approaches undergo the same clinical validation process as traditional medications, and are often developed in collaboration with large drug developers. Examples that are in development or have received FDA (Food and Drug Administration) approval include treatments for SUD, ADHD, schizophrenia, ASD, MDD, PTSD, and generalized anxiety disorder (GAD)²⁶⁰.

While digital therapeutics attempt to scale treatment, telehealth aims to scale the treatment provider network. Mental health treatment is an area where there are many effective treatments but little access to treatment providers²⁶¹. The issue of access became most acute during the emergence of COVID-19 when significant wide-spread mental health needs emerged along with greatly decreased access to care. To address this need, a large array of options have emerged, many reinforced by the emergency COVID-19 Telehealth Act of 2021 that enabled remote patient care²⁶². These services, which are directly accessible to consumers, or more often, provided by a third-party payer, provide access to professionals based on different mobile platforms including text, voice, and video communication. Services vary from mental health coaching provided by certified professionals to mental health assessment provided by psychiatrists and psychologists. While still in their infancy, early evidence has shown that telehealth services can perform at parity with traditional in-person therapy²⁶³.

Limitations. Although digital measurements have been rapidly adopted in mental health applications, most of current commercial AI applications for mental health have not focused on neuroimaging data. This partially because of several reasons. First, the data acquisition, quality, privacy and security issues have not been sufficiently or satisfactorily addressed (see Discussion section), creating a gap between commercial and laboratory settings. Second, the scalability of real-time neuroimaging technology (such as EEG) are not yet quite

ready. Third, the lack of an automatic and efficient (e.g., cloud-based) ML-powered neuroimaging data analysis pipeline also creates a barrier for brain-state monitoring. Future integration of portable EEG recordings and ML-powered platforms would expand the horizon of mobile mental health and teletherapy.

5. Multimodal 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 multimodal data analysis tasks is to explore the relationship between modalities, complementarity, shared versus modality-specific information and other mutual properties. Multimodal data fusion techniques present a framework to infer information how different data modalities interact and can be integrated for improved disease prediction^{264,265,34}. In this section, we will review several data fusion methods in diagnostic analytics (Section 5.1). We will focus on multimodal neuroimaging data (Section 5.2), and then extend the discussion to other modalities including vocal and visual expression data (Section 5.3).

5.1. Popular ML Methods for Multimodal Fusion

In the past decades, numerous research efforts have been dedicated to developing powerful ML methods for multimodal data fusion^{266,267,34,268,269}. Some commonly used approaches are summarized below.

Multivariate Correlation Analysis. Canonical 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 (aka. canonical variates) that have maximized correlation between two different data modalities²⁷⁰. For a more general setting, multi-set/multiway CCA (mCCA) has been developed as an extension of the standard CCA to multimodal fusion by maximizing the overall correlation among latent variables from more than two sets of modalities^{271,34}. Similar to CCA, partial least squares (PLS) and its extensions, i.e., multiway PLS (N-PLS), provide alternative approaches to integrate multimodal data by maximizing the covariance between latent variables from different modalities^{272,273}.

Matrix and Tensor Factorization. Based on matrix and tensor factorization techniques, joint blind source separation (BSS) approaches have been developed and successfully applied to multimodal fusion of biomedical data^{274,267}. As a typical example, joint independent component analysis (jICA) aims to maximize the independence among jointly estimated components from multiple modalities that are assumed to share the same mixing matrix²⁷⁵. The jICA approach involves concatenating modality features alongside each other and then performing ICA on the composite feature matrix²⁷⁶. 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^{267,266}. Coupled matrix and tensor factorization (CMTF) was also developed to simultaneously factorize multiple datasets in the form of matrices and high-orders tensors using tensor decomposition²⁷⁷, showing strength in capturing the potential multilinear structure for multimodal fusion. Besides extracting shared common components, some multimodal fusion tasks are also interested in deriving individual components that are modality-specific. Common and individual feature analysis (CIFA)²⁷⁸ and joint and individual variation explained (JIVE)²⁷⁹ 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 factorization²⁷⁴, which provides an efficient way to perform a multidimensional fusion of multiple data modalities.

Multi-Kernel Learning. Multi-kernel learning (MKL) has won many successful applications in multimodal data fusion due to the full utilization of multiple kernels that enable simultaneous learning from various modalities with heterogeneous data^{280,281}. 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 nonlinear function with specified forms of regularization. MKL may be designed under different ML models including SVM, Gaussian process, and clustering. Among them, MKL-SVM has been most popularly applied to integrate heterogeneous data modalities in studies of mental health^{282,283,284}.

Deep Learning-based Fusion. Cutting-edge deep learning techniques have become increasingly popular for deep multimodal fusion²⁸⁵. Data fusion through deep learning allows integration of multiple modalities based on learned high-level feature representations that are theoretically more comparable to each other and more informative for predicting the targets²⁸⁶. By exploiting cross-modal manifolds as a feature graph, a deep manifold-regularized learning model was recently designed to integrate transcriptomics and electrophysiology data from neuronal cells, and yield promising performance for phenotype prediction²⁸⁷. Graph neural networks (GNNs) show capability in information fusion for multimodal causability by defining causal links between features with graph structures, thereby enhancing the explainability of the derived multimodal feature representation²⁸⁸. By extending GNNs to multimodal structures, deep representation approaches have also been designed for integrating brain networks constructed from diverse modalities^{289,290,291}.

5.2. Multimodal Neuroimaging Studies

Neuroimaging data types are intrinsically dissimilar in nature, having different spatial and temporal resolutions²⁶⁸. Instead of feeding the entire data set 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

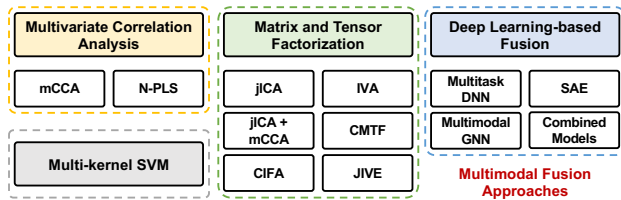


Figure 8: Summary of typical approaches for multimodal data fusion in psychiatry studies.

feature representations from multiple neuroimaging modalities for diagnosis has generally shown to improve performance compared to using a single modality alone²⁹². Multimodal fusion allows for integration of neuroimaging data modalities from different scales of spatial and temporal resolutions. Combining multimodal 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 disorder²⁹³. CMTF has been applied to identify diagnostic biomarkers of schizophrenia by integrating sMRI, fMRI, and EEG²⁹⁴. MKL-SVM has been successfully applied to integrate multimodal structural neuroimaging for predicting differential diagnosis between bipolar and unipolar depression²⁹⁵ and to combine sMRI and fMRI for improved classification of trauma survivors with and without PTSD²⁹⁶. More recently, it also showed efficacy in the diagnosis of early adolescent ADHD by integrating sMRI, fMRI, and DTI²⁹⁷. In learning low-dimensional representations of fMRI and sMRI²⁹⁸, 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 DTI²⁹⁹. A multimodal graph convolutional network (GCN) was designed to integrate functional and structural connectomics data for an improved prediction of phenotypic characterizations in ASD²⁸⁹. By combining multiple typical neural network structures, multimodal deep learning models have also been developed to effectively integrate fMRI connectivity and sMRI features³⁰⁰, and also genomic data³⁰¹ for discovering schizophrenia-associated brain dysfunction. Methods for learning joint representations from neuroimaging and non-neuroimaging data are still in early development³⁰², 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 signals³⁰³.

5.3. Multimodal Fusion of Non-imaging Data

Multimodal 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,

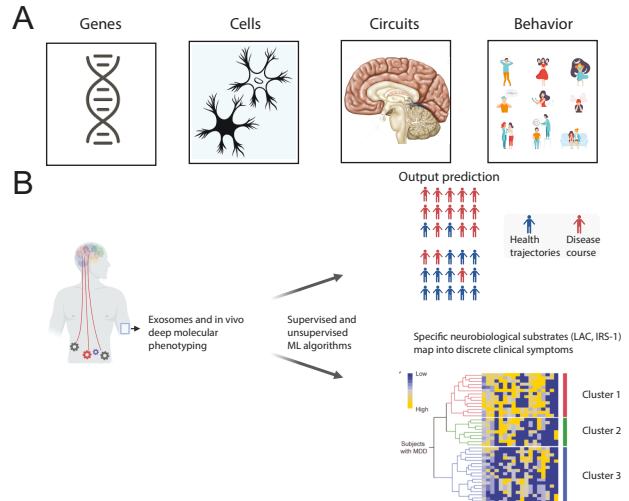


Figure 9: (A) Different levels of interacting variables (genes, cells, circuits) to behaviors in mental illnesses. (B) Combining ML with novel molecular biology technologies (for deep molecular phenotyping of brain plasticity) creates opportunities to develop new mechanistic models for prevention and treatment of clinical endophenotypes of mood and cognitive disorders.

social media, speech, video, and sensor data carry the most phenotypes for mental illnesses. Naturally, it makes sense to combine the information from these data sources. In addition, this enables modeling the inter-dependencies between these data which may not be observable by a human expert at the same time. For example, many features listed in Table 4 are known to be relevant in depression; however, observing them simultaneously can be very challenging for a clinician. MKL and MTL methods have been used to jointly learn from sensor and smartphone usage data to predict subjective well-being³⁰⁴. The success of transformer networks in jointly modeling video, speech, and language data has catalyzed multimodal modeling in mental health³⁰⁵. Multimodal modeling techniques can also be used in characterizing symptoms such as emotion dysregulation³⁰⁶, loneliness³⁰⁷ and sentiment analysis³⁰⁸. In a prognostic study, an SVM-based multimodal ML approach was developed to integrate clinical, neurocognitive, neuroimaging, and genetic information to predict psychosis in patients with clinical high-risk states⁸⁴. Deep autoencoder-based fusion approaches have been designed to integrate dynamics of facial and head movement and vocalization, and successfully applied to the prediction of depression severity³⁰⁹.

6. 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 subject 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^{310,311}. 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 MDD patients as compared to age and sex-matched controls, along with remarkable sex differences in these molecular pathways³¹². 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 cases³¹³. Importantly, these gene expression studies in humans were supported by findings in rodents showing a brain that continually changes with experience³¹⁴. Several studies based on RNA-sequencing 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 stress³¹⁵ but also a target for the responses to next-generation anti-depressants^{316,317}.

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 multimodal data integration^{318,319}. For instance, the software toolkit for single cell genomics, Seurat (<https://satijalab.org/seurat/>), combines unsupervised nonlinear dimensionality reduction, K-nearest neighbor graph analysis for cell-type clustering, and weighted nearest neighbor analysis for multimodal data integration³²⁰. Deep learning approaches, such as the deep autoencoder, provide analytic tools for denoising and dimensionality

reduction^{321,322,323}. 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 transfer this knowledge to cluster new datasets³²⁴. The combination of multimodal data generated from the simultaneous assessment of transcriptomic profiles with regulatory landscape or spatial location in the same single cell^{325,326,327,328} will allow a deeper molecular characterization of discrete cellular states³²⁰.

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 pathways³²⁹, involving the 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 MDD patients, recent work has showed that the specific neurobiological substrates map into discrete clinical symptoms, including anhedonia³³⁰. Furthermore, the integration of multidimensional 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 treatment³³¹. 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^{332,333}.

Another example of ML applications include the integration of multidimensional phenotypic measures to identify those mechanisms that predispose 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 collect more than 2000 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^{334,335}. 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 remain resilient. Additionally, integration of features can improve the classification sensitivity (~80%), which is better than the performance derived from either individual measure alone³³⁶.

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 tar-

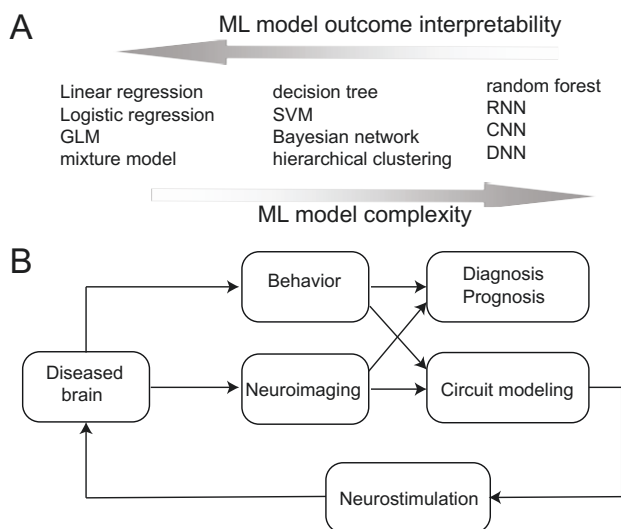


Figure 10: (A) A wide spectrum of interpretability in representative ML models. (B) Schematic of the closed loop of neuroimaging/modeling/neurostimulation.

gets 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.

7. Explainable AI in Psychiatry

Explainable Artificial Intelligence (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^{337,338}. 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^{339,340}. XAI distinguishes the standard AI in two important ways: (i) promote transparency, interpretability, and generalizability; (ii) transform 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 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 methods to fit models with specific interpretations.

7.1. 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 (Figure 10A). Post-hoc interpretability is referred to the application of interpretation methods after model training³⁴¹ (online resource: <https://christophm.github.io/interpretable-ml-book/>). Interpretation may appear in different forms: (i) finite feature summary statistics, (ii) meaningful model parameters, (iii) easy visualization of the model outcome (e.g., feature summary or decision boundary). Interpretability and explainability are two similar concepts and 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 explainability³⁴¹: 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, co-dependence of input variables may make explanations ambiguous; latent variables of probabilistic generative models may face the problem of “explaining away”³⁴². 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 data³⁴³. 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 model³⁴⁴, or constrained mixture models used for clustering³⁴⁵. 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^{346,347}. 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.

7.2. 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 improve understanding, prediction and treatment of mental illness³⁴⁸. The levels of computation range from short to long timescales (minutes/hours/days/weeks) and focus on the changes in brain activity and behavior. The types of computation vary from circuit-level modeling, data-driven analytics, to theory-driven algorithmic development. Two complementary approaches have been proposed in computational psychiatry: (i) data-driven approaches, which apply ML methods to high-dimensional multimodal data to tackle classification and prediction problems (Section 3); (ii) 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.^{349,350,351} 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 populations³⁵². Furthermore, individual brain nodes are coupled together according to the empirically-derived anatomical network³⁵³. 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 resting-state fMRI^{354,355}, 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 induced by neurostimulation. These changes are often, but not always, associated with changes in functional activation of specific brain regions.

8. 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^{356,357}.

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 correlational³⁵⁸. Closed-loop experimental design would help to test the potential causality³⁵⁶. 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 gap³⁵⁹. 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^{358,356}. 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 the brain-behavior mapping³⁶⁰.

Neuromodulation therapies have become increasingly popular in treating various neuropsychiatric and neurological disorders. Temporally precise neurostimulation tools provide a plausible means to perturb or stimulate the brain. Clinically used neuromodulation methods include invasive deep brain stimulation (DBS), noninvasive transcranial magnetic stimulation (TMS), noninvasive 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^{361,362,363}. To date, repetitive TMS (rTMS) has been cleared by the FDA for the treatment of depression and recently used in the studies of neural functioning and behavior^{364,365}. 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 action³⁶⁶, but how to identify the optimal target to achieve effective treatment outcomes remains unexplored. For WHEN and HOW questions, in comparison to open-loop stimulation, closed-loop stimulation can deliver temporally precise stimulation triggered by detected features, symptoms or the user demand. Traditional neurostimulation strategies are designed in an on/off stimulation fashion, where the stimulation is determined by preselected parameters. However, these stimulation parameters may not be optimal. To accommodate an adaptive subject-specific stimula-

tion 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^{367,368,369}. 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 neurostimulation³⁷⁰. 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^{369,371}.

Neurostimulation can not only induce changes in behavior, but 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^{372,373}. 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")³⁷⁴. 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: (i) Can a target node stimulation rewire brain connectivity in evoked and steady-state conditions? (ii) Can the neurostimulation-induced change of evoked or resting-state brain connectivity distinguish a pathological from a healthy brain? (iii) 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 treatment³⁷⁵? (iv) 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 psychiatry³⁷⁶.

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 towards better understanding of brain-behavior causation as well as individualized precision treatment in psychiatry. For instance, looping neuromodulation and DCM may provide a way to test the impact of neurostimulation on neural plasticity that underlies the change in adaptive or maladaptive behaviors.

9. Discussion and Conclusion

9.1. 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^{166,377}, 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., autism spectrum disorder, 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 prone³⁷⁸. 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 pretrained language model²²³. 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 which can differ from clinical criteria, leading to inaccurately labeled samples. Furthermore, collection of properly matched control samples remain difficult, given the diversity of age, gender, race, education, family history and life style among users. Furthermore, there are also intrinsic dataset biases during data collection³⁷⁹.

Data privacy and security. Advances in sensing technologies enable us to collect a large amount of personal data, including the 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 rule and regulation (e.g., internal review board) to assure the ethical use, social media data are collected in 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 to 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 journey³⁸⁰. In addition to biological underpinnings, the domains of data (such as language) also represent social underpinnings³⁸¹, 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 method³⁴⁰, 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 a "explainability-performance" trade-off. For instance, deep learning may outperform traditional methods at the cost of reduced interpretability. Despite the rapid progress in development of XAI and interpretable ML techniques, the challenge of ML explainability still remains 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 speak to the underlying causes and nor have limited implications on treating the causes. ML-based causal inference methods³⁸² may help with "precision" treatment design³⁸³. Recent developments in ultra-high-field neuroimaging with sufficient temporal and spatial resolution³⁸⁴ 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 application³⁸⁵ to the mentally ill patients³⁸⁶. 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 practice³⁸⁷. Additionally, considering a possible disparity of interests between the clinical and ML communities³⁷⁹, 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 treatment¹⁰. 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 GPT-3 (Generative Pre-trained Transformer 3), which should be fully considered before deployment in clinical settings³⁸⁸.

9.2. 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/>), which advocates using good “small” data instead of simply collecting from big but possibly noisy data. The good quality criteria include: (i) consistency; (ii) coverage of important cases; (iii) 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 performance³⁸⁹. 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 confidence³⁹⁰. 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 instances³⁹¹. 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^{392,393,394}. 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 machine learning (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 training³⁹⁵. This has become increasingly important since the volume of social media and multimedia data streams is so overwhelming that even a small effort of human involvement would make the task prohibitive.

Data integration approach. Integration of multimodal data is critical for psychiatric diagnostics and monitoring. Therefore, it is urgently needed to develop weakly supervised, interpretable, multimodal deep learning pipelines to fuse histopathology, genomics, neuroimaging, and behavioral data, as well as to develop multimodal fusion algorithms for speech, video, and EHRs to assist both psychiatrists and patients. Because of the nature of multimodality, not all data can be quan-

tified in the Euclidean space. Graph and geometric deep learning may play a role in this research direction^{396,397}. 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 applications³⁹⁸.

9.3. 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 multimodal 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 subtypes of symptoms and cognitive deficits, and develop model-based phenotyping³⁹⁹. 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 government), and end users (physicians, nurses, patients, friends and family)⁴⁰⁰. 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 need and specific pathology.

Acknowledgments

The research was partially supported from the US National Science Foundation (CBET-1835000 to Z.S.C.), the National Institutes of Health (R01-NS121776 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.

Appendix: List of Acronyms

ADHD Attention-Deficit/Hyperactivity Disorder.

AI Artificial Intelligence.

APA American Psychiatric Association.

ASD Autism Spectrum Disorder.

ASI Anxiety Sensitivity Index.

BMI Brain Machine Interface.

BSS Blind Source Separation.

CBT Cognitive Behavioral Therapy.

CCA Canonical Correlation Analysis.

CDC Centers for Disease Control and Prevention.

CIFA Common and Individual Feature Analysis.

CMTF Coupled Matrix and Tensor Factorization.

CNN Convolutional Neural Network.

CNS Central Nervous System.

DBS Deep Brain Stimulation.

DCM Dynamic Causal Modeling.

dMRI Diffusion Magnetic Resonance Imaging.

DSM Diagnostic and Statistical Manual of Mental Disorders.

EAML Expert-Augmented Machine Learning.

ECoG Electroocorticography.

EDA Electrodermal Activity.

EEG Electroencephalography.

EHR Electronic Health Record.

FDA Food and Drug Administration.

fMRI functional Magnetic Resonance Imaging.

fNIRS functional Near-Infrared Spectroscopy.

GAD Generalized Anxiety Disorder.

GAN Generative Adversarial Network.

GCN Graph Convolutional Network.

GNN Graph Neural Network.

GP Gaussian Process.

GPT-3 Generative Pre-trained Transformer 3.

HIPAA Health Insurance Portability and Accountability Act.

HRV Heart Rate Variability.

ICA Independent Component Analysis.

ICD International Classification of Diseases.

IMS Immediate Mood Scaler.

IVA Independent Vector Analysis.

JICA Joint Independent Component Analysis.

JIVE Joint and Individual Variance Explained.

LCA Latent Class Analysis.

LGMM Latent Growth Mixture Modeling.

LIME Local Interpretable Model-agnostic Explanations.

LSTM Long Short-Term Memory.

LTA Latent Transition Analysis.

MDD Major Depressive Disorder.

MEG Magnetoencephalography.

MKL Multi-Kernel Learning.

ML Machine Learning.

MRI Magnetic Resonance Imaging.

MTL Multi-Task Learning.

NGS Next-Generation Sequencing.

NIMH National Institute of Mental Health.

NLP Natural Language Processing.

NPC Non-Player Characters.

PCA Principal Component Analysis.

PET Positron Emission Tomography.

PFC Prefrontal Cortex.

PHQ Patient Health Questionnaire.

PII Personal Identifiable Information.

PLS Partial Least Squares.

PTSD Post-Traumatic Stress Disorder.

RDoC Reserach Domain Criteria.

RNN Recurrent Neural Network.

rTMS Repetitive Transcranial Magnetic Stimulation.

RVM Relevance Vector Machine.

SAE Stacked AutoEncoder.

SDS Social Defeat Stress.

SELSER Sparse EEG Latent Space Regression.

SSM State Space Model.

SSRI Serotonin Reuptake Inhibitors.

SUD Substance Use Disorder.

SVM Support Vector Machine.

SVR Support Vector Regression.

t-SNE t-distributed Stochastic Neighbor Embedding.

tACS Transcranial Alternating Current Stimulation.

tDCS Transcranial Direct Current Stimulation.

tFUS Transcranial Focused Ultrasound Stimulation.

TMS Transcranial Magnetic Stimulation.

XAI EXplainable Artificial Intelligence.

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.É., Lane, R.I., Petrosky, E., Wiley, J.F., Christensen, A., Njai, R., et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic—United States, June 24–30, 2020. *Morbidity and Mortality Weekly Report* 2020;69(32):1049.
- [2] Greenberg, P.E., Fournier, A.A., Sisitsky, T., Simes, M., Berman, R., Koenigsberg, S.H., et al. The economic burden of adults with major depressive disorder in the United States (2010 and 2018). *Pharmacoeconomics* 2021;39(6):653–665.
- [3] Insel, T.R., Cuthbert, B.N.. Brain disorders? precisely. *Science* 2015;348(6234):499–500.
- [4] Fernandes, B.S., Williams, L.M., Steiner, J., Leboyer, M., Carvalho, A.F., Berk, M.. The new field of ‘precision psychiatry’. *BMC Medicine* 2017;15(1):1–7.
- [5] Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. 2010.
- [6] Insel, T.R.. The nimh research domain criteria (RDoC) project: precision medicine for psychiatry. *American Journal of Psychiatry* 2014;171(4):395–397.
- [7] Bzdok, D., Meyer-Lindenberg, A.. Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2018;3(3):223–230.
- [8] Zhou, Z., Wu, T.C., Wang, B., Wang, H., Tu, X.M., Feng, C.. Machine learning methods in psychiatry: a brief introduction. *General Psychiatry* 2020;33(1):e100171.
- [9] Allen, M., Salmon, A.. Synthesising artificial patient-level data for open science—an evaluation of five methods. *medRxiv preprint* 2020;URL: <https://doi.org/10.1101/2020.10.09.20210138>.
- [10] Burr, C., Morley, J., Taddeo, M., Floridi, L.. Digital psychiatry: Risks and opportunities for public health and wellbeing. *IEEE Transactions on Technology and Society* 2020;1(1):21–33.
- [11] Doraiswamy, P.M., Blease, C., Bodner, K.. Artificial intelligence and the future of psychiatry: Insights from a global physician survey. *Artificial Intelligence in Medicine* 2020;102:101753.
- [12] Rutledge, R.B., Chekroud, A.M., Huys, Q.J.. Machine learning and big data in psychiatry: toward clinical applications. *Current Opinion in Neurobiology* 2019;55:152–159.
- [13] Chandler, C., Foltz, P.W., Elvevåg, B.. Using machine learning in psychiatry: the need to establish a framework that nurtures trustworthiness. *Schizophrenia Bulletin* 2020;46(1):11–14.
- [14] Galatzer-Levy, I.R., Ruggles, K.V., Chen, Z.. Data science in the research domain criteria era: relevance of machine learning to the study of stress pathology, recovery, and resilience. *Chronic Stress* 2018;2:1–14.
- [15] Shatte, A.B., Hutchinson, D.M., Teague, S.J.. Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine* 2019;49(9):1426–1448.
- [16] Su, C., Xu, Z., Pathak, J., Wang, F.. Deep learning in mental health outcome research: a scoping review. *Translational Psychiatry* 2020;10(1):1–26.
- [17] Liu, G.D., Li, Y.C., Zhang, W., Zhang, L.. A brief review of artificial intelligence applications and algorithms for psychiatric disorders. *Engineering* 2020;6(4):462–467.
- [18] Thieme, A., Belgrave, D., Doherty, G.. Machine learning in mental health: A systematic review of the hci literature to support the development of effective and implementable ml systems. *ACM Transactions on Computer-Human Interaction (TOCHI)* 2020;27(5):1–53.
- [19] Durstewitz, D., Koppe, G., Meyer-Lindenberg, A.. Deep neural networks in psychiatry. *Molecular Psychiatry* 2019;24(11):1583–1598.
- [20] Koppe, G., Meyer-Lindenberg, A., Durstewitz, D.. Deep learning for small and big data in psychiatry. *Neuropsychopharmacology* 2021;46(1):176–190.
- [21] Hedderich, D.M., Eickhoff, S.B.. Machine learning for psychiatry: getting doctors at the black box? *Molecular Psychiatry* 2021;26:23–25.
- [22] Bracher-Smith, M., Crawford, K., Escott-Price, V.. Machine learning for genetic prediction of psychiatric disorders: a systematic review. *Molecular Psychiatry* 2021;26:70–79.
- [23] Allen, S.. Artificial intelligence and the future of psychiatry. *IEEE Pulse* 2020;11(3):2–6.
- [24] Thukral, A., Ershad, F., Enan, N., Rao, Z., Yu, C.. Soft ultrathin silicon electronics for soft neural interfaces: A review of recent advances of soft neural interfaces based on ultrathin silicon. *IEEE Nanotechnology Magazine* 2018;12(1):21–34.
- [25] Etkin, A.. A reckoning and research agenda for neuroimaging in psychiatry. *American Journal of Psychiatry* 2019;176(7):507–511.
- [26] Noda, Y., Barr, M.S., ElSalhy, M., Masuda, F., Tarumi, R., Ogyu, K., et al. Neural correlates of delay discount alterations in addiction and psychiatric disorders: a systematic review of magnetic resonance imaging studies. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 2020;99:109822.
- [27] Noggle, C.A., Davis, A.S.. Advances in neuroimaging. In: *Understanding the Biological Basis of Behavior*. Springer; 2021, p. 107–137.
- [28] Keren, H., O’Callaghan, G., Vidal-Ribas, P., Buzzell, G.A., Brotman, M.A., Leibenluft, E., et al. Reward processing in depression: a conceptual and meta-analytic review across fMRI and EEG studies. *American Journal of Psychiatry* 2018;175(11):1111–1120.
- [29] Lukow, P., Kienes, A., Kempton, M., Turkheimer, F., McGuire, P., Modinos, G.. Neural correlates of emotional processing in psychosis risk and onset—a systematic review and meta-analysis of fMRI studies. *Neuroscience & Biobehavioral Reviews* 2021;128:780–788.
- [30] Smith, S.M., Vidaurre, D., Beckmann, C.F., Glasser, M.F., Jenkinson, M., Miller, K.L., et al. Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences* 2013;17(12):666–682.
- [31] Woodward, N.D., Cascio, C.J.. Resting-state functional connectivity in psychiatric disorders. *JAMA Psychiatry* 2015;72(8):743–744.
- [32] Ma, S., Calhoun, V.D., Phlypo, R., Adali, T.. Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis. *NeuroImage* 2014;90:196–206.

- [33] Rolls, E.T., Cheng, W., Feng, J.. Brain dynamics: the temporal variability of connectivity, and differences in schizophrenia and ADHD. *Translational Psychiatry* 2021;11(1):1–11.
- [34] Calhoun, V.D., Sui, J.. Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2016;1(3):230–244.
- [35] Zheng, A., Casari, A.. *Feature Engineering for Machine Learning*. O'Reilly Media, Inc; 2018.
- [36] Hipp, J.F., Hawellek, D.J., Corbetta, M., Siegel, M., Engel, A.K.. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience* 2012;15(6):884–890.
- [37] Siems, M., Pape, A.A., Hipp, J.F., Siegel, M.. Measuring the cortical correlation structure of spontaneous oscillatory activity with eeg and meg. *NeuroImage* 2016;129:345–355.
- [38] Zhang, Y., Wu, W., Toll, R.T., Naparstek, S., Maron-Katz, A., Watts, M., et al. Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nature Biomedical Engineering* 2021;5(4):309–323.
- [39] Feczko, E., Miranda-Dominguez, O., Marr, M., Graham, A.M., Nigg, J.T., Fair, D.A.. The heterogeneity problem: approaches to identify psychiatric subtypes. *Trends in Cognitive Sciences* 2019;23(7):584–601.
- [40] Satterthwaite, T.D., Feczko, E., Kaczkurkin, A.N., Fair, D.A.. Parsing psychiatric heterogeneity through common and unique circuit-level deficits. *Biological Psychiatry* 2020;88(1):4.
- [41] Wang, D., Buckner, R.L., Fox, M.D., Holt, D.J., Holmes, A.J., Stoerklein, S., et al. Parcellating cortical functional networks in individuals. *Nature Neuroscience* 2015;18(12):1853–1860.
- [42] Walter, M., Alizadeh, S., Jamalabadi, H., Lucken, U., Dannlowski, U., Walter, H., et al. Translational machine learning for psychiatric neuroimaging. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 2019;91:113–121.
- [43] Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine* 2017;23(1):28–38.
- [44] Marquand, A.F., Kia, S.M., Zabihi, M., Wolfers, T., Buitelaar, J.K., Beckmann, C.F.. Conceptualizing mental disorders as deviations from normative functioning. *Molecular Psychiatry* 2019;24(10):1415–1424.
- [45] Feczko, E., Balba, N., Miranda-Dominguez, O., Cordova, M., Karalunas, S., Irwin, L., et al. Subtyping cognitive profiles in autism spectrum disorder using a functional random forest algorithm. *NeuroImage* 2018;172:674–688.
- [46] Sargent, K., Chavez-Baldini, U., Master, S.L., Verweij, K.J., Lok, A., Sutherland, A.L., et al. Resting-state brain oscillations predict cognitive function in psychiatric disorders: A transdiagnostic machine learning approach. *NeuroImage: Clinical* 2021;30:102617.
- [47] Barch, D.M.. The neural correlates of transdiagnostic dimensions of psychopathology. *American Journal of Psychiatry* 2017;174(7):613–615.
- [48] Xia, C.H., Ma, Z., Ciric, R., Gu, S., Betzel, R.F., Kaczkurkin, A.N., et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nature Communications* 2018;9(1):1–14.
- [49] Kebets, V., Holmes, A.J., Orban, C., Tang, S., Li, J., Sun, N., et al. Somatosensory-motor dysconnectivity spans multiple transdiagnostic dimensions of psychopathology. *Biological Psychiatry* 2019;86(10):779–791.
- [50] McTeague, L.M., Rosenberg, B.M., Lopez, J.W., Carreon, D.M., Huemer, J., Jiang, Y., et al. Identification of common neural circuit disruptions in emotional processing across psychiatric disorders. *American Journal of Psychiatry* 2020;177(5):411–421.
- [51] Wachinger, C., Nho, K., Saykin, A.J., Reuter, M., Rieckmann, A., Initiative, A.D.N., et al. A longitudinal imaging genetics study of neuroanatomical asymmetry in alzheimer's disease. *Biological Psychiatry* 2018;84(7):522–530.
- [52] Vidal-Ribas, P., Benson, B., Vitale, A.D., Keren, H., Harrewijn, A., Fox, N.A., et al. Bidirectional associations between stress and reward processing in children and adolescents: a longitudinal neuroimaging study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2019;4(10):893–901.
- [53] Roekner, A.R., Oliver, K.I., Lebois, L.A., van Rooij, S.J., Stevens, J.S.. Neural contributors to trauma resilience: a review of longitudinal neuroimaging studies. *Translational Psychiatry* 2021;11(1):1–17.
- [54] Smith, S.M., Nichols, T.E.. Statistical challenges in “big data” human neuroimaging. *Neuron* 2018;97(2):263–268.
- [55] Duncan, N.W., Northoff, G.. Overview of potential procedural and participant-related confounds for neuroimaging of the resting state. *Journal of Psychiatry and Neuroscience* 2013;38(2):84–96.
- [56] Blume, W.T.. Drug effects on eeg. *Journal of Clinical Neurophysiology* 2006;23(4):306–311.
- [57] Linke, A.C., Olson, L., Gao, Y., Fishman, I., Müller, R.A.. Psychotropic medication use in autism spectrum disorders may affect functional brain connectivity. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2017;2(6):518–527.
- [58] Yu, M., Linn, K.A., Cook, P.A., Phillips, M.L., McInnis, M., Fava, M., et al. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fmri data. *Human Brain Mapping* 2018;39(11):4213–4227.
- [59] Zhao, Q., Adeli, E., Pohl, K.M.. Training confounder-free deep learning models for medical applications. *Nature Communications* 2020;11(1):1–9.
- [60] Nielsen, A.N., Barch, D.M., Petersen, S.E., Schlaggar, B.L., Greene, D.J.. Machine learning with neuroimaging: Evaluating its applications in psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2020;5(8):791–798.
- [61] Pelin, H., Ising, M., Stein, F.e.a.. Identification of transdiagnostic psychiatric disorder subtypes using unsupervised learning. *Neuropsychopharmacology* 2021;46:1895–1905.
- [62] Janssen, R.J., Mourão-Miranda, J., Schnack, H.G.. Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2018;3(9):798–808.
- [63] Gao, S., Calhoun, V.D., Sui, J.. Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neuroscience & Therapeutics* 2018;24(11):1037–1052.
- [64] Cearn, M., Opel, N., Clark, S., Kaehler, C., Thalamuthu, A., Heindel, W., et al. Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. *Translational Psychiatry* 2019;9(1):1–9.
- [65] Gueorgieva, R., Chekroud, A.M., Krystal, J.H.. Trajectories of relapse in randomised, placebo-controlled trials of treatment discontinuation in major depressive disorder: an individual patient-level data meta-analysis. *The Lancet Psychiatry* 2017;4(3):230–237.
- [66] Murphy, K.P.. *Machine Learning: A Probabilistic Perspective*. MIT Press; 2012.
- [67] Emmert-Streib, F., Dehmer, M.. Taxonomy of machine learning paradigms: A data-centric perspective. *WIREs Data Mining and Knowledge Discovery* 2022;2022:e1470.
- [68] Zhao, K., Duka, B., Xie, H., Oathes, D.J., Calhoun, V., Zhang, Y.. A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD. *NeuroImage* 2022;246:118774.
- [69] Luo, Y., Alvarez, T.L., Halperin, J.M., Li, X.. Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. *NeuroImage: Clinical* 2020;26:102238.
- [70] Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., et al. Braingnn: Interpretable brain graph neural network for fMRI analysis. *Medical Image Analysis* 2021;74:102233.
- [71] Khosla, M., Jamison, K., Kuceyeski, A., Sabuncu, M.R.. Ensemble learning with 3d convolutional neural networks for functional connectome-based prediction. *NeuroImage* 2019;199:651–662.
- [72] Wu, M.J., Mwangi, B., Bauer, I.E., Passos, I.C., Sanches, M., Zunta-Soares, G.B., et al. Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. *NeuroImage* 2017;145:254–264.
- [73] Zhu, H., Yuan, M., Qiu, C., Ren, Z., Li, Y., Wang, J., et al. Multivariate classification of earthquake survivors with post-traumatic stress disorder based on large-scale brain networks. *Acta Psychiatrica Scandinavica* 2020;141(3):285–298.
- [74] Yan, W., Calhoun, V., Song, M., Cui, Y., Yan, H., Liu, S., et al. Discriminating schizophrenia using recurrent neural network applied on

- time courses of multi-site fMRI data. *EBioMedicine* 2019;47:543–552.
- [75] Mikolas, P., Hlinka, J., Skoch, A., Pitra, Z., Frodl, T., Spaniel, F., et al. Machine learning classification of first-episode schizophrenia spectrum disorders and controls using whole brain white matter fractional anisotropy. *BMC Psychiatry* 2018;18(1):1–7.
 - [76] Uyulan, C., Ergüzel, T.T., Unbol, H., Cebi, M., Sayar, G.H., Nezhad Asad, M., et al. Major depressive disorder classification based on different convolutional neural network models: Deep learning approach. *Clinical EEG and Neuroscience* 2021;52(1):38–51.
 - [77] Heinsfeld, A.S., Franco, A.R., Craddock, R.C., Buchweitz, A., Meneguzzi, F. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: Clinical* 2018;17:16–23.
 - [78] Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., et al. Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease. *Medical Image Analysis* 2018;48:117–130.
 - [79] Maron-Katz, A., Zhang, Y., Narayan, M., Wu, W., Toll, R.T., Narpstek, S., et al. Individual patterns of abnormality in resting-state functional connectivity reveal two data-driven PTSD subgroups. *American Journal of Psychiatry* 2020;177(3):244–253.
 - [80] Lecei, A., van Hulst, B.M., de Zeeuw, P., van der Pluijm, M., Rijks, Y., Durston, S. 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: Clinical* 2019;21:101601.
 - [81] Parkes, L., Moore, T.M., Calkins, M.E., Cook, P.A., Cieslak, M., Roalf, D.R., et al. Transdiagnostic dimensions of psychopathology explain individuals' unique deviations from normative neurodevelopment in brain structure. *Translational Psychiatry* 2021;11(1):1–13.
 - [82] Schmaal, L., Marquand, A.F., Rhebergen, D., van Tol, M.J., Ruhé, H.G., van der Wee, N.J., et al. Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: a multivariate pattern recognition study. *Biological Psychiatry* 2015;78(4):278–286.
 - [83] Ramyeed, A., Studerus, E., Kometer, M., Uttinger, M., Gschwandtner, U., Fuhr, P., et al. Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naïve at-risk patients. *The World Journal of Biological Psychiatry* 2016;17(4):285–295.
 - [84] Koutsouleris, N., Dwyer, D.B., Degenhardt, F., Maj, C., Urquijo-Castro, M.F., Sanfelici, R., et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry* 2021;78(2):195–209.
 - [85] Sheynin, S., Wolf, L., Ben-Zion, Z., Sheynin, J., Reznik, S., Keynan, J.N., et al. Deep learning model of fMRI connectivity predicts PTSD symptom trajectories in recent trauma survivors. *NeuroImage* 2021;238:118242.
 - [86] Nieuwenhuis, M., Schnack, H.G., van Haren, N.E., Lappin, J., Morgan, C., Reinders, A.A., et al. Multi-center MRI prediction models: Predicting sex and illness course in first episode psychosis patients. *NeuroImage* 2017;145:246–253.
 - [87] Smucny, J., Davidson, I., Carter, C.S. Comparing machine and deep learning-based algorithms for prediction of clinical improvement in psychosis with functional magnetic resonance imaging. *Human Brain Mapping* 2021;42(4):1197–1205.
 - [88] Bertocci, M.A., Bebko, G., Versace, A., Iyengar, S., Bonar, L., Forbes, E.E., et al. Reward-related neural activity and structure predict future substance use in dysregulated youth. *Psychological Medicine* 2017;47(8):1357–1369.
 - [89] Zhang, J., Wong, S.M., Richardson, J.D., Jetly, R., Dunkley, B.T. Predicting PTSD severity using longitudinal magnetoencephalography with a multi-step learning framework. *Journal of Neural Engineering* 2020;17(6):066013.
 - [90] Chang, J.C., Lin, H.Y., Lv, J., Tseng, W.Y.I., Gau, S.S.F. Regional brain volume predicts response to methylphenidate treatment in individuals with ADHD. *BMC Psychiatry* 2021;21(1):1–14.
 - [91] Koutsouleris, N., Meisenzahl, E.M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Archives of General Psychiatry* 2009;66(7):700–712.
 - [92] Zhutovsky, P., Thomas, R.M., Olff, M., van Rooij, S.J., Kennis, M., van Wingen, G.A., et al. Individual prediction of psychotherapy outcome in posttraumatic stress disorder using neuroimaging data. *Translational Psychiatry* 2019;9(1):1–10.
 - [93] Cao, B., Cho, R.Y., Chen, D., Xiu, M., Wang, L., Soares, J.C., et al. Treatment response prediction and individualized identification of first-episode drug-naïve schizophrenia using brain functional connectivity. *Molecular Psychiatry* 2020;25(4):906–913.
 - [94] Zhdanov, A., Atluri, S., Wong, W., Vaghei, Y., Daskalakis, Z.J., Blumberger, D.M., et al. Use of machine learning for predicting escitalopram treatment outcome from electroencephalography recordings in adult patients with depression. *JAMA Network Open* 2020;3(1):e1918377–e1918377.
 - [95] Redlich, R., Opel, N., Grotegerd, D., Dohm, K., Zaremba, D., Bürger, C., et al. Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry* 2016;73(6):557–564.
 - [96] Wu, W., Zhang, Y., Jiang, J., Lucas, M.V., Fonzo, G.A., Rolle, C.E., et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nature Biotechnology* 2020;38(4):439–447.
 - [97] Fonzo, G.A., Etkin, A., Zhang, Y., Wu, W., Cooper, C., Chinn-Fatt, C., et al. Brain regulation of emotional conflict predicts antidepressant treatment response for depression. *Nature Human Behaviour* 2019;3(12):1319–1331.
 - [98] Yang, D., Pelphrey, K., Sukhodolsky, D., Crowley, M., Dayan, E., Dvornek, N., et al. Brain responses to biological motion predict treatment outcome in young children with autism. *Translational Psychiatry* 2016;6(11):e948–e948.
 - [99] Reggente, N., Moody, T.D., Morfini, F., Sheen, C., Rissman, J., O'Neill, J., et al. Multivariate resting-state functional connectivity predicts response to cognitive behavioral therapy in obsessive-compulsive disorder. *Proc Natl Acad Sci USA* 2018;115(9):2222–2227.
 - [100] Edgcomb, J., Shaddock, T., Hellemann, G., Brooks III, J.O. High-risk phenotypes of early psychiatric readmission in bipolar disorder with comorbid medical illness. *Psychosomatics* 2019;60(6):563–573.
 - [101] Morel, D., Kalvin, C.Y., Liu-Ferrara, A., Caceres-Suriel, A.J., Kurtz, S.G., Tabak, Y.P. Predicting hospital readmission in patients with mental or substance use disorders: a machine learning approach. *International Journal of Medical Informatics* 2020;139:104136.
 - [102] Cearns, M., Hahn, T., Baune, B.T. Recommendations and future directions for supervised machine learning in psychiatry. *Translational Psychiatry* 2019;9(1):1–12.
 - [103] Grzenda, A., Kraguljac, N.V., McDonald, W.M., Nemeroff, C., Torous, J., Alpert, J.E., et al. Evaluating the machine learning literature: a primer and user's guide for psychiatrists. *American Journal of Psychiatry* 2021;178(8):715–729.
 - [104] Tai, A.M., Albuquerque, A., Carmona, N.E., Subramaniampillai, M., Cha, D.S., Sheko, M., et al. Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artificial Intelligence in Medicine* 2019;99:101704.
 - [105] Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., Aafjes, M. A scoping review of machine learning in psychotherapy research. *Psychotherapy Research* 2021;31(1):92–116.
 - [106] Sui, J., Jiang, R., Bustillo, J., Calhoun, V. Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. *Biological Psychiatry* 2020;88(11):818–828.
 - [107] Zhang, L., Wang, M., Liu, M., Zhang, D. A survey on deep learning for neuroimaging-based brain disorder analysis. *Frontiers in Neuroscience* 2020;14:779.
 - [108] Cho, G., Yim, J., Choi, Y., Ko, J., Lee, S.H. Review of machine learning algorithms for diagnosing mental illness. *Psychiatry Investigation* 2019;16(4):262.
 - [109] Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience* 2015;18(11):1664–1671.
 - [110] Shen, X., Finn, E.S., Scheinost, D., Rosenberg, M.D., Chun, M.M., Papademetris, X., et al. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*

- 2017;12(3):506–518.
- [111] Tipping, M.E.. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* 2001;1(6):211–244.
- [112] Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., Cichocki, A.. Sparse bayesian classification of EEG for brain-computer interface. *IEEE Transactions on Neural Networks and Learning Systems* 2015;27(11):2256–2267.
- [113] Ma, Q., Zhang, T., Zanetti, M.V., Shen, H., Satterthwaite, T.D., Wolf, D.H., et al. Classification of multi-site mr images in the presence of heterogeneity using multi-task learning. *NeuroImage: Clinical* 2018;19:476–486.
- [114] Xiao, L., Stephen, J.M., Wilson, T.W., Calhoun, V.D., Wang, Y.P.. A manifold regularized multi-task learning model for iq prediction from two fMRI paradigms. *IEEE Transactions on Biomedical Engineering* 2019;67(3):796–806.
- [115] Kim, M., Min, E.J., Liu, K., Yan, J., Saykin, A.J., Moore, J.H., et al. Multi-task learning based structured sparse canonical correlation analysis for brain imaging genetics. *Medical Image Analysis* 2022;76:102297.
- [116] Kwak, S., Park, S., Kim, J., Park, S., Lee, J.Y.. Multivariate neuroanatomical correlates of behavioral and psychological symptoms in dementia and the moderating role of education. *NeuroImage: Clinical* 2020;28:102452.
- [117] Kalmady, S.V., Greiner, R., Agrawal, R., Shivakumar, V., Narayanaswamy, J.C., Brown, M.R., et al. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. *NPJ Schizophrenia* 2019;5(1):1–11.
- [118] Wang, J., Zhang, L., Wang, Q., Chen, L., Shi, J., Chen, X., et al. Multi-class asd classification based on functional connectivity and functional correlation tensor via multi-source domain adaptation and multi-view sparse representation. *IEEE Transactions on Medical Imaging* 2020;39(10):3137–3147.
- [119] Elmer, J., Jones, B.L., Nagin, D.S.. Using the beta distribution in group-based trajectory models. *BMC Medical Research Methodology* 2018;18(1):1–5.
- [120] van der Nest, G., Passos, V.L., Candel, M.J., van Breukelen, G.J.. An overview of mixture modelling for latent evolutions in longitudinal data: Modelling approaches, fit statistics and software. *Advances in Life Course Research* 2020;43:100323.
- [121] Ellis, J.D., Rabinowitz, J.A., Wells, J., Liu, F., Finan, P.H., Stein, M.D., et al. Latent trajectories of anxiety and depressive symptoms among adults in early treatment for nonmedical opioid use. *Journal of Affective Disorders* 2022;299:223–232.
- [122] Ulvenes, P., Soma, C.S., Melsom, L., Wampold, B.E.. A latent trajectory analysis of inpatient depression treatment. *Psychotherapy* 2022;59(1):113–124.
- [123] Waizbard-Bartov, E., Ferrer, E., Heath, B., Rogers, S.J., Nordahl, C.W., Solomon, M., et al. Identifying autism symptom severity trajectories across childhood. *Autism Research* 2022;.
- [124] Schultebrucks, K., Shalev, A.Y., Michopoulos, V., Grudzen, C.R., Shin, S.M., Stevens, J.S., et al. A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor. *Nature Medicine* 2020;26:1064–1088.
- [125] Crable, E.L., Drainoni, M.L., Jones, D.K., Walley, A.Y., Hicks, J.M.. Predicting longitudinal service use for individuals with substance use disorders: A latent profile analysis. *Journal of Substance Abuse Treatment* 2022;132:108632.
- [126] Chapelle, O., Scholkopf, B., Zien, A.. *Semi-Supervised Learning*. MIT Press; 2009.
- [127] Yin, W., Li, L., Wu, F.X.. A semi-supervised autoencoder for autism disease diagnosis. *Neurocomputing* 2022;483:140–147.
- [128] Varol, E., Sotiras, A., Davatzikos, C., Initiative, A.D.N., et al. Hydra: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *NeuroImage* 2017;145:346–364.
- [129] Yang, T., Frangou, S., Lam, R.W., Huang, J., Su, Y., Zhao, G., et al. Probing the clinical and brain structural boundaries of bipolar and major depressive disorder. *Translational Psychiatry* 2021;11(1):1–8.
- [130] Honnorat, N., Dong, A., Meisenzahl-Lechner, E., Koutsouleris, N., Davatzikos, C.. Neuroanatomical heterogeneity of schizophrenia revealed by semi-supervised machine learning methods. *Schizophrenia Research* 2019;214:43–50.
- [131] Kaczurkin, A.N., Sotiras, A., Baller, E.B., Barzilay, R., Calkins, M.E., Chand, G.B., et al. Neurostructural heterogeneity in youths with internalizing symptoms. *Biological Psychiatry* 2020;87(5):473–482.
- [132] Beaulieu-Jones, B.K., Greene, C.S., et al. Semi-supervised learning of the electronic health record for phenotype stratification. *Journal of Biomedical Informatics* 2016;64:168–178.
- [133] Yazdavar, A.H., Al-Olimat, H.S., Ebrahimi, M., Bajaj, G., Banerjee, T., Thirunarayan, K., et al. 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. 2017, p. 1191–1198.
- [134] Dong, G., Tang, M., Cai, L., Barnes, L.E., Boukhechba, M.. Semi-supervised graph instance transformer for mental health inference. In: *Proc. 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE; 2021, p. 1221–1228.
- [135] Rutherford, S., Kia, S.M., Wolfers, T., Frazz, C., Zabihi, M., Dinga, R., et al. The normative modeling framework for computational psychiatry. *Nature Protocols* 2022;1–24.
- [136] Marquand, A.F., Rezek, I., Buitelaar, J., Beckmann, C.F. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. *Biological Psychiatry* 2016;80(7):552–561.
- [137] Chamberland, M., Genc, S., Tax, C.M., Shastin, D., Koller, K., Raven, E.P., et al. Detecting microstructural deviations in individuals with deep diffusion MRI tractometry. *Nature Computational Science* 2021;1(9):598–606.
- [138] Guggenmos, M., Schmack, K., Veer, I.M., Lett, T., Maria Sekutowicz, M.S., Garbusow, M., et al. A multimodal neuroimaging classifier for alcohol dependence. *Scientific Reports* 2020;10:298.
- [139] Beaulieu-Jones, B.K., Greene, C.S.. Semi-supervised learning of the electronic health record for phenotype stratification. *Journal of Biomedical Informatics* 2016;64:168–178.
- [140] Shen, D., Wu, G., Suk, H.I.. *Deep learning in medical image analysis. Annual Review of Biomedical Engineering* 2017;19:221–248.
- [141] Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., et al. Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nature communications* 2021;12(1):1–17.
- [142] Quaak, M., van de Mortel, L., Thomas, R.M., van Wingen, G.. Deep learning applications for the classification of psychiatric disorders using neuroimaging data: systematic review and meta-analysis. *NeuroImage: Clinical* 2021;30:102584.
- [143] Chang, M., Womer, F.Y., Gong, X., Chen, X., Tang, L., Feng, R., et al. Identifying and validating subtypes within major psychiatric disorders based on frontal-posterior functional imbalance via deep learning. *Molecular Psychiatry* 2021;26(7):2991–3002.
- [144] Almuqhim, F., Saeed, F.. ASD-SAEtNet: a sparse autoencoder, and deep-neural network model for detecting autism spectrum disorder (ASD) using fMRI data. *Frontiers in Computational Neuroscience* 2021;15:27.
- [145] Pinaya, W.H., Mechelli, A., Sato, J.R.. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. *Human Brain Mapping* 2019;40(3):944–954.
- [146] Aglinskias, A., Hartshorne, J.K., Anzellotti, S.. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science* 2022;376(6597):1070–1074.
- [147] Anwar, S.M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., Khan, M.K.. Medical image analysis using convolutional neural networks: a review. *Journal of Medical Systems* 2018;42(11):1–13.
- [148] Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K.. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging* 2018;9(4):611–629.
- [149] Zhang, J., Li, X., Li, Y., Wang, M., Huang, B., Yao, S., et al. Three dimensional convolutional neural network-based classification of conduct disorder with structural mri. *Brain Imaging and Behavior* 2020;14(6):2333–2340.
- [150] Bessadok, A., Mahjoub, M.A., Rekik, I.. Graph neural networks in network neuroscience. *arXiv preprint arXiv:210603535* 2021;.
- [151] Durstewitz, D., Huys, Q.J., Koppe, G.. Psychiatric illnesses as dis-

- orders of network dynamics. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2021;6(9):865–876.
- [152] Zhao, Y., Li, X., Huang, H., Zhang, W., Zhao, S., Makkie, M., et al. 4D modeling of fMRI data via spatio-temporal convolutional neural networks (ST-CNN). *IEEE Transactions on Cognitive and Developmental Systems* 2020;12(3):451.
- [153] Lashgari, E., Liang, D., Maoz, U.. Data augmentation for deep-learning-based electroencephalography. *Journal of Neuroscience Methods* 2020;346:108885.
- [154] Shang, C., Palmer, A., Sun, J., Chen, K.S., Lu, J., Bi, J.. Vigan: Missing view imputation with generative adversarial networks. In: 2017 IEEE International Conference on Big Data (Big Data). IEEE; 2017, p. 766–775.
- [155] Dvornek, N.C., Li, X., Zhuang, J., Duncan, J.S.. Jointly discriminative and generative recurrent neural networks for learning from fMRI. In: *International Workshop on Machine Learning in Medical Imaging*. Springer; 2019, p. 382–390.
- [156] Zhao, J., Huang, J., Zhi, D., Yan, W., Ma, X., Yang, X., et al. Functional network connectivity (FNC)-based generative adversarial network (GAN) and its applications in classification of mental disorders. *Journal of Neuroscience Methods* 2020;341:108756.
- [157] Wen, Z., Marin, M.F., Blackford, J.U., Chen, Z.S., Milad, M.R.. Fear-induced brain activations distinguish anxious and trauma-exposed brains. *Translational Psychiatry* 2021;11(1):1–10.
- [158] Cutillo, C.M., Sharma, K.R., Foschini, L., Kundu, S., Mackintosh, M., Mandl, K.D.. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digital Medicine* 2020;3(1):1–5.
- [159] Sani, O.G., Yang, Y., Lee, M.B., Dawes, H.E., Chang, E.F., Shanechi, M.M.. Mood variations decoded from multi-site intracranial human brain activity. *Nature Biotechnology* 2018;36(10):954–961.
- [160] Bestseny, O., Gilbert, G., Harris, A., Rost, J.. Telehealth: A quarter-trillion-dollar post-covid-19 reality? 2021. URL: <https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/telehealth-a-quarter-trillion-dollar-post-covid-19-reality>.
- [161] Lee, E.E., Torous, J., De Choudhury, M., Depp, C.A., Graham, S.A., Kim, H.C., et al. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2021;6(9):856–864.
- [162] Moustafa, A.A.. *Big Data in Psychiatry and Neurology*. 2021.
- [163] Graham, S., Depp, C., Lee, E.E., Nebeker, C., Tu, X., Kim, H.C., et al. Artificial intelligence for mental health and mental illnesses: an overview. *Current Psychiatry Reports* 2019;21(11):1–18.
- [164] Abbas, A., Schultebrucks, K., Galatzer-Levy, I.R.. Digital measurement of mental health: challenges, promises, and future directions. *Psychiatric Annals* 2021;51(1):14–20.
- [165] Bickman, L.. Improving mental health services: A 50-year journey from randomized experiments to artificial intelligence and precision mental health. *Administration and Policy in Mental Health and Mental Health Services Research* 2020;47(5):795–843.
- [166] Wilkinson, J., Arnold, K.F., Murray, E.J., van Smeden, M., Carr, K., Sippy, R., et al. Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health* 2020;2(12):e677–e680.
- [167] Barron, D.. *Reading Our minds: The rise of Big data psychiatry*. Columbia Global Reports; 2021.
- [168] Vázquez-Romero, A., Gallardo-Antolín, A.. Automatic detection of depression in speech using ensemble convolutional neural networks. *Entropy* 2020;22(6):688.
- [169] Harati, A., Shriberg, E., Rutowski, T., Chlebek, P., Lu, Y., Oliveira, R.. 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; 2021, p. 7273–7277.
- [170] Huang, Z., Epps, J., Joachim, D.. Investigation of speech landmark patterns for depression detection. *IEEE Transactions on Affective Computing* 2019;.
- [171] Zhu, Y., Shang, Y., Shao, Z., Guo, G.. Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. *IEEE Transactions on Affective Computing* 2017;9(4):578–584.
- [172] Shao, W., You, Z., Liang, L., Hu, X., Li, C., Wang, W., et al. A multi-modal gait analysis-based depression detection system. *IEEE Journal of Biomedical and Health Informatics* 2021;.
- [173] Lu, Y., Harati, A., Rutowski, T., Oliveira, R., Chlebek, P., Shriberg, E.. Robust speech and natural language processing models for depression screening. In: *Proc. IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE; 2020, p. 1–5.
- [174] Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Ungar, L.H., Crutchley, P., Preotjuc-Pietro, D., et al. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences USA* 2018;115(44):11203–11208.
- [175] Sun, H., Liu, J., Chai, S., Qiu, Z., Lin, L., Huang, X., et al. Multi-modal adaptive fusion transformer network for the estimation of depression level. *Sensors* 2021;21(14):4764.
- [176] Weiner, L., Guidi, A., Doignon-Camus, N., Giersch, A., Bertschy, G., Vanello, N.. Vocal features obtained through automated methods in verbal fluency tasks can aid the identification of mixed episodes in bipolar disorder. *Translational Psychiatry* 2021;11(1):1–9.
- [177] Palmius, N., Tsanas, A., Saunders, K.E., Bilderbeck, A.C., Geddes, J.R., Goodwin, G.M., et al. Detecting bipolar depression from geographic location data. *IEEE Transactions on Biomedical Engineering* 2016;64(8):1761–1771.
- [178] Marmar, C.R., Brown, A.D., Qian, M., Laska, E., Siegel, C., Li, M., et al. Speech-based markers for posttraumatic stress disorder in us veterans. *Depression and Anxiety* 2019;36(7):607–616.
- [179] Mallol-Ragolta, A., Dhamija, S., Boulton, T.E.. A multimodal approach for predicting changes in ptsd symptom severity. In: *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 2018, p. 324–333.
- [180] Tahir, Y., Yang, Z., Chakraborty, D., Thalmann, N., Thalmann, D., Maniam, Y., et al. Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. *PLoS One* 2019;14(4):e0214314.
- [181] Abbas, A., Yadav, V., Smith, E., Ramjas, E., Rutter, S.B., Benavidez, C., et al. Computer vision-based assessment of motor functioning in schizophrenia: Use of smartphones for remote measurement of schizophrenia symptomatology. *Digital Biomarkers* 2021;5(1):29–36.
- [182] Birnbaum, M.L., Van Meter, A., Chen, V., Rizvi, A.F., Arenare, E., De Choudhury, M., et al. Utilizing machine learning on internet search activity to support the diagnostic process and relapse detection in young individuals with early psychosis: feasibility study. *JMIR Mental Health* 2020;7(9):e19348.
- [183] Birnbaum, M.L., Abrami, A., Heisig, S., Ali, A., Arenare, E., Arguto, C., et al. Acoustic and facial features from clinical interviews for machine learning-based psychiatric diagnosis: Algorithm development. *JMIR Mental Health* 2022;9(1):e24699.
- [184] Abdullah, S., Choudhury, T.. Sensing technologies for monitoring serious mental illnesses. *IEEE MultiMedia* 2018;25(1):61–75.
- [185] Kraepelin, E.. Manic depressive insanity and paranoia. *The Journal of Nervous and Mental Disease* 1921;53(4):350.
- [186] Low, D.M., Bentley, K.H., Ghosh, S.S.. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology* 2020;5(1):96–116.
- [187] Eyben, F., Scherer, K.R., Schuller, B.W., Sundberg, J., André, E., Busso, C., et al. The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing* 2015;7(2):190–202.
- [188] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T.F.. A review of depression and suicide risk assessment using speech analysis. *Speech Communication* 2015;71:10–49.
- [189] Karam, Z.N., Provost, E.M., Singh, S., Montgomery, J., Archer, C., Harrington, G., et al. Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2014, p. 4858–4862.
- [190] Rutowski, T., Shriberg, E., Harati, A., Lu, Y., Oliveira, R., Chlebek, P.. Cross-demographic portability of deep NLP-based depression models. In: *Proc. IEEE Spoken Language Technology Workshop (SLT)*. IEEE; 2021, p. 1052–1057.
- [191] Kesari, G.. AI can now detect depression from your voice, and it's twice as accurate as human practitioners. 2021. URL:

- <https://www.forbes.com/sites/ganeskesari/2021/05/24/ai-can-now-detect-depression-from-just-your-voice/?sh=7baea16a4c8d>.
- [192] Lovett, L.. Sonde launches voice API to detect mental illness. 2021. URL: <https://www.mobihealthnews.com/news/sonde-launches-voice-api-detect-mental-illness>.
 - [193] Galatzer-Levy, I., Abbas, A., Ries, A., Homan, S., Sels, L., Koesmahargyo, V., et al. Validation of visual and auditory digital markers of suicidality in acutely suicidal psychiatric inpatients: Proof-of-concept study. *Journal of Medical Internet Research* 2021;23(6):e25199.
 - [194] Song, S., Shen, L., Valstar, M.. 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; 2018, p. 158–165.
 - [195] de Belen, R.A.J., Bednarz, T., Sowmya, A., Del Favero, D.. Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019. *Translational Psychiatry* 2020;10(1):1–20.
 - [196] Chen, Q., Chaturvedi, I., Ji, S., Cambria, E.. Sequential fusion of facial appearance and dynamics for depression recognition. *Pattern Recognition Letters* 2021;150:115–121.
 - [197] He, L., Niu, M., Tiwari, P., Marttinen, P., Su, R., Jiang, J., et al. Deep learning for depression recognition with audiovisual cues: A review. *Information Fusion* 2022;80:56–86.
 - [198] Zhou, X., Jin, K., Shang, Y., Guo, G.. Visually interpretable representation learning for depression recognition from facial images. *IEEE Transactions on Affective Computing* 2018;11(3):542–552.
 - [199] Smrke, U., Mlakar, I., Lin, S., Musil, B., Pohl, N., et al. Language, speech, and facial expression features for artificial intelligence-based detection of cancer survivors' depression: Scoping meta-review. *JMIR Mental Health* 2021;8(12):e30439.
 - [200] Abbas, A., Hansen, B.J., Koesmahargyo, V., Yadav, V., Rosenfield, P.J., Patil, O., et al. Facial and vocal markers of schizophrenia measured using remote smartphone assessments: Observational study. *JMIR Formative Research* 2022;6(1):e26276.
 - [201] Castro, V.M., Minnier, J., Murphy, S.N., Kohane, I., Churchill, S.E., Gainer, V., et al. Validation of electronic health record phenotyping of bipolar disorder cases and controls. *American Journal of Psychiatry* 2015;172(4):363–372.
 - [202] Morgan, S.E., Dieren, K., Vértes, P.E., Ip, S.H., Wang, B., Thompson, B., et al. Natural language processing markers in first episode psychosis and people at clinical high-risk. *Translational Psychiatry* 2021;11(1):1–9.
 - [203] Rutowski, T., Shriberg, E., Harati, A., Lu, Y., Chlebek, P., Oliveira, R.. Depression and anxiety prediction using deep language models and transfer learning. In: *Proc. 7th International Conference on Behavioural and Social Computing (BESC)*. IEEE; 2020, p. 1–6.
 - [204] Kim, J., Lee, J., Park, E., Han, J.. A deep learning model for detecting mental illness from user content on social media. *Scientific Reports* 2020;10:11846.
 - [205] Rezaei, N., Wolff, P., Price, B.H.. Natural language processing in psychiatry: the promises and perils of a transformative approach. *The British Journal of Psychiatry* 2022;1–3.
 - [206] Le Glaz, A., Haralambous, Y., Kim-Dufor, D.H., Lenca, P., Billot, R., Ryan, T.C., et al. Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research* 2021;23(5):e15708.
 - [207] National electronic health records survey: 2015 specialty and overall physicians electronic health record adoption summary tables. 2015. URL: https://www.cdc.gov/nchs/data/ahcd/nehrs/2015_nehrs_ehr_by_specialty.pdf.
 - [208] Van Le, D., Montgomery, J., Kirkby, K.C., Scanlan, J.. Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting. *Journal of Biomedical Informatics* 2018;86:49–58.
 - [209] Walsh, C.G., Ribeiro, J.D., Franklin, J.C.. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *Journal of Child Psychology and Psychiatry* 2018;59(12):1261–1270.
 - [210] Rumshisky, A., Ghassemi, M., Naumann, T., Szolovits, P., Castro, V., McCoy, T., et al. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational Psychiatry* 2016;6(10):e921–e921.
 - [211] Stewart, R., Velupillai, S.. Applied natural language processing in mental health big data. *Neuropsychopharmacology* 2021;46(1):252.
 - [212] Ive, J., Viani, N., Kam, J., Yin, L., Verma, S., Puntis, S., et al. Generation and evaluation of artificial mental health records for natural language processing. *NPJ Digital Medicine* 2020;3(1):1–9.
 - [213] Bantilan, N., Malgaroli, M., Ray, B., Hull, T.D.. Just in time crisis response: suicide alert system for telemedicine psychotherapy settings. *Psychotherapy Research* 2021;31(3):289–299.
 - [214] Raveau, M.P., Goñi, J., Rodriguez, J., Paiva, I., Barriga, F., Hermosilla, M.P., et al. Natural language processing of helpline chat data before and during the pandemic revealed significant decrease in self-image appreciation and changes in other traits. *Preprints* 2022;.
 - [215] Bollen, J., Ten Thij, M., Breithaupt, F., Barron, A.T., Rutter, L.A., Lorenzo-Luaces, L., et al. Historical language records reveal a surge of cognitive distortions in recent decades. *Proc Natl Acad Sci USA* 2021;118(30):e2102061118.
 - [216] Bathina, K.C., Ten Thij, M., Lorenzo-Luaces, L., Rutter, L.A., Bollen, J.. Individuals with depression express more distorted thinking on social media. *Nature Human Behaviour* 2021;5(4):458–466.
 - [217] Guntuku, S.C., Buffone, A., Jaidka, K., Eichstaedt, J.C., Ungar, L.H.. Understanding and measuring psychological stress using social media. In: *Proceedings of the International AAI Conference on Web and Social Media*; vol. 13. 2019, p. 214–225.
 - [218] Rissola, E.A., Losada, D.E., Crestani, F.. A survey of computational methods for online mental state assessment on social media. *ACM Transactions on Computing for Healthcare* 2021;2(2):1–31.
 - [219] Hänsel, K., Lin, I.W., Sobolev, M., Muscat, W., Yum-Chan, S., De Choudhury, M., et al. Utilizing instagram data to identify usage patterns associated with schizophrenia spectrum disorders. *Frontiers in Psychiatry* 2021;12:691327.
 - [220] Birnbaum, M.L., Norel, R., Van Meter, A., Ali, A.F., Arenare, E., Eyigöz, E., et al. Identifying signals associated with psychiatric illness utilizing language and images posted to facebook. *NPJ Schizophrenia* 2020;6(1):1–10.
 - [221] El-Ramly, M., Abu-Elyazid, H., Mo'men, Y., Alshaer, G., Adib, N., Eldeen, K.A., et al. CairoDep: Detecting depression in arabic posts using bert transformers. In: *Proc. Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*. 2021, p. 207–212.
 - [222] Martínez-Castaño, R., Htaï, A., Azzopardi, L., Moshfeghi, Y.. BERT-based transformers for early detection of mental health illnesses. In: *Proc. International Conference of the Cross-Language Evaluation Forum for European Languages*. 2021, p. 189–200.
 - [223] Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., Cambria, E.. Mentalbert: Publicly available pretrained language models for mental healthcare. *arXiv preprint: arXiv:211015621* 2021;.
 - [224] Kjell, O.N., Sikström, S., Kjell, K., Schwartz, H.A.. Natural language analyzed with ai-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy. *Scientific Reports* 2022;12(1):1–9.
 - [225] Chancellor, S., De Choudhury, M.. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digital Medicine* 2020;3(1):1–11.
 - [226] Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K.J., Tørresen, J.. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing* 2018;51:1–26.
 - [227] Dagum, P.. Digital biomarkers of cognitive function. *NPJ Digital Medicine* 2018;1(1):1–3.
 - [228] Zulmeta, J., Piscitello, A., Rasic, M., Easter, R., Babu, P., Lange-necker, S.A., et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffect digital phenotyping study. *Journal of Medical Internet Research* 2018;20(7):e9775.
 - [229] Mandryk, R.L., Birk, M.V.. The potential of game-based digital biomarkers for modeling mental health. *JMIR Mental Health* 2019;6(4):e13485.
 - [230] Dechant, M., Frommel, J., Mandryk, R.. Assessing social anxiety through digital biomarkers embedded in a gaming task. In: *Proc. 2021 CHI Conference on Human Factors in Computing Systems*. 2021;.

- [231] Winkler, R.. Apple is working on iphone features to help detect depression, cognitive decline. 2021. URL: <https://www.wsj.com/articles/apple-wants-iphones-to-help-detect-depression-cognitive-decline-11632716601>.
- [232] Seppälä, J., De Vita, I., Jämsä, T., Miettunen, J., Isohanni, M., Rubinstein, K., et al. Mobile phone and wearable sensor-based mhealth approaches for psychiatric disorders and symptoms: systematic review. *JMIR Mental Health* 2019;6(2):e9819.
- [233] Chikersal, P., Doryab, A., Tumminia, M., Villalba, D.K., Dutcher, J.M., Liu, X., et al. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: a machine learning approach with robust feature selection. *ACM Transactions on Computer-Human Interaction (TOCHI)* 2021;28(1):1–41.
- [234] Torous, J., Kiang, M.V., Lorme, J., Onnela, J.P., et al. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health* 2016;3:e5165.
- [235] Mohr, D.C., Zhang, M., Schueller, S.M.. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology* 2017;13:23–47.
- [236] Huckvale, K., Venkatesh, S., Christensen, H.. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *NPJ Digital Medicine* 2019;2(1):1–11.
- [237] Insel, T.R., Cuthbert, B.N.. Digital phenotyping: a global tool for psychiatry. *World Psychiatry* 2018;17(3):276–277.
- [238] Benoit, J., Onyeaka, H., Keshavan, M., Torous, J.. Systematic review of digital phenotyping and machine learning in psychosis spectrum illnesses. *Harvard Review of Psychiatry* 2020;28(5):296–304.
- [239] Mendes, J.P.M., Moura, I.R., Van de Ven, P., Viana, D., Silva, F.J.S., Coutinho, L.R., et al. Sensing apps and public data sets for digital phenotyping of mental health: Systematic review. *Journal of Medical Internet Research* 2022;24(2):e28735.
- [240] Prochaska, J.J., Vogel, E.A., Chieng, A., Kendra, M., Baiocchi, M., Pajarito, S., et al. A therapeutic relational agent for reducing problematic substance use (woebot): development and usability study. *Journal of Medical Internet Research* 2021;23(3):e24850.
- [241] Bohn, D.. Amazon announces halo, a fitness band and app that scans your body and voice. 2020. URL: <https://www.theverge.com/2020/8/27/21402493/amazon-halo-band-health-fitness-body-scan-tone-emotion-activity-tracking>.
- [242] fitbit, . Understand your stress so you can manage it. 2022. URL: <https://www.fitbit.com/global/us/technology/stress>.
- [243] Kintsugi, . Kintsugi for health plans. 2022. URL: <https://kintsugihello.com/for-health-plans>.
- [244] Auxier, B., Bucaille, A., Westcott, K.. Mental health goes mobile: The mental health app market will keep on growing. 2021. URL: <https://www2.deloitte.com/us/en/insights/industry/technology/technology-media-and-telecom-predictions/2022/mental-health-app-market.html>.
- [245] Torous, J., Bucci, S., Bell, L.H., Kessing, L.V., et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* 2021;20(3):318–335.
- [246] Onnela, J.P., Rauch, S.L.. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 2016;41(7):1691–1696.
- [247] Stein, D.J., Fineberg, N.A., Chamberlain, S.R.. *Mental health in a digital world*. Elsevier; 2021.
- [248] Abbas, A., Schultebrucks, K., Galatzer-Levy, I.R.. Digital measurement of mental health: Challenges, promises, and future directions. *Psychiatric Annals* 2021;51(1):14–20.
- [249] Galatzer-Levy, I.R., Bryant, R.A.. 636,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science* 2013;8(6):651–662.
- [250] Jacobs, B.L.. Serotonin, motor activity and depression-related disorders. *American Scientist* 1994;82(5):456–463.
- [251] Gigliucci, V., O'Dowd, G., Casey, S., Egan, D., Gibney, S., Harkin, A.. Ketamine elicits sustained antidepressant-like activity via a serotonin-dependent mechanism. *Psychopharmacology* 2013;228(1):157–166.
- [252] Galatzer-Levy, I., Abbas, A., Ries, A., Homan, S., Sels, L., Koesmahargyo, V., et al. Validation of visual and auditory digital markers of suicidality in acutely suicidal psychiatric inpatients: Proof-of-concept study. *Journal of Medical Internet Research* 2021;23(6):e25199.
- [253] Abbas, A., Sels, L., Koesmahargyo, V., Aghjayan, A., Marecki, S., et al. Remote digital measurement of facial and vocal markers of major depressive disorder severity and treatment response: a pilot study. *Frontiers in Digital Health* 2021;3:28.
- [254] Zhang, L., Koesmahargyo, V., Galatzer-Levy, I.. Estimation of clinical tremor using spatio-temporal adversarial autoencoder. In: *Proc. 25th International Conference on Pattern Recognition (ICPR'20)*. IEEE; 2021, p. 8259–8266.
- [255] Ewbank, M.P., Cummins, R., Tablan, V., Bateup, S., Catarino, A., Martin, A.J., et al. Quantifying the association between psychotherapy content and clinical outcomes using deep learning. *JAMA Psychiatry* 2020;77(1):35–43.
- [256] Economides, M., Martman, J., Bell, M.J., Sanderson, B.. Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. *Mindfulness* 2018;9(5):1584–1593.
- [257] Kunkle, S., Yip, M., Hunt, J., Watson, X., Udall, D., Arean, P., et al. Association between care utilization and anxiety outcomes in an on-demand mental health system: Retrospective observational study. *JMIR Formative Research* 2021;5(1):e24662.
- [258] Ewbank, M., Cummins, R., Tablan, V., Catarino, A., Buchholz, S., Blackwell, A.. Understanding the relationship between patient language and outcomes in internet-enabled cognitive behavioural therapy: A deep learning approach to automatic coding of session transcripts. *Psychotherapy Research* 2021;31(3):300–312.
- [259] Flenotomos, N., Martinez, V.R., Chen, Z., Creed, T.A., Atkins, D.C., Narayanan, S.. Automated quality assessment of cognitive behavioral therapy sessions through highly contextualized language representations. *PLoS One* 2021;16(10):e0258639.
- [260] Patel, N.A., Butte, A.J.. Characteristics and challenges of the clinical pipeline of digital therapeutics. *NPJ Digital Medicine* 2020;3(1):1–5.
- [261] Insel, T.R.. Bending the curve for mental health: technology for a public health approach. *American Journal of Public Health* 2019;109(S3):S168–S170.
- [262] Folk, J.B., Schiel, M.A., Oblath, R., Feuer, V., Sharma, A., Khan, S., et al. The transition of academic mental health clinics to telehealth during the covid-19 pandemic. *Journal of the American Academy of Child & Adolescent Psychiatry* 2022;61(2):277–290.
- [263] Wagner, B., Horn, A.B., Maercker, A.. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. *Journal of Affective Disorders* 2014;152:113–121.
- [264] Lahat, D., Adali, T., Jutten, C.. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE* 2015;103(9):1449–1477.
- [265] Croitor-Sava, A., Martinez-Bisbal, M., Laudadio, T., Piquer, J., Celda, B., Heerschap, A., et al. Fusing in vivo and ex vivo nmr sources of information for brain tumor classification. *Measurement Science and Technology* 2011;22(11):114012.
- [266] Adali, T., Levin-Schwartz, Y., Calhoun, V.D.. Multimodal data fusion using source separation: Two effective models based on ICA and IVA and their properties. *Proceedings of the IEEE* 2015;103(9):1478–1493.
- [267] Adali, T., Levin-Schwartz, Y., Calhoun, V.D.. Multimodal data fusion using source separation: Application to medical imaging. *Proceedings of the IEEE* 2015;103(9):1494–1506.
- [268] Calhoun, V., Sui, J., Qi, S.. Multimodal fusion signature as transdiagnostic psychiatric biomarker. *Biological Psychiatry* 2020;87(9):S37.
- [269] Zhang, Y.D., Dong, Z., Wang, S.H., Yu, X., Yao, X., Zhou, Q., et al. Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation. *Information Fusion* 2020;64:149–187.
- [270] Correa, N.M., Adali, T., Li, Y.O., Calhoun, V.D.. Canonical correlation analysis for data fusion and group inferences. *IEEE Signal Processing Magazine* 2010;27(4):39–50.
- [271] de Cheveigné, A., Di Liberto, G.M., Arzounian, D., Wong, D.D., Hjortkjaer, J., Fuglsang, S., et al. Multiway canonical correlation analysis of brain data. *NeuroImage* 2019;186:728–740.
- [272] Chen, X., Wang, Z.J., McKeown, M.. Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods.

- IEEE Signal Processing Magazine 2016;33(3):86–107.
- [273] Silva, R.F., Plis, S.M.. How to integrate data from multiple biological layers in mental health? In: *Personalized Psychiatry*. Springer; 2019, p. 135–159.
- [274] Zhou, G., Zhao, Q., Zhang, Y., Adali, T., Xie, S., Cichocki, A.. Linked component analysis from matrices to high-order tensors: Applications to biomedical data. *Proceedings of the IEEE* 2016;104(2):310–331.
- [275] Calhoun, V.D., Adali, T.. Feature-based fusion of medical imaging data. *IEEE Transactions on Information Technology in Biomedicine* 2008;13(5):711–720.
- [276] Calhoun, V.D., Liu, J., Adali, T.. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage* 2009;45(1):S163–S172.
- [277] Acar, E., Bro, R., Smilde, A.K.. Data fusion in metabolomics using coupled matrix and tensor factorizations. *Proceedings of the IEEE* 2015;103(9):1602–1620.
- [278] Zhou, G., Cichocki, A., Zhang, Y., Mandic, D.P.. Group component analysis for multiblock data: Common and individual feature extraction. *IEEE Transactions on Neural Networks and Learning Systems* 2015;27(11):2426–2439.
- [279] Lock, E.F., Hoadley, K.A., Marron, J.S., Nobel, A.B.. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *The Annals of Applied Statistics* 2013;7(1):523.
- [280] Rakotomamonjy, A., Bach, F., Canu, S., Grandvalet, Y.. Simplemkl. *Journal of Machine Learning Research* 2008;9:2491–2521.
- [281] Mariette, J., Villa-Vialaneix, N.. Unsupervised multiple kernel learning for heterogeneous data integration. *Bioinformatics* 2018;34(6):1009–1015.
- [282] Squarcina, L., Castellani, U., Bellani, M., Perlini, C., Lasalvia, A., Dusi, N., et al. Classification of first-episode psychosis in a large cohort of patients using support vector machine and multiple kernel learning techniques. *NeuroImage* 2017;145:238–245.
- [283] Dyrba, M., Grothe, M., Kirste, T., Teipel, S.J.. Multimodal analysis of functional and structural disconnection in a Alzheimer's disease using multiple kernel svm. *Human Brain Mapping* 2015;36(6):2118–2131.
- [284] Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., Initiative, A.D.N., et al. Multimodal classification of alzheimer's disease and mild cognitive impairment. *NeuroImage* 2011;55(3):856–867.
- [285] Ramachandram, D., Taylor, G.W.. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine* 2017;34(6):96–108.
- [286] Zhou, T., Thung, K.H., Zhu, X., Shen, D.. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human Brain Mapping* 2019;40(3):1001–1016.
- [287] Nguyen, N.D., Huang, J., Wang, D.. A deep manifold-regularized learning model for improving phenotype prediction from multi-modal data. *Nature Computational Science* 2022;2(1):38–46.
- [288] Holzinger, A., Malle, B., Saranti, A., Pfeifer, B.. Towards multi-modal causability with graph neural networks enabling information fusion for explainable ai. *Information Fusion* 2021;71:28–37.
- [289] Dsouza, N.S., Nebel, M.B., Crocetti, D., Robinson, J., Mostofsky, S., Venkataraman, A.. M-GCN: A multimodal graph convolutional network to integrate functional and structural connectomics data to predict multidimensional phenotypic characterizations. In: *Medical Imaging with Deep Learning*. PMLR; 2021, p. 119–130.
- [290] Zhang, W., Zhan, L., Thompson, P., Wang, Y.. Deep representation learning for multimodal brain networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2020, p. 613–624.
- [291] Kong, Z., Sun, L., Peng, H., Zhan, L., Chen, Y., He, L.. Multiplex graph networks for multimodal brain network analysis. *arXiv preprint arXiv:210800158* 2021;.
- [292] Tulay, E.E., Metin, B., Tarhan, N., Arkan, M.K.. Multimodal neuroimaging: basic concepts and classification of neuropsychiatric diseases. *Clinical EEG and Neuroscience* 2019;50(1):20–33.
- [293] Sui, J., Pearson, G., Caprihan, A., Adali, T., Kiehl, K.A., Liu, J., et al. Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model. *NeuroImage* 2011;57(3):839–855.
- [294] Acar, E., Schenker, C., Levin-Schwartz, Y., Calhoun, V.D., Adali, T.. Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data. *Frontiers in Neuroscience* 2019;13:416.
- [295] Vai, B., Parenti, L., Bollettini, I., Cara, C., Verga, C., Melloni, E., et al. Predicting differential diagnosis between bipolar and unipolar depression with multiple kernel learning on multimodal structural neuroimaging. *European Neuropsychopharmacology* 2020;34:28–38.
- [296] Zhang, Q., Wu, Q., Zhu, H., He, L., Huang, H., Zhang, J., et al. Multimodal mri-based classification of trauma survivors with and without post-traumatic stress disorder. *Frontiers in Neuroscience* 2016;10:292.
- [297] Zhou, X., Lin, Q., Gui, Y., Wang, Z., Liu, M., Lu, H.. Multimodal MR images-based diagnosis of early adolescent attention-deficit/hyperactivity disorder using multiple kernel learning. *Frontiers in Neuroscience* 2021;15.
- [298] Geenjaar, E., Lewis, N., Fu, Z., Venkatdas, R., Plis, S., Calhoun, V.. Fusing multimodal neuroimaging data with a variational autoencoder. In: *Proc. 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE; 2021, p. 3630–3633.
- [299] Liu, J., Wang, X., Zhang, X., Pan, Y., Wang, X., Wang, J.. Mmm: classification of schizophrenia using multi-modality multi-atlas feature representation and multi-kernel learning. *Multimedia Tools and Applications* 2018;77(22):29651–29667.
- [300] Plis, S.M., Amin, M.F., Chekroud, A., Hjelm, D., Damaraju, E., Lee, H.J., et al. 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* 2018;181:734–747.
- [301] Rahaman, M.A., Chen, J., Fu, Z., Lewis, N., Iraj, A., Calhoun, V.D.. 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; 2021, p. 3267–3272.
- [302] Akhonda, M., Levin-Schwartz, Y., Calhoun, V.D., Adali, T.. Association of neuroimaging data with behavioral variables: A class of multivariate methods and their comparison using multi-task fMRI data. *Sensors* 2022;22(3):1224.
- [303] Wang, Y., Jiang, W.B., Li, R., Lu, B.L.. 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; 2021, p. 1575–1578.
- [304] Jaques, N., Taylor, S., Sano, A., Picard, R.. Multi-task, multi-kernel learning for estimating individual wellbeing. In: *Proc. NIPS Workshop on Multimodal Machine Learning*, Montreal, Quebec; vol. 898. 2015, p. 3.
- [305] Lam, G., Dongyan, H., Lin, W.. Context-aware deep learning for multimodal depression detection. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE; 2019, p. 3946–3950.
- [306] Parra, F., Benezeth, Y., Yang, F.. Automatic assessment of emotion dysregulation in american, french, and tunisian adults and new developments in deep multimodal fusion: Cross-sectional study. *JMIR Mental Health* 2022;9(1):e34333.
- [307] Doryab, A., Villalba, D.K., Chikersal, P., Dutcher, J.M., Tumminia, M., Liu, X., et al. 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 and uHealth* 2019;7(7):e13209.
- [308] He, J., Mai, S., Hu, H.. A unimodal reinforced transformer with time squeeze fusion for multimodal sentiment analysis. *IEEE Signal Processing Letters* 2021;28:992–996.
- [309] Dibeklioğlu, H., Hammal, Z., Cohn, J.F.. Dynamic multimodal measurement of depression severity using deep autoencoding. *IEEE Journal of Biomedical and Health Informatics* 2017;22(2):525–536.
- [310] Herzog, D.P., Beckmann, H., Lieb, K., Ryu, S., Müller, M.B.. Understanding and predicting antidepressant response: Using animal models to move toward precision psychiatry. *Frontiers in Psychiatry* 2018;9:512.
- [311] Bale, T.L., Abel, T., Akil, H., Jr., W.A.C., Moghaddam, B., Nestler, E.J., et al. The critical importance of basic animal research for neuropsychiatric disorders. *Neuropsychopharmacology* 2019;44(8):1349–1353.
- [312] Labonté, B., Engmann, O., Purushothaman, I., Menard, C., Wang, J.,

- Tan, C., et al. Sex-specific transcriptional signatures in human depression. *Nature Medicine* 2017;23(9):1102–1111.
- [313] Nagy, C., Maitra, M., Tanti, A., Suderman, M., Thérout, J.F., Davoli, M.A., et al. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nature Neuroscience* 2020;23(6):771–781.
- [314] McEwen, B.S., Bowles, N.P., Gray, J.D., Hill, M.N., Hunter, R.G., Karatsoreos, I.N., et al. Mechanisms of stress in the brain. *Nat Neurosci* 2015;18(10):1353–63.
- [315] Floriou-Servou, A., von Ziegler, L., Stalder, L., Sturman, O., Privitera, M., Rassi, A., et al. Distinct proteomic, transcriptomic, and epigenetic stress responses in dorsal and ventral hippocampus. *Biological Psychiatry* 2018;84(7):531–541.
- [316] Bigio, B., Mathe, A.A., Sousa, V.C., Zelli, D., Svenningsson, P., McEwen, B.S., et al. Epigenetics and energetics in ventral hippocampus mediate rapid antidepressant action: Implications for treatment resistance. *Proc Natl Acad Sci USA* 2016;113(28):7906–7911.
- [317] Flight, M.H. Antidepressant epigenetic action. *Nat Rev Neurosci* 2013;14(4):226.
- [318] Stuart, T., Satija, R. Integrative single-cell analysis. *Nature Reviews Genetics* 2019;20(5):257–272.
- [319] Petegrosso, R., Li, Z., Kuang, R. Machine learning and statistical methods for clustering single-cell rna-sequencing data. *Briefings in Bioinformatics* 2020;21(4):1209–1223.
- [320] Hao, Y., Hao, S., Andersen-Nissen, E., Mauck William M., I., Zheng, S., Butler, A., et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184(13):3573–3587.
- [321] Amodio, M., van Dijk, D., Srinivasan, K., Chen, W.S., Mohsen, H., Moon, K.R., et al. Exploring single-cell data with deep multitasking neural networks. *Nature Methods* 2019;16(11):1139–1145.
- [322] Eraslan, G., Simon, L.M., Mircea, M., Mueller, N.S., Theis, F.J. Single-cell rna-seq denoising using a deep count autoencoder. *Nat Commun* 2019;10(1):390.
- [323] Wang, D., Gu, J. Vasc: Dimension reduction and visualization of single-cell rna-seq data by deep variational autoencoder. *Genomics Proteomics Bioinformatics* 2018;16(5):320–331.
- [324] Wang, J., Agarwal, D., Huang, M., Hu, G., Zhou, Z., Ye, C., et al. Data denoising with transfer learning in single-cell transcriptomics. *Nat Methods* 2019;16(9):875–878.
- [325] Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 2018;361(6409):1380–1385.
- [326] Chen, S., Lake, B.B., Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol* 2019;37(12):1452–1457.
- [327] Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods* 2019;16(10):987–990.
- [328] Rodrigues, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019;363(6434):1463–1467.
- [329] Nasca, C., Rasgon, N., McEwen, B. An emerging epigenetic framework of systemic and central mechanisms underlying stress-related disorders. *Neuropsychopharmacology* 2019;44(1):235–236.
- [330] Nasca, C., Dobbin, J., Bigio, B., Watson, K., de Angelis, P., Kautz, M., et al. Insulin receptor substrate in brain-enriched exosomes in subjects with major depression: on the path of creation of biosignatures of central insulin resistance. *Molecular Psychiatry* 2021;26(9):5140–5149.
- [331] Nasca, C., Barnhill, O., DeAngelis, P., Watson, K., Lin, J., Beasley, J., et al. Multidimensional predictors of antidepressant responses: Integrating mitochondrial, genetic, metabolic and environmental factors with clinical outcomes. *Neurobiol Stress* 2021;15:100407.
- [332] Dean, K.R., Hammamieh, R., Mellon, S.H., Abu-Amara, D., Flory, J.D., Guffanti, G., et al. Multi-omic biomarker identification and validation for diagnosing warzone-related post-traumatic stress disorder. *Mol Psychiatry* 2020;25(12):3337–3349.
- [333] Schultebrucks, K., Qian, M., Abu-Amara, D., Dean, K., Laska, E., Siegel, C., et al. Pre-deployment risk factors for ptsd in active-duty personnel deployed to afghanistan: a machine-learning approach for analyzing multivariate predictors. *Mol Psychiatry* 2021;26(9):5011–5022.
- [334] Lorsch, Z.S., Ambesi-Impiombato, A., Zenowich, R., Morganstern, I., Leahy, E., Bansal, M., et al. Computational analysis of multidimensional behavioral alterations after chronic social defeat stress. *Biol Psychiatry* 2021;89(9):920–928.
- [335] Alexandrov, V., Brunner, D., Hanania, T., Leahy, E. High-throughput analysis of behavior for drug discovery. *European Journal of Pharmacology* 2015;750:82–89.
- [336] Nasca, C., Menard, C., Hodes, G., Bigio, B., Pena, C., Lorsch, Z., et al. Multidimensional predictors of susceptibility and resilience to social defeat stress. *Biol Psychiatry* 2019;86(6):483–491.
- [337] Gunning, D., Aha, D. DARPA's explainable artificial intelligence (XAI) program. *AI Magazine* 2019;40(2):44–58.
- [338] Roessner, V., Rothe, J., Kohls, G., Schomerus, G., Ehrlich, S., Beste, C. Taming the chaos?! using eXplainable Artificial Intelligence (XAI) to tackle the complexity in mental health research. 2021.
- [339] Thorsen-Meyer, H.C., Nielsen, A.B., Nielsen, A.P., Kaas-Hansen, B.S., Toft, P.P., Schierbeck, J., et al. 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. *The Lancet Digital Health* 2020;2(4):e179–e191.
- [340] Sheu, Y.h. Illuminating the black box: Interpreting deep neural network models for psychiatric research. *Frontiers in Psychiatry* 2020;1091.
- [341] Molnar, C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2022. URL: <https://www.amazon.com/Interpretable-Machine-Learning-Making-Explainable/dp/B09TMWHVB4>.
- [342] Pearl, J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Elsevier; 1988.
- [343] Letham, B., Rudin, C., McCormick, T.H., Madigan, D. Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. *Annals of Applied Statistics* 2015;9(3):1350–1371.
- [344] Li, Y., Murias, M., Major, S., Dawson, G., Dzirasa, K., Carin, L., et al. Targeting EEG/LFP synchrony with neural nets. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., et al., editors. *Advances in Neural Information Processing Systems*; vol. 30. Curran Associates, Inc.; 2017.
- [345] Zou, J.Y., Adams, R.P. Priors for diversity in generative latent variable models. In: *Advances in Neural Information Processing Systems (NIPS)* 25. 2012.
- [346] Nair, R., Mattetti, M., Daly, E., Wei, D., Alkan, O., Zhang, Y. What changed? interpretable model comparison. In: *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI'21)*. 2021.
- [347] Daly, E., Mattetti, M., Alkan, O., Nair, R. User driven model adjustment via boolean rule explanation. In: *Proc. 35th Conf. Artificial Intelligence (AAAI'21)*. 2021, p. 5896–5904.
- [348] Huys, Q.J., Maia, T.V., Frank, M.J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience* 2016;19(3):404–413.
- [349] Breakspear, M. Dynamic models of large-scale brain activity. *Nature Neuroscience* 2017;20(3):340–352.
- [350] Murray, J.D., Demirtaş, M., Anticevic, A. Biophysical modeling of large-scale brain dynamics and applications for computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2018;3(9):777–787.
- [351] Papadopoulos, L., Lynn, C.W., Battaglia, D., Bassett, D. Relations between large-scale brain connectivity and effects of regional stimulation depend on collective dynamical state. *PLoS Computational Biology* 2020;e1008144.
- [352] Wilson, H.R., Cowan, J.D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal* 1972;12:1–24.
- [353] Chaudhuri, R., Knoblauch, K., Gariel, M.A., Kennedy, H., Wang, X.J. A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron* 2015;88(2):419–431.
- [354] Friston, K., Harrison, L., Penny, W. Dynamical causal modeling. *NeuroImage* 2003;19:1273–1302.
- [355] Friston, K., Kahan, J., Biswal, B., Razi, A. A DCM for resting state fMRI. *NeuroImage* 2014;94:396–407.

- [356] Chen, Z.S., Pesaran, B.. Improving scalability in systems neuroscience. *Neuron* 2021;109(11):1776–1790.
- [357] Nasr, K., Haslacher, D., Dayan, E., Censor, N., Cohen, L.G., Soekadar, S.R.. Breaking the boundaries of interacting with the human brain using adaptive closed-loop stimulation. *Progress in Neurobiology* 2022;216:102311.
- [358] Jazayeri, M., Afraz, A.. Navigating the neural space in search of the neural code. *Neuron* 2017;93(5):1003–1014.
- [359] Etkin, A.. Addressing the causality gap in human psychiatric neuroscience. *JAMA Psychiatry* 2018;75(1):3–4.
- [360] Fellous, J.M., Sapiro, G., Rossi, A., Mayberg, H., Ferrante, M.. Explainable artificial intelligence for neuroscience: behavioral neurostimulation. *Frontiers in Neuroscience* 2019;13:1346.
- [361] Lewis, P.M., Thomson, R.H., Rosenfeld, J.V., Fitzgerald, P.B.. Brain neuromodulation techniques: a review. *The Neuroscientist* 2016;22(4):406–421.
- [362] Romei, V., Thut, G., Silvanto, J.. Information-based approaches of noninvasive transcranial brain stimulation. *Trends in Neurosciences* 2016;39(11):782–795.
- [363] Lo, M.C., Widege, A.. Closed-loop neuromodulation systems: next generation treatments for psychiatric illness. *Int Rev Psych* 2017;29:191–204.
- [364] Chen, A., Oathes, D.J., Chang, C., Bradley, T., Zhou, Z., Williams, L., et al. Causal interactions between fronto-parietal central executive and default-mode networks in humans. *Proc Natl Acad Sci USA* 2013;110(49):19944–19949.
- [365] Hobot, J., Klineciewicz, M., Sandberg, K., Wierzcioń, M.. Causal inferences in repetitive transcranial magnetic stimulation research: challenges and perspectives. *Frontiers in Human Neuroscience* 2021;14:574.
- [366] Akhtar, H., Bukhari, F., Nazir, M., Anwar, M.N., Shahzad, A.. Therapeutic efficacy of neurostimulation for depression: techniques, current modalities, and future challenges. *Neuroscience Bulletin* 2016;32(1):115–126.
- [367] Pineau, J., Guez, A., Vincent, R., Panuccio, G., Avoli, M.. Treating epilepsy via adaptive neurostimulation: a reinforcement learning approach. *International Journal of Neural Systems* 2009;19(4):227–240.
- [368] Tafazoli, S., MacDowell, C.J., Che, Z., Letai, K.C., Steinhart, C.R., Buschman, T.J.. Learning to control the brain through adaptive closed-loop patterned stimulation. *Journal of Neural Engineering* 2020;17(5):056007.
- [369] Gao, Q., Naumann, M., Jovanov, I., Lesi, V., Kamaravelu, K., Grill, W.M., et al. 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 (ICCPs)*. IEEE; 2020;doi:10.1109/ICCPs48487.2020.00018.
- [370] Hebb, A.O., Zhang, J.J., Mahoor, M.H., Tsiokos, C., Matlack, C., Chizeck, H.J., et al. Creating the feedback loop: closed-loop neurostimulation. *Neurosurgery Clinics of North America* 2014;25(1):187–204.
- [371] Bauer, R., Gharabaghi, A.. Reinforcement learning for adaptive threshold control of restorative brain-computer interfaces: a Bayesian simulation. *Frontiers in Neuroscience* 2015;9:36.
- [372] Tang, E., Bassett, D.S.. Control of dynamics in brain networks. *Reviews of Modern Physics* 2018;90(3):031003.
- [373] Srivastava, P., Nozari, E., Kim, J.Z., Ju, H., Zhou, D., Becker, C., et al. Models of communication and control for brain networks: distinctions, convergence, and future outlook. *Network Neuroscience* 2020;4(4):1122–1159.
- [374] Zhang, X., Braun, U., Tost, H., Bassett, D.S.. Data-driven approaches to neuroimaging analysis to enhance psychiatric diagnosis and therapy. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2020;5(8):780–790.
- [375] Fox, M.D., Buckner, R.L., Liu, H., Chakravarty, M.M., Lozano, A.M., Pascual-Leone, A.. Resting-state networks link invasive and noninvasive brain stimulation across diverse psychiatric and neurological diseases. *Proc Natl Acad Sci USA* 2014;111(41):E4367–E4375.
- [376] Zandvakili, A., Philip, N.S., Jones, S.R., Tyrka, A.R., Greenberg, B.D., Carpenter, L.L.. 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. *Journal of Affective Disorders* 2019;252:47–54.
- [377] Vollmer, S., Mateen, B.A., Bohner, G., Király, F.J., Ghani, R., Jonsson, P., et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ* 2020;368.
- [378] Kessler, R.C., Chiu, W.T., Demler, O., Walters, E.E.. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry* 2005;62(6):617–627.
- [379] Born, J., Beymer, D., Rajan, D., Coy, A., Mukherjee, V.V., Manica, M., et al. On the role of artificial intelligence in medical imaging of COVID-19. *Patterns* 2021;2(6):100269.
- [380] Pendse, S.R., Nkemelu, D., Bidwell, N.J., Jadhav, S., Pathare, S., De, M., et al. From treatment to healing: Envisioning a decolonial digital mental health. In: *Proc. CHI Conference on Human Factors in Computing Systems (CHI'22)*. 2022..
- [381] Palaniyappan, L.. More than a biomarker: could language be a biosocial marker of psychosis? *NPJ Schizophrenia* 2021;7(1):1–5.
- [382] Luo, Y., Peng, J., Ma, J.. When causal inference meets deep learning. *Nature Machine Intelligence* 2020;2(8):426–427.
- [383] Prosperi, M., Guo, Y., Sperrin, M., Koopman, J.S., Min, J.S., He, X., et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence* 2020;2(7):369–375.
- [384] Neuner, I., Veselinović, T., Ramkiran, S., Rajkumar, R., Schnellbaecher, G.J., Shah, N.J.. 7T ultra-high-field neuroimaging for mental health: an emerging tool for precision psychiatry? *Translational Psychiatry* 2022;12(1):1–10.
- [385] Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J.. AI in health and medicine. *Nature Medicine* 2022;28:1–8.
- [386] Grande, D., Mitra, N., Iyengar, R., Merchant, R.M., Asch, D.A., Sharma, M., et al. Consumer willingness to share personal digital information for health-related uses. *JAMA Network Open* 2022;5(1):e2144787–e2144787.
- [387] Davidson, B.I.. The crossroads of digital phenotyping. *General Hospital Psychiatry* 2022;74:126–132.
- [388] Korngiebel, D.M., Mooney, S.D.. Considering the possibilities and pitfalls of generative pre-trained transformer 3 (gpt-3) in healthcare delivery. *NPJ Digital Medicine* 2021;4(1):1–3.
- [389] Polyzotis, N., Zaharia, M.. What can data-centric ai learn from data and ml engineering? *arXiv preprint: arXiv:211206439* 2021;URL: <https://arxiv.org/pdf/2112.06439.pdf>.
- [390] Northcutt, C., Jiang, L., Chuang, L.. Confident learning: Estimating uncertainty in dataset labels. *Journal of Artificial Intelligence Research* 2021;70:1373–1411.
- [391] Chen, R.J., Lu, M.Y., Chen, T.Y., Williamson, D.F., Mahmood, F.. Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering* 2021;5(6):493–497.
- [392] Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., et al. Generative adversarial networks and its applications in biomedical informatics. *Frontiers in Public Health* 2020;8:164.
- [393] Geng, D., Alkhachroum, A., Bicchi, M.A.M., Jagid, J.R., Cajigas, I., Chen, Z.S.. Deep learning for robust detection of interictal epileptiform discharges. *Journal of Neural Engineering* 2021;18(5):056015.
- [394] Weldon, J., Ward, T., Brophy, E.. Generation of synthetic electronic health records using a federated gan. *arXiv preprint arXiv:210902543* 2021;.
- [395] Chen, L., Xia, C., Sun, H.. Recent advances of deep learning in psychiatric disorders. *Precision Clinical Medicine* 2020;3(3):202–213.
- [396] Zhang, Z., Cui, P., Zhu, W.. Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering* 2022;34:249–270.
- [397] Bronstein, M.M., Bruna, J., Cohen, T., Velicković, P.. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:210413478* 2021;.
- [398] Gennatas, E.D., Friedman, J.H., Ungar, L.H., et al. Expert-augmented machine learning. *Proc Natl Acad Sci USA* 2020;117:4571–4577.
- [399] Habtewold, T.D., Rodijk, L.H., Liemburg, E.J., Sidorenkov, G., Boezen, H.M., Bruggeman, R., et al. A systematic review and narrative synthesis of data-driven studies in schizophrenia symptoms and cognitive deficits. *Translational Psychiatry* 2020;10(1):1–24.
- [400] Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V.X., Finale Doshi-Velez, K.J., et al. Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine* 2019;25:1337–1340.