

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,*}

^a*Department of Psychiatry, New York University Grossman School of Medicine, New York, NY 10016, USA*

^b*Department of Neuroscience and Physiology, New York University Grossman School of Medicine, New York, NY 10016, USA*

^c*The Neuroscience Institute, New York University Grossman School of Medicine, New York, NY 10016, USA*

^d*Department of Biomedical Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201, USA*

^e*Headspace Health, San Francisco, CA 94102, USA*

^f*Meta Reality Lab, New York, NY, USA*

^g*Department of Bioengineering, Lehigh University, PA 18015, USA*

^h*Department of Electrical and Computer Engineering, Lehigh University, PA 18015, USA*

Abstract

The bigger picture: Machine learning (ML) and artificial intelligence (AI) have become increasingly popular in analyzing complex patterns of neural and behavioral data for medicine and psychiatry. We provide a comprehensive review of ML methodologies and applications in precision psychiatry. We argue that advances in ML-powered modern technologies have revolutionized the current practice in diagnosis, prognosis, monitoring and treatment of various mental illnesses. We discuss conceptual and practical challenges in precision psychiatry and highlight future research in ML.

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 review, we provide a comprehensive review of the ML methodologies and applications by combining neuroimaging, neuromodulation, and advanced mobile technologies in psychiatry practice. Additionally, we review the role of ML in molecular phenotyping, cross-species biomarker identification in precision psychiatry. We further discuss explainable AI (XAI) and causality testing 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 is epidemic in the United States and the world. According to the National Institute of Mental Health (NIMH), nearly one in five American adults suffer from a form of mental illness or psychiatric disorder (www.nimh.nih.gov/health/statistics/). According to the Centers for Disease Control and Prevention (CDC), The COVID-19 pandemic has witnessed a significant impact on our lifestyle, and considerably elevated adverse mental health conditions caused by fear, worry and uncertainty [1]. 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. For instance, the economic burden of depression alone is estimated to be at least \$210 billion annually (www.workplacementalhealth.org/mental-health-topics/depression/). Precision medicine (or personalized medicine) is an innovative approach to tailoring disease prevention, diagnosis, and treatment that account 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 eval-

*Corresponding authors:

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

uation. Meanwhile, precision psychiatry has increasingly received its deserved attention [2, 3]. 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 [4]. The current gold standard for diagnosis and treatment outcome in mental disorders—the DSM (Diagnostic and Statistical Manual of Mental Disorders) maintained by the American Psychiatric Association (APA), are 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 [5, 6, 7]. 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 issues [8]. A recent global survey has indicated that psychiatrists were somewhat skeptical that AI could replace human empathy, but many predicted that 'man and machine' would increasingly collaborate in undertaking clinical decisions, and psychiatrists were optimistic that AI might improve efficiencies and access to mental care, and reduce costs [9].

The past two decades have witnessed substantial growth of ML applications for psychiatry in the literature, reflected in many applications and reviews [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. 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 technological and methodological developments in ML, multimodal neuroimaging, neuromodulation, large-scale circuit modeling and human-machine interface. It is noteworthy that our reviewed literature is by no means exhaustive due to space limitation. 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 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 neurostimulation. The integration of ML with advanced neuroimaging can potentially help us identify and validate biomarkers in diagnosis and treatment of mental illnesses. Third, there is an increasing demand of psychiatrists in the US, and the shortage is even more acute in poorer countries [21]. ML/AI technologies may change the practice of psychiatry for both clinicians and patients. Finally, advanced technologies such as social media, multimedia (speech and vision), 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. From now on, we will use ML and AI interchangeably throughout the paper.

2. Background of Neuroimaging

2.1. Mind vs. Brain

The brain is a physical organ, which provides the center that supports all cognitive functions. It can be considered as the hardware of the human body. In contrast, the mind is abstract; it creates emotions and enables consciousness, perception, thinking, judgment, and memory. The World Health Organization (WHO) defines mental health as "*a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community.*" Psychiatry is seeking to measure the mind, and relies on the quantification of how people feel under specific tasks or behavioral conditions. Psychiatric disorders are often described as disorders of the mind, which disrupt the brain's ability to function normally to complete certain processes. For instance, anxiety, stress, depression, autism, obsessive-compulsive disorder (OCD), and post-traumatic stress disorder (PTSD) are categorized by varying degrees of psychomotor, cognitive, affective, and volitional impairment. Since the brain and mind are internally interleaved, the syndromes of mental disorders are associated with dysfunctions of neural, cognitive, and behavioral systems [2]. To unravel mysteries of the mind, we need to understand the brain based on modern neuroimaging techniques.

2.2. 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).

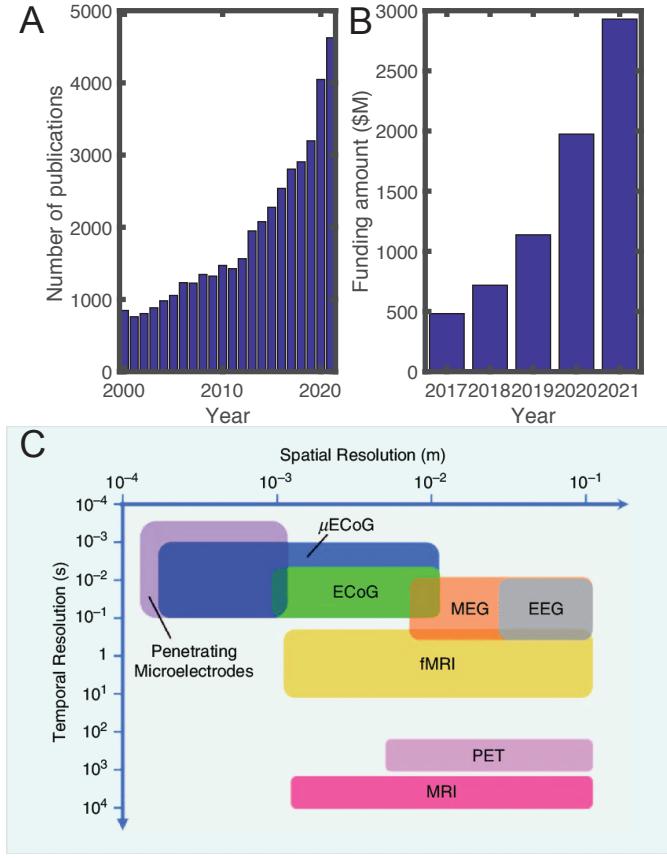


Figure 1: (A) The number of PubMed publications with keywords "machine Learning or AI" and "psychiatry or mental health" in the title or abstract (Year 2000-2021). (B) Growth of mental health tech funding in the US market (Year 2017-2021, data source: cbinsights.com). (C) Human neuroimaging at various spatial and temporal resolution (© IEEE, figure reproduced from [22] with permission).

Our understanding of brain and behavior relationships has expanded exponentially over the last few decades. While this improvement may be attributed to a multitude of factors, advancement in neuroimaging has played a prominent role [23]. Ranging from increased utilization of structural neuroimaging techniques to the significant scientific advancements brought about by the increased availability of functional neuroimaging, these technologies have provided significant benefits to improved understanding of neural correlates and discovery of biomarkers in psychiatric disorders [24, 25]. Some of the most common neuroimaging methods for probing brain function include the utilization of magnetic resonance imaging (MRI), functional MRI (fMRI), diffusion tensor imaging (DTI), electroencephalography (EEG), Magnetoencephalography (MEG), electrocorticography (ECoG), functional near-infrared spectroscopy (fNIRS), and positron emission tomography (PET).

- **MRI.** MRI is a non-invasive imaging technology that produces 3-D anatomical images and has been widely used in clinics for diagnosis, staging, and follow-up of brain disease [26].
- **DTI.** DTI is a technological advancement that allows re-

searchers and clinicians to assess human white matter pathways *in vivo* [27]. DTI has shown its promise in characterizing neuroanatomical connectivity underlying white matter tracts, which is hard to be captured by MRI.

- **fMRI.** fMRI is a hemodynamic technique to characterize functional activity of the brain by measuring blood oxygenation level-dependent (BOLD) signals [28]. Due to its good spatial resolution, fMRI has been extensively used for studying neurobiological basis of various psychiatric disorders.
- **PET.** PET is a functional imaging technique that uses radioactive substances known as radiotracers to visualize and measure changes in metabolic processes, and other physiological activities including blood flow, regional chemical composition, and absorption. The successful application of PET has been reported in biomarker studies for psychotic disorders [29].
- **EEG.** EEG measures electrical activity from neuronal populations via electrodes on the scalp. Due to its low cost and ease of data collection, EEG provides a practical tool to be used in a variety of clinical environments [30].
- **MEG.** MEG uses superconducting quantum interference devices to measure the magnetic fields produced by electrical activity in the brain. MEG is less affected by secondary currents and demonstrates superior spatial resolution compared with EEG [31].
- **ECoG.** ECoG can be viewed as invasive intracranial EEG that measures brain signals from the cortical surface. Because the subdural electrodes are placed directly on the cortical surface, the ECoG signals have better signal-to-noise ratio (SNR) and excellent spatial and spectral resolution compared with EEG [32]. ECoG has a long history of use in clinical neurosurgery since its initial application in epilepsy surgery, and has been recently used in research studies of mental disorders, such as depression and mood disorders [33, 34]. Micro-ECoG (μ ECoG) utilizes micro-scale electrodes with contact site diameters many orders of magnitude smaller than traditional clinical ECoG electrode sites and minimized inter-electrode spacing, allowing for even greater spatial resolution of measured brain signals.
- **fNIRS.** fNIRS is an optical brain monitoring technique that uses near-infrared spectroscopy to measure cortical hemodynamic activities for functional neuroimaging [35]. Alongside EEG, fNIRS can also be used in portable settings for studying psychiatric disorders [36].

To date, EEG and fMRI are two most commonly used modalities for precision psychiatry. Specifically, EEG is low-cost and easy-to-operate, making it more appealing for psychiatric practice or home use.

2.3. Neuroimaging Analysis

The 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, which have focused on the following aspects.

- **Task-related brain activation.** Under specifically designed cognitive paradigms, the collected neuroimaging data enable us to examine brain activities associated with certain experimental tasks and study their relationship with cognitive dysfunctions. Typical measures of task-related brain activities include event-related potential (ERP) and event-related spectral perturbation, and reward or emotional processing-related functional activation [37, 38].
- **Brain structural and functional connectivity.** A promising direction for probing brain function using neuroimaging is to investigate brain connectivity (or connectome) [39]. Studying the resting-state brain connectome provides a promising way to characterize the complex brain architecture and uncover brain dysfunctions in intrinsic brain networks [40].
- **Brain dynamics.** Increasing neuroimaging studies suggest that functional connectivity may fluctuate, rather than being stationary during an entire session of data collection [41]. Studies examining spatiotemporal dynamics of brain networks have recently received increasing attention and may reveal meaningful brain states associated with different psychiatric conditions [42].
- **Multimodal neuroimaging.** Another promising approach to establish robust biomarkers for psychiatry is to combine multiple neuroimaging modalities, which offers opportunities to exploit cross-modality complementary information that a single modality approach may not capture [43].

To fully understand the mind, 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. What and How ML Can Help Psychiatry?

3.1. *What Makes Psychiatry Different from Other Medicine Disciplines?*

The nature and etiology of mental illnesses remain unclear and challenging to study. Psychiatric disorders are typically diagnosed according to a combination of clinical symptoms based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Classification of Diseases (ICD). 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. This framework largely relies on clinically derived diagnostic labels, assuming that a biomarker may differ enough between healthy people and patients. The symptom-based diagnosis covered hundreds of thousands of different symptom combinations, which has caused extensive clinical heterogeneity [44, 45]. It is

increasingly recognized that existing clinical diagnostic categories could misrepresent the causes underlying mental disturbance. The traditional case-control design has limited strengths in delineating the significant clinical and neurobiological heterogeneity of psychiatric disorders, thereby hindering the understanding of psychopathology and the search for biomarkers. On the other hand, previous studies have broadly explored the group effects of neurobiology to explain its connection to behavior and disease. 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 limited in uncovering the neural correlates of comorbidity among multiple disorders and identifying the dimensions of neural circuits and behavioral phenotypes.

Distinct from the traditional case-control design, the NIMH's RDoC aims to address the heterogeneity and comorbidity in psychiatry by linking symptom dimensions with biological systems, cutting across the diagnostic spectrum [46]. The ultimate goal of RDoC is to find "new ways of classifying psychiatric diseases based on multiple dimensions of biology and behavior" [47]. These newly defined disease dimensions could further be utilized to discover neurobiological phenotypes and clarify the causal mechanism underlying the associated brain dysfunctions. To achieve this important vision, new analytical approaches are urgently needed. Thanks to the advancement in cutting-edge ML/AI techniques, psychiatrists and investigators can benefit from a deep understanding of complex patterns in brain, behavior, and genes [5]. Combining these analysis techniques with rich multimodal data from increasing large-scale multi-center cohorts holds significant promise in advancing a biologically grounded redefinition of psychiatric disorders.

Despite rapid progress in psychiatric studies, several areas appear highly underexplored but may carry the substantial potential for achieving major breakthroughs toward precision psychiatry. The capacity to dissect inter- and intra-individual variability is crucial for better understanding the neural basis of variation in human cognition and behavior [48]. Studies focusing on the level of the individual may find greater success over conventional group-level analyses. Translational study-orientated approaches for psychiatric neuroimaging may further enhance the ability to find statistically significant effect sizes that can be used in individuals [49]. On the other hand, identifying subgroups (i.e., subtypes) in psychiatric disorders offers a promising way to delineate disease heterogeneity. Increasing evidence suggests that data-driven subtyping may drive novel neurobiological phenotypes associated with distinctive behavior and cognitive functioning [50]. These stratified phenotypes may further show improved predictability for clinical outcomes than DSM/ICD diagnoses and serve as potential markers for

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	CNN, Deep autoencoder, GCN, RNN, LSTM, GAN	A large class of generic learning problems

treatment selection [51]. Another promising area focuses on transdiagnostic approaches to uncover neural correlates of specific domains, such as cognition, arousal, and emotion regulation, which are implicated in psychopathology across the diagnostic spectrum [52]. Recent ML efforts have been dedicated to identifying transdiagnostic brain dysfunctions and dimensions of psychopathology [53, 54, 55, 56]. 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 studies of psychiatric disorders [57, 58, 59]. Such longitudinal studies may help reveal the neural mechanism underlying the disease progression, and provide new insights for the development of timely interventions. These new frontiers in studying psychiatric disorders can be substantially empowered by ML methodologies summarized in Table 1, including stratifying patients into clinically meaningful subtypes, discovering novel transdiagnostic disease dimensions, and tailoring treatment decisions to individual patients. The research outcome can deliver a significant promise in promoting the development of objective biomarkers-based precision psychiatry.

The applications of ML in psychiatry can be mainly categorized as four types: 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. 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. In this section, 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.

3.2. 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 [96, 23, 97]. 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 dis-

ease [10, 98, 99]. These promising applications in psychiatric disorders have been enabled by leveraging the powerful strength of modern ML techniques [13, 100, 101, 102].

Supervised Learning. Supervised learning, as the most popularly used category, has been widely applied to neuroimaging-based predictive modeling tasks for psychiatric disorders [103]. Classic supervised learning algorithms include logistic regression, support vector machine, and random forest. Given the high-dimensional nature of neuroimaging data, these approaches are commonly accompanied by a feature selection step to obtain low-dimensionality representations. Connectome-based predictive modeling [104, 105] 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 the accurate prediction of a specific brain disease or clinical outcome [90]. Relevance vector machine (RVM) builds upon a probabilistic framework by leveraging automatic relevance determination to learn a sparse solution and penalize unnecessary complexity in the model [106, 107]. RVM has recently demonstrated its strength in quantifying neuroimaging biomarkers for PTSD diagnosis [65] and for treatment outcome prediction in depression [84]. As an extension of the conventional single-task methods, multi-task learning approaches have been increasingly employed to exploit complementary features jointly from multiple views of neuroimaging data [108, 109, 110].

Due to the complex nature of brain 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 recently developed to identify antidepressant-responsive brain signatures in major depression [83]. 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. 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

Table 2: Representative ML applications in psychiatry based on neuroimaging and clinical data.

Application	Method	Disease	Data type	Reference
Diagnosis	Classification (dynamic GCN)	ADHD	rs-fMRI + Phenotypic data	[60]
	Classification (Ensemble learning)	ADHD	Multimodal	[61]
	Classification (GCN)	ASD	Task fMRI	[62]
	Classification (Ensemble learning + GCN)	ASD	rs-fMRI	[63]
	Classification (PCA + LASSO)	Bipolar	DWI + Cognitive data	[64]
	Classification (RVM)	PTSD	rs-fMRI	[65]
	Classification (ICA + LSTM)	Schizophrenia	fMRI	[66]
	Classification (SVM)	Schizophrenia	sMRI	[67]
	Classification (CNN)	Depression	rs-EEG	[68]
	Classification (Autoencoder + MLP)	ASD	rs-fMRI	[69]
	Classification (GNN)	ASD	rs-fMRI + Phenotypic data	[70]
	Subtyping (Normative modeling + clustering)	PTSD	rs-fMRI	[71]
	Subtyping (CCA + Hierarchical clustering)	Depression	rs-fMRI	[50]
	Subtyping (Sparse K-means)	PTSD and Depression	rs-EEG	[51]
	Subtyping (Latent class analysis)	ADHD	Task fMRI	[72]
Prognosis	Transdiagnostic (Normative modeling + GP regression)	Multiple disorders	rs-fMRI	[73]
	Transdiagnostic (Sparse CCA)	Multiple disorders	rs-fMRI	[54]
	Transdiagnostic (PLS)	Multiple disorders	rs-fMRI	[55]
	SVM	Psychosis, Depression	multimodal	[74]
	LASSO	Psychosis	rs-EEG	[75]
	SVM	Depression	rs-EEG	[76]
	GP classifier	Depression	Task fMRI	[77]
	LASSO	Substance use	MR/task fMRI	[78]
	LSTM	PTSD	MEG	[79]
	DNN	PTSD	rs-fMRI / task fMRI	[80]
Treatment prediction	SVM	Schizophrenia	sMRI	[81]
	MLP	Schizophrenia	Task fMRI	[82]
	Latent space learning	Depression	rs-EEG	[83]
	RVM	Depression	Task fMRI	[84]
	SVM	Psychosis	sMRI	[85]
	SVM + GP classifier	Depression	sMRI	[86]
Readmission	SVM	ADHD	sMRI	[87]
	GP classifier	PTSD	MR/rs-fMRI	[88]
	MVPA regression	ASD	Task fMRI	[89]
	LASSO	Anxiety	rs-fMRI	[90]
	SVM	Schizophrenia	rs-fMRI	[91]
Readmission	SVM	Depression	multimodal	[92]
	Classification tree	Bipolar	EHR	[93]
	Ensemble learning	Substance use	Phenotypic data	[94]
	Growth mixture modeling	Depression	Clinical data	[95]

features have been successfully identified by applying canonical correlation analysis (CCA) to resting-state fMRI (rs-fMRI) connectivity and clinical symptoms [50]. 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 captured with the distinct neuroanatomical basis of common and mood-specific factors of the symptoms [111]. A sparse CCA approach has been applied to reveal linked dimensions of psychopathology and functional connectivity in brain networks for psychiatric disorders [54]. This approach successfully identified interpretable dimensions, involving mood, psychosis, fear, and externalizing behavior, guided by 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 connectivity [55]. The latent components defined distinct dimensions with dissociable brain functional signatures, thus providing potential intermediate phenotypes spanning diagnostic categories. These dimensional analytics hold great promise in uncovering novel transdiagnostic phenotypes for the development of targeted interventions.

Ensemble Learning. Though 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 or bias, ensemble learning improves prediction performance over a single model and has proven successful in the robust discovery of biomarkers for psychiatric disorders. For instance, multi-atlas ensemble-learning algorithms have been proposed for improved schizophrenia detection [112] and ASD diagnosis [63]. By utilizing multimodal neuroimaging including sMRI, fMRI, and DTI, a bagging-based SVM was devised to yield significant improvement in the prediction of adult outcomes in childhood-onset ADHD [61]. Based on the selective ensemble algorithm, a sparse multi-view prediction model has been designed with rs-fMRI connectivity for ASD diagnosis [113]. The 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 with the increasing model complexity. We will discuss the interpretable ML methods in more detail later (Section 7). Additionally, labeled data require the knowledge of the ground truth, which is not always accurate or reliable 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 for exploratory data 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 disorders [44]. With rs-fMRI, hierarchical clustering was applied to successfully identify four subtypes of functional connectivity in depression [50]. 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 psychotherapy [51]. As a non-distance probability-based clustering approach, latent class analysis has also been applied to discover subgroups in psychiatric disorders. A proof-of-concept study was conducted using latent class analysis to identify ADHD subtypes from fMRI activation profiles [72] 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 modeling [23]. 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 [114, 115]. 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 been growing in studying psychiatric disorders, such as depression, anxiety, and ASD. They offer flexible ways to identify latent subpopulations that manifest heterogeneous symptom trajectories [116, 117, 118]. LGMM approaches have also been successfully used to predict PTSD course among the population at risk [119]. As an extension of latent class analysis to longitudinal data, latent transition analysis has been applied to predict longitudinal service use for individuals with substance use disorder [120]. Together, these approaches provide powerful tools to delineate longitudinal heterogeneity and the corresponding distinctive phenotypes during the course of psychiatric disorders.

Semi-supervised Learning. Semi-supervised learning is an ML approach that combines supervised learning and unsupervised learning. Popular semi-supervised learning techniques include self-training, mixture models, co-training and multi-view learning, graph-based methods, and semi-supervised clustering [121]. These methods have been increasingly applied to psychiatric studies. By unifying autoencoder and classification, a semi-supervised model was developed for ASD diagnosis [122]. A semi-supervised classification has been devised using graph convolutional networks and applied to the population graph-based diagnosis of ASD [70]. A semi-supervised clustering has also been designed by extending SVM with implicit clustering driven by a convex polytope to form a method called heterogeneity through discriminative analysis, which can achieve joint disease subtyping and diagnosis [123]. This ap-

proach has shown strength in delineating neurostructural heterogeneity in bipolar and major depressive disorders (MDDs) [124], schizophrenia [125], as well as in youth with internalizing symptoms [126]. Additionally, semi-supervised learning has gained increasing mental health applications in digital data from electronic health records (EHRs), social media and mobile phones [127, 128, 129]. See Section 4 for a detailed discussion.

Semi-supervised learning also provides an elegant way through normative modeling [130] to characterize 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 [131]. With this tool, an association was successfully discovered between transdiagnostic dimensions of psychopathology and individual's unique deviations from normative neurodevelopment in brain structure [73]. By combining tolerance interval-based normative modeling and clustering analysis, individual abnormalities in rs-fMRI were accurately quantified to define two stable subtypes in patients with PTSD [71]. 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.

3.3. Deep Learning

Deep learning consists of a collection of methods that use artificial neural networks for machine learning 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 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 [134, 135, 18, 136].

Deep autoencoder. The deep autoencoder, also known as stacked autoencoder, 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 fMRI [137]. Clustering analysis with the latent features uncovered by deep autoencoder further identified two subtypes within major psychiatric disorders including schizophre-

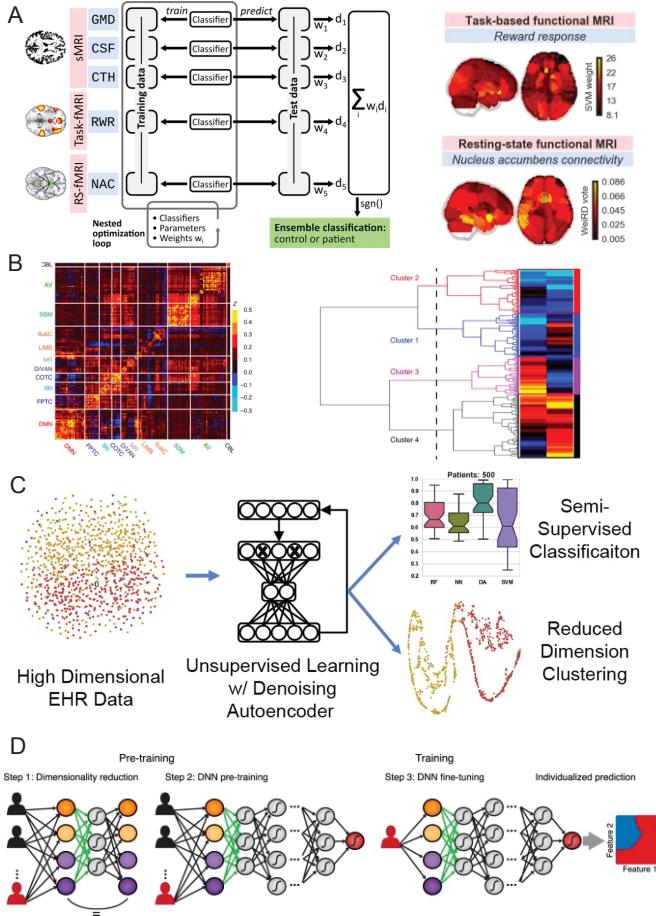


Figure 2: Various ML models for mental health applications. (A) Left: Multi-modal supervised classification scheme. Three modality-specific factors are optimized on the training data: classifier types, parameters and weights. The final diagnostic classification is based on a weighted sum of decision values, where weights correspond to those estimated during training. Right: Feature importance maps of functional neuroimaging modalities ([132]). (B) Unsupervised learning. Left: whole-brain functional-connectivity matrix averaged across all subjects. z = Fischer transformed correlation coefficient. Right: Hierarchical clustering analysis. (© Springer Nature, [50]. Figure reproduced with permission). (C) Semi-supervised learning pipeline for phenotype stratification based on EHRs ([133], 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 ([18], Creative Commons licenses 4.0).

nia, 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 diagnosis [138]. Deep autoencoder has also been applied to implement normative modeling with structural MRI for the quantification of individual abnormalities in neuropsychiatric disorders, including schizophrenia and ASD [139]. 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

variation.¹⁴⁰

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 [141, 102]. Due to its strength in utilizing neighborhood information to learn hierarchies of features [142], CNN has been one of the most successful deep learning models applied in various medical applications. A diagnosis model was established through EEG-based image construction coupled with the CNN for accurate detection of MDD [68]. 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 disorder [143].

Graph neural networks (GNNs). Though deep learning models have shown strengths in capturing complex neuroimaging patterns, they may not generalize well to non-Euclidean data types (e.g., brain networks). In contrast, GNNs provide a clever way of learning the deep graph structure of non-Euclidean data, leading to enhanced performance in various network neuroscience tasks [144]. For instance, a framework based on graph convolutional networks has been designed for the diagnosis of ASD [70]. By building a population graph that integrates rs-fMRI data as node features and phenotypic measures as edges, the designed model outperformed other state-of-the-art methods. An inductive GNN model was also devised to embed the graphs containing different properties of task fMRI and drive interpretable connectome biomarkers for ASD detection [62]. More recently, a novel GNN model was developed to incorporate dynamic graph computation and feature aggregation of 2-hop neighbor nodes into graph convolution for brain network modeling [60]. 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, RNN has the ability to learn features and long-term dependencies from sequential and time-series data. Long-short-term memory (LSTM) model is the most popular RNN and has shown its advantage in capturing temporal dynamic information of neuroimaging data for various psychiatric disorder studies [145]. An LSTM-based RNN architecture was built with the time course of fMRI-independent components to exploit the temporal information, which yielded an improved diagnosis of schizophrenia [66]. By combining RNN 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 network [146]. LSTM has also been applied to incorporate multi-stage neuroimaging data into longitudinal analytic frameworks for modeling the trajectories of psychopathology development in various psychiatric disorders. A recent LSTM-based model was

built with MEG data to achieve accurate longitudinal tracking of pathological brain states and prediction of clinical outcomes in PTSD [79].

Generative Adversarial Networks (GANs). GAN is a type of generative model, which has gained considerable attention in computer vision and natural language processing and also become increasingly popular in neuroimaging analysis [102]. GAN consists 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, GAN has been increasingly applied in data augmentation to enhance the sample size for model training [147]. Moreover, GAN has been used to impute missing values in multimodal datasets, a common problem in psychiatric studies, rather than discarding an entire multivariate data point [148]. 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 LSTM [149]. By integrating GAN with group ICA, a functional connectivity-based deep learning model was developed for the diagnosis of MDD and schizophrenia [150]. 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.

The strength of deep learning algorithms is that they can learn complex predictor-response mappings, but the power also comes at the cost of requiring a very large sample size for model optimization. This poses potential overfitting and interpretability challenges in psychiatric applications [18].

3.4. Key ML Concepts for Precision Psychiatry

Regardless of the ML paradigms in psychiatric applications, there are some common themes that distinguish between human intelligence and automated or human-in-the-loop machine intelligence. In a recently published white paper “*Machine intelligence for healthcare*”, four important features are emphasized for ML systems [151]. These concepts are broadly applicable to precision psychiatry [11].

- **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 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 8.

3.5. Case Studies

To help 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. A robust neurobiological signature for an antidepressant-responsive phenotype is still lacking to determine which patients will benefit from medications. To address the challenge, Wu et al. [83] 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. [51] developed a data-driven framework to achieve simultaneous feature selection and subtyping on the high-dimensional

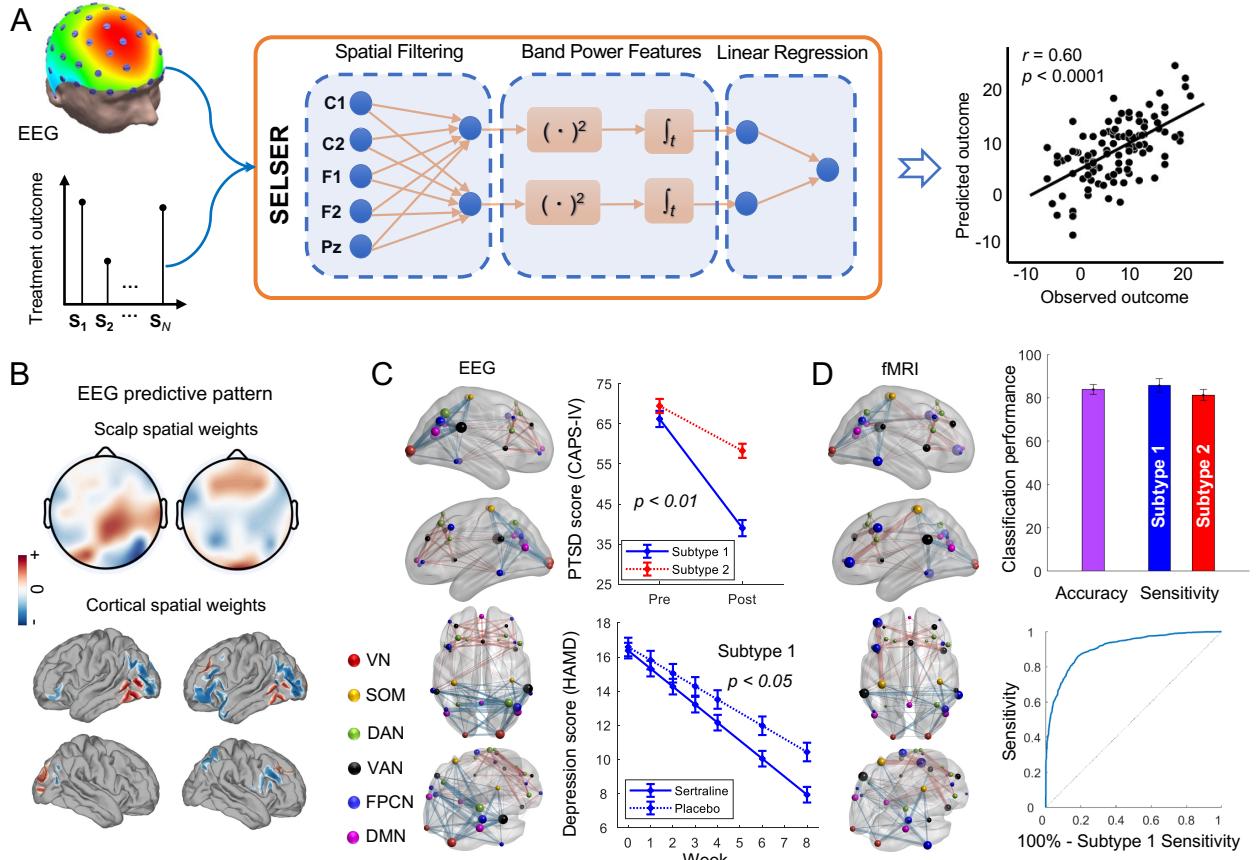


Figure 3: Concepts and major findings in case studies 1 and 2. (A) Illustration of the sparse EEG latent space regression (SELSER) framework in Case Study 1 for treatment outcome prediction. (B) Interpretive cortical pattern derived from the scalp pattern (© Springer Nature, figures are modified from [83] 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 from [51] with permission).

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. [152] 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 multiple days in seven epilepsy patients, Sani et al. [33] 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 gen-

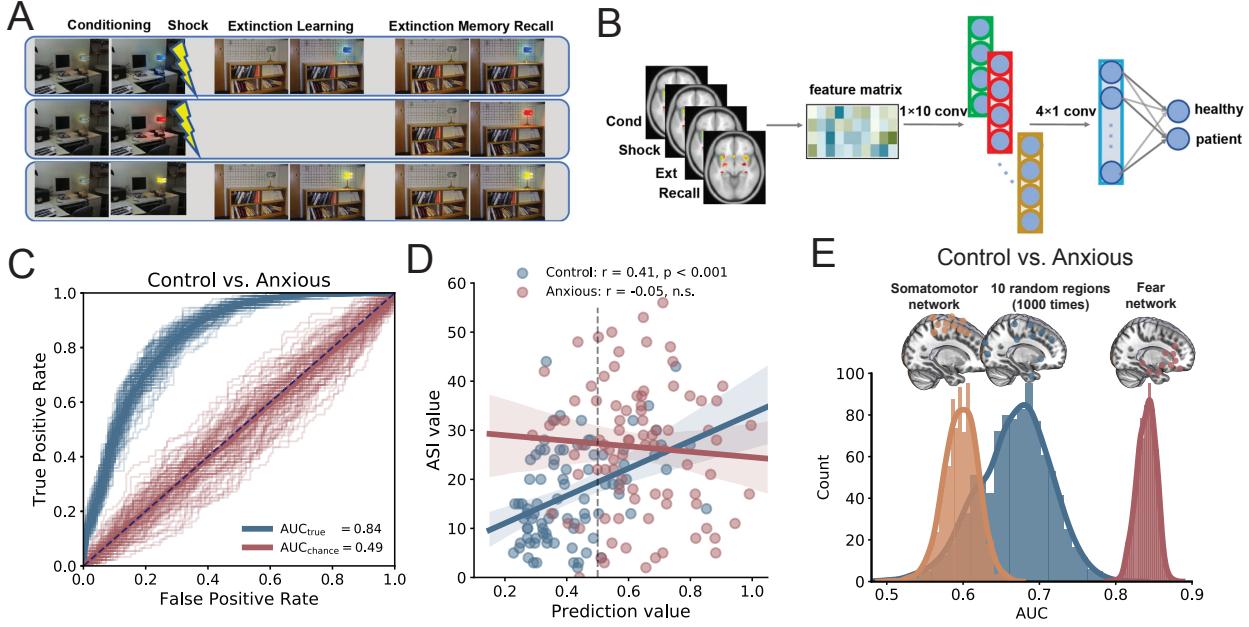


Figure 4: Illustrations of concepts and major findings in case study 3. (A) Experimental paradigm. (B) Schematic of the CNN. (C) AUC curves produced by CNN vs. chance level. (D) The prediction score positively correlated with the anxiety sensitivity index (ASI) for the control group ($r = 0.41$, $p < 0.001$), 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 [152] with permission).

eralized 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 and their underlying ML technologies for precision psychiatry.

A recent McKinsey study showed that use of telehealth has increased by 38-fold as compared to the pre-COVID baseline [153]. With a steep increase in teletherapy demand and consumption, many companies (such as Talkspace and Headspace Health) provide services on chat conversations with licensed mental health professionals. The definition of teletherapy 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, speech, and video analysis technologies, combined with the ML tools, have generated numerous innovations in the

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 [9].

ML can be applied across all stages of a patient's journey [154, 155]: risk assessment, diagnosis, prognosis, treatment, and remission in a variety of disorders [156], where the analytics can be applied to natural language, speech, facial expressions, body language, social media, as well as traditional clinical surveys and neuroimaging data [14, 157]. 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 [5], as opposed to traditional models that are only optimized for group effects. Furthermore, given the inter and intra-disease variability between clinical diagnosis and symptoms, ML can be used to model the differential diagnosis between disease categories using methods like multi-task learning. All of these mentioned ML applications can be considered to be the first level of precision added to ML-powered psychiatry.

However, the amount of precision that can be modeled using ML is far beyond the first level [158, 159]. 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 in a very short time, define what "normal" looks like for the patient, and identify deviations from that normal. This is often done by asking the patient questions and examining

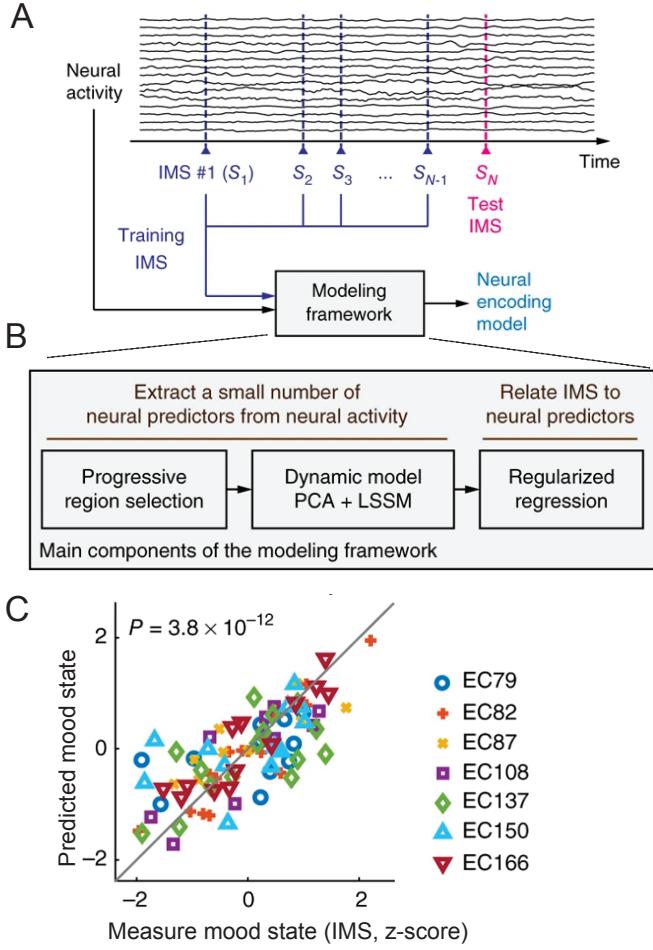


Figure 5: Illustrations of concepts and major findings in case study 4. (A) Schematic of cross-validation. An IMS point (e.g., S_N) is left out as the test IMS to be predicted. The other IMS points (i.e., training IMS, using S_1 to S_{N-1}) and the associated neural activity are used within the modeling 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 [33] with permission).

their speech, body language, and behavioral responses. It is very challenging and almost unrealistic to expect psychiatrists to build an accurate baseline model of the patient's entire life in such a short time span whilst interacting with the patient in a compromised psychological state. ML can help by building baseline models specific to each patient before their visit and present the bounds for various observations as a reference to psychiatrists during the exam [160]. This can be viewed the second level of precision in psychiatry that can be made possible by ML (Figure 6A). Take MDD as an example, Figure 6B shows how ML can be applied at different stages of a patient's journey. Similar applications have also been developed in studies of other disorders.

In the following subsections, we describe how ML technologies 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 smart phones, smart watches and other wearable sensing devices have enabled us to access more information of our physical and mental health than ever [177]. 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: phone on/off status, sleep duration, sleep staging

These signals have different implications and relevance to mental illnesses. Although none of single signals is indicative of mental disorders, 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 and on their mental health applications.

4.2. Speech and Video Analyses

Voice and visual (video of facial expressions and body language behaviors) data have recently gained increasing attention in the studies of mental disorders. ML technologies using speech samples obtained in the clinic or accessed remotely may help identify biomarkers to improve diagnosis and treatment. In the early stage of practice, psychologists have already used auditory and visual cues to assist the mental illness diagnosis [178]. Furthermore, speech and video are not only the readily available in traditional teletherapy settings but also the most 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 health disorders [171, 179], including speech analysis in patients with depression, bipolar, and schizophrenia. Table 4 lists some commonly used acoustic features in the analysis of mental illnesses [180]. These categories have enabled standardization and interpretation of ML-analyzed speech data in clinical applications.

It has been shown that models built from speech-based features are effective in predicting the diagnosis of depression and suicidality [181]. Applications for depression include predicting the presence, severity, or score [161, 169]. 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 between healthy, depressed, and suicidal speech.

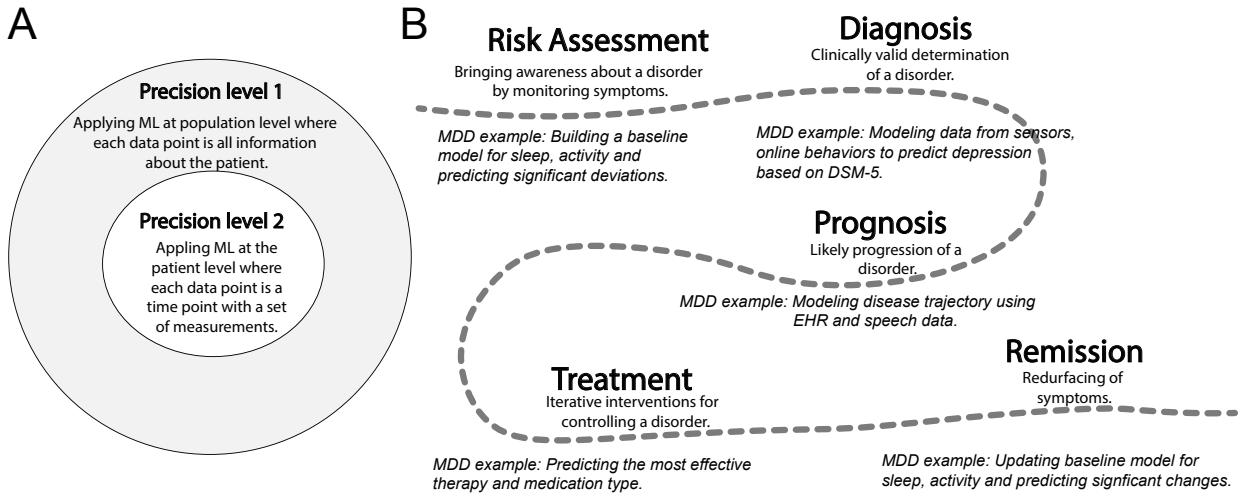


Figure 6: (A) Two levels of precision in applying ML for mental health. (B) Examples of ML applications at various stages of a patient’s journey in case of major depressive disorder (MDD).

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				
[161]	Audio - clinical interviews	Diagnosis	CNN ensemble	47 speakers
[162]	Audio - answers to personal questions	Diagnosis	Transfer learning	3078 speakers
[163]	Audio - clinical interviews	Diagnosis	SVM w/ speech landmark features	47 speakers
[164]	Face video -reading & personal questions	Diagnosis	CNN	50 videos
[165]	Gait-only video - casual walking in a corridor	Diagnosis	LSTM+CNN weighted fusion	40 videos
[166]	Language - answers to personal questions	Diagnosis	LSTM fine-tuned w/ health forum data	2425 subjects
[167]	Language - Facebook posts	Risk assessment	Logistic regression	68 patients
[168]	Audio, video - clinical interviews	Diagnosis	Transformer + multimodal fusion	56 subjects
Bipolar spectrum				
[169]	Audio - verbal fluency tasks	Remission	SVM	56 subjects
[170]	Sensor - GPS	Diagnosis	Linear regression	36 subjects
PTSD				
[171]	Audio - clinical interviews	Diagnosis	Random forest	43 veterans
[172]	Audio, video, skin conductance	Remission	SVM	110 subjects
Schizophrenia spectrum				
[173]	Audio - clinical interviews	Diagnosis	SVM	70 subjects
[174]	Video - neutral open-ended questions	Diagnosis	Logistic regression	16 subjects
[175]	Language - internet search queries	Remission	Random forest	23 subjects
[176]	Audio, video - clinical interviews	Diagnosis	Gradient boosting	17 subjects

One key challenge in applying speech-based models in clinical practice is the lack of longitudinal validation of data acquired in real-world settings. However, this issue is starting to get addressed in recent studies [182], which detect manic and depressive speech from recordings of outgoing speech from phone conversations of consenting participants. Another remaining challenge is the lack of large labeled datasets for evaluating performance across various methods. To this end, it is noted that recent efforts backed by companies like Ellipsis Health [183, 162], have used deep learning and transfer learning to predict depression and anxiety scores with high accuracy based on a large labelled dataset of over 10,000 unique speakers. Human-level accuracy in detecting depression using only 20-30 seconds of audio clip has been reported in some commercial applications [184, 185].

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 been targeted to suicidal ideation [186], depression [164, 187, 165], schizophrenia [174] and autism spectrum disorders [188]. Features derived from overall facial expression, eyes, gait, and posture (Table 4) have been found to be relevant across all of these 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 [189] have also involved fusion of video features derived from each frame that are used to train a sequential deep network and most have used pre-training to account for the relatively small size of depression datasets [190]. 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,

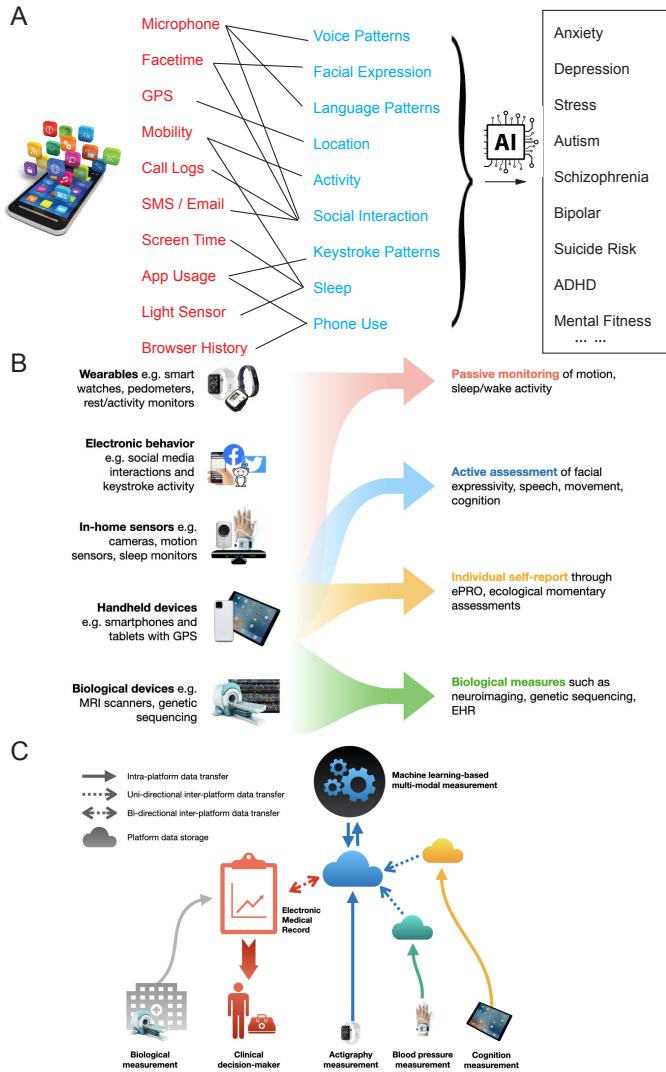


Figure 7: Illustrations of ML-powered technologies for mental health. (A) ML applications in mobile health. (B) Different types of data collection strategies for digital measurement tools. (C) A technological infrastructure for the integration of digital measurement tools. Independent platforms for measurement 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 from [157] with permission).

few researchers have built depression activation maps to highlight the facial areas corresponding to depression severity as learned by the model [191]. Meanwhile, using predefined features has been most successful in providing interpretable results [192, 190].

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. With NLP, mental health professionals can evaluate patterns in language to help identify and predict psychiatric illness in patients (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 [193, 194], but also more abundantly available compared to speech data. For example, social media platforms contain a large quantity of real-world language data, whereas the matched scale of speech data is limited. There are two sorts of NLP applications for detecting specific mental health symptoms. The first type of applications is directly applied to patients, varying from predicting the risk of suicide and early psychiatric readmission to identifying phenotypes and comorbidities. The second type of applications is indirectly applied to EHR and clinical records (tests, transcripts), which can be used for automating chart reviews, clustering patients into phenotype subtypes, predicting patient-specific outcomes. 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 clinical notes.

Massive EHR datasets have provided opportunities to adapt computational 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 EHRs [195]. The EHR language is at least one level abstracted from the patient's symptoms, consisting of clinical notes by physicians. However, the unique advantage of EHR data is the ease with which demographic and socioeconomic features can be combined with language data. Symptoms derived from the free text in EHR data can be used to predict a diagnosis for bipolar disorder [196], situational aggression [197], and suicidal ideation [198], with reported performances comparable to clinicians. Furthermore, discharge summaries from EHRs have also been used to predict remission [199]. Aside from symptoms, a variety of relevant mental health data such as intervention status and physical health comorbidities can be routinely extracted from EHRs using NLP methods [200]. Privacy concerns around EHR data sharing remain one of the key challenges in validating generalization of NLP methods. Encouragingly, there has been growing interests in using transformers for generating artificial mental health clinical notes to mitigate this issue [200, 201].

The advances in text-based mental health interventions (e.g., Talkspace and CrisisTextLine) have made transcripts of clinical sessions easily amenable via NLP. Aside from developing models for detecting suicide ideation [202], NLP can also be applied to these datasets to understand the population-level trend, such as the increase in anxiety and decrease in quality of personal relationships during the COVID-19 pandemic [203]. Since language data are ubiquitous, one of the challenges in applying NLP for advancing mental health 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 speech [204]. On the other hand, data collected "free-speech" samples for diagnostic purposes has been found to be highly effective in developing a language-

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

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

based depression screening that generalizes well across various age groups [166, 205, 183].

4.4. Social Media

To date, social media companies have collected a wide variety and a large amount of language data which may contain clinically-relevant information (Table 5). This information can not only be extracted on a population level, such as the notable rise in cognitive distortions over time [206], but also be attributed on an individual level [207], making 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 diagnosis at hospital [167]. Models applied to Facebook and other social platforms (e.g., Twitter and Reddit) have been successful in predicting diagnosis of psychosis, anorexia, anxiety, and stress levels [208, 209, 210]. Aside from language present in the user posts and comments, ML models often process media data such as Instagram images [211], or integrate images and text to infer the user’s state-of-the-mind [212, 211]. 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 build NLP models generalizable across languages and apply them to social media data, for example, to detect depression or self-harm [213, 214]. Furthermore, specialized language representations that were trained on mental health specific conversations and became publicly available [215], have been shown to improve performance compared to non-specific representations. Finally, transformer embedding can be applied to pre-identified sections in language which correspond to responses to standard clinical assessments such as the subjective well-being scales [216], supporting the high-accuracy prediction of standard survey scale responses without directly running the survey.

While social media solves the scale issue with millions of samples available, most social media data lack clinically-valid labels [217]. Most work has relied on using self-disclosure of

Table 5: Social media data access (OAuth 2.0 is the industry-standard protocol for authorization).

Platform	Examples of accessible data	OAuth 2.0 required
Instagram	Media, captions, total posts	Yes
TikTok	Videos, descriptions, likes	Yes
SnapChat	Bitmoji, public stories/media	Yes
Twitter	Tweets, timestamps, likes, retweets	No
Pinterest	Pins, boards	Yes
Kik	Messages through a bot	Yes
Tumblr	Blogs, profile, likes	Yes
Reddit	Posts, post timing	No
Discord	Conversations on a server	Yes
Facebook	Posts, media	Yes
YouTube	Watch history, comments	Yes
Twitch	Chat history	Yes
Google	Query Strings	Yes

mental illness as the labels, which tend to be noisy and may bring the additional issue 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 connected devices equipped with ambient sensors (Figure 7B) are increasingly capable of recording physiological measurements that are known to affect mental health [218]. In addition, some of the less obvious measurements (such as keystroke usage patterns) have been shown to be implicated by mental illness [219, 220]. 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 [221, 222].

These measurements from mobile sensors (Table 6) form valuable sources of mental health data, and can be useful at various levels of granularity—from raw sensor data (e.g., the accelerometer) to derived high-level features (such as psychomotor activity). This has inspired many corporations to invent technologies for detecting depression and cognitive decline based on data collected from their wearable devices [223].

Table 6: 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 7: Commercial and research platforms and services for mental health applications.

Platform	Primary data source	Mental health appl.
WoeBot [229]	Language	Depression, Anxiety
Mindstrong [219]	Keystrokes	Serious mental illness
Sonde Health [163]	Voice	Mental fitness
Ellipsis Health [162]	Voice	Stress
Amazon Halo [230]	Voice	Emotion detection
Apple Watch [185]	Mobility, Sleep	Depression
Alphabet Fitbit [231]	Skin conductance	Stress
Kintsugi [232]	Voice	Depression, Anxiety
Bewie [226]	Raw data from smartphone	Multiple
MindDoc	Language (ask daily question)	Depression
Clarigent Health	Voice	Suicide risk

Sensor-based measurements are found to be correlated with high-stress levels and a variety of ailments including depression, anxiety, psychosis, and bipolar disorder [224, 225]. Since sensor-based data are widespread and readily available, they offer an opportunity to build baseline models for individual users; these baseline models can be used to identify significant physiological changes in users and further inform clinical interventions.

Digital phenotyping for individuals is based on data acquired from mobile device collected in real time. Devices that collect data streams from patients, such as surveys, cognitive tests, social medial interactions, GPS coordinates, and behavioral patterns (e.g., keyboard typing), have great potentials for monitoring, managing and predicting the individual’s mental health [226, 227]. Overall, continuous quantification of these data streams may result in clinically useful 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 [228].

4.6. Commercial and Research Platforms and Services

While studies have demonstrated promising results in using ML to support the patient’s journey in mental health, the applicability in clinical practice remains limited. Table 7 lists examples of platforms and services that use ML for mental health. While most platforms focus on developing risk assessment based on single modality, the initial commercial viability of these platforms is promising for the success of using ML in mental health since they enable collection of large amounts of data which can be used in further development of biomarkers.

Many areas of mental health technology development focus on scalability in clinical research. For example, there are between 10,000 to 20,000 smartphone apps that digitize mindfulness or cognitive behavioral therapy techniques [233], allowing

the user to engage in psychotherapy on their own 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 inter-related areas of development: digital measurements and digital interventions.

Digital measurement applications. We are entering a new era of digital psychiatry [8, 234]. In 2016, the Harvard professor Jukka-Pekka Onnela coined the term *digital phenotype* [235], 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 that were developed in laboratory settings such as direct quantification of motor (i.e. movement, muscle activation) and physiological activity (i.e. heart rate, electrodermal 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 [236, 237, 238]. 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-depressants such as ketamine and psilocybin/psilocin 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/psilocin, and ketamine, act on serotonin receptors that ultimately impact peripheral motor and physiological activity [239, 240]. 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 risk [241], SSRI response in MDD [242], negative symptomatology in schizophrenia [243], and Parkinsonian tremor [244]. Such approaches are now being commercialized for all phases of drug development from proof-of-concept to direct measurement 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 can provide a sensitive numeric value.

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 increase in physiology 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 and rarely sufficiently transparent in their approach to be used for regulatory approval.

Digital interventions. The other rapidly emerging area of mental health technology are digital approaches to clinical care. We will briefly outline some of the leading approaches. Importantly, digital approaches to clinical care are often aligned with a digital measurement approach as these approaches are "blind" without some sort of remote data. A number of companies such as Mindstrong Health [219], IesoTrigger Health [245], and Headspace Health [246, 247], 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 development of models that can identify patterns in patient's and clinician's language that are markers of improved outcomes [248], which can be further used to measure success of various therapy modalities, treatment design, as well as to improve care quality [249].

Digital therapeutics proposes the use of mobile devices to offer automated cognitive behavioral therapy (CBT), mindfulness, or other validated psychotherapy in an automated fashion. These app-based approaches undergo the same clinical validation process and 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 substance use disorder (SUD), ADHD, schizophrenia, ASD, MDD, PTSD, and generalized anxiety disorder (GAD) [250].

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 providers [251]. 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 care [252]. These services, which are accessible directly to consumers, or more often, provided by a third-party payer, provide access to coaches and clinicians via different mobile platforms including text, voice, and video commu-

nication. Services vary from mental health coaching provided by lay-professionals to psychiatric and psychological services. While still in their infancy, early evidence has shown that tele-health services can perform at parity with traditional in person therapy [253]. Therefore, teletherapy is likely to be a dominant form of mental health treatment in the future.

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, the 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 [254, 255, 43]. 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 [256, 257, 43, 258, 259]. 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 modalities [260]. For a more general setting, multiset/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 [261, 43]. 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 [262, 263].

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 [264, 257]. 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 matrix [265]. The jICA approach involves concatenating modality features alongside each other and then performing ICA on the composite feature matrix [266]. 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 [257, 256]. 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 decomposition [267], 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) [268] and joint and individual variation explained (JIVE) [269] models have been proposed to achieve this goal. By jointly decomposing multiple feature matrices, CIFA and JIVE are able to simultaneously estimate common and individual feature subspaces. A further extension of CIFA has been achieved by leveraging high-order tensor factorization [264], 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 [270, 271]. 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 [272, 273, 274].

Deep Learning-based Fusion. Empowered by cutting-edge deep learning techniques, emerging methods have been increasingly developed for deep multimodal fusion [275]. Data fusion through deep learning allows integrating multiple modalities based on learned high-level feature representations that are theoretically more comparable to each other and more informative for predicting the targets [276]. 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 prediction [277]. Graph neural networks 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 representation [278]. By extending graph neural networks to multimodal structures, deep representation approaches have also been designed for integrating brain networks constructed from diverse modalities [279, 280, 281].

5.2. Multimodal Neuroimaging Studies

Neuroimaging data types are intrinsically dissimilar in nature, having different spatial and temporal resolutions [258].

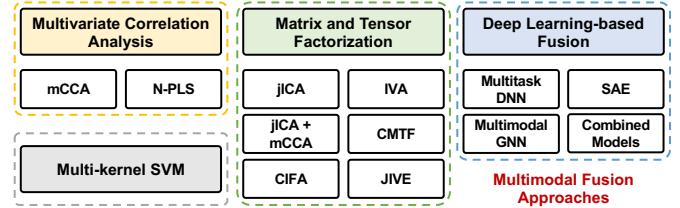


Figure 8: Summary of typical approaches for multimodal data fusion in psychiatry studies.

Instead of entering the entire data sets into a combined analysis, an alternate approach is to reduce each modality to low-dimensional (latent) features of selected brain activity or structure and then explore associations across these feature sets through variations across individuals. Exploiting such latent feature representations from multiple neuroimaging modalities for diagnosis has generally shown to improve performance compared to using a single modality alone [282]. Multimodal fusion allows for the 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, treatment, prognosis, and planning in many applications in medicine.

A combination of mCCA and jICA was successfully applied to fMRI and DTI fusion in the diagnosis of schizophrenia and bipolar disorder [283]. CMTF has been applied to identify diagnostic biomarkers of schizophrenia by integrating sMRI, fMRI, and EEG [284]. MKL-SVM has been successfully applied to integrate multimodal structural neuroimaging for predicting differential diagnosis between bipolar and unipolar depression [285] and to combine sMRI and fMRI for improved classification of trauma survivors with and without PTSD [286]. More recently, it also showed efficacy in the diagnosis of early adolescent ADHD by integrating sMRI, fMRI, and DTI [287]. In learning low dimensional representations of functional and structural MRI [288], the functional MRI can be split into several independent component networks, each treated as a separate modality along with the structural scan for learning using autoencoders. Furthermore, MKL methods have been used for diagnosing schizophrenia by combining markers from MRI and DTI [289]. A multimodal graph convolutional network was designed to integrate functional and structural connectomics data for an improved prediction of phenotypic characterizations in ASD [279]. By combining multiple typical neural network structures, multimodal deep learning models have also been developed to effectively integrate fMRI connectivity and sMRI features [290], and also genomic data [291] for discovering schizophrenia-associated brain dysfunction. Methods for learning joint representations from neuroimaging and non-neuroimaging data are still in early development [292] and there is an opportunity for ML methods to evolve for this task. For example, transformer networks with late fusion can be used to learn joint representations from various modalities such as EEG and eye movement signals [293].

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, social media, speech, video, sensor data carries the most phenotypes for mental illness, it only makes sense to combine the information from these data sources. In addition, this also 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 multi-task learning methods have been used to jointly learn from sensor and smartphone usage data to predict subjective well-being [294]. The success of transformer networks in jointly modeling video, speech, and language data has catalyzed multimodal modeling in mental health [295]. Multimodal modeling techniques can also be used in modeling symptoms such as emotion dysregulation [296], loneliness [297] and sentiment analysis [298]. 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 states [74]. 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 severity [299].

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 therefore is a vulnerable organ subject to changes in multiple interacting biological networks at the systems level. ML methods have been contributing enormously to capturing the complexities of interacting variables within and across multiple levels (Figure 9A) – particularly at the molecular level – for identifying mechanistic-based phenotyping models as new targets for the prevention and treatment of mood and cognitive disorders. The advent of unbiased next-generation sequencing (NGS) prompted the development of advanced 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, there is increasing recognition of the utility of ML methods to integrate these multi-level molecular datasets with clinical characteristics to map specific neurobiological substrates into the complexity of symptom clusters, which may aid the classification of diseases, prediction of treatment outcomes, and selection of personalized treatment. Emerging avenues in molecular neuroscience and computational psychiatry for new mechanistic models for the prevention and treatment of CNS disorders.

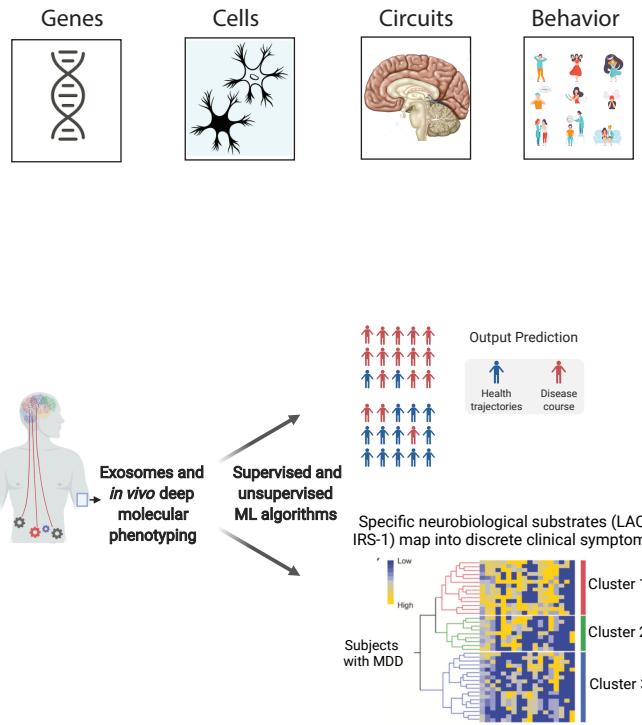


Figure 9: (A) Different levels of interacting variables (genes, cells, circuits) to behaviors in mental illnesses. (B) Combining ML with novel molecular biology technologies (for deep molecular phenotyping of brain plasticity) opens up opportunities to develop new mechanistic models for prevention and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

Animal models have been playing a vital role towards precision psychiatry in understanding mechanisms of diseases and predicting treatment responses [300, 301]. To bridge the scientific knowledge gap from animals to humans, gene expression studies offer an example of how integrating basic neuroscience, ML, and bioinformatics approaches can contribute to advancing understanding of the molecular basis of MDD. Using RNA sequencing (RNA-seq) 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, such as the ventromedial prefrontal cortex (PFC) and dorsolateral PFC among others, in post-mortem tissues of subjects suffering from MDD as compared to age and sex-matched controls with remarkable sex differences in these molecular pathways [302]. Recent work using RNA-seq assays at single nucleus resolution (snRNA-seq) and t-distributed stochastic neighbor (t-SNE) embedding analyses showed cell-type specific transcriptomic profiles in the post-mortem dorsolateral PFC that are differentially regulated in MDD cases as compared to respective controls, with the greatest gene expression changes in deep layer excitatory neurons and immature oligodendrocyte precursor cells [303]. Importantly, these gene expression studies in humans were supported by findings in rodents showing a brain that continually changes with experience [304]. Several groups using RNA-seq assays and bioinformatic analyses 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 particularly sensitive not only to the effects of stress [305] but also a target for the responses to next-generation antidepressants [306, 307].

The expansion of NGS to single-cell resolution assays requires more advanced bioinformatics analyses, which utilize ML to analyze these large datasets, including denoising and dimensionality reduction, cell-type classification, gene regulatory network inference, and multimodal data integration [308, 309]. Bioinformatic approaches, such as Seurat, combine unsupervised nonlinear dimensionality reduction, K-nearest neighbor graph analysis for cell-type clustering, and weighted nearest neighbor analysis for multimodal data integration [310]. Deep learning approaches, such as the autoencoder, provide another example of tools used for denoising and dimensionality reduction with computational scalability [311, 312, 313]. Autoencoders can also be used in a supervised manner for transfer learning across datasets, e.g., to learn the embedding from a larger, already annotated dataset and transfer this knowledge to cluster new datasets [314]. The combination of multimodal data generated from the simultaneous assessment of transcriptomic profiles with regulatory landscape or spatial location in the same single cell [315, 316, 317, 318] will allow a deeper molecular characterization of discrete cellular states [310].

Integrating multidimensional factors for new mechanistic treatment models. It has been increasingly recognized that mood and cognitive disorders are unlikely to be only brain-based diseases. Additionally, growing evidence suggest that they are system-level disorders affecting multiple interacting biological pathways [319], 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 showed that, at least in these cases, the specific neurobiological substrates map into discrete clinical symptoms, including anhedonia [320]. Furthermore, the integration of multidimensional factors spanning mitochondrial metabolism, cellular aging, metabolic function, and childhood trauma has been shown to provide more detailed signatures to predict longitudinal changes in depression severity in response to the metabolic agents used as antidepressant treatment than individual factors [321]. Use of multi-omics approaches and random forest classifier has been shown to achieve 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, with certain mitochondrial metabolites among the most important predictors [322, 323].

An additional ML application example 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. Using a high-throughput unbiased automated phenotyping platform that collects more than 2000 behavioral features and supervised ML that minimizes Bayesian misclassification probability, recent work showed that a rich set of behavioral alterations distinguish susceptible versus resilient phenotypes after exposure to social defeat stress (SDS) in rodents

[324, 325]. At the individual level, a ML classifier integrated *a priori* constructs, including measures of anxiety and immune system function, predicted if a given animal developed SDS-induced social withdrawal, or remained resilient, with 80% sensitivity, which is better than the categorization power based on either individual measure alone [326].

To develop personalized psychiatry strategies for better diagnosing and treating mental disorders, it is essential to meet the demand for ML enforced by the recent advent of molecular biology protocols that opened up the opportunity to capture central nervous system nanovesicles (known as *exosomes*) for examining specific neurobiological substrates, including transcriptomic profiles, dynamically and temporally *in vivo*. The development of advanced ML methods for dynamic network analyses will permit to link brain molecular targets and signaling pathways with other levels of analyses (e.g., functional connectivity) *in vivo* and to incorporate complex relationships between the brain and the rest of the body to redefine thinking about the modifiable mechanisms throughout the complex clinical disease course.

7. Explainable AI and Causality Testing 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 and make it easily understood by the end users. XAI has found emergent applications in medicine, finance, economy, security, and defense [327, 328]. In psychiatry, the mission of XAI is to help clarify the link between neural circuits to behavior, and to improve our understanding of therapeutic strategies to enhance cognitive, affective, and social functions [329, 330]. 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 can be explainable to the point of providing hidden mechanistic insights into the way brains execute a particular function or complex behaviors. For instance, a classification function learned by the machine to predict a disease outcome would not only need to report a probability outcome but also need to address additional questions for the end-user: why is this outcome instead of the alternative? How reliable is the outcome? When does it fail if something is missing or misrepresented? When and why the prediction is wrong? Accordingly, a model with improved interpretability is often accompanied with parameter/structure/connectivity constraints and some prior domain knowledge. The models can be continuously adapted such that an iterative process may be required to force ML methods to fit models with specific interpretations. From the treatment perspective, an improved understanding of brain dynamics responsible for dysfunctional cognitive functions and/or maladaptive behaviors in mental illnesses is also critical. To find the hidden cause, it is useful to discuss the concept of “causality”.

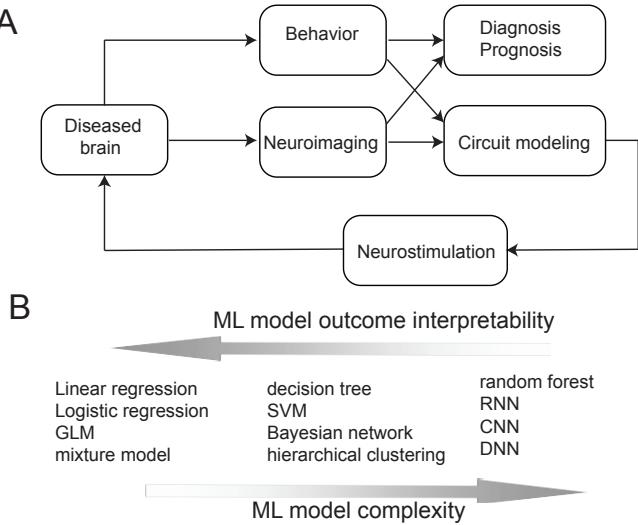


Figure 10: (A) Schematic of the closed loop of neuroimaging/modeling/neurostimulation. (B) A wide spectrum of interpretability in representative ML models.

Neuroimaging provides a passive sensing approach to observe the (correlational) brain-behavior relationship. However, correlation is different from causation. Correlational dependencies describe associations of measurements that experiments do not control, whereas causal dependencies link a dependent variable to an experimentally controlled variable. The key concept in causal inference is to introduce randomization to perturb the mapping. The relationship between every dependable variable and the randomized variable is causal, whereas the relationship between non-randomized variables and behavior, remains correlational [331]. Closed-loop experimental design would help to test the potential causality [332]. In human experiments, we classify closed-loop testing into two categories: one being fully automated, and the other being closed-human-in-the-loop.

In what follows, we will systematically review several important topics along these lines. We will present some “glass box” ML models in the literature (Section 7.1), then introduce neuroimaging-based circuit-level modeling (Section 7.2) and adaptive neuromodulation (Section 7.3). We argue that combining these efforts in the loop (“Neuroimaging → Circuit modeling → Neurostimulation → Observing behaviors → Revising models”, Figure 10A) will provide an effective way towards interpretable neuropsychiatry in understanding brain-behavior causation.

7.1. Explanation and Constraints in Interpretable ML Models

In terms of taxonomy, intrinsic interpretability refers to ML models that are considered interpretable due to their simple structures, such as short decision trees or sparse linear models (Figure 10B). Post-hoc interpretability refers to the application of interpretation methods after model training [333]. Interpretation may appear with different forms: (i) finite feature summary statistics, (ii) meaningful model parameters, (iii) ease of visualization of model outcome (e.g., feature summary or decision boundary). An ML model that is highly interpretable

because it has a few or more of the following properties [333]: high expressive power, low translucency, high portability, low algorithmic complexity and informative constraints. Generally, there is a trade-off between explainability and performance. For instance, a constrained linear or bilinear model will 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 model may face the problem of “explaining away” [334]. Here we briefly mention several classes of interpretable ML models.

Hybrid rule-based ML models: This type of ML models can be used for generating rules, such as a decision rule set: IF (condition) THEN (outcome 1) ELSE (outcome 2) statement, where the conditional clause will be learned from the data [335]. This type of model has more expressive power but less portability.

Constrained ML models: This type of ML models imposes parameter constraints to either avoid overfitting or enhance interpretability. Examples of such include the constrained convolutional filters in the CNN [336], or constrained mixture models [337]. 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 constraints in the optimization problem [338, 339]. The feedback can appear as a form of rule sets, which can be either known or unknown in advance. Iterating the feedback-rule optimization 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

Sharing the same goal of XAI, computational psychiatry tries to combine multiple levels and types of computation with multiple types of data in an effort to improve understanding, prediction and treatment of mental illness [340]. Two complementary approaches are used in computational psychiatry: (i) the data-driven approaches apply ML methods to high-dimensional data, multimodal to tackle classification and prediction problems (Section 3); (ii) the theory-driven approaches develop empirical or mechanistic models to test hypothesis.

For mechanistic modeling approaches, circuit-level modeling of macroscopic or mesoscopic brain dynamics represents an important research topic in computational models for psychiatric disorders [341, 342, 343]. A common strategy is to first use a biologically-inspired model to simulate neural activity based on a network of interacting neural masses, and then within each brain area, to model the neuronal population activity as the Wilson-Cowan neural mass model, with each consisting of excitatory and inhibitory populations [344]. Furthermore, individual brain nodes are coupled according to the empirically-derived anatomical network [345]. The computational model can be driven by 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 [346, 347], where the model parameter estimation and inference is done by unsupervised learning. By incorporating prior knowledge or hypotheses of network connections, DCM may reveal important brain mechanisms and offer experimental predictions. One of potential applications of DCM is to characterize the neural plasticity in the human brain, especially the change in functional connectivity informed by neuroimaging studies. The functional connectivity can either change gradually during the course of tasks or behavioral protocols, or induced by perturbation or neurostimulation. These changes are often, but not always, associated with changes in functional activations of specific brain regions. Combination of 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.

7.3. Closing the Loop for Testing Causality

One big challenge in human psychiatric neuroscience is the causality gap [348]. 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 important to close the loop in experiments by manipulating or perturbing the brain circuits and measuring its outcome, as commonly done in animal experiments [331, 332]. Unfortunately, a rigorous and causal grounding of clinical symptoms and behavior in specific neural circuit alternations is still missing. Specifically, since the clinical symptoms are diverse, how to define the dimension of brain function that defines one or few clinical symptoms and effectively manipulate them remains unknown.

Temporally precise neurostimulation tools provide a plausible means to perturb or stimulate the brain. To date, human 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 [349, 350, 351]. 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 [352, 353].

The brain connectivity and dynamics can be studied from a network communication and control perspective [354, 355]. The distinction between a healthy and a pathological brain can be characterized by their different efficiency to route the information between distributed brain nodes, control/modulate the target node under specific constraints, or influence its behavior in order to perform specific tasks ("cognitive control") [356]. Brain connectivity and neurostimulation can be studied by applying the network and graph theory. Specifically, the control-theoretic models have also been applied to quantify the response of brain networks to exogenous and endoge-

nous perturbations. Within the XAI-neuromodulation framework, it is convenient to formulae a mathematical framework for important research questions: (i) can a target node stimulation rewire the 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 an input constraint (such as the energy of neurostimulation), what is the optimal neurostimulation policy? Can alternate or simultaneous neurostimulations at multiple sites more effectively influence the network connectivity or bring additional benefit in treatment [357]?

XAI for neurostimulation in mental health can be seen as an extension in the design of human brain-machine interface (BMI) [358]. Specifically, XAI can search and identify behaviorally activated targets through active and scheduled stimulation strategies. Traditional neurostimulation strategies are designed in an on/off stimulation fashion triggered by predetermined stimulation parameters. However, these stimulation parameters will not be optimal. To accommodate an adaptive subject-specific stimulation strategy, adaptive stimulation uses neurofeedback to adjust the stimulation parameters or control policy for achieving various optimality criteria; *reinforcement learning* and *active learning* (two other important ML topics not reviewed here) can play a guiding role in online adaptive experiments [359, 360]. Additionally, simultaneous or post-stimulation neuroimaging provides a window of examining the change in brain network connectivity patterns. Can the induced brain patterns or changes in network connectivity be used to predict the treatment outcome? ML may address such a question by providing individualized treatment-response likelihood in precision psychiatry [361].

8. Discussion and Conclusion

8.1. Challenge and Opportunities

The past few decades have witnessed growing interests and rapid developments in ML methods toward precision psychiatry. However, caution has been raised regarding the unrealistic hope for ML applications in clinical practice [159, 362], 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," which implies a level of neurobiological and physiological understanding that is lacking in psychiatry. This makes it very difficult to build clinical inference models for mental disorders since the neurological and physiological mechanisms are not always observable with sufficient precision. As a result, it is not yet feasible to develop treatments that target underlying physiological risk factors similar to how it is possible in other areas of medicine (e.g., treating hypertension in heart diseases). Furthermore, each mental disorder has various types of overlapping symptoms with varying degrees, bringing an additional challenge to uniquely define the psychiatric disorder (unlike a clear cut in cardiology or oncology).

Second, many disorders are presented as a spectrum, for example, autism spectrum disorder, generalized anxiety spectrum,

and schizophrenia spectrum, and vary across different patients, creating a wide range of subtypes and subject variability diagnosed with the same mental disorder. Third, due to various genetic, biochemical, neuropathological factors, the same mental disorder may have different causes and symptoms in different populations. Fourth, many mental disorders have overlapping symptoms with other physical or mental disorders [363]. 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. This makes it difficult to accurately diagnose mental disorders. One of the challenges of precision psychiatry is to fully dissect the mechanisms and reveal the one-to-one relationship. This can be catalyzed by rigorous and continuous measurements of factors relevant to mental health using novel data sources as described in this review.

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

Sample size. Datasets used in many ML applications have very small sample sizes, especially by the standard of other speech/image/video applications. Neuroimaging data collection from mental health patients is limited to one-shot examples, which brings in high signal variability in addition to the intrinsic heterogeneity or disease comorbidity. Recent developments in foundation models and their applications to mental health domain can help overcome this challenge to some degree—for example, by sharing pretrained language models [215]. However, further caution is needed to ensure appropriate validation methods on the problem-specific data. For example, studies which do not hold out all data from specific patients for validation can lead to incorrect conclusions.

Data quality. There is a lack of standardization in data acquisition and uneven data quality, bringing up the issues of rigor and reproducibility. For example, in many studies using social media data, mental disorders self-identified by users are used as labels in the absence of clinical labels. This can lead to a post containing: “I am depressed”, being labeled as a positive class for depression regardless of the underlying clinical symptoms. Terms such as “depressed” or “anxious” have colloquial uses which can differ from clinical usage, leading to a highly noisy dataset. Furthermore, such studies also make it very difficult to generate data for a healthy control class since a lack of mention about a disorder does not rule out its presence. Because of these reasons, the interpretation of many results reported today should be cautioned.

Data privacy and security. Advances in sensing technologies can collect a large amount of personal and sensitive data, including the location data, face images, speech conversations, and social interactions. How to collect, store and process these data without the leakage risk of privacy information remains an important challenge for advancing ML research. While research studies have oversights like internal review boards to assure the ethical use of such data, this type of data is collected in the most massive scale by large tech and social media companies. Existing regulations do not treat this data as personally identifiable information (PII) which can be used to inform the user’s health. This is a major challenge in securing identifi-

able user data. If overcome, this will enable large-scale mental health research based on this data. To make this possible, U.S regulations such as the Health Insurance Portability and Accountability Act (HIPAA) can be used to govern PII acquired by all commercial entities or social media companies.

Social implications and environmental factors. Factors such as gender and race are a critical determinant of mental health. According to WHO, mental disorders have a history of gender bias. First, in terms of gender risk factors, females are more likely to suffer from depression and anxiety; whereas there is prevalence of autism in males. Second, in terms of gender treatment bias, doctors are more likely to diagnose depression in women compared with men, and women are more likely to be prescribed with mood altering psychotropic drugs. ML can play a role in uncovering gender or race risk factors and minimize the diagnosis or treatment bias related to these social factors.

Generalizability. The classic ML generalization issue applies very deeply to mental health applications, especially due to small sample sizes and poor data quality. Most studies use cross validation methods to avoid overfitting but do not go out to collect new validation data “in the wild” to assess generalizability. Furthermore, very few studies test generalization across data sources and research institutions. For example, it is important to test how well ML models built based on speech from clinical interviews apply to non-clinical speech data.

Algorithmic bias. Digital mental health inherits the long history of bias in psychiatry, which is present at all stages of a patient journey [364]. This poses a major challenge in developing ML applications for mental health, making the effect of algorithmic bias possibly worse than other fields of medicine. In addition to biological underpinnings, the domains of data (such as language) also represent social underpinnings [365], and thus it is important to consider how socioeconomic factors are influencing measurements. Using training and validation sets that are representative across all demographics including race, gender, and age can not only help address some of the issues, but also uncover new symptom expressions in various groups. This is even more important for ML approaches that inherit bias 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 by mental health professionals as well as for feeling the trust by the users of mental health technology. This is also an important dimension to improve “precision” in mental health. The choice of the interpretation method [330], like 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 learn about model functioning, it is important to note that the interpretation results can only be trusted as long as the challenges of generalizability and data quality are addressed. In other words, model interpretation methods can produce erratic results with insufficient or poor-quality data.

Causal inference. Most ML applications in mental health

have focused on integrating more information from various data sources (as compared to a mental health professional) and reaching a diagnosis decision faster. However, diagnosing a mental disorder, even with a highly interpretable model, does not speak to the underlying causes and thus has limited implications on treating the causes. Causal inference methods supported by ML models [366] can help with “precision” treatment design [367], which is the next step in the patient’s journey after precision diagnosis. Recent developments in ultra-high-field neuroimaging [368] may provide a pathway for developing inference models for mental disorders by observing their underlying neural mechanisms with sufficient temporal and spatial precision.

Clinical integration. Many ML-based studies have deployed limited experimental datasets. While structuring experiments and challenges for developing ML applications in mental health, it is important to consider the clinical need from a user experience perspective as well as consider key factors such as sources of data and size of held out datasets. From a user experience point of view, it is important to consider both the mental health professionals using the application [369] and their patients [370]. Some of this work, such as running user research in various demographics, lies outside of the ML domain; however, such cross-functional research can inform best practices in developing ML models. This type of thinking with the end goal in mind is important for successful translation of precision psychiatry research to widespread clinical practice [371].

Ethical considerations. ML applications in mental health raises several important ethical considerations. For example, ML models for risk assessment can lead to early screening that can in theory help with starting treatment early [8]. However, when screening techniques are available outside clinical settings, it also creates the risk of misinterpretation by patients, which can 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 using ML approaches that use foundation models like GPT-3 (Generative Pre-trained Transformer 3), which should be fully considered before deployment in clinical settings [372].

8.2. Applications of New ML Technologies

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

Data-centric approach. In the data-driven ML view (“ML system = model/algorith + 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, 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/algorith (while fixing the data) to improve the performance, a data-centric ML approach

(<https://datacentricai.org/>) involves building ML systems with quality data, with a goal to systematically process/augment the data (while fixing the model) to improve the ML performance [373]. The modification of the available data may include data regeneration, data augmentation, and label refinement strategies to improve the data consistency. The process of two approaches can be iterated to bootstrap the system performance.

Data augmentation approach. To deal with the small sample size issue in the medical field, another independent ML approach is to create synthetic data (as a data augmentation strategy) for ML [374]. Deep learning methods such as GAN and its variants have proved a powerful tool to generate synthetic brain images, speech, video, physiological data, and EHRs [375, 376, 377].

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 training [378]. This has become increasingly important as the volume of social media and multimedia data streams is overwhelming and prohibitive for human effort.

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 multimodality, not all data can be quantified in the Euclidean space, graph or geometric deep learning can play a role towards this research direction [379, 380].

8.3. 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 can be used for various stages in data analytics: detection/diagnosis, treatment selection/optimization, outcome monitoring/tracking, and relapse prevention. We predict that the multimodal integration of modernized neuroimaging, ML, genetics, behavioral neuroscience, and mobile health will open doors for new method development and technology inventions. First, making brain scans more accessible and more accurate 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 be applied to identify subtypes of its symptoms and cognitive deficits, and develop model-based phenotyping [381]. Third, ML methods, combined with large electronic health databases, could enable a personalized approach to psychiatry through improved diagnosis and prediction of individual responses to therapies. Fourth, when developing ML-powered technologies for psychiatry, it is important 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).³⁸² Finally, a combination of medications, wearable devices, mobile health apps, social support, and online education will be essential to improve mental health or assist therapeutic outcomes in the new era of digital psychiatry. Future precision psychiatry will leverage ML and new technologies to provide individualized custom packages that are built upon patient need and specific neural 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.). We thank Robert MacKay for English proofreading.

Declaration of interests

The authors declare no competing financial interests.

References

- ¹ Mark É Czeisler, Rashon I Lane, Emiko Petrosky, Joshua F Wiley, Aleta Christensen, Rashid Njai, Matthew D Weaver, Rebecca Robbins, Elise R Facer-Childs, Laura K Barger, 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*, 69(32):1049, 2020.
- ² Thomas R Insel and Bruce N Cuthbert. Brain disorders? precisely. *Science*, 348(6234):499–500, 2015.
- ³ Brisa S Fernandes, Leanne M Williams, Johann Steiner, Marion Leboyer, André F Carvalho, and Michael Berk. The new field of ‘precision psychiatry’. *BMC Medicine*, 15(1):1–7, 2017.
- ⁴ Thomas Insel, Bruce Cuthbert, Marjorie Garvey, Robert Heinssen, Daniel S Pine, Kevin Quinn, Charles Sanislow, and Philip Wang. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders, 2010.
- ⁵ Danilo Bzdok and Andreas Meyer-Lindenberg. Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3):223–230, 2018.
- ⁶ Zhirou Zhou, Tsung-Chin Wu, Bokai Wang, Hongyue Wang, Xin M Tu, and Changyong Feng. Machine learning methods in psychiatry: a brief introduction. *General Psychiatry*, 33(1):e100171, 2020.
- ⁷ Michael Allen and Andrew Salmon. Synthesising artificial patient-level data for open science—an evaluation of five methods. *medRxiv preprint*, 2020.
- ⁸ Christopher Burr, Jessica Morley, Mariarosaria Taddeo, and Luciano Floridi. Digital psychiatry: Risks and opportunities for public health and wellbeing. *IEEE Transactions on Technology and Society*, 1(1):21–33, 2020.
- ⁹ P Murali Doraiswamy, Charlotte Bleasie, and Kaylee Bodner. Artificial intelligence and the future of psychiatry: Insights from a global physician survey. *Artificial Intelligence in Medicine*, 102:101753, 2020.
- ¹⁰ Robb B Rutledge, Adam M Chekroud, and Quentin JM Huys. Machine learning and big data in psychiatry: toward clinical applications. *Current Opinion in Neurobiology*, 55:152–159, 2019.
- ¹¹ Chelsea Chandler, Peter W Foltz, and Brita Elvevåg. Using machine learning in psychiatry: the need to establish a framework that nurtures trustworthiness. *Schizophrenia Bulletin*, 46(1):11–14, 2020.
- ¹² Isaac R Galatzer-Levy, Kelly V Ruggles, and Zhe Chen. Data science in the research domain criteria era: relevance of machine learning to the study of stress pathology, recovery, and resilience. *Chronic Stress*, 2:1–14, 2018.
- ¹³ Adrian BR Shatte, Delyse M Hutchinson, and Samantha J Teague. Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine*, 49(9):1426–1448, 2019.
- ¹⁴ Chang Su, Zhenxing Xu, Jyotishman Pathak, and Fei Wang. Deep learning in mental health outcome research: a scoping review. *Translational Psychiatry*, 10(1):1–26, 2020.
- ¹⁵ Guang-Di Liu, Yu-Chen Li, Wei Zhang, and Le Zhang. A brief review of artificial intelligence applications and algorithms for psychiatric disorders. *Engineering*, 6(4):462–467, 2020.
- ¹⁶ Anja Thieme, Danielle Belgrave, and Gavin Doherty. 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)*, 27(5):1–53, 2020.
- ¹⁷ Daniel Durstewitz, Georgia Koppe, and Andreas Meyer-Lindenberg. Deep neural networks in psychiatry. *Molecular Psychiatry*, 24(11):1583–1598, 2019.
- ¹⁸ Georgia Koppe, Andreas Meyer-Lindenberg, and Daniel Durstewitz. Deep learning for small and big data in psychiatry. *Neuropsychopharmacology*, 46(1):176–190, 2021.
- ¹⁹ Dennis M. Hedderich and Simon B. Eickhoff. Machine learning for psychiatry: getting doctors at the black box? *Molecular Psychiatry*, 26:23–25, 2021.
- ²⁰ Matthew Bracher-Smith, Karen Crawford, and Valentina Escott-Price. Machine learning for genetic prediction of psychiatric disorders: a systematic review. *Molecular Psychiatry*, 26:70–79, 2021.
- ²¹ Summer Allen. Artificial intelligence and the future of psychiatry. *IEEE Pulse*, 11(3):2–6, 2020.
- ²² Anish Thukral, Faheem Ershad, Nada Enan, Zhoulyu Rao, and Cunjiang Yu. Soft ultrathin silicon electronics for soft neural interfaces: A review of recent advances of soft neural interfaces based on ultrathin silicon. *IEEE Nanotechnology Magazine*, 12(1):21–34, 2018.
- ²³ Amit Etkin. A reckoning and research agenda for neuroimaging in psychiatry. *American Journal of Psychiatry*, 176(7):507–511, 2019.
- ²⁴ Yoshihiro Noda, Mera S Barr, Muhammad ElSalhy, Fumi Masuda, Ryosuke Tarumi, Kamiyu Ogyu, Masataka Wada, Sakiko Tsugawa, Takahiro Miyazaki, Shinichiro Nakajima, et al. Neural correlates of delay discount alterations in addiction and psychiatric disorders: a systematic review of magnetic resonance imaging studies. *Progress in Neuropsychopharmacology and Biological Psychiatry*, 99:109822, 2020.
- ²⁵ Chad A Noggle and Andrew S Davis. Advances in neuroimaging. In *Understanding the Biological Basis of Behavior*, pages 107–137. Springer, 2021.
- ²⁶ Bayanne Olabi, Ian Ellison-Wright, Andrew M McIntosh, Stephen J Wood, Ed Bullmore, and Stephen M Lawrie. Are there progressive brain changes in schizophrenia? a meta-analysis of structural magnetic resonance imaging studies. *Biological Psychiatry*, 70(1):88–96, 2011.
- ²⁷ Piotr Podwalski, Krzysztof Szczygiel, Ernest Tyburski, Leszek Sagan, Błażej Misiak, and Jerzy Samochowiec. Magnetic resonance diffusion tensor imaging in psychiatry: A narrative review of its potential role in diagnosis. *Pharmacological Reports*, 73(1):43–56, 2021.
- ²⁸ Stephen M Smith, Christian F Beckmann, Jesper Andersson, Edward J Auerbach, Janine Bijsterbosch, Gwenaëlle Douaud, Eugene Duff, David A Feinberg, Ludovica Griffanti, Michael P Harms, et al. Resting-state fMRI in the human connectome project. *NeuroImage*, 80:144–168, 2013.
- ²⁹ Jennifer M Coughlin, Andrew G Horti, and Martin G Pomper. Opportunities in precision psychiatry using pet neuroimaging in psychosis. *Neurobiology of Disease*, 131:104428, 2019.
- ³⁰ Holly K Hamilton, Alison K Boos, and Daniel H Mathalon. Electroencephalography and event-related potential biomarkers in individuals at clinical high risk for psychosis. *Biological Psychiatry*, 88(4):294–303, 2020.
- ³¹ Nitin Williams and Richard N Henson. Recent advances in functional neuroimaging analysis for cognitive neuroscience. *Brain and Neuroscience Advances*, 2:1–4, 2018.
- ³² Josef Parvizi and Sabine Kastner. Human intracranial EEG: Promises and limitations. *Nature Neuroscience*, 21(4):474–483, 2018.
- ³³ Omid G Sani, Yuxiao Yang, Morgan B Lee, Heather E Dawes, Edward F Chang, and Maryam M Shanechi. Mood variations decoded from multi-site intracranial human brain activity. *Nature Biotechnology*, 36(10):954–961, 2018.
- ³⁴ Katherine W. Scangos, Ankit N. Khambhati, Patrick M. Daly, Ghassan S. Makhloul, Leo P. Sugrue, Hashem Zamanian, Tony X. Liu, Vikram R. Rao, Kristin K. Sellers, Heather E. Dawes, Philip A. Starr, Andrew D. Krystal,

- and Edward F Chang. Closed-loop neuromodulation in an individual with treatment-resistant depression. *Nature Medicine*, 27:1696–1700, 2021.
- ³⁵ Marco Ferrari and Valentina Quaresima. A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application. *NeuroImage*, 63(2):921–935, 2012.
- ³⁶ Louisa K Gossé, Sarah W Bell, and SM Hosseini. Functional near-infrared spectroscopy in developmental psychiatry: a review of attention deficit hyperactivity disorder. *European Archives of Psychiatry and Clinical Neuroscience*, pages 1–18, 2021.
- ³⁷ Hanna Keren, Georgia O’Callaghan, Pablo Vidal-Ribas, George A Buzzell, Melissa A Brotman, Ellen Leibenluft, Pedro M Pan, Liana Meffert, Ariela Kaiser, Selina Wolke, et al. Reward processing in depression: a conceptual and meta-analytic review across fMRI and EEG studies. *American Journal of Psychiatry*, 175(11):1111–1120, 2018.
- ³⁸ PB Lukow, Amanda Kiemes, MJ Kempton, FE Turkheimer, Philip McGuire, and Gemma Modinos. Neural correlates of emotional processing in psychosis risk and onset—a systematic review and meta-analysis of fMRI studies. *Neuroscience & Biobehavioral Reviews*, 128:780–788, 2021.
- ³⁹ Stephen M Smith, Diego Vidaurre, Christian F Beckmann, Matthew F Glasser, Mark Jenkinson, Karla L Miller, Thomas E Nichols, Emma C Robinson, Gholamreza Salimi-Khorshidi, Mark W Woolrich, et al. Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, 17(12):666–682, 2013.
- ⁴⁰ Neil D Woodward and Carissa J Cascio. Resting-state functional connectivity in psychiatric disorders. *JAMA Psychiatry*, 72(8):743–744, 2015.
- ⁴¹ Sai Ma, Vince D Calhoun, Ronald Phlypo, and Tülay Adali. Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis. *NeuroImage*, 90:196–206, 2014.
- ⁴² Edmund T Rolls, Wei Cheng, and Jianfeng Feng. Brain dynamics: the temporal variability of connectivity, and differences in schizophrenia and ADHD. *Translational Psychiatry*, 11(1):1–11, 2021.
- ⁴³ Vince D Calhoun and Jing Sui. Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1(3):230–244, 2016.
- ⁴⁴ Eric Feczkó, Oscar Miranda-Dominguez, Mollie Marr, Alice M Graham, Joel T Nigg, and Damien A Fair. The heterogeneity problem: approaches to identify psychiatric subtypes. *Trends in Cognitive Sciences*, 23(7):584–601, 2019.
- ⁴⁵ Theodore D Satterthwaite, Eric Feczkó, Antonia N Kaczkurkin, and Damien A Fair. Parsing psychiatric heterogeneity through common and unique circuit-level deficits. *Biological Psychiatry*, 88(1):4, 2020.
- ⁴⁶ Thomas R Insel. The nimb research domain criteria (RDoC) project: precision medicine for psychiatry. *American Journal of Psychiatry*, 171(4):395–397, 2014.
- ⁴⁷ Bruce N Cuthbert. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, 13(1):28–35, 2014.
- ⁴⁸ Danhong Wang, Randy L Buckner, Michael D Fox, Daphne J Holt, Avram J Holmes, Sophia Stoecklein, Georg Langs, Ruiqi Pan, Tianyi Qian, Kuncheng Li, et al. Parcellating cortical functional networks in individuals. *Nature Neuroscience*, 18(12):1853–1860, 2015.
- ⁴⁹ Martin Walter, Sarah Alizadeh, Hamidreza Jamalabadi, Ulrike Lueken, Udo Dannlowski, Henrik Walter, Sebastian Olbrich, Lejla Colic, Joseph Kambeitz, Nikolaos Koutsouleris, et al. Translational machine learning for psychiatric neuroimaging. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 91:113–121, 2019.
- ⁵⁰ Andrew T Drysdale, Logan Grosenick, Jonathan Downar, Katharine Dunlop, Farrokh Mansouri, Yue Meng, Robert N Fetcho, Benjamin Zebley, Desmond J Oathes, Amit Etkin, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, 23(1):28–38, 2017.
- ⁵¹ Yu Zhang, Wei Wu, Russell T Toll, Sharon Naparstek, Adi Maron-Katz, Mallissa Watts, Joseph Gordon, Jisoo Jeong, Laura Astolfi, Emmanuel Shpigel, et al. Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nature Biomedical Engineering*, 5(4):309–323, 2021.
- ⁵² Kaia Sargent, UnYoung Chavez-Baldini, Sarah L Master, Karin JH Verweij, Anja Lok, Arjen L Sutterland, Nienke C Vulink, Damiaan Denys, Dirk JA Smit, and Dorien H Nieman. Resting-state brain oscillations predict cognitive function in psychiatric disorders: A transdiagnostic machine learning approach. *NeuroImage: Clinical*, 30:102617, 2021.
- ⁵³ Deanna M Barch. The neural correlates of transdiagnostic dimensions of psychopathology. *American Journal of Psychiatry*, 174(7):613–615, 2017.
- ⁵⁴ Cedric Huchuan Xia, Zongming Ma, Rastko Ciric, Shi Gu, Richard F Betzel, Antonia N Kaczkurkin, Monica E Calkins, Philip A Cook, Angel García de la Garza, Simon N Vandekar, et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nature Communications*, 9(1):1–14, 2018.
- ⁵⁵ Valeria Kebets, Avram J Holmes, Csaba Orban, Siyi Tang, Jingwei Li, Nanbo Sun, Ru Kong, Russell A Poldrack, and BT Thomas Yeo. Somatosensory-motor dysconnectivity spans multiple transdiagnostic dimensions of psychopathology. *Biological Psychiatry*, 86(10):779–791, 2019.
- ⁵⁶ Lisa M McTeague, Benjamin M Rosenberg, James W Lopez, David M Carreon, Julia Huemer, Ying Jiang, Christina F Chick, Simon B Eickhoff, and Amit Etkin. Identification of common neural circuit disruptions in emotional processing across psychiatric disorders. *American Journal of Psychiatry*, 177(5):411–421, 2020.
- ⁵⁷ Christian Wachinger, Kwangsik Nho, Andrew J Saykin, Martin Reuter, Anna Rieckmann, Alzheimer’s Disease Neuroimaging Initiative, et al. A longitudinal imaging genetics study of neuroanatomical asymmetry in Alzheimer’s disease. *Biological Psychiatry*, 84(7):522–530, 2018.
- ⁵⁸ Pablo Vidal-Ribas, Brenda Benson, Aria D Vitale, Hanna Keren, Anita Harrewijn, Nathan A Fox, Daniel S Pine, and Argyris Stringaris. Bidirectional associations between stress and reward processing in children and adolescents: a longitudinal neuroimaging study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(10):893–901, 2019.
- ⁵⁹ Alyssa R Roeckner, Katelyn I Oliver, Lauren AM Lebois, Sanne JH van Rooij, and Jennifer S Stevens. Neural contributors to trauma resilience: a review of longitudinal neuroimaging studies. *Translational Psychiatry*, 11(1):1–17, 2021.
- ⁶⁰ Kanhai Zhao, Boris Duka, Hua Xie, Desmond J Oathes, Vince Calhoun, and Yu Zhang. A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD. *NeuroImage*, 246:118774, 2022.
- ⁶¹ Yuyang Luo, Tara L Alvarez, Jeffrey M Halperin, and Xiaobo Li. Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. *NeuroImage: Clinical*, 26:102238, 2020.
- ⁶² Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Jun-tang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. Braingnn: Interpretable brain graph neural network for fMRI analysis. *Medical Image Analysis*, 74:102233, 2021.
- ⁶³ Meenakshi Khosla, Keith Jamison, Amy Kuceyeski, and Mert R Sabuncu. Ensemble learning with 3d convolutional neural networks for functional connectome-based prediction. *NeuroImage*, 199:651–662, 2019.
- ⁶⁴ Mon-Ju Wu, Benson Mwangi, Isabelle E Bauer, Ives C Passos, Marsal Sanches, Giovana B Zunta-Soares, Thomas D Meyer, Khader M Hasan, and Jair C Soares. Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. *NeuroImage*, 145:254–264, 2017.
- ⁶⁵ Hongru Zhu, Minlan Yuan, Changjian Qiu, Zhengjia Ren, Yuchen Li, Jian Wang, Xiaoqi Huang, Su Lui, Qiyong Gong, Wei Zhang, et al. Multivariate classification of earthquake survivors with post-traumatic stress disorder based on large-scale brain networks. *Acta Psychiatrica Scandinavica*, 141(3):285–298, 2020.
- ⁶⁶ Weizheng Yan, Vince Calhoun, Ming Song, Yue Cui, Hao Yan, Shengfeng Liu, Lingzhong Fan, Nianming Zuo, Zhengyi Yang, Kaibin Xu, et al. Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site fMRI data. *EBioMedicine*, 47:543–552, 2019.
- ⁶⁷ Pavol Mikolas, Jaroslav Hlinka, Antonin Skoch, Zbynek Pitra, Thomas Frodl, Filip Spaniel, and Tomas Hajek. Machine learning classification of first-episode schizophrenia spectrum disorders and controls using whole brain white matter fractional anisotropy. *BMC Psychiatry*, 18(1):1–7, 2018.
- ⁶⁸ Caglar Uyulan, Türker Tekin Ergüzel, Huseyin Unubol, Merve Cebi, Gokben Hizli Sayar, Mahdi Nezhad Asad, and Nevzat Tarhan. Major depressive disorder classification based on different convolutional neural network models: Deep learning approach. *Clinical EEG and Neuroscience*, 52(1):38–51, 2021.
- ⁶⁹ Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R Cameron Craddock, Au-

- gusto Buchweitz, and Felipe Meneguzzi. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: Clinical*, 17:16–23, 2018.
- ⁷⁰ Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrero, Ben Glocker, and Daniel Rueckert. Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease. *Medical Image Analysis*, 48:117–130, 2018.
- ⁷¹ Adi Maron-Katz, Yu Zhang, Manjari Narayan, Wei Wu, Russell T Toll, Sharon Naparstek, Carlo De Los Angeles, Parker Longwell, Emmanuel Shpigel, Jennifer Newman, et al. Individual patterns of abnormality in resting-state functional connectivity reveal two data-driven PTSD subgroups. *American Journal of Psychiatry*, 177(3):244–253, 2020.
- ⁷² Aleksandra Lecei, Branko M van Hulst, Patrick de Zeeuw, Marieke van der Pluijm, Yvonne Rijks, and Sarah Durston. 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*, 21:101601, 2019.
- ⁷³ Linden Parkes, Tyler M Moore, Monica E Calkins, Philip A Cook, Matthew Cieslak, David R Roalf, Daniel H Wolf, Ruben C Gur, Raquel E Gur, Theodore D Satterthwaite, et al. Transdiagnostic dimensions of psychopathology explain individuals' unique deviations from normative neurodevelopment in brain structure. *Translational Psychiatry*, 11(1):1–13, 2021.
- ⁷⁴ Nikolaos Koutsouleris, Dominic B Dwyer, Franziska Degenhardt, Carlo Maj, Maria Fernanda Urquijo-Castro, Rachele Sanfelici, David Popovic, Oemer Oeztuerk, Shalaila S Haas, Johanna Weiske, et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry*, 78(2):195–209, 2021.
- ⁷⁵ Avinash Ramyead, Erich Studerus, Michael Komter, Martina Uttinger, Ute Gschwandtner, Peter Fuhr, and Anita Riecher-Rössler. Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naïve at-risk patients. *The World Journal of Biological Psychiatry*, 17(4):285–295, 2016.
- ⁷⁶ Andrey Zhdanov, Sravya Atluri, Willy Wong, Yasaman Vaghei, Zafiris J Daskalakis, Daniel M Blumberger, Benicio N Frey, Peter Giacobbe, Raymond W Lam, Roumen Milev, et al. Use of machine learning for predicting escitalopram treatment outcome from electroencephalography recordings in adult patients with depression. *JAMA Network Open*, 3(1):e1918377–e1918377, 2020.
- ⁷⁷ Lianne Schmaal, Andre F Marquand, Didi Rhebergen, Marie-José van Tol, Henricus G Ruhé, Nic JA van der Wee, Dick J Veltman, and Brenda WJH Penninx. Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: a multivariate pattern recognition study. *Biological Psychiatry*, 78(4):278–286, 2015.
- ⁷⁸ Michele A Bertocci, Genna Bebko, Amelia Versace, Satish Iyengar, Lisa Bonar, Erika E Forbes, Jorge RC Almeida, Susan B Perlman, Claudiu Schirda, MJ Travis, et al. Reward-related neural activity and structure predict future substance use in dysregulated youth. *Psychological Medicine*, 47(8):1357–1369, 2017.
- ⁷⁹ Jing Zhang, Simeon M Wong, J Don Richardson, Rakesh Jetly, and Benjamin T Dunkley. Predicting ptsd severity using longitudinal magnetoencephalography with a multi-step learning framework. *Journal of Neural Engineering*, 17(6):066013, 2020.
- ⁸⁰ Shelly Sheynin, Lior Wolf, Ziv Ben-Zion, Jony Sheynin, Shira Reznik, Jackob Nimrod Keynan, Roei Admon, Arie Shalev, Talma Hendler, and Israel Liberzon. Deep learning model of fMRI connectivity predicts PTSD symptom trajectories in recent trauma survivors. *NeuroImage*, 238:118242, 2021.
- ⁸¹ Mireille Nieuwenhuis, Hugo G Schnack, Neeltje E van Haren, Julia Lapin, Craig Morgan, Antje A Reinders, Diana Gutierrez-Tordesillas, Roberto Roiz-Santiañez, Maristela S Schaufelberger, Pedro G Rosa, et al. Multi-center mri prediction models: Predicting sex and illness course in first episode psychosis patients. *NeuroImage*, 145:246–253, 2017.
- ⁸² Jason Smucny, Ian Davidson, and Cameron S Carter. Comparing machine and deep learning-based algorithms for prediction of clinical improvement in psychosis with functional magnetic resonance imaging. *Human Brain Mapping*, 42(4):1197–1205, 2021.
- ⁸³ Wei Wu, Yu Zhang, Jing Jiang, Molly V Lucas, Gregory A Fonzo, Camarin E Rolle, Crystal Cooper, Cherise Chin-Fatt, Nosalie Krepel, Carena A Cornelissen, et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nature Biotechnology*, 38(4):439–447, 2020.
- ⁸⁴ Gregory A Fonzo, Amit Etkin, Yu Zhang, Wei Wu, Crystal Cooper, Cherise Chin-Fatt, Manish K Jha, Joseph Trombello, Thilo Deckersbach, Phil Adams, et al. Brain regulation of emotional conflict predicts antidepressant treatment response for depression. *Nature Human Behaviour*, 3(12):1319–1331, 2019.
- ⁸⁵ Nikolaos Koutsouleris, Eva M Meisenzahl, Christos Davatzikos, Ronald Bottlender, Thomas Frodl, Johanna Scheuerecker, Gisela Schmitt, Thomas Zetsche, Petra Decker, Maximilian Reiser, 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*, 66(7):700–712, 2009.
- ⁸⁶ Ronny Redlich, Nils Opel, Dominik Grotzgerd, Katharina Dohm, Dario Zaremba, Christian Bürger, Sandra Mühlmann, Patricia Wahl, Walter Heindel, et al. Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry*, 73(6):557–564, 2016.
- ⁸⁷ Jung-Chi Chang, Hsiang-Yuan Lin, Jinglei Lv, Wen-Yih Issac Tseng, and Susan Shur-Fen Gau. Regional brain volume predicts response to methylphenidate treatment in individuals with ADHD. *BMC Psychiatry*, 21(1):1–14, 2021.
- ⁸⁸ Paul Zhutovsky, Rajat M Thomas, Miranda Off, Sanne JH van Rooij, Mitzy Kennis, Guido A van Wingen, and Elbert Geuze. Individual prediction of psychotherapy outcome in posttraumatic stress disorder using neuroimaging data. *Translational Psychiatry*, 9(1):1–10, 2019.
- ⁸⁹ D Yang, KA Pelphrey, DG Sukhodolsky, MJ Crowley, E Dayan, NC Dvornek, A Venkataraman, J Duncan, L Staib, and P Ventola. Brain responses to biological motion predict treatment outcome in young children with autism. *Translational Psychiatry*, 6(11):e948–e948, 2016.
- ⁹⁰ Nicco Reggente, Teena D Moody, Francesca Morfini, Courtney Sheen, Jesse Rissman, Joseph O'Neill, and Jamie D Feusner. Multivariate resting-state functional connectivity predicts response to cognitive behavioral therapy in obsessive-compulsive disorder. *Proc. Natl. Acad. Sci. USA*, 115(9):2222–2227, 2018.
- ⁹¹ Bo Cao, Raymond Y Cho, Dachun Chen, Meihong Xiu, Li Wang, Jair C Soares, and Xiang Yang Zhang. Treatment response prediction and individualized identification of first-episode drug-naïve schizophrenia using brain functional connectivity. *Molecular Psychiatry*, 25(4):906–913, 2020.
- ⁹² Micah Cearns, Nils Opel, Scott Clark, Claas Kaehler, Anupalam Thalamuthu, Walter Heindel, Theresa Winter, Henning Teismann, Heike Minnerup, Udo Dannowski, et al. Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. *Translational Psychiatry*, 9(1):1–9, 2019.
- ⁹³ Juliet Edgcomb, Trevor Shaddox, Gerhard Hellermann, and John O Brooks III. High-risk phenotypes of early psychiatric readmission in bipolar disorder with comorbid medical illness. *Psychosomatics*, 60(6):563–573, 2019.
- ⁹⁴ Didier Morel, C Yu Kalvin, Ann Liu-Ferrara, Ambiorix J Caceres-Suriel, Stephan G Kurtz, and Ying P Tabak. Predicting hospital readmission in patients with mental or substance use disorders: a machine learning approach. *International Journal of Medical Informatics*, 139:104136, 2020.
- ⁹⁵ Ralitza Gueorguieva, Adam M Chekroud, and John H Krystal. 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*, 4(3):230–237, 2017.
- ⁹⁶ Micah Cearns, Tim Hahn, and Bernhard T Baune. Recommendations and future directions for supervised machine learning in psychiatry. *Translational Psychiatry*, 9(1):1–12, 2019.
- ⁹⁷ Adrienne Grzenda, Nina V Kraguljac, William M McDonald, Charles Neuneroff, John Torous, Jonathan E Alpert, Carolyn I Rodriguez, and Alik S Wigde. Evaluating the machine learning literature: a primer and user's guide for psychiatrists. *American Journal of Psychiatry*, 178(8):715–729, 2021.
- ⁹⁸ Andy MY Tai, Alcides Albuquerque, Nicole E Carmona, Mehala Subramanieapillai, Danielle S Cha, Margarita Sheko, Yena Lee, Rodrigo Mansur, and Roger S McIntyre. Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. *Artificial Intelligence in Medicine*, 99:101704, 2019.
- ⁹⁹ Katie Aafjes-van Doorn, Céline Kamsteeg, Jordan Bate, and Marc Aafjes. A scoping review of machine learning in psychotherapy research. *Psychotherapy Research*, 31(1):92–116, 2021.

- ¹⁰⁰ Ashley N Nielsen, Deanna M Barch, Steven E Petersen, Bradley L Schlaggar, and Deanna J Greene. Machine learning with neuroimaging: Evaluating its applications in psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(8):791–798, 2020.
- ¹⁰¹ Jing Sui, Rongtao Jiang, Juan Bustillo, and Vince Calhoun. Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. *Biological Psychiatry*, 88(11):818–828, 2020.
- ¹⁰² Li Zhang, Mingliang Wang, Mingxia Liu, and Daoqiang Zhang. A survey on deep learning for neuroimaging-based brain disorder analysis. *Frontiers in Neuroscience*, 14:779, 2020.
- ¹⁰³ Gyeongcheol Cho, Jinyeong Yim, Younyoung Choi, Jungmin Ko, and Seoung-Hwan Lee. Review of machine learning algorithms for diagnosing mental illness. *Psychiatry Investigation*, 16(4):262, 2019.
- ¹⁰⁴ Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11):1664–1671, 2015.
- ¹⁰⁵ Xilin Shen, Emily S Finn, Dustin Scheinost, Monica D Rosenberg, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, 12(3):506–518, 2017.
- ¹⁰⁶ Michael E Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1(6):211–244, 2001.
- ¹⁰⁷ Yu Zhang, Guoxu Zhou, Jing Jin, Qibin Zhao, Xingyu Wang, and Andrzej Cichocki. Sparse bayesian classification of EEG for brain-computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 27(11):2256–2267, 2015.
- ¹⁰⁸ Qiongmin Ma, Tianhao Zhang, Marcus V Zanetti, Hui Shen, Theodore D Satterthwaite, Daniel H Wolf, Raquel E Gur, Yong Fan, Dewen Hu, Gerardo F Busatto, et al. Classification of multi-site mr images in the presence of heterogeneity using multi-task learning. *NeuroImage: Clinical*, 19:476–486, 2018.
- ¹⁰⁹ Li Xiao, Julia M Stephen, Tony W Wilson, Vince D Calhoun, and Yu-Ping Wang. A manifold regularized multi-task learning model for iq prediction from two fmri paradigms. *IEEE Transactions on Biomedical Engineering*, 67(3):796–806, 2019.
- ¹¹⁰ Mansu Kim, Eun Jeong Min, Kefei Liu, Jingwen Yan, Andrew J Saykin, Jason H Moore, Qi Long, and Li Shen. Multi-task learning based structured sparse canonical correlation analysis for brain imaging genetics. *Medical Image Analysis*, 76:102297, 2022.
- ¹¹¹ Seyul Kwak, Soowon Park, Jeongsim Kim, Seho Park, and Jun-Young Lee. Multivariate neuroanatomical correlates of behavioral and psychological symptoms in dementia and the moderating role of education. *NeuroImage: Clinical*, 28:102452, 2020.
- ¹¹² Sunil Vasu Kalnady, Russell Greiner, Rimjhim Agrawal, Venkataram Shivakumar, Janardhanan C Narayanaswamy, Matthew RG Brown, Andrew J Greenshaw, Serdar M Dursun, and Ganeshan Venkatasubramanian. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. *NPJ Schizophrenia*, 5(1):1–11, 2019.
- ¹¹³ Jun Wang, Lichi Zhang, Qian Wang, Lei Chen, Jun Shi, Xiaobo Chen, Zuoyong Li, and Dinggang Shen. Multi-classasd classification based on functional connectivity and functional correlation tensor via multi-source domain adaptation and multi-view sparse representation. *IEEE Transactions on Medical Imaging*, 39(10):3137–3147, 2020.
- ¹¹⁴ Jonathan Elmer, Bobby L Jones, and Daniel S Nagin. Using the beta distribution in group-based trajectory models. *BMC Medical Research Methodology*, 18(1):1–5, 2018.
- ¹¹⁵ Gavin van der Nest, Valéria Lima Passos, Math JJM Candel, and Gerard JP van Breukelen. An overview of mixture modelling for latent evolutions in longitudinal data: Modelling approaches, fit statistics and software. *Advances in Life Course Research*, 43:100323, 2020.
- ¹¹⁶ Jennifer D Ellis, Jill A Rabinowitz, Jonathan Wells, Fangyu Liu, Patrick H Finan, Michael D Stein, Denis G Antoine II, Gregory J Hobelmann, and Andrew S Huhn. Latent trajectories of anxiety and depressive symptoms among adults in early treatment for nonmedical opioid use. *Journal of Affective Disorders*, 299:223–232, 2022.
- ¹¹⁷ Pål Ulvenes, Christina S Soma, Linne Melsom, and Bruce E Wampold. A latent trajectory analysis of inpatient depression treatment. *Psychotherapy*, 59(1):113–124, 2022.
- ¹¹⁸ Einat Waizbard-Bartov, Emilio Ferrer, Brianna Heath, Sally J Rogers, Christine Wu Nordahl, Marjorie Solomon, and David G Amaral. Identifying autism symptom severity trajectories across childhood. *Autism Research*, 2022.
- ¹¹⁹ Katharina Schultebraucks, Arie Y. Shalev, Vasiliki Michopoulos, Corita R. Grudzen, Soo-Min Shin, Jennifer S. Stevens, Jessica L. Maples-Keller, Tanja Jovanovic, George A. Bonanno, Barbara O. Rothbaum, Charles R. Marmar, Charles B. Nemerooff, Kerry J. Ressler, and Isaac R. Galatzer-Levy. A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor. *Nature Medicine*, 26:1064–1088, 2020.
- ¹²⁰ Erika L Crable, Mari-Lynn Drainoni, David K Jones, Alexander Y Walley, and Jacqueline Milton Hicks. Predicting longitudinal service use for individuals with substance use disorders: A latent profile analysis. *Journal of Substance Abuse Treatment*, 132:108632, 2022.
- ¹²¹ Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. *Semi-Supervised Learning*. MIT Press, 2009.
- ¹²² Wutao Yin, Longhai Li, and Fang-Xiang Wu. A semi-supervised autoencoder for autism disease diagnosis. *Neurocomputing*, 483:140–147, 2022.
- ¹²³ Erdem Varol, Aristeidis Sotiras, Christos Davatzikos, Alzheimer's Disease Neuroimaging Initiative, et al. Hydra: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *NeuroImage*, 145:346–364, 2017.
- ¹²⁴ Tao Yang, Sophia Frangou, Raymond W Lam, Jia Huang, Yousong Su, Guoqing Zhao, Ruizhi Mao, Na Zhu, Rubai Zhou, Xiao Lin, et al. Probing the clinical and brain structural boundaries of bipolar and major depressive disorder. *Translational Psychiatry*, 11(1):1–8, 2021.
- ¹²⁵ Nicolas Honnorat, Aoyan Dong, Eva Meisenzahl-Lechner, Nikolaos Koutsouleris, and Christos Davatzikos. Neuroanatomical heterogeneity of schizophrenia revealed by semi-supervised machine learning methods. *Schizophrenia Research*, 214:43–50, 2019.
- ¹²⁶ Antonia N Kaczkurkin, Aristeidis Sotiras, Erica B Baller, Ran Barzilay, Monica E Calkins, Ganesh B Chand, Zaixu Cui, Guray Erus, Yong Fan, Raquel E Gur, et al. Neurostructural heterogeneity in youths with internalizing symptoms. *Biological Psychiatry*, 87(5):473–482, 2020.
- ¹²⁷ Brett K Beaulieu-Jones, Casey S Greene, et al. Semi-supervised learning of the electronic health record for phenotype stratification. *Journal of Biomedical Informatics*, 64:168–178, 2016.
- ¹²⁸ Amir Hossein Yazdavar, Hussein S Al-Olimat, Monireh Ebrahimi, Goonmeet Bajaj, Tanvi Banerjee, Krishnaprasad Thirunarayan, Jyotishman Pathak, and Amit Sheth. 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*, pages 1191–1198, 2017.
- ¹²⁹ Guimin Dong, Mingyue Tang, Lihua Cai, Laura E Barnes, and Mehdi Boukhechba. Semi-supervised graph instance transformer for mental health inference. In *Proc. 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1221–1228. IEEE, 2021.
- ¹³⁰ Andre F Marquand, Iead Rezek, Jan Buitelaar, and Christian F Beckmann. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. *Biological Psychiatry*, 80(7):552–561, 2016.
- ¹³¹ Saige Rutherford, Seyed Mostafa Kia, Thomas Wolfers, Charlotte Fraza, Mariam Zabih, Richard Dinga, Pierre Berthet, Amanda Worker, Serena Verdi, Henricus G Ruhe, et al. The normative modeling framework for computational psychiatry. *bioRxiv*, 2021.
- ¹³² Matthias Guggenmos, Katharina Schmack, Ilya M. Veer, Tristram Lett, Miriam Sebold Maria Sekutowicz, Maria Garbusow, Christian Sommer, Hans-Ulrich Wittchen, Ulrich S. Zimmermann, Henrik Walter Michael N. Smolka, Andreas Heinz, and Philipp Sterzer. A multimodal neuroimaging classifier for alcohol dependence. *Scientific Reports*, 10:298, 2020.
- ¹³³ Brett K. Beaulieu-Jones and Casey S. Greene. Semi-supervised learning of the electronic health record for phenotype stratification. *Journal of Biomedical Informatics*, 64:168–178, 2016.
- ¹³⁴ Dinggang Shen, Guorong Wu, and Heung-II Suk. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19:221–248, 2017.
- ¹³⁵ Anees Abrol, Zening Fu, Mustafa Salman, Rogers Silva, Yuhui Du, Sergey Plis, and Vince Calhoun. Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nature communications*, 12(1):1–17, 2021.

- ¹³⁶ Mirjam Quaak, Laurens van de Mortel, Rajat Mani Thomas, and Guido van Wingen. Deep learning applications for the classification of psychiatric disorders using neuroimaging data: systematic review and meta-analysis. *NeuroImage: Clinical*, 30:102584, 2021.
- ¹³⁷ Miao Chang, Fay Y Womer, Xiaohong Gong, Xi Chen, Lili Tang, Ruiqi Feng, Shuai Dong, Jia Duan, Yifan Chen, Ran Zhang, et al. Identifying and validating subtypes within major psychiatric disorders based on frontal-posterior functional imbalance via deep learning. *Molecular Psychiatry*, 26(7):2991–3002, 2021.
- ¹³⁸ Fahad Almuqhim and Fahad Saeed. ASD-SAENet: a sparse autoencoder, and deep-neural network model for detecting autism spectrum disorder (ASD) using fMRI data. *Frontiers in Computational Neuroscience*, 15:27, 2021.
- ¹³⁹ Walter HL Pinaya, Andrea Mechelli, and João R Sato. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. *Human Brain Mapping*, 40(3):944–954, 2019.
- ¹⁴⁰ Aidas Aglinskas, Joshua K. Hartshorne, and Stefano Anzellotti. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science*, 376(6597):1070–1074, 2022.
- ¹⁴¹ Syed Muhammad Anwar, Muhammad Majid, Adnan Qayyum, Muhammad Awais, Majdi Alnowami, and Muhammad Khurram Khan. Medical image analysis using convolutional neural networks: a review. *Journal of Medical Systems*, 42(11):1–13, 2018.
- ¹⁴² Rikiya Yamashita, Mizuho Nishio, Richard Kinsh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9(4):611–629, 2018.
- ¹⁴³ Jianing Zhang, Xuechen Li, Yuexiang Li, Mingyu Wang, Bingsheng Huang, Shuqiao Yao, and Linlin Shen. Three dimensional convolutional neural network-based classification of conduct disorder with structural mri. *Brain Imaging and Behavior*, 14(6):2333–2340, 2020.
- ¹⁴⁴ Alaa Bessadok, Mohamed Ali Mahjoub, and Islem Rekik. Graph neural networks in network neuroscience. *arXiv preprint arXiv:2106.03535*, 2021.
- ¹⁴⁵ Daniel Durstewitz, Quentin JM Huys, and Georgia Koppe. Psychiatric illnesses as disorders of network dynamics. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(9):865–876, 2021.
- ¹⁴⁶ Yu Zhao, Xiang Li, Heng Huang, Wei Zhang, Shijie Zhao, Milad Makkie, Mo Zhang, Quanzheng Li, and Tianming Liu. 4D modeling of fMRI data via spatio-temporal convolutional neural networks (ST-CNN). *IEEE Transactions on Cognitive and Developmental Systems*, 12(3):451, 2020.
- ¹⁴⁷ Elnaz Lashgari, Dehua Liang, and Uri Maoz. Data augmentation for deep-learning-based electroencephalography. *Journal of Neuroscience Methods*, 346:10885, 2020.
- ¹⁴⁸ Chao Shang, Aaron Palmer, Jiangwen Sun, Ko-Shin Chen, Jin Lu, and Jinbo Bi. Vigan: Missing view imputation with generative adversarial networks. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 766–775. IEEE, 2017.
- ¹⁴⁹ Nicha C Dvornek, Xiaoxiao Li, Juntang Zhuang, and James S Duncan. Jointly discriminative and generative recurrent neural networks for learning from fMRI. In *International Workshop on Machine Learning in Medical Imaging*, pages 382–390. Springer, 2019.
- ¹⁵⁰ Jianlong Zhao, Jinjie Huang, Dongmei Zhi, Weizheng Yan, Xiaohong Ma, Xiao Yang, Xianbin Li, Qing Ke, Tianzi Jiang, Vince D Calhoun, et al. Functional network connectivity (FNC)-based generative adversarial network (GAN) and its applications in classification of mental disorders. *Journal of Neuroscience Methods*, 341:108756, 2020.
- ¹⁵¹ Christine M Cutillo, Karlie R Sharma, Luca Foschini, Shinjini Kundu, Maxine Mackintosh, and Kenneth D Mandl. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digital Medicine*, 3(1):1–5, 2020.
- ¹⁵² Zhenfu Wen, Marie-France Marin, Jennifer Urbano Blackford, Zhe Sage Chen, and Mohammed R Milad. Fear-induced brain activations distinguish anxious and trauma-exposed brains. *Translational Psychiatry*, 11(1):1–10, 2021.
- ¹⁵³ Oleg Bestsennyy, Greg Gilbert, Alex Harris, and Jennifer Rost. Telehealth: A quarter-trillion-dollar post-covid-19 reality?, 2021.
- ¹⁵⁴ Ellen E Lee, John Torous, Mumunun De Choudhury, Colin A Depp, Sarah A Graham, Ho-Cheol Kim, Martin P Paulus, John H Krystal, and Dilip V Jeste. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(9):856–864, 2021.
- ¹⁵⁵ Ahmed A Moustafa. *Big Data in Psychiatry and Neurology*. 2021.
- ¹⁵⁶ Sarah Graham, Colin Depp, Ellen E Lee, Camille Nebeker, Xin Tu, Ho-Cheol Kim, and Dilip V Jeste. Artificial intelligence for mental health and mental illnesses: an overview. *Current Psychiatry Reports*, 21(11):1–18, 2019.
- ¹⁵⁷ Anzar Abbas, Katharina Schultebraucks, and Isaac R Galatzer-Levy. Digital measurement of mental health: challenges, promises, and future directions. *Psychiatric Annals*, 51(1):14–20, 2021.
- ¹⁵⁸ Leonard Bickman. 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*, 47(5):795–843, 2020.
- ¹⁵⁹ Jack Wilkinson, Kellyn F Arnold, Eleanor J Murray, Maarten van Smeden, Kareem Carr, Rachel Sippy, Marc de Kamps, Andrew Beam, Stefan Konigorski, Christoph Lippert, et al. Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health*, 2(12):e677–e680, 2020.
- ¹⁶⁰ Daniel Barron. *Reading Our minds: The rise of Big data psychiatry*. Columbia Global Reports, 2021.
- ¹⁶¹ Adrián Vázquez-Romero and Ascensión Gallardo-Antolín. Automatic detection of depression in speech using ensemble convolutional neural networks. *Entropy*, 22(6):688, 2020.
- ¹⁶² Amir Harati, Elizabeth Shriberg, Tomasz Rutowski, Piotr Chlebek, Yang Lu, and Ricardo Oliveira. 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)*, pages 7273–7277. IEEE, 2021.
- ¹⁶³ Zhaocheng Huang, Julien Epps, and Dale Joachim. Investigation of speech landmark patterns for depression detection. *IEEE Transactions on Affective Computing*, 2019.
- ¹⁶⁴ Yu Zhu, Yuanyuan Shang, Zhuhong Shao, and Guodong Guo. Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. *IEEE Transactions on Affective Computing*, 9(4):578–584, 2017.
- ¹⁶⁵ Wei Shao, Zhiyang You, Lesheng Liang, Xiping Hu, Chengming Li, Wei Wang, and Bin Hu. A multi-modal gait analysis-based depression detection system. *IEEE Journal of Biomedical and Health Informatics*, 2021.
- ¹⁶⁶ Yang Lu, Amir Harati, Tomasz Rutowski, Ricardo Oliveira, P Chlebek, and E Shriberg. Robust speech and natural language processing models for depression screening. In *Proc. IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–5. IEEE, 2020.
- ¹⁶⁷ Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preotiuc-Pietro, David A Asch, and H Andrew Schwartz. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences USA*, 115(44):11203–11208, 2018.
- ¹⁶⁸ Hao Sun, Jiaqing Liu, Shurong Chai, Zhaolin Qiu, Lanfen Lin, Xinyin Huang, and Yenwei Chen. Multi-modal adaptive fusion transformer network for the estimation of depression level. *Sensors*, 21(14):4764, 2021.
- ¹⁶⁹ Luisa Weiner, Andrea Guidi, Nadège Doignon-Camus, Anne Giersch, Gilles Bertschy, and Nicola Vanello. Vocal features obtained through automated methods in verbal fluency tasks can aid the identification of mixed episodes in bipolar disorder. *Translational Psychiatry*, 11(1):1–9, 2021.
- ¹⁷⁰ Niclas Palmius, Athanasios Tsanas, Kate EA Saunders, Amy C Bilderbeck, John R Geddes, Guy M Goodwin, and Maartens De Vos. Detecting bipolar depression from geographic location data. *IEEE Transactions on Biomedical Engineering*, 64(8):1761–1771, 2016.
- ¹⁷¹ Charles R Marmar, Adam D Brown, Meng Qian, Eugene Laska, Carole Siegel, Meng Li, Duna Abu-Amara, Andreas Tsiantas, Colleen Richéy, Jennifer Smith, et al. Speech-based markers for posttraumatic stress disorder in us veterans. *Depression and Anxiety*, 36(7):607–616, 2019.
- ¹⁷² Adria Mallol-Ragolta, Svatia Dhamija, and Terrance E Boult. A multimodal approach for predicting changes in ptsd symptom severity. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pages 324–333, 2018.
- ¹⁷³ Yasir Tahir, Zixu Yang, Debsubhra Chakraborty, Nadia Thalmann, Daniel Thalmann, Yogeswary Maniam, Nur Amirah binte Abdul Rashid, Bhing-Leet Tan, Jimmy Lee Chee Keong, and Justin Dauwels. Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. *PLoS One*, 14(4):e0214314, 2019.
- ¹⁷⁴ Anzar Abbas, Vijay Yadav, Emma Smith, Elizabeth Ramjas, Sarah B Rut-

- ter, Caridad Benavidez, Vidya Koesmahargyo, Li Zhang, Lei Guan, Paul Rosenfield, et al. Computer vision-based assessment of motor functioning in schizophrenia: Use of smartphones for remote measurement of schizophrenia symptomatology. *Digital Biomarkers*, 5(1):29–36, 2021.
- ¹⁷⁵ Michael Leo Birnbaum, Anna Van Meter, Victor Chen, Asra F Rizvi, Elizabeth Arenare, Munmun De Choudhury, John M Kane, 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*, 7(9):e19348, 2020.
- ¹⁷⁶ Michael L Birnbaum, Avner Abrami, Stephen Heisig, Asra Ali, Elizabeth Arenare, Carla Agurto, Nathaniel Lu, John M Kane, and Guillermo Cecchi. Acoustic and facial features from clinical interviews for machine learning-based psychiatric diagnosis: Algorithm development. *JMIR Mental Health*, 9(1):e24699, 2022.
- ¹⁷⁷ Saeed Abdullah and Tanzeem Choudhury. Sensing technologies for monitoring serious mental illnesses. *IEEE MultiMedia*, 25(1):61–75, 2018.
- ¹⁷⁸ Emil Kraepelin. Manic depressive insanity and paranoia. *The Journal of Nervous and Mental Disease*, 53(4):350, 1921.
- ¹⁷⁹ Daniel M Low, Kate H Bentley, and Satrajit S Ghosh. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*, 5(1):96–116, 2020.
- ¹⁸⁰ Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7(2):190–202, 2015.
- ¹⁸¹ Nicholas Cummins, Stefan Scherer, Jarek Krajewski, Sebastian Schnieder, Julien Epps, and Thomas F Quatieri. A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, 71:10–49, 2015.
- ¹⁸² Zahra N Karam, Emily Mower Provost, Satinder Singh, Jennifer Montgomery, Christopher Archer, Gloria Harrington, and Melvin G McInnis. 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)*, pages 4858–4862. IEEE, 2014.
- ¹⁸³ Tomek Rutowski, Elizabeth Shriberg, Amir Harati, Yang Lu, Ricardo Oliveira, and Piotr Chlebek. Cross-demographic portability of deep NLP-based depression models. In *Proc. IEEE Spoken Language Technology Workshop (SLT)*, pages 1052–1057. IEEE, 2021.
- ¹⁸⁴ Ganes Kesari. AI can now detect depression from your voice, and it's twice as accurate as human practitioners, 2021.
- ¹⁸⁵ Laura Lovett. Sonde launches voice API to detect mental illness, 2021.
- ¹⁸⁶ Isaac Galatzer-Levy, Anzar Abbas, Anja Ries, Stephanie Homan, Laura Sels, Vidya Koesmahargyo, Vijay Yadav, Michael Colla, Hanne Scheerer, Stefan Vetter, 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*, 23(6):e25199, 2021.
- ¹⁸⁷ Siyang Song, Linlin Shen, and Michel Valstar. 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)*, pages 158–165. IEEE, 2018.
- ¹⁸⁸ Ryan Anthony J de Belen, Tomasz Bednarz, Arcot Sowmya, and Dennis Del Favero. Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019. *Translational Psychiatry*, 10(1):1–20, 2020.
- ¹⁸⁹ Qian Chen, Iti Chaturvedi, Shaoxiong Ji, and Erik Cambria. Sequential fusion of facial appearance and dynamics for depression recognition. *Pattern Recognition Letters*, 150:115–121, 2021.
- ¹⁹⁰ Lang He, Mingyue Niu, Prayag Tiwari, Pekka Marttinen, Rui Su, Jiewei Jiang, Chenguang Guo, Hongyu Wang, Songtao Ding, Zhongmin Wang, et al. Deep learning for depression recognition with audiovisual cues: A review. *Information Fusion*, 80:56–86, 2022.
- ¹⁹¹ Xizhuang Zhou, Kai Jin, Yuanyuan Shang, and Guodong Guo. Visually interpretable representation learning for depression recognition from facial images. *IEEE Transactions on Affective Computing*, 11(3):542–552, 2018.
- ¹⁹² Urška Smrke, Izidor Mlakar, Simon Lin, Bojan Musil, Nejc Plohl, et al. Language, speech, and facial expression features for artificial intelligence-based detection of cancer survivors' depression: Scoping meta-review. *JMIR Mental Health*, 8(12):e30439, 2021.
- ¹⁹³ Neguine Rezaii, Phillip Wolff, and Bruce H Price. Natural language processing in psychiatry: the promises and perils of a transformative approach. *The British Journal of Psychiatry*, pages 1–3, 2022.
- ¹⁹⁴ Aziliz Le Glaz, Yannis Haralambous, Deok-Hee Kim-Dufor, Philippe Lenca, Romain Billot, Taylor C Ryan, Jonathan Marsh, Jordan Devylder, Michel Walter, Sofian Berrouiguet, et al. Machine learning and natural language processing in mental health: Systematic review. *Journal of Medical Internet Research*, 23(5):e15708, 2021.
- ¹⁹⁵ National electronic health records survey: 2015 specialty and overall physicians electronic health record adoption summary tables, 2015.
- ¹⁹⁶ Victor M Castro, Jessica Minnier, Shawn N Murphy, Isaac Kohane, Susanne E Churchill, Vivian Gainer, Tianxi Cai, Alison G Hoffnagle, Yael Dai, Stefanie Block, et al. Validation of electronic health record phenotyping of bipolar disorder cases and controls. *American Journal of Psychiatry*, 172(4):363–372, 2015.
- ¹⁹⁷ Duy Van Le, James Montgomery, Kenneth C Kirkby, and Joel Scanlan. Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting. *Journal of Biomedical Informatics*, 86:49–58, 2018.
- ¹⁹⁸ Colin G Walsh, Jessica D Ribeiro, and Joseph C Franklin. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *Journal of Child Psychology and Psychiatry*, 59(12):1261–1270, 2018.
- ¹⁹⁹ Anna Rumshisky, Marzyeh Ghassemi, Tristan Naumann, Peter Szolovits, VM Castro, TH McCoy, and RH Perlis. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational Psychiatry*, 6(10):e921–e921, 2016.
- ²⁰⁰ Robert Stewart and Sumithra Velupillai. Applied natural language processing in mental health big data. *Neuropsychopharmacology*, 46(1):252, 2021.
- ²⁰¹ Julia Ive, Natalia Viani, Joyce Kam, Lucia Yin, Somain Verma, Stephen Puntis, Rudolf N Cardinal, Angus Roberts, Robert Stewart, and Sumithra Velupillai. Generation and evaluation of artificial mental health records for natural language processing. *NPJ Digital Medicine*, 3(1):1–9, 2020.
- ²⁰² Niels Bantilan, Matteo Malgaroli, Bonnie Ray, and Thomas D Hull. Just in time crisis response: suicide alert system for telemedicine psychotherapy settings. *Psychotherapy Research*, 31(3):289–299, 2021.
- ²⁰³ Maria Paz Raveau, Julian Goñi, José Rodriguez, Isidora Paiva, Fernanda Barriga, María Paz Hermosilla, Claudio Fuentes, and Susana Eyheramendy. Natural language processing of helpline chat data before and during the pandemic revealed significant decrease in self-image appreciation and changes in other traits. 2022.
- ²⁰⁴ Sarah E Morgan, Kelly Diederen, Petra E Vértes, Samantha HY Ip, Bo Wang, Bethany Thompson, Arsime Demjaha, Andrea De Micheli, Dominic Oliver, Maria Liakata, et al. Natural language processing markers in first episode psychosis and people at clinical high-risk. *Translational Psychiatry*, 11(1):1–9, 2021.
- ²⁰⁵ Tomasz Rutowski, Elizabeth Shriberg, Amir Harati, Yang Lu, Piotr Chlebek, and Ricardo Oliveira. Depression and anxiety prediction using deep language models and transfer learning. In *Proc. 7th International Conference on Behavioural and Social Computing (BESC)*, pages 1–6. IEEE, 2020.
- ²⁰⁶ Johan Bollen, Marijn Ten Thij, Fritz Breithaupt, Alexander TJ Barron, Lauren A Rutter, Lorenzo Lorenzo-Luaces, and Marten Scheffer. Historical language records reveal a surge of cognitive distortions in recent decades. *Proc. Natl. Acad. Sci. USA*, 118(30), 2021.
- ²⁰⁷ Krishna C Bathina, Marijn Ten Thij, Lorenzo Lorenzo-Luaces, Lauren A Rutter, and Johan Bollen. Individuals with depression express more distorted thinking on social media. *Nature Human Behaviour*, 5(4):458–466, 2021.
- ²⁰⁸ Jina Kim, Jieon Lee, Eunil Park, and Jinyoung Han. A deep learning model for detecting mental illness from user content on social media. *Scientific Reports*, 10:11846, 2020.
- ²⁰⁹ Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes C Eichstaedt, and Lyle H Ungar. Understanding and measuring psychological stress using social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 214–225, 2019.
- ²¹⁰ Esteban A Ríssola, David E Losada, and Fabio Crestani. A survey of computational methods for online mental state assessment on social media. *ACM Transactions on Computing for Healthcare*, 2(2):1–31, 2021.
- ²¹¹ Katrin Hänsel, Inna Wanxin Lin, Michael Sobolev, Whitney Muscat, Sabrina Yum-Chan, Munmun De Choudhury, John M Kane, and Michael L Birnbaum. Utilizing instagram data to identify usage patterns associated with schizophrenia spectrum disorders. *Frontiers in Psychiatry*, 12:691327, 2021.

- ²¹² Michael L Birnbaum, Raquel Norel, Anna Van Meter, Asra F Ali, Elizabeth Arenare, Elif Eyigoz, Carla Agurto, Nicole Germano, John M Kane, and Guillermo A Cecchi. Identifying signals associated with psychiatric illness utilizing language and images posted to facebook. *NPJ Schizophrenia*, 6(1):1–10, 2020.
- ²¹³ Mohammad El-Ramly, Hager Abu-Elyazid, Yousef Mo'men, Gameel Alshaer, Nardine Adib, Kareem Alaa Eldeen, and Mariam El-Shazly. CairoDep: Detecting depression in arabic posts using bert transformers. In *Proc. Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pages 207–212, 2021.
- ²¹⁴ Rodrigo Martínez-Castaño, Amal Htait, Leif Azzopardi, and Yashar Moshfeghi. BERT-based transformers for early detection of mental health illnesses. In *Proc. International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 189–200, 2021.
- ²¹⁵ Shaoxiong Ji, Tianlin Zhang, Luna Ansari, Jie Fu, Prayag Tiwari, and Erik Cambria. Mentalbert: Publicly available pretrained language models for mental healthcare. *arXiv preprint arXiv:2110.15621*, 2021.
- ²¹⁶ Oscar NE Kjell, Sverker Sikström, Katarina Kjell, and H Andrew Schwartz. Natural language analyzed with ai-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy. *Scientific Reports*, 12(1):1–9, 2022.
- ²¹⁷ Stevie Chancellor and Munmun De Choudhury. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digital Medicine*, 3(1):1–11, 2020.
- ²¹⁸ Enrique Garcia-Ceja, Michael Riegler, Tine Nordgreen, Petter Jakobsen, Ketil J Oedegaard, and Jim Tørresen. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*, 51:1–26, 2018.
- ²¹⁹ Paul Dagum. Digital biomarkers of cognitive function. *NPJ Digital Medicine*, 1(1):1–3, 2018.
- ²²⁰ John Zulueta, Andrea Piscitello, Mladen Rasic, Rebecca Easter, Pallavi Babu, Scott A Langenecker, Melvin McInnis, Olusola Ajilore, Peter C Nelson, Kelly Ryan, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: a biaffect digital phenotyping study. *Journal of Medical Internet Research*, 20(7):e9775, 2018.
- ²²¹ Regan Lee Mandryk and Max Valentin Birk. The potential of game-based digital biomarkers for modeling mental health. *JMIR Mental Health*, 6(4):e13485, 2019.
- ²²² M.J. Dechant, J. Frommel, and R. Mandryk. Assessing social anxiety through digital biomarkers embedded in a gaming task. In *Proc. 2021 CHI Conference on Human Factors in Computing Systems*, 2021.
- ²²³ Rolfe Winkler. Apple is working on iphone features to help detect depression, cognitive decline, 2021.
- ²²⁴ Jussi Seppälä, Ilaria De Vita, Timo Jäämsä, Jouko Miettunen, Matti Isohanni, Katya Rubinstein, Yoram Feldman, Eva Grasa, Iluminada Corripio, Jesus Berdun, et al. Mobile phone and wearable sensor-based mhealth approaches for psychiatric disorders and symptoms: systematic review. *JMIR Mental Health*, 6(2):e9819, 2019.
- ²²⁵ Prerna Chikarsal, Afsaneh Doryab, Michael Tummminia, Daniella K Villalba, Janine M Dutcher, Xinwen Liu, Sheldon Cohen, Kasey G Creswell, Jennifer Mankoff, J David Creswell, 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)*, 28(1):1–41, 2021.
- ²²⁶ John Torous, Mathew V Kiang, Jeanette Lorme, Jukka-Pekka Onnela, et al. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. *JMIR Mental Health*, 3:e5165, 2016.
- ²²⁷ David C Mohr, Mi Zhang, and Stephen M Schueller. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual Review of Clinical Psychology*, 13:23–47, 2017.
- ²²⁸ Kit Huckvale, Svetha Venkatesh, and Helen Christensen. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *NPJ Digital Medicine*, 2(1):1–11, 2019.
- ²²⁹ Judith J Prochaska, Erin A Vogel, Amy Chieng, Matthew Kendra, Michael Baiocchi, Sarah Pajarito, and Athena Robinson. A therapeutic relational agent for reducing problematic substance use (woebot): development and usability study. *Journal of Medical Internet Research*, 23(3):e24850, 2021.
- ²³⁰ Dieter Bohn. Amazon announces halo, a fitness band and app that scans your body and voice, 2020.
- ²³¹ fitbit. Understand your stress so you can manage it, 2022.
- ²³² Kintsugi. Kintsugi for health plans, 2022.
- ²³³ Brooke Auxier, Ariane Bucaille, and Kevin Westcott. Mental health goes mobile: The mental health app market will keep on growing, 2021.
- ²³⁴ John Torous, Sandra Bucci, Imogen H Bell, Lars V Kessing, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*, 20(3):318–335, 2021.
- ²³⁵ Jukka-Pekka Onnela and Scott L Rauch. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*, 41(7):1691–1696, 2016.
- ²³⁶ Dan J Stein, Naomi A Fineberg, and Samuel R Chamberlain. *Mental health in a digital world*. Elsevier, 2021.
- ²³⁷ Anzar Abbas, Katharina Schultebraucks, and Isaac R. Galatzer-Levy. Digital measurement of mental health: Challenges, promises, and future directions. *Psychiatric Annals*, 51(1):14–20, 2021.
- ²³⁸ Isaac R Galatzer-Levy and Richard A Bryant. 636,120 ways to have post-traumatic stress disorder. *Perspectives on Psychological Science*, 8(6):651–662, 2013.
- ²³⁹ Barry L Jacobs. Serotonin, motor activity and depression-related disorders. *American Scientist*, 82(5):456–463, 1994.
- ²⁴⁰ Valentina Gigliucci, Grainne O'Dowd, Sheena Casey, Danielle Egan, Sinead Gibney, and Andrew Harkin. Ketamine elicits sustained antidepressant-like activity via a serotonin-dependent mechanism. *Psychopharmacology*, 228(1):157–166, 2013.
- ²⁴¹ Isaac Galatzer-Levy, Anzar Abbas, Anja Ries, Stephanie Homan, Laura Sels, Vidya Koesmargyo, Vijay Yadav, Michael Colla, Hanne Scheerer, Stefan Vetter, Erich Seifritz, Urte Scholz, and Birgit Kleim. Validation of visual and auditory digital markers of suicidality in acutely suicidal psychiatric inpatients: Proof-of-concept study. *Journal of Medical Internet Research*, 23(6):e25199, 2021.
- ²⁴² Anzar Abbas, Colin Sauder, Vijay Yadav, Vidya Koesmargyo, Allison Aghayan, Serena Marecki, Miriam Evans, and Isaac R Galatzer-Levy. Remote digital measurement of facial and vocal markers of major depressive disorder severity and treatment response: a pilot study. *Frontiers in Digital Health*, 3:28, 2021.
- ²⁴³ Anzar Abbas, Bryan J Hansen, Vidya Koesmargyo, Vijay Yadav, Paul J Rosenfield, Omkar Patil, Marissa F Dockendorf, Matthew Moyer, Lisa A Shipley, M Mercedes Perez-Rodriguez, and Isaac R Galatzer-Levy. Facial and vocal markers of schizophrenia measured using remote smartphone assessments: Observational study. *JMIR Formative Research*, 6(1):e26276, 2022.
- ²⁴⁴ Li Zhang, Vidya Koesmargyo, and Isaac Galatzer-Levy. Estimation of clinical tremor using spatio-temporal adversarial autoencoder. In *Proc. 25th International Conference on Pattern Recognition (ICPR'20)*, pages 8259–8266. IEEE, 2021.
- ²⁴⁵ Michael P Ebwank, Ronan Cummins, Valentin Tablan, Sarah Bateup, Ana Catarino, Alan J Martin, and Andrew D Blackwell. Quantifying the association between psychotherapy content and clinical outcomes using deep learning. *JAMA Psychiatry*, 77(1):35–43, 2020.
- ²⁴⁶ Marcos Economides, Janis Martman, Megan J Bell, and Brad Sanderson. Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. *Mindfulness*, 9(5):1584–1593, 2018.
- ²⁴⁷ Sarah Kunkle, Manny Yip, Justin Hunt, E Watson, Dana Udall, Patricia Arean, Andrew Nierenberg, John A Naslund, et al. Association between care utilization and anxiety outcomes in an on-demand mental health system: Retrospective observational study. *JMIR Formative Research*, 5(1):e24662, 2021.
- ²⁴⁸ MP Ebwank, R Cummins, V Tablan, A Catarino, S Buchholz, and AD Blackwell. 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*, 31(3):300–312, 2021.
- ²⁴⁹ Nikolaos Flemotomas, Victor R Martinez, Zhuohao Chen, Torrey A Creed, David C Atkins, and Shrikanth Narayanan. Automated quality assessment of cognitive behavioral therapy sessions through highly contextualized language representations. *PLoS One*, 16(10):e0258639, 2021.
- ²⁵⁰ Nisarg A Patel and Atul J Butte. Characteristics and challenges of the clinical pipeline of digital therapeutics. *NPJ Digital Medicine*, 3(1):1–5, 2020.
- ²⁵¹ Thomas R Insel. Bending the curve for mental health: technology for a public health approach. *American Journal of Public Health*, 109(S3):S168–

- S170, 2019.
- ²⁵² Johanna B Folk, Marissa A Schiel, Rachel Oblath, Vera Feuer, Aditi Sharma, Shabana Khan, Bridget Doan, Chetana Kulkarni, Ujjwal Ramtekkar, Jessica Hawks, 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*, 61(2):277–290, 2022.
- ²⁵³ Birgit Wagner, Andrea B Horn, and Andreas Maercker. Internet-based versus face-to-face cognitive-behavioral intervention for depression: a randomized controlled non-inferiority trial. *Journal of Affective Disorders*, 152:113–121, 2014.
- ²⁵⁴ Dana Lahat, Tülay Adali, and Christian Jutten. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9):1449–1477, 2015.
- ²⁵⁵ AR Croitor-Sava, MC Martinez-Bisbal, T Laudadio, J Piquer, B Celda, A Heerschap, DM Sima, and Sabine Van Huffel. Fusing in vivo and ex vivo nmr sources of information for brain tumor classification. *Measurement Science and Technology*, 22(11):114012, 2011.
- ²⁵⁶ Tülay Adali, Yuri Levin-Schwartz, and Vince D Calhoun. Multimodal data fusion using source separation: Two effective models based on ICA and IVA and their properties. *Proceedings of the IEEE*, 103(9):1478–1493, 2015.
- ²⁵⁷ Tülay Adali, Yuri Levin-Schwartz, and Vince D Calhoun. Multimodal data fusion using source separation: Application to medical imaging. *Proceedings of the IEEE*, 103(9):1494–1506, 2015.
- ²⁵⁸ Vince Calhoun, Jing Sui, and Shile Qi. Multimodal fusion signature as trans-diagnostic psychiatric biomarker. *Biological Psychiatry*, 87(9):S37, 2020.
- ²⁵⁹ Yu-Dong Zhang, Zhengchao Dong, Shui-Hua Wang, Xiang Yu, Xujing Yao, Qinghua Zhou, Hua Hu, Min Li, Carmen Jiménez-Mesa, Javier Ramirez, et al. Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation. *Information Fusion*, 64:149–187, 2020.
- ²⁶⁰ Nicolle M Correa, Tulay Adali, Yi-Ou Li, and Vince D Calhoun. Canonical correlation analysis for data fusion and group inferences. *IEEE Signal Processing Magazine*, 27(4):39–50, 2010.
- ²⁶¹ Alain de Cheveigné, Giovanni M Di Liberto, Dorothée Arzouanian, Daniel DE Wong, Jens Hjortkjaer, Søren Fuglsang, and Lucas C Parra. Multitask canonical correlation analysis of brain data. *NeuroImage*, 186:728–740, 2019.
- ²⁶² Xun Chen, Z Jane Wang, and Martin McKeown. Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods. *IEEE Signal Processing Magazine*, 33(3):86–107, 2016.
- ²⁶³ Rogers F Silva and Sergey M Plis. How to integrate data from multiple biological layers in mental health? In *Personalized Psychiatry*, pages 135–159. Springer, 2019.
- ²⁶⁴ Guoxu Zhou, Qibin Zhao, Yu Zhang, Tülay Adali, Shengli Xie, and Andrzej Cichocki. Linked component analysis from matrices to high-order tensors: Applications to biomedical data. *Proceedings of the IEEE*, 104(2):310–331, 2016.
- ²⁶⁵ Vince D Calhoun and Tulay Adali. Feature-based fusion of medical imaging data. *IEEE Transactions on Information Technology in Biomedicine*, 13(5):711–720, 2008.
- ²⁶⁶ Vince D Calhoun, Jingyu Liu, and Tülay Adali. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*, 45(1):S163–S172, 2009.
- ²⁶⁷ Evrím Acar, Rasmus Bro, and Åge K Smilde. Data fusion in metabolomics using coupled matrix and tensor factorizations. *Proceedings of the IEEE*, 103(9):1602–1620, 2015.
- ²⁶⁸ Guoxu Zhou, Andrzej Cichocki, Yu Zhang, and Danilo P Mandic. Group component analysis for multiblock data: Common and individual feature extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 27(11):2426–2439, 2015.
- ²⁶⁹ Eric F Lock, Katherine A Hoadley, James Stephen Marron, and Andrew B Nobel. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *The Annals of Applied Statistics*, 7(1):523, 2013.
- ²⁷⁰ Alain Rakotomamonjy, Francis Bach, Stéphane Canu, and Yves Grandvalet. Simplemk1. *Journal of Machine Learning Research*, 9:2491–2521, 2008.
- ²⁷¹ Jérôme Mariette and Nathalie Villa-Vialaneix. Unsupervised multi-pkernel learning for heterogeneous data integration. *Bioinformatics*, 34(6):1009–1015, 2018.
- ²⁷² Letizia Squarcina, Umberto Castellani, Marcella Bellani, Cinzia Perlini, Antonio Lasalvia, Nicola Dusi, Chiara Bonetto, Doriana Cristofalo, Sarah Tosato, Gianluca Rambaldelli, et al. Classification of first-episode psychosis in a large cohort of patients using support vector machine and multiple kernel learning techniques. *NeuroImage*, 145:238–245, 2017.
- ²⁷³ Martin Dyrba, Michel Grothe, Thomas Kirste, and Stefan J Teipel. Multimodal analysis of functional and structural disconnection in a Alzheimer's disease using multiple kernel svm. *Human Brain Mapping*, 36(6):2118–2131, 2015.
- ²⁷⁴ Daoqiang Zhang, Yaping Wang, Luping Zhou, Hong Yuan, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Multimodal classification of alzheimer's disease and mild cognitive impairment. *NeuroImage*, 55(3):856–867, 2011.
- ²⁷⁵ Dhanesh Ramachandram and Graham W Taylor. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine*, 34(6):96–108, 2017.
- ²⁷⁶ Tao Zhou, Kim-Han Thung, Xiaofeng Zhu, and Dinggang Shen. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human Brain Mapping*, 40(3):1001–1016, 2019.
- ²⁷⁷ Nam D Nguyen, Jiawei Huang, and Daifeng Wang. A deep manifold-regularized learning model for improving phenotype prediction from multimodal data. *Nature Computational Science*, 2(1):38–46, 2022.
- ²⁷⁸ Andreas Holzinger, Bernd Malle, Anna Saranti, and Bastian Pfeifer. Towards multi-modal causability with graph neural networks enabling information fusion for explainable ai. *Information Fusion*, 71:28–37, 2021.
- ²⁷⁹ Niharika Shimona Dsouza, Mary Beth Nebel, Deana Crocetti, Joshua Robinson, Stewart Mostofsky, and Archana Venkataraman. 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*, pages 119–130. PMLR, 2021.
- ²⁸⁰ Wen Zhang, Liang Zhan, Paul Thompson, and Yalin Wang. Deep representation learning for multimodal brain networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 613–624. Springer, 2020.
- ²⁸¹ Zhaoming Kong, Lichao Sun, Hao Peng, Liang Zhan, Yong Chen, and Li-fang He. Multiplex graph networks for multimodal brain network analysis. *arXiv preprint arXiv:2108.00158*, 2021.
- ²⁸² Emine Elif Tulay, Barış Metin, Nevzat Tarhan, and Mehmet Kemal Arıkan. Multimodal neuroimaging: basic concepts and classification of neuropsychiatric diseases. *Clinical EEG and Neuroscience*, 50(1):20–33, 2019.
- ²⁸³ Jing Sui, Godfrey Pearson, Arvind Caprihan, Tülay Adali, Kent A Kiehl, Jingyu Liu, Jeremy Yamamoto, and Vince D Calhoun. Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model. *NeuroImage*, 57(3):839–855, 2011.
- ²⁸⁴ Evrím Acar, Carla Schenker, Yuri Levin-Schwartz, Vince D Calhoun, and Tülay Adali. Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data. *Frontiers in Neuroscience*, 13:416, 2019.
- ²⁸⁵ Benedetta Vai, Lorenzo Parenti, Irene Bollettini, Cristina Cara, Chiara Verga, Elisa Melloni, Elena Mazza, Sara Poletti, Cristina Colombo, and Francesco Benedetti. Predicting differential diagnosis between bipolar and unipolar depression with multiple kernel learning on multimodal structural neuroimaging. *European Neuropsychopharmacology*, 34:28–38, 2020.
- ²⁸⁶ Qiongmin Zhang, Qizhu Wu, Hongru Zhu, Ling He, Hua Huang, Junran Zhang, and Wei Zhang. Multimodal mri-based classification of trauma survivors with and without post-traumatic stress disorder. *Frontiers in Neuroscience*, 10:292, 2016.
- ²⁸⁷ Xiaocheng Zhou, Qingmin Lin, Yuanyuan Gui, Zixin Wang, Manhua Liu, and Hui Lu. Multimodal MR images-based diagnosis of early adolescent attention-deficit/hyperactivity disorder using multiple kernel learning. *Frontiers in Neuroscience*, 15, 2021.
- ²⁸⁸ Eloy Geenjaar, Noah Lewis, Zening Fu, Rohan Venkatdas, Sergey Plis, and Vince Calhoun. Fusing multimodal neuroimaging data with a variational autoencoder. In *Proc. 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3630–3633. IEEE, 2021.
- ²⁸⁹ Jin Liu, Xiang Wang, Xiangrong Zhang, Yi Pan, Xiaosheng Wang, and Jianxin Wang. Mmm: classification of schizophrenia using multi-modality multi-atlas feature representation and multi-kernel learning. *Multimedia Tools and Applications*, 77(22):29651–29667, 2018.
- ²⁹⁰ Sergey M Plis, Md Faijul Amin, Adam Chekroud, Devon Hjelm, Eswar Damaraju, Hyo Jong Lee, Juan R Bustillo, KyungHyun Cho, Godfrey D Pearson, and Vince D Calhoun. Reading the (functional) writing on the

- (structural) wall: Multimodal fusion of brain structure and function via a deep neural network based translation approach reveals novel impairments in schizophrenia. *NeuroImage*, 181:734–747, 2018.
- ²⁹¹ Md Abdur Rahaman, Jiayu Chen, Zening Fu, Noah Lewis, Armin Iraji, and Vince D Calhoun. 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)*, pages 3267–3272. IEEE, 2021.
- ²⁹² MABS Akhonda, Yuri Levin-Schwartz, Vince D Calhoun, and Tülay Adalı. Association of neuroimaging data with behavioral variables: A class of multivariate methods and their comparison using multi-task fMRI data. *Sensors*, 22(3):1224, 2022.
- ²⁹³ Yiting Wang, Wei-Bang Jiang, Rui Li, and Bao-Liang Lu. 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)*, pages 1575–1578. IEEE, 2021.
- ²⁹⁴ Natasha Jaques, Sara Taylor, Akane Sano, and Rosalind Picard. Multi-task, multi-kernel learning for estimating individual wellbeing. In *Proc. NIPS Workshop on Multimodal Machine Learning, Montreal, Quebec*, volume 898, page 3, 2015.
- ²⁹⁵ Genevieve Lam, Huang Dongyan, and Weisi Lin. Context-aware deep learning for multi-modal depression detection. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3946–3950. IEEE, 2019.
- ²⁹⁶ F Parra, Y Benzezeth, and F Yang. Automatic assessment of emotion dysregulation in american, french, and tunisian adults and new developments in deep multimodal fusion: Cross-sectional study. *JMIR Mental Health*, 9(1):e34333, 2022.
- ²⁹⁷ Afsaneh Doryab, Daniella K Villalba, Prerna Chikarsal, Janine M Dutcher, Michael Tumminia, Xinwen Liu, Sheldon Cohen, Kasey Creswell, Jennifer Mankoff, John D Creswell, 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*, 7(7):e13209, 2019.
- ²⁹⁸ Jiaxuan He, Sijie Mai, and Haifeng Hu. A unimodal reinforced transformer with time squeeze fusion for multimodal sentiment analysis. *IEEE Signal Processing Letters*, 28:992–996, 2021.
- ²⁹⁹ Hamdi Dibeklioğlu, Zakia Hammal, and Jeffrey F Cohn. Dynamic multimodal measurement of depression severity using deep autoencoding. *IEEE Journal of Biomedical and Health Informatics*, 22(2):525–536, 2017.
- ³⁰⁰ David P. Herzog, Holger Beckmann, Klaus Lieb, Soojin Ryu, and Marianne B. Müller. Understanding and predicting antidepressant response: Using animal models to move toward precision psychiatry. *Frontiers in Psychiatry*, 9:512, 2018.
- ³⁰¹ Tracy L. Bale, Ted Abel, Huda Akil, William A. Carlezon Jr., Bita Moghadam, Eric J. Nestler, Kerry J. Ressler, and Scott M. Thompson. The critical importance of basic animal research for neuropsychiatric disorders. *Neuropsychopharmacology*, 44(8):1349–1353, 2019.
- ³⁰² B. Labonté, O. Engmann, I. Purushothaman, C. Menard, J. Wang, C. Tan, J. R. Scarpa, G. Moy, Y. E. Loh, M. Cahill, Z. S. Lorsch, P. J. Hamilton, E. S. Calipari, G. E. Hodes, O. Issler, H. Kronman, M. Pfau, A. L. J. Obradovic, Y. Dong, R. L. Neve, S. Russo, A. Kazarskis, C. Tamminga, N. Mechawar, G. Turecki, B. Zhang, L. Shen, and E. J. Nestler. Sex-specific transcriptional signatures in human depression. *Nat Med*, 23(9):1102–1111, 2017.
- ³⁰³ Corina Nagy, Malosree Maitra, Arnaud Tanti, Matthew Suderman, Jean-Francois Thérioux, Maria Antonietta Davoli, Kelly Perlman, Volodymyr Yerko, Yu Chang Wang, Shreejoy J. Tripathy, Paul Pavlidis, Naguib Mechawar, Jiannis Ragoussis, and Gustavo Turecki. Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nature Neuroscience*, 23(6):771–781, 2020.
- ³⁰⁴ B. S. McEwen, N. P. Bowles, J. D. Gray, M. N. Hill, R. G. Hunter, I. N. Karatsoreos, and C. Nasca. Mechanisms of stress in the brain. *Nat Neurosci*, 18(10):1353–63, 2015.
- ³⁰⁵ Amalia Floriou-Servou, Lukas von Ziegler, Luzia Stalder, Oliver Sturman, Mattia Privitera, Anahita Rossi, Alessio Cremonesi, Beat Thöny, and Johannes Bohacek. Distinct proteomic, transcriptomic, and epigenetic stress responses in dorsal and ventral hippocampus. *Biological Psychiatry*, 84(7):531–541, 2018.
- ³⁰⁶ B. Bigio, A. A. Mathe, V. C. Sousa, D. Zelli, P. Svenningsson, B. S. McEwen, and C. Nasca. Epigenetics and energetics in ventral hippocampus mediate rapid antidepressant action: Implications for treatment resistance. *Proc. Natl. Acad. Sci. USA*, 113(28):7906–7911, 2016.
- ³⁰⁷ M. H. Flight. Antidepressant epigenetic action. *Nat Rev Neurosci*, 14(4):226, 2013.
- ³⁰⁸ Tim Stuart and Rahul Satija. Integrative single-cell analysis. *Nature Reviews Genetics*, 20(5):257–272, 2019.
- ³⁰⁹ Raphael Petegrosso, Zhiliu Li, and Rui Kuang. Machine learning and statistical methods for clustering single-cell rna-sequencing data. *Briefings in Bioinformatics*, 21(4):1209–1223, 2020.
- ³¹⁰ Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, III Mauck, William M., Shiwei Zheng, Andrew Butler, Maddie J. Lee, Aaron J. Wilk, Charlotte Darby, Michael Zager, Paul Hoffman, Marlon Stoeckius, Efthymia Papalexi, Eleni P. Mimitou, Jaison Jain, Avi Srivastava, Tim Stuart, Lamar M. Fleming, Bertrand Yeung, Angela J. Rogers, Juliana M. McElrath, Catherine A. Blish, Raphael Gottardo, Peter Smibert, and Rahul Satija. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021.
- ³¹¹ Matthew Amodio, David van Dijk, Krishnan Srinivasan, William S. Chen, Hussein Mohsen, Kevin R. Moon, Allison Campbell, Yujiao Zhao, Xiaomei Wang, Manjunatha Venkataswamy, Anita Desai, V. Ravi, Priti Kumar, Ruth Montgomery, Guy Wolf, and Smita Krishnaswamy. Exploring single-cell data with deep multitasking neural networks. *Nature Methods*, 16(11):1139–1145, 2019.
- ³¹² G. Eraslan, L. M. Simon, M. Mircea, N. S. Mueller, and F. J. Theis. Single-cell rna-seq denoising using a deep count autoencoder. *Nat Commun*, 10(1):390, 2019.
- ³¹³ D. Wang and J. Gu. Vasc: Dimension reduction and visualization of single-cell rna-seq data by deep variational autoencoder. *Genomics Proteomics Bioinformatics*, 16(5):320–331, 2018.
- ³¹⁴ J. Wang, D. Agarwal, M. Huang, G. Hu, Z. Zhou, C. Ye, and N. R. Zhang. Data denoising with transfer learning in single-cell transcriptomics. *Nat Methods*, 16(9):875–878, 2019.
- ³¹⁵ J. Cao, D. A. Cusanovich, V. Ramani, D. Aghamirzaie, H. A. Pliner, A. J. Hill, R. M. Daza, J. L. McFaline-Figueroa, J. S. Packer, L. Christiansen, F. J. Steemers, A. C. Adey, C. Trapnell, and J. Shendure. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*, 361(6409):1380–1385, 2018.
- ³¹⁶ S. Chen, B. B. Lake, and K. Zhang. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat Biotechnol*, 37(12):1452–1457, 2019.
- ³¹⁷ Sanja Vickovic, Gökcen Eraslan, Fredrik Salmén, Johanna Klughammer, Linnea Stenbeck, Denis Schapiro, Tarmo Äijö, Richard Bonneau, Ludvig Bergenstråhl, José Fernández Navarro, et al. High-definition spatial transcriptomics for *in situ* tissue profiling. *Nat Methods*, 16(10):987–990, 2019.
- ³¹⁸ S. G. Rodrigues, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, and E. Z. Macosko. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434):1463–1467, 2019.
- ³¹⁹ C. Nasca, N. Rasgon, and B. McEwen. An emerging epigenetic framework of systemic and central mechanisms underlying stress-related disorders. *Neuropsychopharmacology*, 44(1):235–236, 2019.
- ³²⁰ C. Nasca, J. Dobbin, B. Bigio, K. Watson, P. de Angelis, M. Kautz, A. Cochran, A. A. Mathé, J. H. Kocsis, F. S. Lee, J. W. Murrough, B. S. McEwen, and N. Rasgon. 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*, 26(9):5140–5149, 2021.
- ³²¹ C. Nasca, O. Barnhill, P. DeAngelis, K. Watson, J. Lin, J. Beasley, S. P. Young, A. Myoraku, J. Dobbin, B. Bigio, B. McEwen, and N. Rasgon. Multidimensional predictors of antidepressant responses: Integrating mitochondrial, genetic, metabolic and environmental factors with clinical outcomes. *Neurobiol Stress*, 15:100407, 2021.
- ³²² K. R. Dean, R. Hammamieh, S. H. Mellon, D. Abu-Amara, J. D. Flory, G. Guffanti, K. Wang, Jr. Daigle, B. J., A. Gautam, I. Lee, R. Yang, L. M. Almli, F. S. Bersani, N. Chakraborty, D. Donohue, K. Kerley, T. K. Kim, E. Laska, M. Young Lee, D. Lindqvist, A. Lori, L. Lu, B. Misganaw, S. Muhie, J. Newman, N. D. Price, S. Qin, V. I. Reus, C. Siegel, P. R. Somvanshi, G. S. Thakur, Y. Zhou, L. Hood, K. J. Ressler, O. M. Wolkowitz, R. Yehuda, M. Jett, F. J. Doyle, and C. Marmar. Multi-omic biomarker identification and validation for diagnosing warzone-related post-traumatic stress disorder. *Mol Psychiatry*, 25(12):3337–3349, 2020.

- ³²³ K. Schultebraucks, M. Qian, D. Abu-Amara, K. Dean, E. Laska, C. Siegel, A. Gautam, G. Guffanti, R. Hammamieh, B. Misganaw, S. H. Mellon, O. M. Wolkowitz, E. M. Blessing, A. Etkin, K. J. Ressler, 3rd Doyle, F. J., M. Jett, and C. R. Marmar. Pre-deployment risk factors for ptsd in active-duty personnel deployed to afghanistan: a machine-learning approach for analyzing multivariate predictors. *Mol Psychiatry*, 26(9):5011–5022, 2021.
- ³²⁴ Z. S. Lorsch, A. Ambesi-Impiombato, R. Zenowich, I. Morganstern, E. Leahy, M. Bansal, E. J. Nestler, and T. Hanania. Computational analysis of multidimensional behavioral alterations after chronic social defeat stress. *Biol Psychiatry*, 89(9):920–928, 2021.
- ³²⁵ Vadim Alexandrov, Dani Brunner, Taleen Hanania, and Emer Leahy. High-throughput analysis of behavior for drug discovery. *European Journal of Pharmacology*, 750:82–89, 2015.
- ³²⁶ C. Nasca, C. Menard, G. Hodes, B. Bigio, C. Pena, Z. Lorsch, D. Zelli, A. Ferris, V. Kana, I. Purushothaman, J. Dobbin, M. Nassim, P. DeAngelis, M. Merad, N. Rasgon, M. Meaney, E. J. Nestler, B. S. McEwen, and S. J. Russo. Multidimensional predictors of susceptibility and resilience to social defeat stress. *Biol Psychiatry*, 86(6):483–491, 2019.
- ³²⁷ David Gunning and David Aha. DARPA’s explainable artificial intelligence (XAI) program. *AI Magazine*, 40(2):44–58, 2019.
- ³²⁸ Veit Roessner, Josefine Rothe, Gregor Kohls, Georg Schomerus, Stefan Ehrlich, and Christian Beste. Taming the chaos?! using explainable artificial intelligence (xai) to tackle the complexity in mental health research, 2021.
- ³²⁹ Hans-Christian Thorsen-Meyer, Annelaura B Nielsen, Anna P Nielsen, Benjamin Skov Kaas-Hansen, Prof Palle Toft, Jens Schierbeck, Thomas Strøm, Piotr J Chmura, Marc Heimann, Lars Dybdahl, Lasse Spangsege, Patrick Hulsen, Kirstine Bellring, Prof Søren Brunak, and Prof Anders Perner. 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*, 2(4):e179–e191, 2020.
- ³³⁰ Yi-han Sheu. Illuminating the black box: Interpreting deep neural network models for psychiatric research. *Frontiers in Psychiatry*, page 1091, 2020.
- ³³¹ Mehrdad Jazayeri and Arash Afraz. Navigating the neural space in search of the neural code. *Neuron*, 93(5):1003–1014, 2017.
- ³³² Zhe Sage Chen and Bijan Pesaran. Improving scalability in systems neuroscience. *Neuron*, 109(11):1776–1790, 2021.
- ³³³ Christoph Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2022.
- ³³⁴ Jueda Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Elsevier, 1988.
- ³³⁵ Benjamin Letham, Cynthia Rudin, Tyler H. McCormick, and David Madigan. Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. *Annals of Applied Statistics*, 9(3):1350–1371, 2015.
- ³³⁶ Yitong Li, Michael Murias, Samantha Major, Geraldine Dawson, Kafui Dzirasa, Lawrence Carin, and David E Carlson. Targeting EEG/LFP synchrony with neural nets. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- ³³⁷ James Y. Zou and Ryan P. Adams. Priors for diversity in generative latent variable models. In *Advances in Neural Information Processing Systems (NIPS) 25*, 2012.
- ³³⁸ Rahul Nair, Massimiliano Mattetti, Elizabeth Daly, Dennis Wei, Oznur Alkan, and Yufeng Zhang. What changed? interpretable model comparison. In *Proc. Int. Joint Conf. Artificial Intelligence (IJCAI’21)*, 2021.
- ³³⁹ E.M. Daly, M Mattetti, O Alkan, and R Nair. User driven model adjustment via boolean rule explanation. In *Proc. 35th Conf. Artificial Intelligence (AAAI’21)*, pages 5896–5904, 2021.
- ³⁴⁰ Quentin JM Huys, Tiago V Maia, and Michael J Frank. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, 19(3):404–413, 2016.
- ³⁴¹ Michael Breakspear. Dynamic models of large-scale brain activity. *Nature Neuroscience*, 20(3):340–352, 2017.
- ³⁴² John D Murray, Murat Demirtaş, and Alan Anticevic. Biophysical modeling of large-scale brain dynamics and applications for computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(9):777–787, 2018.
- ³⁴³ L Papadopoulos, C. W. Lynn, D Battaglia, and D.S. Bassett. Relations between large-scale brain connectivity and effects of regional stimulation depend on collective dynamical state. *PLoS Computational Biology*, page e1008144, 2020.
- ³⁴⁴ H. R. Wilson and Jack D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12:1–24, 1972.
- ³⁴⁵ Rishidev Chaudhuri, Kenneth Knoblauch, Marie-Alice Gariel, Henry Kennedy, and Xiao-Jing Wang. A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron*, 88(2):419–431, 2015.
- ³⁴⁶ K.J. Friston, L. Harrison, and W Penny. Dynamica causal modeling. *NeuroImage*, 19:1273–1302, 2003.
- ³⁴⁷ K.J. Friston, J. Kahan, B. Biswal, and A Razi. A DCM for resting state fMRI. *NeuroImage*, 94:396–407, 2014.
- ³⁴⁸ Amit Etkin. Addressing the causality gap in human psychiatric neuroscience. *JAMA Psychiatry*, 75(1):3–4, 2018.
- ³⁴⁹ Philip M Lewis, Richard H Thomson, Jeffrey V Rosenfeld, and Paul B Fitzgerald. Brain neuromodulation techniques: a review. *The Neuroscientist*, 22(4):406–421, 2016.
- ³⁵⁰ Vincenzo Romei, Gregor Thut, and Juha Silvanto. Information-based approaches of noninvasive transcranial brain stimulation. *Trends in Neurosciences*, 39(11):782–795, 2016.
- ³⁵¹ M-C. Lo and A.S. Widege. Closed-loop neuromodulation systems: next generation treatments for psychiatric illness. *Int. Rev. Psych.*, 29:191–204, 2017.
- ³⁵² A.C. Chen, D J Oathes, C Chang, T Bradley, Z.W. Zhou, L.M. Williams, G.H. Glover, K Deisseroth, and A Etkin. Causal interactions between fronto-parietal central executive and default-mode networks in humans. *Proc. Natl. Acad. Sci. USA*, 110(49):19944–19949, 2013.
- ³⁵³ Justyna Hobot, Michał Klincewicz, Kristian Sandberg, and Michał Wierzchoń. Causal inferences in repetitive transcranial magnetic stimulation research: challenges and perspectives. *Frontiers in Human Neuroscience*, 14:574, 2021.
- ³⁵⁴ Evelyn Tang and Danielle S Bassett. Control of dynamics in brain networks. *Reviews of Modern Physics*, 90(3):031003, 2018.
- ³⁵⁵ Pragya Srivastava, Erfan Nozari, Jason Z Kim, Harang Ju, Dale Zhou, Cassiano Becker, Fabio Pasqualetti, George J Pappas, and Danielle S Bassett. Models of communication and control for brain networks: distinctions, convergence, and future outlook. *Network Neuroscience*, 4(4):1122–1159, 2020.
- ³⁵⁶ Xiaolong Zhang, Urs Braun, Heike Tost, and Danielle S Bassett. Data-driven approaches to neuroimaging analysis to enhance psychiatric diagnosis and therapy. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(8):780–790, 2020.
- ³⁵⁷ Michael D Fox, Randy L Buckner, Hesheng Liu, M Mallar Chakravarty, Andres M Lozano, and Alvaro Pascual-Leone. Resting-state networks link invasive and noninvasive brain stimulation across diverse psychiatric and neurological diseases. *Proc. Natl. Acad. Sci. USA*, 111(41):E4367–E4375, 2014.
- ³⁵⁸ Jean-Marc Fellous, Guillermo Sapiro, Andrew Rossi, Helen Mayberg, and Michele Ferrante. Explainable artificial intelligence for neuroscience: behavioral neurostimulation. *Frontiers in Neuroscience*, 13:1346, 2019.
- ³⁵⁹ Joelle Pineau, Arthur Guez, Robert Vincent, Gabriella Panuccio, and Massimo Avoli. Treating epilepsy via adaptive neurostimulation: a reinforcement learning approach. *International Journal of Neural Systems*, 19(4):227–240, 2009.
- ³⁶⁰ Sina Tafazoli, Camden J MacDowell, Zongda Che, Katherine C Letai, Cynthia R Steinhardt, and Timothy J Buschman. Learning to control the brain through adaptive closed-loop patterned stimulation. *Journal of Neural Engineering*, 17(5):056007, 2020.
- ³⁶¹ Amin Zandvakili, Noah S Philip, Stephanie R Jones, Audrey R Tyrka, Benjamin D Greenberg, and Linda L Carpenter. 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*, 252:47–54, 2019.
- ³⁶² Sebastian Vollmer, Bilal A Mateen, Gergo Bohner, Franz J Király, Rayid Ghani, Pall Jonsson, Sarah Cumbers, Adrian Jonas, Katherine SL McAlister, Puja Myles, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ*, 368, 2020.
- ³⁶³ Ronald C Kessler, Wai Tat Chiu, Olga Demler, and Ellen E Walters. Prevalence, severity, and comorbidity of 12-month dsm-iv disorders in the na-

- tional comorbidity survey replication. *Archives of General Psychiatry*, 62(6):617–627, 2005.
- ³⁶⁴ Sachin R Pendse, Daniel Nkemelu, Nicola J Bidwell, Sushrut Jadhav, Soumitra Pathare, Munmun De, and Neha Kumar. From treatment to healing: Envisioning a decolonial digital mental health. In *Proc. CHI Conference on Human Factors in Computing Systems (CHI'22)*, 2022.
- ³⁶⁵ Lena Palaniyappan. More than a biomarker: could language be a biosocial marker of psychosis? *NPJ Schizophrenia*, 7(1):1–5, 2021.
- ³⁶⁶ Yunan Luo, Jian Peng, and Jianzhu Ma. When causal inference meets deep learning. *Nature Machine Intelligence*, 2(8):426–427, 2020.
- ³⁶⁷ Mattia Prosperi, Yi Guo, Matt Sperrin, James S Koopman, Jae S Min, Xing He, Shannan Rich, Mo Wang, Iain E Buchan, and Jiang Bian. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7):369–375, 2020.
- ³⁶⁸ Irene Neuner, Tanja Veselinović, Shukti Ramkiran, Ravichandran Rajkumar, Geron Johannes Schnellbaecher, and N Jon Shah. 7T ultra-high-field neuroimaging for mental health: an emerging tool for precision psychiatry? *Translational Psychiatry*, 12(1):1–10, 2022.
- ³⁶⁹ Pranav Rajpurkar, Emma Chen, Oishi Banerjee, and Eric J Topol. AI in health and medicine. *Nature Medicine*, 28:1–8, 2022.
- ³⁷⁰ David Grande, Nandita Mitra, Raghuram Iyengar, Raina M Merchant, David A Asch, Meghana Sharma, and Carolyn C Cannuscio. Consumer willingness to share personal digital information for health-related uses. *JAMA Network Open*, 5(1):e2144787–e2144787, 2022.
- ³⁷¹ Brittany I Davidson. The crossroads of digital phenotyping. *General Hospital Psychiatry*, 74:126–132, 2022.
- ³⁷² Diane M Korngiebel and Sean D Mooney. Considering the possibilities and pitfalls of generative pre-trained transformer 3 (gpt-3) in healthcare delivery. *NPJ Digital Medicine*, 4(1):1–3, 2021.
- ³⁷³ N Polyzotis and M. Zaharia. What can data-centric ai learn from data and ml engineering? *arXiv preprint*, 2021.
- ³⁷⁴ Richard J Chen, Ming Y Lu, Tiffany Y Chen, Drew FK Williamson, and Faisal Mahmood. Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, 5(6):493–497, 2021.
- ³⁷⁵ Lan Lan, Lei You, Zeyang Zhang, Zhiwei Fan, Weiling Zhao, Nianyin Zeng, Yidong Chen, and Xiaobo Zhou. Generative adversarial networks and its applications in biomedical informatics. *Frontiers in Public Health*, 8:164, 2020.
- ³⁷⁶ David Geng, Ayham Alkhachroum, Manuel A Melo Bicchi, Jonathan R Jagid, Iahn Cajigas, and Zhe Sage Chen. Deep learning for robust detection of interictal epileptiform discharges. *Journal of Neural Engineering*, 18(5):056015, 2021.
- ³⁷⁷ John Weldon, Tomas Ward, and Eoin Brophy. Generation of synthetic electronic health records using a federated gan. *arXiv preprint arXiv:2109.02543*, 2021.
- ³⁷⁸ L Chen, C Xia, and H Sun. Recent advances of deep learning in psychiatric disorders. *Precision Clinical Medicine*, 3(3):202–213, 2020.
- ³⁷⁹ Ziwei Zhang, Peng Cui, and Wenwu Zhu. Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 34:249–270, 2022.
- ³⁸⁰ Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.
- ³⁸¹ Tesfa Dejenie Habtewold, Lyan H Rodijk, Edith J Liemburg, Grigory Sidorenkov, H Marike Boezen, Richard Bruggeman, and Behrooz Z Alizadeh. A systematic review and narrative synthesis of data-driven studies in schizophrenia symptoms and cognitive deficits. *Translational Psychiatry*, 10(1):1–24, 2020.
- ³⁸² Jenna Wiens, Suchi Saria, Mark Sendak, Marzyeh Ghassemi, Vincent X. Liu, Kenneth Jung Finale Doshi-Velez, Katherine Heller, David Kale, Mohammed Saeed, Pilar N. Ossorio, Sonoo Thadaney-Israni, and Anna Goldenberg. Do no harm: a roadmap for responsible machine learning for health care. *Nature Medicine*, 25:1337–1340, 2019.