EXPERT REVIEW   OPEN

Practical AI application in psychiatry: historical review and future directions

Jie Sun[1,2,13], Tangsheng Lu[3,13], Xuexiao Shao[4], Ying Han[3], Yu Xia[5], Yongbo Zheng[2,6], Yongxiang Wang[4], Xinmin Li[7], Arun Ravindran[8,9], Lizhou Fan[10], Yin Fang[11], Xiujun Zhang[11], Nisha Ravindran[8,9], Yumei Wang[4,6✉], Xiaoxing Liu[2✉] and Lin Lu[2,3,4,12✉]

The integration of artificial intelligence (AI) in mental healthcare holds promise for enhancing diagnostic precision, treatment efficacy, and personalized care. Despite AI's potential to analyze vast datasets and identify subtle patterns, its clinical adoption in psychiatry remains limited. This review critically examines the emerging role of AI in psychiatry, elucidating its utility, challenges, and implications for clinical practice. Through an extensive analysis of the existing literature and empirical evidence, we seek to inform psychiatric stakeholders about both opportunities and obstacles that are presented by AI. We evaluate AI's potential to improve diagnostic accuracy, prognostic performance, and therapeutic interventions. Our pragmatic approach bridges the gap between theoretical advancements and practical implementation, providing valuable insights and actionable recommendations for psychiatric professionals. This article highlights the supportive role of AI, advocating for its judicious integration to enhance patient outcomes while maintaining the human-centric essence of psychiatric practice. By addressing these challenges and fostering collaboration, AI can significantly advance mental healthcare, reduce clinical burdens, and improve patient outcomes.

INTRODUCTION

Mental health is a global priority, with such disorders as anxiety and depression affecting 29% of the global population during their lifetime, as recognized by the World Health Organization [1, 2]. Beyond personal impacts, these disorders incur an estimated USD $1 trillion in global economic costs annually [2]. Traditionally, mental health diagnosis and treatment have relied heavily on the experiential knowledge and judgments of human physicians. However, the increasing complexity of mental healthcare, compounded by a proliferation of diagnostic tests, a growing body of biomedical and clinical evidence, and the challenges of managing comorbidities in aging populations, necessitates novel approaches [3]. Artificial intelligence (AI) and machine learning (ML), conceptually suggested by Alan Turing and coined by Arthur Samuel in the 1950s, have the potential to transform mental healthcare by enhancing diagnostic and treatment processes through advanced data analysis [4].

Artificial intelligence systems can process and analyze vast datasets, identify subtle patterns, and generate predictions that might elude human clinicians [5]. This technology has the potential to facilitate earlier detection, more accurate diagnosis, individualized interventions, and streamlined service delivery through automation (Fig. 1). However, the inherent complexity of mental health conditions, nuances of patient interactions, subjective judgement and the empathy to integrate to psychiatric practice, pose significant difficulties with AI replication. This discrepancy highlights a significant gap between theoretical advancements and practical implementation. Despite a quickly growing body of research in AI application to health care, the potential of AI to fully replace the role of clinicians in diagnosing and making treatment decisions, remains currently unrealistic. Consequently, the most efficient and impactful approach to integrating AI into clinical practice is to employ it as a supportive tool to enhance clinical decision-making. By offering data-driven insights, predictive analytics, and the automation of routine tasks, AI can assist clinicians in making more informed decisions, thereby improving patient outcomes while allowing physicians to concentrate on humanistic care.

Realizing the potential benefits of AI for clinical practice necessitates a comprehensive understanding of its merits and

[1]Pain Medicine Center, Peking University Third Hospital, Peking University, Beijing 100191, China. [2]Peking University Sixth Hospital, Peking University Institute of Mental Health, NHC Key Laboratory of Mental Health (Peking University), and National Clinical Research Center for Mental Disorders (Peking University Sixth Hospital), Peking University, Beijing 100191, China. [3]National Institute on Drug Dependence and Beijing Key Laboratory of Drug Dependence, Peking University, Beijing 100191, China. [4]Institute of Brain Science and Brain-inspired Research, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, Shandong 250117, China. [5]School of Basic Medical Sciences, Peking University, Beijing 100871, China. [6]Department of Psychology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong 250021, China. [7]Department of Psychiatry, Faculty of Medicine, 1E1 Walter Mackenzie Health Sciences Centre, University of Alberta, Edmonton, AB T6G 2B7, Canada. [8]Institute of Medical Science, Toronto, ON M5S 1A8, Canada. [9]Department of Psychiatry, University of Toronto, Toronto, ON M5T 1R8, Canada. [10]Artificial Intelligence in Medicine Program, Mass General Brigham, Harvard Medical School, Boston, MA 02115, USA. [11]North China University of Science and Technology, Tangshan, Hebei 063210, China. [12]Peking-Tsinghua Center for Life Sciences and PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China. [13]These authors contributed equally: Jie Sun, Tangsheng Lu. ✉email: wangyumei19771021@hotmail.com; liuxiaoxing@pku.edu.cn; linlu@bjmu.edu.cn
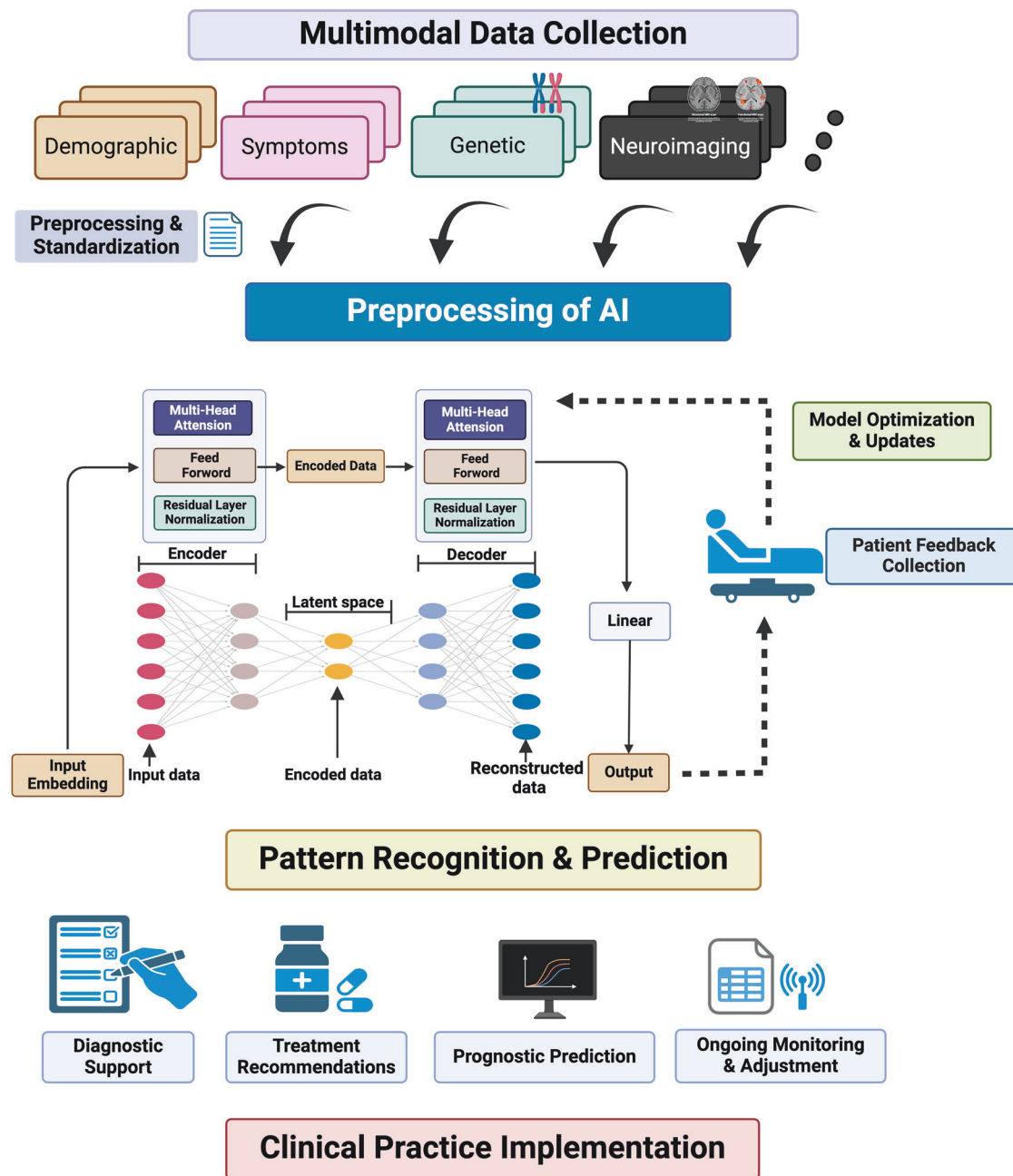
Fig. 1 Artificial intelligence model application in psychiatry. Data flow from multimodal sources through preprocessing and AI-based modeling using a transformer architecture. The model consists of input embedding, encoding, and decoding stages utilizing multi-head attention and feed-forward layers, residual layer normalization, and a linear output layer. Patient feedback informs model optimization and updates, and pattern recognition and prediction support clinical practice through diagnostic support, treatment recommendations, prognostic prediction, and ongoing monitoring.

limitations within clinical settings [6]. The dichotomy between AI's potential and its practical applications presents a formidable challenge, underscoring the need for physicians to have a deep understanding of its mechanisms and potential clinical applications, as well as related clinical measurements and performance metrics. Psychiatric practitioners and researchers must thoroughly grasp these tasks to effectively assign, monitor, and evaluate AI-assisted interventions. The present review aims to elucidate the current landscape of AI applications in mental health from the perspective of psychiatry physicians. We assess AI's potential to enhance diagnostic precision, prognostic performance, treatment efficacy, and the development of interventions, such as precision

treatment, chatbots, and digital monitoring. By bridging the gap between research and clinical application, this review seeks to illuminate new possibilities for clinicians who are unfamiliar with AI. We illustrate how AI can alleviate clinical burdens and ultimately foster more effective and efficient mental healthcare (Fig. 2).

APPLICATIONS OF AI TO IMPROVE CLINICAL DIAGNOSTIC SYSTEM
The accuracy of AI diagnostics in psychiatric diseases across various population datasets has shown promising yet variable
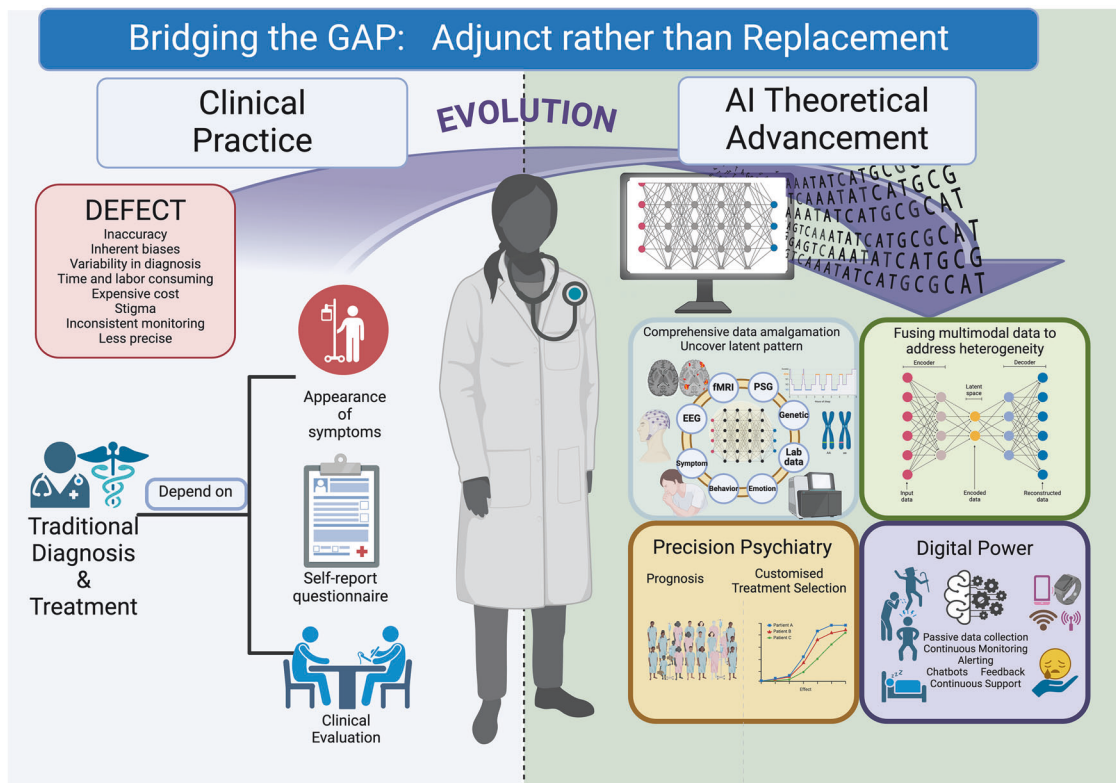
Fig. 2 Synergy between clinical expertise and analytical capabilities of AI for optimized patient care.

results. The efficacy of AI for psychiatric diagnosis processes is contingent upon the quality and the extent of the data that are utilized, as well as the individual psychiatric condition. Current research indicates that ML models can reach diagnostic accuracies of 48.1–62.0% when applied to comprehensive multivariate neuroimaging data and polygenic risk scores [7]. AI models may at times outperform random guessing for diagnostic accuracy, but only by a modest margin. Thus, their true strength appears to be in their ability to process and analyze vast amounts of multimodal data, revealing latent patterns that are not easily discernible to human clinicians [8]. While AI may enhance diagnostic processes and support clinical decision-making, it currently remains an adjunct rather than a replacement for traditional diagnostic methods.

AI UNCOVERS LATENT PATTERNS WITHIN MULTIMODAL DATASETS TO SUPPORT DIAGNOSIS

Currently, the primary methods for diagnosing mental illness predominantly rely on self-reported questionnaires and clinical assessments, as outlined in the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition [9]. However, psychiatric illnesses are inherently heterogeneous and not manifest uniformly across subjects. Thus, patients with the same condition can exhibit widely divergent symptom profiles [10]. Also, the subjective nature of information often introduces limitations, such as inaccuracies and inherent biases. AI algorithms are effective in revealing latent patterns within complex datasets and furthermore own the ability for continual learning and adaptation through exposure to new data. This can contribute to enhancing diagnostic accuracy and efficacy over time, enabling a deeper understanding of neurobiological substrates of psychiatric disorders and refining diagnostic precision [11].

ML algorithms could help identify biomarkers that offer greater insight into pathophysiological processes, revealing variables that

conventional methodologies may overlook. In turn, it leads to more refined taxonomy. For example, supervised deep learning has shown potential to surpass expert clinicians in psychiatric diagnostics through enhanced diagnostic precision utilizing diverse data source amalgamation [12]. Nienke et al. have demonstrated the use of an analysis framework using natural language processing (NLP) and computational modeling to analyze the complex symptomatology and heterogeneity of neuropsychiatric disorders [13]. Other studies have shown that Large Language Models (LLMs)—such as GPT-4, Llama 2, and the more recent Gemini—can significantly outperform physicians in diagnostic accuracy, in some cases achieving nearly double the accuracy [12]. By incorporating diverse data streams that encompass genetic, neuroimaging, behavioral, and laboratory-based data, comprehensive data amalgamation can be used to elucidate dimensional nuances within diagnoses and often overlapping symptomatology across diverse psychiatric disorders [14]. This framework shows new clinical subtypes and reveals patterns of misdiagnosis, thereby enhancing our understanding of these diseases and providing a valuable resource for future research in psychiatry.

CLUSTER ANALYSIS FOR SUBTYPING HETEROGENEITY ENHANCES DIAGNOSTIC PRECISION IN MENTAL HEALTH

Cluster analysis has been utilized extensively in mental health research as a tool to reduce interindividual heterogeneity through the identification of homogeneous subgroups within the population under study. In the field of psychiatric research, various clustering techniques have been employed. This includes center-based partitioning (e.g., K-means and K-medoids), hierarchical clustering, density-based clustering, and model-based clustering (e.g., finite mixture model and latent class analysis) [15]. Notable advancements include the development of a multi-view clustering approach based on a nonparametric Bayesian mixture model for

depression subtyping [16], the integration of fuzzy C-means and Gaussian Mixture Models to identify ambiguous data points in schizophrenia subtyping [17], and the combination of deep autoencoders with clustering ensembles [18]. Resampling techniques, such as cross-validation, are employed to select optimal parameters and evaluate generalization capabilities. Clustering algorithms can be combined with clinical variables to identify subtypes, and the resulting clusters can be evaluated for meaningful associations with clinical outcomes [15]. Additionally, AI-driven clustering methods enhance the identification of antipsychotic side effects by grouping medications based on receptor binding profiles [19]. By refining drug classifications to reflect side-effect profiles, AI promotes more personalized prescribing, particularly for patients sensitive to certain adverse reactions, improving treatment outcomes in psychiatric care. However, because of the unsupervised nature of clustering, the true performance of these methods is often uncertain.

The Dimensional Neuroimaging Endophenotype (DNE) framework, supported by the AI clustering of neuroimaging data, has been shown to effectively capture disease heterogeneity, enabling the identification of distinct subtypes as well as delineation of continuous dimensions of disease progression [20]. These findings have been validated across several psychiatric disorders, including Alzheimer's disease [21, 22], schizophrenia [17, 23–26], major depressive disorder [27–29], and autism spectrum disorder [30–37]. These investigations have consistently identified biologically valued subtypes and dimensions. Furthermore, DNEs have been linked to genetic underpinnings and clinical outcomes, demonstrating its potential value as a biomarker for early detection, treatment selection and outcome. Although confounding factors and the sparseness of longitudinal and multi-omics data integration remain challenging, DNE is proposed as a promising tool to advance precision medicine in psychiatry, warranting further research.

While the outcomes of using multimodal data are encouraging, it is crucial to approach the integration of genomic data into AI models with caution. Predictive models that utilize multimodal genomic data are at risk of "double dipping" and overfitting. "Double dipping" occurs when data is inadvertently reused in either feature selection or model training, potentially biasing performance metrics. If differential expression analysis is conducted immediately after unsupervised clustering without preventing data reuse, it may lead to over-clustering and an exaggerated assessment of the model's efficacy. A recent study revealed that 46% of studies employed feature selection techniques prone to data leakage [38]. The frequency of hyperparameter tuning, the occurrence of data leakage from feature selection, the type of model used, and the modeling of autoimmune disorders were all significantly correlated with higher reported model performance.

There is optimism that computational models, through their adeptness at integrating heterogeneous, multimodal data, will deconvolve the heterogeneity of psychiatric taxonomy. They can generate novel, objective, and biologically based measures, thereby overcoming the limitations of traditional diagnostic methods that rely solely on descriptive symptomatology. The standalone accuracy of AI may be modest, but its potential to support clinical decision-making by uncovering hidden patterns and improving the granularity of diagnostic categories is being increasingly recognized (Fig. 3).

THE CRYSTAL BALL OF AI: PREDICTING MENTAL HEALTH OUTCOMES WITH MULTIMODAL DATA INTEGRATION

Psychiatry diverges from other medical specialties in its emphasis on clinical observations rather than anatomy and physiological based interventions. The benefit of early prediction and intervention for mental illness is now well recognized. As such, clinical management

of mental illness necessitates a pragmatic evaluation of prognosis beyond mere diagnosis [39]. AI models may significantly contribute to this process by condensing and aggregating data, as well as identifying meaningful patterns of significant variations. This capability aids in forecasting clinically pertinent developments in the course of the illness [40]. Furthermore, AI can continuously learn and adapt from new data, refining its predictive models over time. This dynamic learning capability enhances the accuracy and reliability of predictions. In contrast, traditional methods may rely more on historical data and clinician intuition, which may not capture the full spectrum of influential factors with the same granularity and consistency. Therefore, although a human-centric approach and contextual judgment remain the strength of traditional psychiatric practice, AI offers a complementary tool to enhance predictive accuracy and individualized care.

While discriminative capability has been well demonstrated, recent studies indicate its potential to enhance predictive accuracy utilizing amalgamating data with other clinical measurements, including clinician evaluations, genetic profiles, and neuropsychological assessments. Koutsouleris et al. employed various algorithms, including random forest (RF), gradient boosting, and deep neural networks, and achieved high sensitivity (76.0–88.0%) and balanced accuracy (85.5%) in predicting the transition to psychosis [41]. Dwyer et al. utilized nonnegative matrix factorization and multiclass supervised ML techniques, yielding stable cluster solutions and transferable predictions that were validated across subgroups [42]. Furthermore, Basaraba et al. utilized random probability forest models, achieving areas under the curve (AUCs) that ranged from 0.68–0.88 for education/occupational status and trajectories of hospitalization [43]. Similarly, Garcia-Argibay et al. employed deep neural networks, yielding an AUC of 0.75 and balanced accuracy of 0.69 in predicting the onset of attention-deficit/hyperactivity disorder [44]. Zhu et al. used multisite structural magnetic resonance imaging to develop a model to predict psychosis onset, demonstrating promising predictive performance when considering adolescent brain development, age, and sex differences [45].

The ML models for genetic data show considerable variability with altered conditions. Schizophrenia models achieved the highest AUCs (0.54–0.95), with XGBoost with exome sequencing data reaching an AUC of 0.95 [46, 47]. Autism spectrum disorder models showed moderate performance, whereas those that used Copy Number Variation data often performed poorly. Models for anorexia nervosa demonstrated moderate predictive accuracy [47]. Multimodal analyses, such as DeepGAMI, enhance phenotype prediction by prioritizing disease-associated variants and regulatory networks. DeepGAMI outperformed the Polygenic Risk Score for schizophrenia (AUC = 0.61) with dual (AUC = 0.895) and single (AUC = 0.867) models, indicating that its hidden layers effectively capture covariate effects [48].

Longitudinal validation methods, in which a validation sample is drawn from the same population over time, offer a pragmatic approach to avoid overgeneralization from one clinical setting to another. The Accelerating Medicines Partnership Schizophrenia (AMP SCZ) project is a robust and internationally coordinated endeavor that is aimed at developing and validating predictive biomarkers that are pertinent to individuals who exhibit a clinically high risk of psychosis. The AMP SCZ project utilizes comprehensive data collection methods, including behavioral, neurobiological, and genetic approaches, to establish reliable predictive models. These models can delineate the clinical trajectory of psychiatric conditions with greater precision at the individual level, enabling personalized therapeutic strategies. The development of predictive biomarkers has significant implications for clinical practice, allowing more nuanced and individualized treatment strategies based on patients' unique biopsychosocial profiles.

To be clinically actionable and impactful in real-world scenarios, prediction models must be more than just highly predictive. In
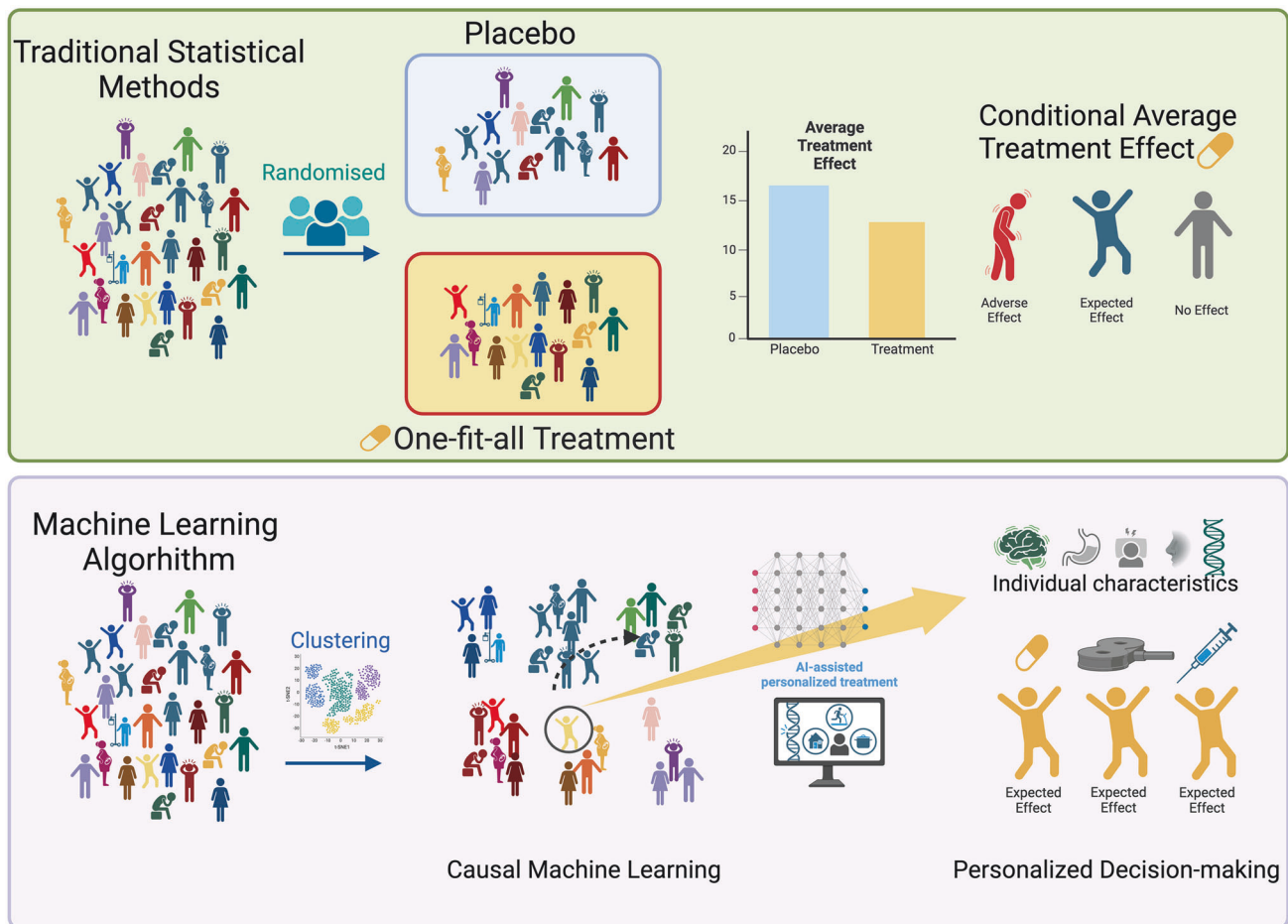
Fig. 3 From traditional statistical methods to machine learning in psychiatry. This figure illustrates the shift from traditional statistical methods to advanced machine learning techniques in identifying different treatment benefits in psychiatry. Traditional statistical approaches typically yield results that reflect average treatment effects across a population, which may not accurately represent actual treatment effects in individual patients with heterogeneous characteristics. These methods, which test variations in treatment effects across individual patient characteristics using statistical significance cutoffs, are susceptible to false discoveries and false-negative results. In contrast, causal machine learning methods provide robust alternatives that are capable of more effectively identifying heterogeneous treatment effects. These methods offer a granular understanding of when treatments are beneficial or harmful, thereby enabling personalized decision-making in patient care that is tailored to individual patient profiles.

psychiatry, given the probabilistic nature of data related to syndromes, outcomes, disorders, and their determinants, interpretability is of utmost importance. Uninterpretable (black-box) methods can lead to unintended, improper use and unfair clinical decisions [49]. Thus, physicians need to be able to understand and debug model predictions, and models should contribute to medical knowledge to avoid misuse [50]. Explainable AI (XAI) holds promise for building trust and facilitating better AI application. Feature-based interpretability methods like LIME, SHAP, and integrated gradients clarify predictions through feature contributions, while example-based methods such as SimplEx do so via similar training-set instances [50, 51]. Model-based explanations employ secondary models, known as meta-models, to transform a black-box model into an alternative format. Counterfactual model provides explanations for model predictions by creating artificial examples that are closely related to the original instance but lead to distinct predictions. Nevertheless, there are limitations, especially in explaining complex models and causal mechanisms. Future efforts should focus on creating more comprehensive and interpretable models, integrating XAI with domain knowledge, and bridging the interpretability gap to promote trust and wider adoption of AI in vital applications.

Application of AI to diverse data sources fosters better diagnostic and predictive performance and has promise of improved

interpretability, thus facilitating a deeper comprehension of the ways in which various sources. Furthermore, it provides insights into how mechanism-driven features may illuminate mechanism-agnostic characteristics, thus potentially enhancing diagnostic precision and measurement efficacy in clinical practice. As the field evolves, AI holds potential in mitigating errors and advancing psychiatric practice toward higher precision and reliability.

AI-POWERED ANALYTICS WILL ENHANCE PRECISION IN TREATMENT SELECTION AND IMPROVE PATIENT MANAGEMENT

A clinician's ability to predict the likelihood of a patient benefiting from a particular treatment is an essential aspect in clinical practice. Psychiatry has developed various therapeutic interventions, but a substantial proportion of patients will not benefit from these treatments [52]. Finding effective interventions is often a "trial and error" process. Computational models hold promises for improving treatment predictions, helping clinicians choose between therapeutic strategies (e.g., psychotherapy *vs.* medication) or specific class/agents among medications (e.g., selective serotonin reuptake inhibitors *vs.* selective norepinephrine reuptake inhibitors).

ML methodologies, particularly those that are grounded in causal inference, facilitate the development of personalized

predictive models for individual patients by examining their unique attributes, such as sociodemographic factors, genetic data, and brain function, among others, that are extracted from a range of unstructured datasets. A proposed application includes subgroups vulnerable to self-harm behavior and helped or harmed by hospitalization [53]. Variability in the response to treatments, such as interventions (e.g. repetitive transcranial stimulation (rTMS) [54] vs. antidepressants [55]) can be more effectively deciphered using predictive information from clinical, sociodemographic, genetic, and investigational data with AI. For example, algorithms were reported to consistently achieve over 90% accuracy in distinguishing between bipolar disorder and control groups and contribute to medication selection (e.g. lithium using genetic data) [55]. The application of ML in predicting adverse effects of psychiatric medications has the potential to significantly enhance precision treatment. By integrating pharmacokinetic data with electronic health records and utilizing machine learning models [56], clinicians can better predict side effects, which in turn allow for more individualized dosing strategies.

Additionally, ML can leverage both experimental and observational data sources, such as randomized controlled trials and electronic health records [57], to generate clinical information. This is particularly valuable in psychiatry, where conducting randomized trials is often challenging because of ethical concerns (e.g. informed consent) and difficulties in blinding. Unlike traditional statistical methods that provide population-level average treatment effects, ML can capture treatment heterogeneity across patient subgroups.

However, several deficiencies in AI models have been reported. For example, Adam et al. analyzed the performance of AI models across multiple independent clinical trials that involved antipsychotic medication for schizophrenia [58]. Models that performed excellently in one sample often failed to generalize to other populations. Moreover, combining data from multiple trials did not improve predictive accuracy for those trials that were not included in the aggregation. Many research investigations of ML-based predictive models have been compromised by poor methodological rigor and significant potential for bias. Many proposed predictive model findings lack independent validation samples. Contributory factors include limited sample sizes and the need to effectively address missing data. It has been proposed that shifting from studies based on linear and static concepts of disease to prospective research that accounts for contextual factors, as well as the dynamic and complex nature of mental health, and employs rigorous methodologies, will significantly advance AI research in this field.

THE UTILITY OF AI-DRIVEN CONTINUOUS DIGITAL MONITORING IN MENTAL HEALTH
Inadequate resources, such as underfunding and understaffing, as well as societal stigma, contribute to suboptimal mental health care, exacerbating unmet needs and delayed interventions. It has been proposed that the adoption of AI can revolutionize mental healthcare by processing born-digital records from social media platforms, mobile applications, and chatbots, alleviating long-standing inequalities and enhancing access to services, especially for marginalized populations [59]. Utilizing a diverse range of digital technologies can enable valid multidimensional measurements of mental health-related metrics and serve to narrow the mental health gap globally. Several potential applications spanning from documentation, communication, and prevention of relapses have been proposed. By leveraging AI to alleviate repetitive and mechanical tasks, clinicians can free up time to focus on the human-centric component of psychiatric practice.

Smartphones and social media platforms are promising avenues for collecting extensive real-world data from individuals. Innovative approaches including geolocation tracking, sleep monitoring, and analyses of smartphone usage patterns that have been utilized with good compliance including longitudinal data collection [60, 61]. AI-driven wearable have significantly improved the reliability and ease of the diagnosis and monitoring of sleep disorders [62, 63]. Furthermore, through continuous data collection from wearable devices and integration with multimodal data inputs, such as neuroimaging and polysomnography, AI can detect subtle patterns of sleep disturbances that traditional methods may miss [64]. This development has led to an improved understanding of the interconnections between forms of various mental illnesses and sleep architectures, enabling a more targeted and precise therapeutic approach to the management of sleep disorders [65–67]. As well, AI facilitates real-time monitoring, allowing for adaptive interventions and more precise predictions of the outcome of conditions in the interface of sleep disorders and mental illness [64, 68, 69].

Psychomotor changes are common symptoms in mental illness. Psychomotor agitation may present as pacing, fidgeting, or shaking, and can be studied using movement data utilizing smartphones. Altered psychomotor activity patterns are often seen in other conditions including unipolar [70] and bipolar [71] depression, as well as dementia [72]. Smartphone accelerometer data has been shown to improve the predictive accuracy of ML models by approximately 5%, outperforming models that are based solely on clinical data and typing dynamics [71]. Integrating device sensor data, such as smartphone geolocation and wearable smartwatch activity patterns, has shown greater accuracy compared to traditional clinical data for modeling chronic health conditions.

With advancements in NLP techniques, AI algorithms can analyze vast textual data from such platforms as X (formerly Twitter) and Facebook, enabling early identification of mental illness and emotional distress. This, in turn, can enhance prevention, including reducing the risk of self-harm. Studies have demonstrated the efficacy of ML algorithms, such as support vector machines, RF, naive Bayes, recurrent neural networks, and attention-based networks, enabling the identification of mental illness and distress, as well as maladaptive behavioral patterns [73]. Furthermore, monitoring early warning signs can lead to quicker initiation of interventions and thus, have the potential to reduce the risk of relapses. The integration of smartphone technologies into clinical practice represents a promising advancement, as supported by the findings from the EMPOWER study, which demonstrated their efficacy in preventing relapses in schizophrenia by combining smartphone technology with clinical triage and peer support [74].

CHATBOTS FACILITATE MENTAL HEALTH SUPPORT BY REDUCING STIGMA AND BRIDGING DISPARITIES
Psychiatry is unique within medical disciplines, being language-centric, with subjective reports, symptoms, and greater emphasis on personalized individualized interventions. A meta-analysis found that mental health smartphone apps yield modest, albeit measurable, improvements in depression and generalized anxiety disorder, claimed to parallel outcomes of conventional first-line interventions, such as psychotherapy and pharmacotherapy [59]. Observations from the integration of chatbots into psychiatric practice suggest that their features make them particularly well-suited for broader implementation in clinical settings. Digital tools provide a way to offer performance-based feedback to healthcare providers, facilitating improvement in high-quality care. The advent of chatbots has been suggested as a significant milestone in the intersection of AI and healthcare, with optimum benefits for mental health care including being a discreet and accessible platform to share mental health concerns, receive continuous support, and engage in personalized tailored interventions [75–77].

Originating from seminal work by Joseph Weizenbaum, who developed ELIZA in the 1960s, chatbot technology has undergone exponential evolution, leveraging advances in computing power and access to vast repositories of textual data [78]. Large Language
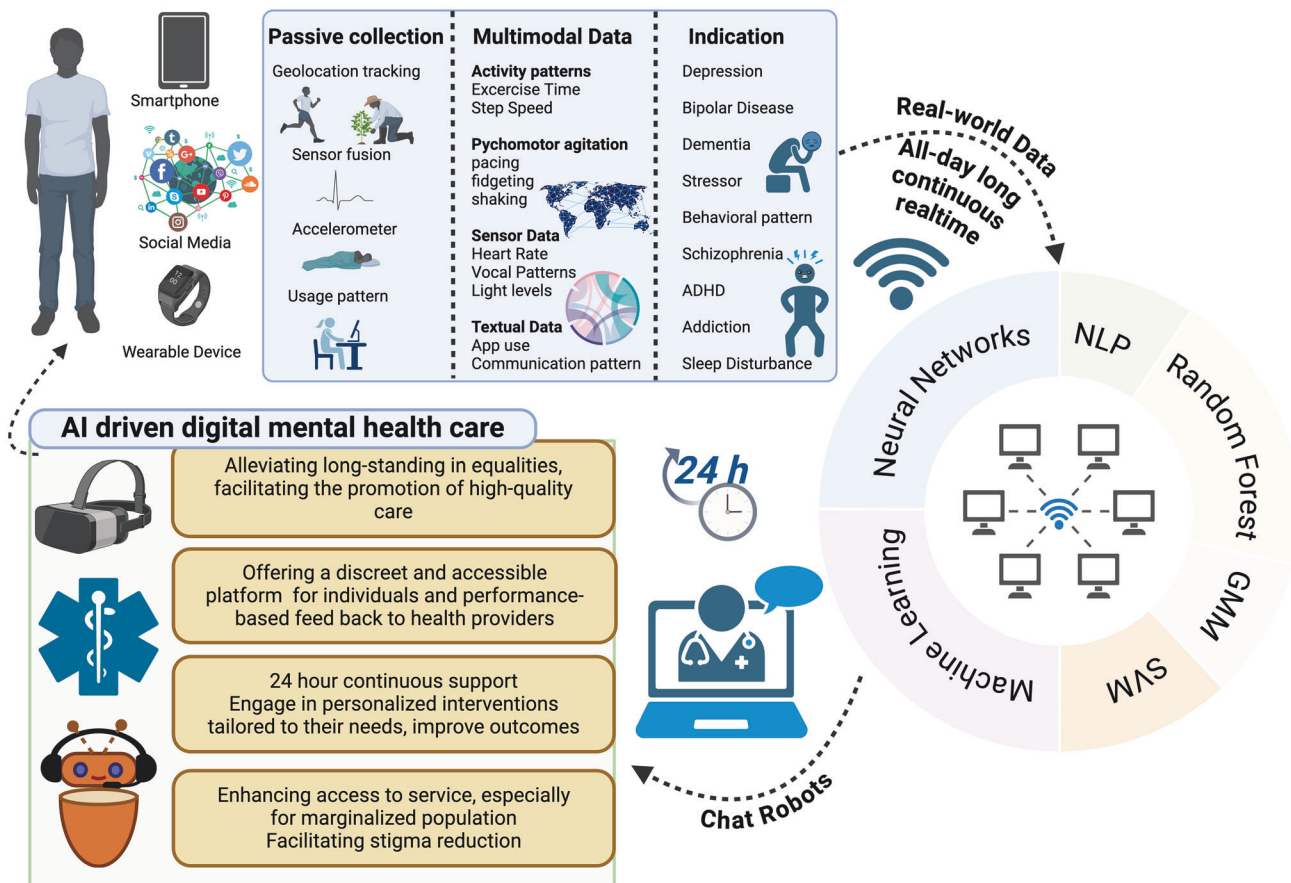
Fig. 4 Artificial intelligence-driven digital tools, leveraging internet-connected devices and machine learning advancements, offer transformative opportunities to promote mental health in the digital age. ADHD attention-deficit/hyperactivity disorder, SVM support vector machine, GMM Gaussian Mixture Model.

Models, such as PaLM and Flan-PaLM, have shown state-of-the-art performance in the MultiMedQA benchmark, achieving high accuracy across multiple medical question-answering datasets [79]. Among these innovations, the introduction of Chat Generative Pre-trained Transformer (ChatGPT) by OpenAI stands out [80], offering unprecedented capabilities in NLP and conversational AI. Personalized self-referral chatbots currently in use in mental health services, such as NHS Talking Therapies in the UK [81], utilize AI-powered NLP algorithms to enable patient self-referral and streamline access to mental health support [76]. These apps are reported to be particularly effective for depression [70], with potential as early treatment or as an alternative for individuals without access to standard care. However, the benefit of the application of such apps to anxiety conditions has less evidence, with existing data constrained by methodological flaws. The inclusion of mood tracking and chatbot features has been observed to increase personalization and user engagement, augmenting therapeutic impact. Reduced stigma has been noted to increase help-seeking, and access to care and reduce treatment disparities. Furthermore, through dynamic adaptation and continuous monitoring, chatbots can enhance patient engagement, foster therapeutic rapport, and contribute to a comprehensive approach to mental health care [82]. (Fig. 4).

DISCUSSION

The integration of AI into psychiatric practice represents a transformative shift with significant potential to enhance diagnostic accuracy, treatment personalization, improved access, and overall quality of care [83]. As AI evolves in clinical practice, it is imperative for clinicians to recognize its benefits and integrate its use judiciously.

This review provides a comprehensive review of the benefits of ML algorithms and provides a synopsis of the AI applications in supporting clinical decision-making processes. An overarching goal of AI in mental health care is to synthesize data into valid diagnostic categories that accurately reflect the underlying pathophysiology of mental illnesses [84]. The individualized ML approach offers substantial improvements in the identification of patterns and correlations that are subtle and as such, are often missed by traditional methods. This review provides practical recommendations for integrating AI into clinical practice, alongside an overview of theoretical advancements and potential applications, and emphasizes the importance of closely aligning these two components.

Foremost among ML's benefits, is its use to craft personalized predictive models that can be tailored to individual patients based on their unique attributes [85]. ML algorithms can yield precise estimations of treatment effects for each patient by assimilating patient-specific multimodal data [86–89]. The scalability and versatility of AI techniques make them valuable tools for addressing complex clinical questions in psychiatry [90]. This individualized approach enhances the efficacy of therapeutic interventions and mitigates adverse outcomes by discerning patient subgroups that may derive benefits or encounter detriments from a particular intervention. Thus, it can serve as a personalized risk assessment tool to provide individualized care.

However, the application of AI in psychiatry has several limitations. One common constraint is the restricted sample sizes that are attributable to the high cost of data collection. As well, the translation of AI models into clinical settings poses a considerable challenge. Clinical populations are heterogenous, and these models are typically not equipped to handle the heightened heterogeneity

and noise that are inherent to real-world clinical samples which may compromise efficacy and generalizability in clinical practice. Finally, AI models have inherited human biases and need specialized fine-tuning for medical diagnoses [84, 91, 92], further highlighting the need for continued adaptation.

The successful deployment of AI in psychiatry requires the careful consideration of several factors. Despite the ability of ML algorithms to enhance predictive accuracy, their performance varies significantly across different contexts and patient populations [93]. The limitation of current AI and ML systems in psychiatric care settings lies in their rigidity post-deployment, which hampers their ability to adapt and localize to specific clinical environments [94]. This fixed model approach not only limits opportunities for adaption and learning but also overlook the invaluable contribution of clinical expertise and patient feedback. To enhance individual-level accuracy and transparency in AI applications for psychiatry, proposed strategies can be employed. These include the adaptation of techniques like conformal prediction [95], integration of diverse data sources, and development of interpretable AI models in collaboration with clinicians [96]. By fostering more dynamic, locally adapted systems, AI can evolve with clinical experience, ensuring that they remain sustainable and relevant in diverse psychiatric care settings. The instruments and conditions for this continuous assessment must be repeated frequently, nearing a state of perpetual monitoring, and the assessments should be conducted within the actual clinical setting where the AI is employed [97]. This approach would not only enhance the precision and effectiveness of AI interventions but also promote a more inclusive and patient-centered approach to psychiatric care, thereby supporting the sustainable integration of AI into clinical practice. Additionally, promoting high ethical considerations and regulatory standards is crucial to guarantee patient safety and privacy [41, 98].

The presentations and course of psychiatric illness is influenced by the interactions between biological, psychological, and social factors. This can pose a significant challenge for AI models that are primarily designed for single-task applications. Current AI systems often lack the expressiveness and interactive capabilities that are necessary to fully capture the multifaceted nature of mental health conditions. Certain adaptations have been proposed to overcome this deficiency, leading to more effective and personalized interventions. These include the integration of AI with complementary methodologies that are grounded in logic and reasoning, alongside values-based practice [99], which have been proposed as tools to bridge this gap. Adept utilization of AI tools can help clinicians in navigating the complex terrain of emotional disorders by synthesizing information on patient-specific values, thereby fostering an adaptive therapeutic alliance and preserving patient autonomy. Values-based practice emphasizes the importance of incorporating patient values and preferences into clinical decision-making. Extending this principle to AI-driven interventions ensures that such technologies are not only empirically robust but also ethically aligned with the foundations of patient-centered care.

The integration of AI methodology into psychiatric practice also necessitates addressing several biases and ethical concerns about misconceptions often expressed about AI. AI algorithms may inadvertently perpetuate biases that are present in training data, leading to inaccuracies being interested as disparities in treatment outcomes across patient groups. Therefore, developing transparent and interpretable AI models is crucial for fostering trust and acceptability.

It has been proposed that "big data" sets from large prospective cohorts, such as the UK Biobank, Human Connectome Project, and China Brain Project, may offer a significant opportunity to address many current limitations of AI integration in mental health [100, 101]. Firstly, the analysis of longitudinal data they provide will enable tracking of disease progression and course of illness, and thus, predict long-term outcomes. Secondly, the integration of multimodal data (e.g. demographics, biological and behavioral

factors, among others) will enable AI to create more comprehensive models that better reflect the complexity of mental illness. Finally, data from diverse global populations will lead to greater generalizability of the models, by minimizing bias and ensuring applicability across patient populations and health systems. The large sample sizes and longitudinal long-term data will enable robust validation and enhance clinical utility.

In practice, application of AI in clinical settings has often faced resource and infrastructure challenges. Clinicians require adequate training and support to effectively utilize AI tools, highlight the need for a workforce equipped with digital skills and an understanding of AI's potential and limitations. The development of interactive visualization tools and incorporation of user feedback can further enhance the usability and clinical utility of AI predictions.

Despite these challenges, the evolving landscape of digital technologies, including the ubiquitous presence of internet-connected devices and advancements in ML, offers unprecedented opportunities for digital psychiatry [102]. Various tools, such as chatbots, virtual reality, and smartphone applications, have demonstrated their utility in expanding access to mental healthcare, particularly in contexts where traditional in-person services are limited. However, ethical practice necessitates a balanced approach that prioritizes patient safety, autonomy, and privacy.

In conclusion, AI holds significant promise for transforming psychiatric practice, but its clinical application needs careful evaluation of scientific validity, cost and benefits, as well as ethical implications. A collaborative and transparent dialogue among researchers, clinicians, and patients will contribute towards harnessing the potential of AI to advance the field of mental healthcare.

REFERENCES

1. Ornell F, Borelli WV, Benzano D, Schuch JB, Moura HF, Sordi AO, et al. The next pandemic: impact of COVID-19 in mental healthcare assistance in a nationwide epidemiological study. Lancet Reg Health Am. 2021;4:100061.
2. The Lancet Global Health. Mental health matters. Lancet Glob Health. 2020;8:e1352.
3. Abbasi J, Hswen Y. One day, AI could mean better mental health for all. JAMA. 2024;331:1691–4.
4. Babic B, Gerke S, Evgeniou T, Cohen IG. Beware explanations from AI in health care. Science. 2021;373:284–6.
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44–56.
6. Nature Medicine Editorial Team. How to support the transition to AI-powered healthcare. Nat Med. 2024;30:609–10.
7. Winter NR, Blanke J, Leenings R, Ernsting J, Fisch L, Sarink K, et al. A systematic evaluation of machine learning–based biomarkers for major depressive disorder. JAMA Psychiatry. 2024;81:386–95.
8. Hauser TU, Skvortsova V, Choudhury MD, Koutsouleris N. The promise of a model-based psychiatry: building computational models of mental ill health. Lancet Digit Health. 2022;4:e816–e828.
9. Haque A, Milstein A, Fei-Fei L. Illuminating the dark spaces of healthcare with ambient intelligence. Nature. 2020;585:193–202.
10. Dhamala E, Yeo BTT, Holmes AJ. One size does not fit all: methodological considerations for brain-based predictive modeling in psychiatry. Biol Psychiatry. 2023;93:717–28.
11. Montag C, Quintana DS. Digital phenotyping in molecular psychiatry—a missed opportunity? Mol Psychiatry. 2023;28:6–9.
12. Topol EJ. Toward the eradication of medical diagnostic errors. Science. 2024;383:eadn9602.
13. Mekkes NJ, Groot M, Hoekstra E, de Boer A, Dagkesamanskaia E, Bouwman S, et al. Identification of clinical disease trajectories in neurodegenerative disorders with natural language processing. Nat Med. 2024;30:1143–53.
14. Chen J, Patil KR, Yeo BTT, Eickhoff SB. Leveraging machine learning for gaining neurobiological and nosological insights in psychiatric research. Biol Psychiatry. 2023;93:18–28.
15. Gao CX, Dwyer D, Zhu Y, Smith CL, Du L, Filia KM, et al. An overview of clustering methods with guidelines for application in mental health research. Psychiatry Res. 2023;327:115265.
16. Tokuda T, Yoshimoto J, Shimizu Y, Okada G, Takamura M, Okamoto Y, et al. Identification of depression subtypes and relevant brain regions using a data-driven approach. Sci Rep. 2018;8:14082.

17. Voineskos AN, Jacobs GR, Ameis SH. Neuroimaging heterogeneity in psychosis: neurobiological underpinnings and opportunities for prognostic and therapeutic innovation. Biol Psychiatry. 2020;88:95–102.

18. Chang M, Womer FY, Gong X, Chen X, Tang L, Feng R, et al. Identifying and validating subtypes within major psychiatric disorders based on frontal–posterior functional imbalance via deep learning. Mol Psychiatry. 2021;26:2991–3002.

19. McCutcheon RA, Harrison PJ, Howes OD, McGuire PK, Taylor DM, Pillinger T. Data-driven taxonomy for antipsychotic medication: a new classification system. Biol Psychiatry. 2023;94:561–8.

20. Wen J, Antoniades M, Yang Z, Hwang G, Skampardoni I, Wang R, et al. Dimensional neuroimaging endophenotypes: neurobiological representations of disease heterogeneity through machine learning. Biol Psychiatry. 2024;96:564–84.

21. Sirkis DW, Bonham LW, Johnson TP, La Joie R, Yokoyama JS. Dissecting the clinical heterogeneity of early-onset Alzheimer's disease. Mol Psychiatry. 2022;27:2674–88.

22. Ashton NJ, Janelidze S, Mattsson-Carlgren N, Binette AP, Strandberg O, Brum WS, et al. Differential roles of Aβ42/40, p-tau231 and p-tau217 for Alzheimer's trial selection and disease monitoring. Nat Med. 2022;28:2555–62.

23. Dwyer DB, Chand GB, Pigoni A, Khuntia A, Wen J, Antoniades M, et al. Psychosis brain subtypes validated in first-episode cohorts and related to illness remission: results from the PHENOM consortium. Mol Psychiatry. 2023;28:2008–17.

24. Chand GB, Singhal P, Dwyer DB, Wen J, Erus G, Doshi J, et al. Schizophrenia imaging signatures and their associations with cognition, psychopathology, and genetics in the general population. Am J Psychiatry. 2022;179:650–60.

25. Habtewold TD, Rodijk LH, Liemburg EJ, Sidorenkov G, Boezen HM, Bruggeman R, et al. A systematic review and narrative synthesis of data-driven studies in schizophrenia symptoms and cognitive deficits. Transl Psychiatry. 2020;10:244.

26. Chand GB, Dwyer DB, Erus G, Sotiras A, Varol E, Srinivasan D, et al. Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. Brain. 2020;143:1027–38.

27. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat Med. 2017;23:28–38.

28. Kim Y-K, Park S-C. An alternative approach to future diagnostic standards for major depressive disorder. Prog Neuropsychopharmacol Biol Psychiatry. 2021;105:110133.

29. Wen J, Fu CHY, Tosun D, Veturi Y, Yang Z, Abdulkadir A, et al. Characterizing heterogeneity in neuroimaging, cognition, clinical symptoms, and genetics among patients with late-life depression. JAMA Psychiatry. 2022;79:464–74.

30. Shan X, Uddin LQ, Ma R, Xu P, Xiao J, Li L, et al. Disentangling the individual-shared and individual-specific subspace of altered brain functional connectivity in autism spectrum disorder. Biol Psychiatry. 2024;95:870–80.

31. Hwang G, Wen J, Sotardi S, Brodkin ES, Chand GB, Dwyer DB, et al. Assessment of neuroanatomical endophenotypes of autism spectrum disorder and association with characteristics of individuals with schizophrenia and the general population. JAMA Psychiatry. 2023;80:498–507.

32. Hwang G. Autism spectrum disorder: time to notice the individuals more than the group. Biol Psychiatry. 2022;92:606–8.

33. Aglinskas A, Hartshorne JK, Anzellotti S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. Science. 2022;376:1070–4.

34. Tang S, Sun N, Floris DL, Zhang X, Di Martino A, Yeo BTT. Reconciling dimensional and categorical models of autism heterogeneity: a brain connectomics and behavioral study. Biol Psychiatry. 2020;87:1071–82.

35. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The heritability of autism spectrum disorder. JAMA. 2017;318:1182–4.

36. Hazlett HC, Gu H, Munsell BC, Kim SH, Styner M, Wolff JJ, et al. Early brain development in infants at high risk for autism spectrum disorder. Nature. 2017;542:348–51.

37. Ecker C, Bookheimer SY, Murphy DGM. Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. Lancet Neurol. 2015;14:1121–34.

38. Barnett EJ, Onete DG, Salekin A, Faraone SV. Genomic machine learning meta-regression: insights on associations of study features with reported model performance. IEEE/ACM Trans Comput Biol Bioinform. 2024;21:169–77.

39. Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW. The science of prognosis in psychiatry: a review. JAMA Psychiatry. 2018;75:1289.

40. Raket LL, Jaskolowski J, Kinon BJ, Brasen JC, Jönsson L, Wehnert A, et al. Dynamic electronic health record detection (DETECT) of individuals at risk of a first episode of psychosis: a case-control development and validation study. Lancet Digit Health. 2020;2:e229–e239.

41. Koutsouleris N, Dwyer DB, Degenhardt F, Maj C, Urquijo-Castro MF, Sanfelici R, et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. JAMA Psychiatry. 2021;78:195–209.

42. Dwyer DB, Buciuman M-O, Ruef A, Kambeitz J, Sen Dong M, Stinson C, et al. Clinical, brain, and multilevel clustering in early psychosis and affective stages. JAMA Psychiatry. 2022;79:677–89.

43. Basaraba CN, Scodes JM, Dambreville R, Radigan M, Dachepally P, Gu G, et al. Prediction tool for individual outcome trajectories across the next year in first-episode psychosis in coordinated specialty care. JAMA Psychiatry. 2023;80:49–56.

44. Garcia-Argibay M, Zhang-James Y, Cortese S, Lichtenstein P, Larsson H, Faraone SV. Predicting childhood and adolescent attention-deficit/hyperactivity disorder onset: a nationwide deep learning approach. Mol Psychiatry. 2023;28:1232–9.

45. Zhu Y, Maikusa N, Radua J, Sämann PG, Fusar-Poli P, Agartz I, et al. Using brain structural neuroimaging measures to predict psychosis onset for individuals at clinical high-risk. Mol Psychiatry. 2024;29:1465–77.

46. Trakadis YJ, Sardaar S, Chen A, Fulginiti V, Krishnan A. Machine learning in schizophrenia genomics, a case-control study using 5090 exomes. Am J Med Genet B Neuropsychiatr Genet. 2019;180:103–12.

47. Bracher-Smith M, Crawford K, Escott-Price V. Machine learning for genetic prediction of psychiatric disorders: a systematic review. Mol Psychiatry. 2021;26:70–9.

48. Chandrashekar PB, Alatkar S, Wang J, Hoffman GE, He C, Jin T, et al. DeepGAMI: deep biologically guided auxiliary learning for multimodal integration and imputation to improve genotype–phenotype prediction. Genome Med. 2023;15:88.

49. Bienefeld N, Boss JM, Lüthy R, Brodbeck D, Azzati J, Blaser M, et al. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. NPJ Digit Med. 2023;6:1–7.

50. Imrie F, Davis R, van der Schaar M. Multiple stakeholders drive diverse interpretability requirements for machine learning in healthcare. Nat Mach Intell. 2023;5:824–9.

51. Kessler RC, Bauer MS, Bishop TM, Bossarte RM, Castro VM, Demler OV, et al. Evaluation of a model to target high-risk psychiatric inpatients for an intensive postdischarge suicide prevention intervention. JAMA Psychiatry. 2023;80:230–40.

52. Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA. The heterogeneity problem: approaches to identify psychiatric subtypes. Trends Cogn Sci. 2019;23:584–601.

53. Ross EL, Bossarte RM, Dobscha SK, Gildea SM, Hwang I, Kennedy CJ, et al. Estimated average treatment effect of psychiatric hospitalization in patients with suicidal behaviors: a precision treatment analysis. JAMA Psychiatry. 2024;81:135–43.

54. Dong MS, Rokicki J, Dwyer D, Papiol S, Streit F, Rietschel M, et al. Multimodal workflows optimally predict response to repetitive transcranial magnetic stimulation in patients with schizophrenia: a multisite machine learning analysis. Transl Psychiatry. 2024;14:1–11.

55. Mizrahi L, Choudhary A, Ofer P, Goldberg G, Milanesi E, Kelsoe JR, et al. Immunoglobulin genes expressed in lymphoblastoid cell lines discern and predict lithium response in bipolar disorder patients. Mol Psychiatry. 2023;28:4280–93.

56. Poweleit EA, Vaughn SE, Desta Z, Dexheimer JW, Strawn JR, Ramsey LB. Machine learning-based prediction of escitalopram and sertraline side effects with pharmacokinetic data in children and adolescents. Clin Pharmacol Ther. 2024;115:860–70.

57. Luo Y, Eran A, Palmer N, Avillach P, Levy-Moonshine A, Szolovits P, et al. A multidimensional precision medicine approach identifies an autism subtype characterized by dyslipidemia. Nat Med. 2020;26:1375–9.

58. Chekroud AM, Hawrilenko M, Loho H, Bondar J, Gueorguieva R, Hasan A, et al. Illusory generalizability of clinical prediction models. Science. 2024;383:164–7.

59. Linardon J, Torous J, Firth J, Cuijpers P, Messer M, Fuller-Tyszkiewicz M. Current evidence on the efficacy of mental health smartphone apps for symptoms of depression and anxiety. A meta-analysis of 176 randomized controlled trials. World Psychiatry. 2024;23:139–49.

60. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med. 2020;3:1–11.

61. Heller AS, Shi TC, Ezie CEC, Reneau TR, Baez LM, Gibbons CJ, et al. Association between real-world experiential diversity and positive affect relates to hippocampal–striatal functional connectivity. Nat Neurosci. 2020;23:800–4.

62. Kaveh R, Schwendeman C, Pu L, Arias AC, Muller R. Wireless ear EEG to monitor drowsiness. Nat Commun. 2024;15:6520.

63. Le-Dong N-N, Martinot J-B, Coumans N, Cuthbert V, Tamisier R, Bailly S, et al. Machine learning–based sleep staging in patients with sleep apnea using a single mandibular movement signal. Am J Respir Crit Care Med. 2021;204:1227–31.

64. An P, Zhao J, Du B, Zhao W, Zhang T, Yuan Z. Amplitude-time dual-view fused EEG temporal feature learning for automatic sleep staging. IEEE Trans Neural Netw Learn Syst. 2024;35:6492–506.

65. Yang Y, Yuan Y, Zhang G, Wang H, Chen Y-C, Liu Y, et al. Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. Nat Med. 2022;28:2207–15.

66. Watson NF, Fernandez CR. Artificial intelligence and sleep: advancing sleep medicine. Sleep Med Rev. 2021;59:101512.

67. Shahin M, Ahmed B, Hamida ST-B, Mulaffer FL, Glos M, Penzel T. Deep learning and insomnia: assisting clinicians with their diagnosis. IEEE J Biomed Health Inform. 2017;21:1546–53.

68. Yue H, Chen Z, Guo W, Sun L, Dai Y, Wang Y, et al. Research and application of deep learning-based sleep staging: data, modeling, validation, and clinical practice. Sleep Med Rev. 2024;74:101897.

69. Levy J, Álvarez D, Del Campo F, Behar JA. Deep learning for obstructive sleep apnea diagnosis based on single channel oximetry. Nat Commun. 2023;14:4881.

70. Balliu B, Douglas C, Seok D, Shenhav L, Wu Y, Chatzopoulou D, et al. Personalized mood prediction from patterns of behavior collected with smartphones. NPJ Digit Med. 2024;7:49.

71. Bennett CC, Ross MK, Baek E, Kim D, Leow AD. Smartphone accelerometer data as a proxy for clinical data in modeling of bipolar disorder symptom trajectory. NPJ Digit Med. 2022;5:181.

72. Bringas S, Salomón S, Duque R, Lage C, Montaña JL. Alzheimer's disease stage identification using deep learning models. J Biomed Inform. 2020;109:103514.

73. Mentis A-FA, Lee D, Roussos P. Applications of artificial intelligence−machine learning for detection of stress: a critical overview. Mol Psychiatry. 2024;29:1882–94.

74. Gumley AI, Bradstreet S, Ainsworth J, Allan S, Alvarez-Jimenez M, Aucott L, et al. The EMPOWER blended digital intervention for relapse prevention in schizophrenia: a feasibility cluster randomised controlled trial in Scotland and Australia. Lancet Psychiatry. 2022;9:477–86.

75. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023;183:589–96.

76. Habicht J, Viswanathan S, Carrington B, Hauser TU, Harper R, Rollwage M. Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot. Nat Med. 2024;30:595–602.

77. Sin J. An AI chatbot for talking therapy referrals. Nat Med. 2024;30:350–1.

78. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med. 2023;388:1233–9.

79. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620:172–80.

80. OpenAI. ChatGPT (Mar 14 version) [Large language model]. 2023. https://chatgpt.com. Accessed 21 October 2024.

81. NHS England» Digitally enabled therapies assessment criteria. https://www.england.nhs.uk/mental-health/adults/nhs-talking-therapies/digital/assessment-criteria/. Accessed 6 June 2024.

82. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry. 2021;20:318–35.

83. Sun J, Dong Q-X, Wang S-W, Zheng Y-B, Liu X-X, Lu T-S, et al. Artificial intelligence in psychiatry research, diagnosis, and therapy. Asian J Psychiatr. 2023;87:103705.

84. Torous J, Blease C. Generative artificial intelligence in mental health care: potential benefits and current challenges. World Psychiatry. 2024;23:1–2.

85. Bone C, Simmonds-Buckley M, Thwaites R, Sandford D, Merzhvynska M, Rubel J, et al. Dynamic prediction of psychological treatment outcomes: development and validation of a prediction model using routinely collected symptom data. Lancet Digit Health. 2021;3:e231–e240.

86. Koch E, Pardinas AF, O'Connell KS, Selvaggi P, Camacho Collados J, Babic A, et al. How real-world data can facilitate the development of precision medicine treatment in psychiatry. Biol Psychiatry. 2024;96:543–51.

87. Desai RJ, Glynn RJ, Solomon SD, Claggett B, Wang SV, Vaduganathan M. Individualized treatment effect prediction with machine learning — salient considerations. NEJM Evid. 2024;3:EVIDoa2300041.

88. Subbiah V. The next generation of evidence-based medicine. Nat Med. 2023;29:49–58.

89. Yarnell CJ, Fralick M. Heterogeneity of treatment effect — an evolution in subgroup analysis. NEJM Evid. 2024;3:EVIDe2400054.

90. Koutsouleris N, Hauser TU, Skvortsova V, De Choudhury M. From promise to practice: towards the realisation of AI-informed mental health care. Lancet Digit Health. 2022;4:e829–e840.

91. Young AL, Oxtoby NP, Garbarino S, Fox NC, Barkhof F, Schott JM, et al. Data-driven modelling of neurodegenerative disease progression: thinking outside the black box. Nat Rev Neurosci. 2024;25:111–30.

92. Adler-Milstein J, Redelmeier DA, Wachter RM. The limits of clinician vigilance as an AI safety bulwark. JAMA. 2024;331:1173–4.

93. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med. 2023;388:1201–8.

94. Youssef A, Pencina M, Thakur A, Zhu T, Clifton D, Shah NH. External validation of AI models in health should be replaced with recurring local validation. Nat Med. 2023;29:2686–7.

95. Banerji CRS, Chakraborti T, Harbron C, MacArthur BD. Clinical AI tools must convey predictive uncertainty for each individual patient. Nat Med. 2023;29:2996–8.

96. Lu T, Liu X, Sun J, Bao Y, Schuller BW, Han Y, et al. Bridging the gap between artificial intelligence and mental health. Sci Bull. 2023;68:1606–10.

97. Warraich HJ, Tazbaz T, Califf RM. FDA perspective on the regulation of artificial intelligence in health care and biomedicine. JAMA. 2025;333:241–7.

98. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med. 2022;28:1773–84.

99. Womersley K, Fulford KBill, Peile E, Koralus P, Handa A. Hearing the patient's voice in AI-enhanced healthcare. BMJ. 2023;383:p2758.

100. Allen NE, Lacey B, Lawlor DA, Pell JP, Gallacher J, Smeeth L, et al. Prospective study design and data analysis in UK biobank. Sci Transl Med. 2024;16:eadf4428.

101. Liu X, Gao T, Lu T, Bao Y, Schumann G, Lu L. China brain project: from bench to bedside. Sci Bull. 2023;68:444–7.

102. Health TLD. Mental health in the digital age. Lancet Digit Health. 2022;4:e765.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS
JS and TL performed the literature search, wrote the initial draft of the manuscript, and edited the manuscript. XS, YF, and YX performed the literature search and study selection with support from YZ, YXW, XML, and AR provided critical insights into the interpretation of the literature and assisted with figure interpretation. LF, XZ, YH, NR, and YMW contributed to the critical revision of the manuscript for important intellectual content. LL, XL, and YMW, as corresponding authors, oversaw the entire review process, ensuring its accuracy and completeness, and were responsible for the final submission. All authors critically reviewed the manuscript, and approved the final submitted version.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Correspondence and requests for materials should be addressed to Yumei Wang, Xiaoxing Liu or Lin Lu.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.