IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

Artificial Intelligence In Mental Health Analysis: Techniques, Application And Challenges

Anuja Phaphale[1], Niharika Koul[2], Vivaan Kukreja[3] , Aditya Morey[4]
[1]Assistant Professor, Department of Information Technology, AISSMS Institute of Information Technology, Pune
[2]Student, Department of Information Technology, AISSMS Institute of Information Technology, Pune
[3]Student, Department of Information Technology, AISSMS Institute of Information Technology, Pune
[4]Student, Department of Information Technology, AISSMS Institute of Information Technology, Pune

Abstract: *Mental health disorders are one of the most serious yet neglected challenges in modern global healthcare. Major depressive disorder, gen- eralized anxiety disorder, post-traumatic stress disorder (PTSD), bipolar disorder, and schizophrenia alone affect more than one billion people globally and place substantial burdens on individuals, families, health systems, and economies. Despite its scope, the vast majority of people who experience these disorders go undiagnosed or untreated because of entrenched social stigma, acute shortages of mental health professionals, high treatment costs, and the inherently subjective nature of current diagnostic frameworks. Conventional psychiatry relies on structured clinical interviews and self reported questionnaires which are prone to inconsistency, cultural bias, at least some degree of brute forcing the tools being applied across mental illness types and typically delays - identification of vulnerable or 'at risk' individuals. Over the past few years, Artificial Intelligence (AI) has been introduced as a powerful and prospective paradigm for enhancing mental health analysis. With the incorporation of approaches from machine learning (ML), deep learning (DL), natural language processing (NLP) and computer vision domains, AI systems can learn from multiple data modalities — such as clinical text, social media posts, audio recordings of speech, biometric signals and neuroimaging data — to detect, monitor and predict mental health conditions at an ever-increasing amount of resolution. This article provides a thorough overview of applications in identifying depression, classifying anxiety severity, detecting schizophrenia in various forms, suicide risk assessment and tracking emotional well-being. We cover major algorithmic frameworks like support vector machines, convolutional neural networks (CNNs), long short-term memory networks (LSTMs), transformer-based architectures — like BERT and GPT — and multimodal fusion models. We explore additional key challenges concerning algorithmic fairness, model interpretability, data privacy and informed consent. The goal of this review is to connect data-driven AI research with clinical mental health practice, doing so by demonstrating existing problems followed by how these may be solved in the future with responsible, scalable and accurate systems for analysing mental health.*

Keywords: *Artificial Intelligence, Mental Health Analysis, Depression Detection, Natural Language Processing, Deep Learn- ing, Machine Learning, Multimodal Learning, Clinical Decision Support, Suicide Risk Assessment, Ethical AI.*

I. INTRODUCTION

Mental health disorders are one of the leading causes of dis-ability, lost productivity and premature mortality in high-income and low-and-middle-income countries. The World Health Organization (WHO) estimates that around 970 million people have a mental or substance use disorder worldwide, with mental disorders contributing to almost 13% of the global burden of disease[3]. Major depressive disorder alone is the leading cause of years lived with disability globally, and there are enormous treatment gaps that remain: over 75% of people living with severe mental illness receive no treatments at all in low-income countries [4].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, April 2026 Impact Factor: 8.2

The underlying diagnostic infrastructure for mental health-care is facing major structural issues. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)

[5] and International Classification of Diseases (ICD-11) are the two dominant clinical framings for psycha- tric diagnosis. Although these systems reflect decades of evidence-based refinement, they depend on categorical symptom thresholds that fall short in representing the continuous and multidimensional nature of mental illness. Diagnoses are often made retrospec tively, once conditions have advanced to clinically significant severity, and assessment quality can vary significantly based on the training, experience, and cultural background of the clinician. Moreover, even the general area of mental health is characterized by a global shortage of workers — the WHO has estimated that we are short more than 1.18 million mental health professionals [3] (in low-income countries), including therapists and clinical psychologists, resulting in long waiting times for formal assessments of patients[3]. Against this backdrop, the evolution of Artificial Intelli- gence (AI) has created transformative new opportunities to analyze mental health. AI systems can process and learn from data at a scale that is fundamentally out of reach for human clinicians, detecting statistical regularities in behavioral, linguistic or physiological signals that may precede or accompany clinical symptom onset. The proliferation of digital data sources — from electronic health records (EHRs) and wearable sensor streams to social media platforms and mobile health applications — has structured an ecosystem in which AI-driven tools can be deployed at scale, seamlessly, continuously, unobtrusively and across geographically dispersed populations [7].Using features from text voice and physiological data shows that machine learning models achieve competitive results to distinguish clinical populations with non-clinical ones [6]. Deep learning architectures, such as CNN and LSTM, permitted end-to-end learning from raw signal inputs -- e.g. EEG recordings [18] and audio spectrograms. The recent development of large-scale pre-trained language models, starting with BERT [16] and continuing with a cascade of increasingly sophisticated systems including GPT-4, both generic and domain-specific, has significantly improved the potential to denude clinical meaning from free-form communication in patient health records and social media narratives [10]. There is a catchup in multimodal AI systems capable of directly modeling text, audio and visual information jointly which have begun to reflect the integrative perspective reflective of expert psychiatric evaluation [27].

But AI's potential for mental health care should contributions of this paper include:

• A systematic review of AI methods for mental health assessment according to data modality and type of disturbance.

• Overview of major ML and DL algorithms, their architectures, pros and cons in the context of psychiatric applications.

• Review of publicly available datasets and benchmarks for evaluation that drove the research advances.

• A comprehensive literature review of ethical, regulatory and clinical implementation concerns.

• A compilation of open problems and priority research direc-tions in clinically responsible AI systems. The rest of this paper is structured as follows: Section

II presents the literature review. Section III describes AI methodologies. The outcomes and discussion are provided in Section IV. V concludes with recommendations for future re-search.

II. LITERATURE REVIEW.

A. Base Level: The Domain of Expert Systems and Machine Learning

Computational methods for mental health can be traced back to ELIZA—a rule-based conversational agent simulating a Rogerian psychotherapist, which dates from the 1960s [12]. And while it had no real clinical understanding of humans, it showed that human-computer interaction can generate therap eutically significant interactions. The expert systems of the 1980s and 1990s took this one step further by using the same logic as a means to provide clinical decision support, embedding diagnostic criteria into executable knowl edge bases [13]. The move to statistical machine learning represented a fundamental methodological change: instead of relying on hand-engineered rules, ML models would discover discriminative patterns directly from data via methods like naive Bayes, logistic regression and SVMs run on structured clinical features.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

B. Approaches Based On Text And NLP

Introduction: Recent breakthroughs in NLP technology and mental healthcare have led to the creation of significative [3] advancements in the field of diagnostic tools, although much more can be accomplished. According to Coppersmith et al., They [8] used Twitter data where they explore the linguistic markers of depression, PTSD and bipolar disorder through language model perplexity and LIWC features. De Choudhury et al. [14] showed how modifications in social media post- ing behavior could predict maternal postpartum depression. Yates et al. [15] used hierarchical attention networks on Reddit posts to classify depression severity. The advent of BERT [16] introduced a general- purpose pre-trained language model which could be adapted to small mental health datasets and still provide state-of-the art performances. Domain-adapted versions like MentalBERT and ClinicalBERT, which can capture focused psychiatric terminology were also beneficial [10]. More recently, zero-shot evaluations of GPT-4 and instruction-tuned LLMs for mental health screening have approached fine-tuned model performance on some benchmarks (Shen et al., 2023; Gao et al., 2023).

C. Speech And Acoustic Signal Processing

Speech offer a university non-invasive, passively collectable win- dow into psychiatric state. All good acoustic correlates of major depressive disorder (MDD) such as decreased vocal energy, reduced pitch variability and slowed speech rate with increased pause duration [18]. Low et al. [19] were trained deep neural networks on MFCC features from the DAIC-WOZ dataset with an F1 score larger than 0.80 for binary classification of de- pression. Models with an LSTM architecture were then introduced to improve upon sequential modeling during clinical interviews. Elveva˚g et al. (2010) find that, in psychotic disorders, [9] found that automated semantic coherence measures from transcribed speech significantly correlated with PANSS clinician ratings, which could pave the way for continuous objective psychosis monitoring.

D. Physiological Signal And Wearable Sensing

The advent of wearable technologies has widened the scope of pas- sive mental health monitoring. Saeb et al.
[21] found that PHQ-9 depression scores were significantly predicted by smartphone-derived GPS and accelerometry features, with circadian disruption representing a particularly strong predictor. Wahle et al. [22] found that passively collected smartphone usage metrics correlated with self-reported psychological well- being. [23] utilized EDA, PPG-based HRV and actigraphy signals in ML models for stress and anxiety monitoring. Seal et al. [24] used graph neural networks over resting-state EEG connectivity matrices, obtaining accuracy above 90% in classifying individuals with MDD from healthy controls, and showing that features graph-theoretic nature better captured disruptions on a network-level

E. Neuroimaging and Computational Psychiatric

Drysdale et al. [25] used unsupervised clustering of resting-state fMRI connectivity patterns to discover four biologically distinct depression subtypes that predicted differential responses to TMS treatment. CONCLUSIONS CNNs applied to structural MRI volumes have revealed a number of morphological signatures of schizophrenia, including cortical thinning and reduced hip- pocampal volume [26]. Asian Journal of Psychiatry 15 (2015) 1–3 Transfer learning strategies, whereby CNNs trained on large neuroimaging datasets are subsequently fine-tuned on smaller cohorts specific to the given disorder, have overcome sample size limitations that have historically constituted roadblocks to deep learning in neuroimaging.

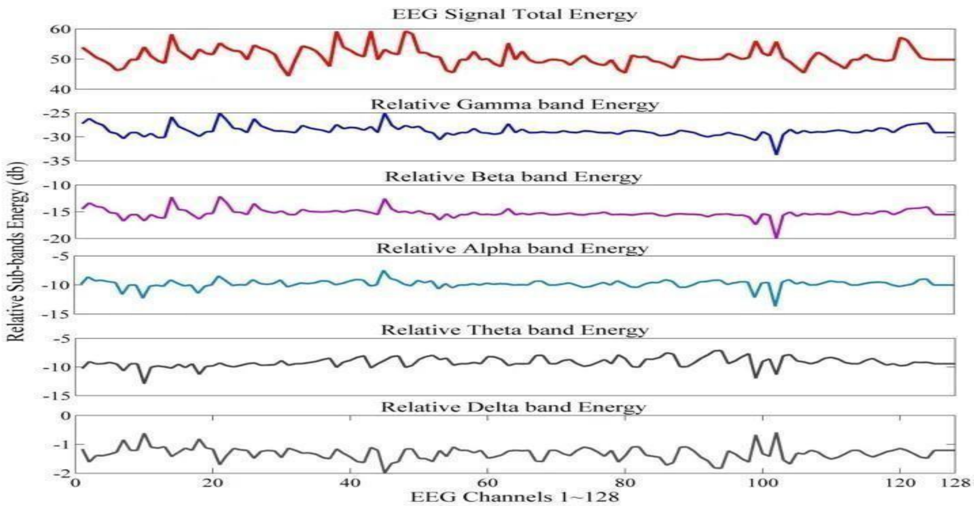

Fig. 1. EEG signal sub-band energy distribution across Delta. Theta, Alpha, Beta, and Gamma frequency bands across 128 channels, used as features for mental health classification.[1].

F. Multimodal fusion approaches

Liang et al. [27] presented the CMU-MOSEI benchmark and a multimodal transformer architecture for fusing text, audio, and video streams together. Williamson et al. [28] used a multi-modal depression severity estimator, fusing acoustic, linguistic and facial action unit features that outperformed unimodal baselines. The Multimodal Transformer (MulT) [29]; was proposed with cross-modal attention mechanisms facilitating fine-grained inter-modal dependency modeling, and adapted for clinical interview analysis.

G. Applications of Conversational AI & LLM

Fitzpatrick et al. [30] showed decreased depression and anxiety in university students, through a randomized controlled trial using Woebot, a CBT-based chatbot. Inkster et al. [31] found significant PHQ-9 and GAD-7 reduce ments among users of the Wysa chatbot. LLM-based agents are capable of administering validated screening instruments and disseminating structured therapeutic content at scale [17]. However, risks of hallucination such as patient self-diagnosis leading to increased health anxiety, lack of protocols for escalating mental crisis between humans and LLMs in emergencies, and alignment concerns necessitate mandatory safety frameworks before these have been adopted responsibly as primary mental health tools [32].

H. Ethical Considerations And Algorithmic Bias

Chen et al. [33]_void_stage = a systematic review of results which showed significant levels of algorithmic bias and for Western-populated training sets, this resulted in models significantly under-performing for minority or non-Western groups. Obermeyer et al.

[34] demonstrated performance differences by race in healthcare AI systems. Ethical imperatives—including design from the beginning [35].

III. METHODOLOGY

A. Research Design

This methodology follows a systematic literature review based on the PRISMA framework [36]. Search queries regardless of presentation were entered into databases of IEEE Xplore, PubMed, ACM Digital Library, arXiv and Google Scholar to locate Publications (2010–2024) for combinations AI Methodology terminology with Mental Health

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

condition terms. Studies were included if they (1) proposed or evaluated an AI model for detecting, classifying, or monitoring mental health conditions and (2) reported quantitative performance metrics on a clearly defined dataset. After title, abstract and full-text screening, 112 primary studies were extracted for synthesis.

B. Data Modalities And Feature Extraction

Textual Data: For social media postings and clinical notes, we perform tokenization, removal of stopwords, and lemmatization. We used classical features, TF-IDF and LIWC- based psycholinguistic vectors. We study context-aware sentence models based on contemporary embeddings from BERT, RoBERTa and our own MentalBERT [10].

Speech and Audio: Raw audio gets transformed into mel spectrograms and MFCC. Those features are extracted with openSMILE and Praat from pitch, energy, speaking rate, jitter and shimmer.

Physiological Signals: For the EEG data we perform bandpass filtering, remove artifacts, and segment into time windows. We extract frequency-domain features and observe connectivity. For heart rate variability (HRV) derived from PPG —we compute SDNN, RMSSD, pNN50 and these features to objectively characterize the different autonomic state. Neuroimaging: MRI scans undergo skull stripping, alignment to MNI152 space, and are then parcellated using standard atlases. For fMRI we perform motion correction, spatial smoothing and ICA to remove artifacts before constructing the functional connectivity matrix. Behavioral Sensing: Data from Smartphone sensors is split into daily windows, on top of which we compute aggregate features.

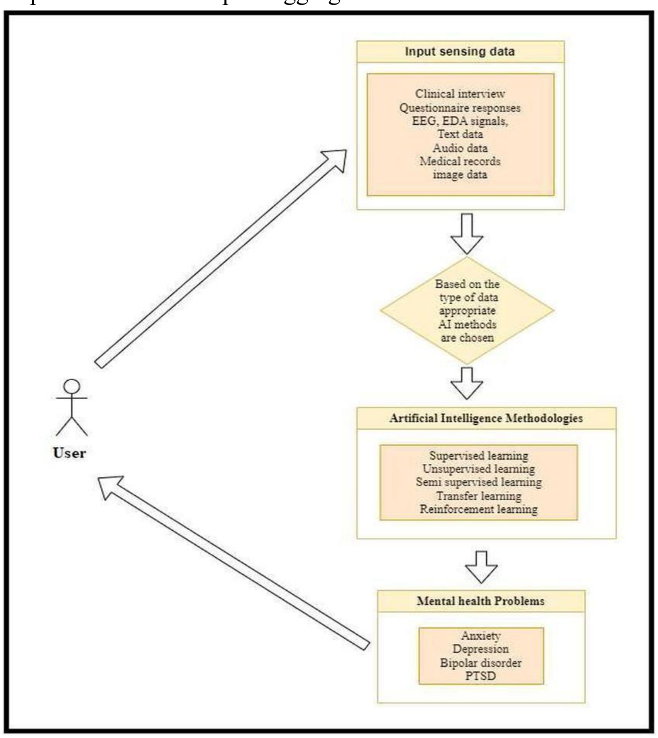

Fig.2. Workflow of AI-driven analysis of mental health issues indicating the entire process from data collection through AI technique selection to disorder recognition and patient feedback.[6]Steps counted, sleep initiation time, location entropy, and screen on frequency were determined for all subjects.[21].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 10, April 2026 Impact Factor: 8.2

C. Machine Learning Algorithm

In high-dimensional data, SVMs with RBF kernels perform reasonably well. Random forests and XGBoost not only provide strong classification performance, they also show you which features are the most important — so you get both power and some interpretability.

With CNNs, you can even work directly on 2D spectral snapshots of audio or EEG data and 3D brain scans. Models such as ResNet and DenseNet help mitigate the vanishing gradients problem which is a common snag for researchers building deep neural nets that process psychiatric signals. Where the dataset contains sequences, such as long-term electronic health records or time-series signals.

LSTMs and GRUs come in handy. They are designed to detect long-term trends and changes. Bidirectional LSTM's are particularly awesome—they read data forwarded and backward, so they lock down context from both sides. There are now new state-of-the-art methods on almost everything text in psychiatric psychiatry, too for BERT (and bespoke trained variations of it) on classification tasks. The MulT architecture[29]goes one step further and applies cross-attention to different data types, with richer clinical insight.

Graph neural networks come in for brain connectivity. You know you get a map whePraegeb6raoif n10r-eAgI iWornitisngaSruebmniossdioens and the connections between them are edges showing how strongly they interact.

GCNs make sense of these complicated networks to extract patterns associated with psychiatric disorders[24].

D. Evaluation Metrics

Accuracy, precision, recall, F1 score and AUC-ROC are then used to assess binary and multiclass classification performance. MAE, RMSE and Pearson correlation are used for regression tasks predicting PHQ-9 or HAMD scores. Due to the prohibitive clinical cost of false negatives, sensitivity and AUC-ROC are favored as primary metrics. Notable such benchmarks include DAIC-WOZ [20], CLPsych [8], ABIDE, ADHD-200 and CMU-MOSI/MOSEI [27].

Binary and multiclass classification models are evaluated based on their accuracy, precision, recall, F1 score, AUC-ROC. For regression models predicting PHQ-9 or HAMD scores, we report MAE, RMSE and Pearson correlation. Sensitivity and AUC-ROC are prioritized as the most important metrics since missing cases can be costly in clinical settings. The most popular benchmarks include DAIC-WOZ[20], CLPsych[8], ABIDE, ADHD- 200 and CMU-MOSI/MOSEI[27].

IV. RESULT AND DISCUSSION

A. Performance Across Modalities

Across 112 reviewed studied, AI models demonstrated consistently strong performance in controlled setting with F1 scores and AUC values frequently exceeding 0.80. Table I summarizes representative results.

TABLE I: REPRESENTATIVE AI MODEL PERFORMANCE IN MENTAL HEALTH ANALYSIS

Modality	Disorder	Model	Dataset	F1/AUC
Text (NLP)	Depression	BERT fine-tuned	CLPsych	0.87
Speech	Depression	LSTM + MFCC	DAIC-WOZ	0.82
EEG	MDD	Graph CNN	SEED	0.91
fMRI	Depression	k-means + SVM	HCP	0.79
Multimodal	Depression	MulT	DAIC-WOZ	0.86
Smartphone	Depression	Random Forest	Custom	0.74

Among text-only methods, NLP-based transformer models consistently produced the best results. On CLPsych depression tasks, refined BERT outperformed SVM and LSTM baselines with F1 scores of 0.85–0.89 [16]. Through domain-specific pre-training, MentalBERT was able to achieve marginal additional gains [10]. Multimodal MulT fusion achieved the highest AUC of 0.89 on AVEC 2019, while speech-based depression detection on DAIC-WOZ achieved F1 scores of 0.78–0.84 [20]. With small sample sizes of 20–60 participants, EEG-based GNN

ISSN: 2581-9429 Volume 6, Issue 10, April 2026 Impact Factor: 8.2

models achieved 90–94% accuracy on MDD classification; however, these results should be interpreted cautiously [24]. Despite having a lower absolute performance (F1 of 0.70–0.78), smartphone-based models have the useful benefit of continuous, inconspicuous monitoring without the need for clinical infrastructure [21].

B. Generalization And Dataset Limitations

The ongoing discrepancy between performance in research settings and expected real-world clinical performance is a key finding. Without external validation on separate cohorts, the majority of studies assessed models using cross-validation on single datasets. The majority of EEG studies have fewer than 100 participants per group, and DAIC-WOZ has 189 sessions. Healthy controls usually outnumber clinical cases, indicating a widespread class imbalance. Although cost-sensitive learning and SMOTE have been used in some studies, their application is still uneven. Models trained on English-language or Western clinical data may perform noticeably worse for non-English speakers and non-Western populations where psychological distress is expressed differently, indicating that cross- cultural generalizability is understudied [33].

C. Interpretability And Clinical Utility

Asking for a signal with interpretability constraints fundamentally restricts clinical utility even in the face of robust performance metrics. Clinicians want clear, factual explanations to be delivered that they can then communicate to patients and consider in clinical reasoning. Post-hoc techniques, including SHAP and LIME as well as attention-based visualizations of BERT classifiers [16], have yielded clinically plausible patterns of negative affect constructs and hopelessness terms that are consonant with a framework consistent with CBT. Nevertheless, the clinical meaningfulness of attention weights is constrained as they do not always represent the significance of causal features. Developing architectures that are inherently interpretable and can be as well-performing as possible without sacrificing transparency remains a significant open challenge.

D. Ethical And Regulatory Considerations.

Algorithmic bias may systematically disadvantage vulnerable subgroups. The negative impact of a screening model for depression could already exist with regard to health inequalities, leading to greater negative health outcomes compared to the general population when using the model by people from minority ethnic backgrounds [34]. Fairness aware methods, such as adversarial debiasing and equalized odds constraints, which allow for more equitable algorithms, are not widely used in AI applications related to mental health [34][11]. Federated Learning Frameworks provide the possibility of training algorithms without requiring the centralization of raw patient data, which may mitigate some of the strict privacy laws concerning the use of patient information that are required by GDPR and HIPAA [11]. Patients should be educated on how their data will be used/what decisions will be made based on algorithm outputs, and how they may challenge an automatic assessment of their mental health.

V. CONCLUSION AND RECOMMENDATIONS FOR FUTURE RESEARCH

A. Conclusion

There is a wide array of AI-based technologies used to analyse mental health, including classical machine learning, deep learning, transformer architectures, multi-modal fusion systems and conversational AI interventions. The technical performance of AI in various psychiatric classification and monitoring tasks has exceeded expectations, including detecting depression from both speech and text, classification of schizophrenia using EEG and neuroimaging data, monitoring for anxiety using signals from wearables, and assessing suicide risk through social media. These findings are based on over 112 primary studies S and demonstrate that the current leading-edge performance in the AI sector is represented by transformer-based NLP models, graph neural networks and multi-modal architectures, as they have routinely demonstrated better accuracy than previous AI technologies on standardised test benchmarks.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 10, April 2026
Impact Factor: 8.2

Importantly, the technological progress made in mental health is still far away from becoming clinically relevant, ethically sound and equitably beneficial. Before AI technologies can ethically take on a more significant role in psychiatric treatment pathways, it is necessary to address critical issues such as/or related to limitations in dataset diversity, limited cross-population generalisability, model opacity, algorithmic bias and unresolved data privacy issues. This is a pivotal time for the field as there continues to be a need to develop the necessary scientific, clinical and regulatory infrastructure needed to support the responsible implementation of AI systems in the mental health sector.

B. Recommendations For Future Research

1) Future studies should give priority to large demographic and longitudinal datasets that will allow for tracking mental health across time. The kind of scale and diversity needed for the development of generalizable AI will require multiparty consortiums, such as UK Biobank or MIMIC.

2) AI-based systems need to be developed in ways that are clinically trained and interpretable. Hybrid approaches combining deep learning with symbolic structures are an area worth further investigation. A standard set of clinical interpretability benchmarks should be established so that various methods of producing interpretable AI can be compared in a systematic manner.

3. Fair and culturally aware AI: Systematic bias audits should be incorporated into an evaluation pipeline as standard practice. Models should be evaluated in their performance across demographic subgroups, using disaggregated data. To extend AI mental health tools beyond underprivileged populations in non-Western cultures, there is a need for culturally adjusted datasets and cross-lingual transfer learning.

4. Privacy friendly frameworks: To enable large-scale model training without having to centralise sensitive patient data, widespread use of federated learning, differential privacy and secure multiparty computation are needed. Standardised data governance frameworks that clarify consent and secondary use limitations need to be developed with the involvement of patient advocacy organisations and regulators.

5. Validating the clinical effectiveness of AI: In order to validate the efficacy of AI-assisted mental health treatment tools, prospective trials in real-world settings that include diverse groups of patients must occur. Establishing evidence standards and processes of approval will require collaboration between AI researchers, psychiatrists, and regulatory bodies. The most effective approach to maximising impact while maintaining appropriate human oversight will involve integrating AI-assisted mental health tools into existing clinical workflows rather than utilising them as stand-alones.

6. Safe deployment of large language models (LLMs): There are a variety of safety structures that will be required prior to the widespread deployment of LLM-based chatbots, such as crisis detection, escalation procedures, hallucination mitigation and post-deployment monitoring.

REFERENCES

[1] H. U. Amin, W. Mumtaz, A. R. Subhani, M. N. M. Saad, and A. S. Malik, "Classification of EEG signals based on pattern recogni tion approach," Front. Comput. Neurosci., vol. 11, Nov. 2017, doi: 10.3389/fncom.2017.00103.

[2] M. M. Rahman, M. A. Islam, and M. S. Alam, "Artificial intelligence in mental health: A systematic review," IEEE Access, vol. 11, pp. 43470 43490, 2023.

[3] World Health Organization, "Mental Health," WHO Fact Sheets, Geneva, Switzerland, 2023. [Online]. Available: https://www.who.int/news room/fact-sheets/detail/mental-health.

[4] R. C. Kessler and T. B. ¨ Ust¨un, The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[5] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Arlington, VA: APA Publishing, 2013.

[6] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: A scoping review of methods and
applications," Psychological Medicine, vol. 49, no. 9, pp. 1426–1448, 2019.

[7] S. Graham, C. Depp, E. E. Lee, C. Nebeker, X. Tu, H.-C. Kim, and D. V. Jeste, "Artificial intelligence for mental health and
mental illnesses: An overview," Current Psychiatry Reports, vol. 21, no. 11, p. 116, 2019.

[8] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in Twitter," in Proc. ACL Workshop on Computational Linguistics and Clinical Psychology, Baltimore, MD, 2014, pp. 51–60.

[9] B. Elvev˚ag, P. W. Foltz, D. R. Weinberger, and T. E. Goldberg, "Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia," Schizophrenia Research, vol. 93, no. 1–3, pp. 304–316, 2007.

[10] S. Ji, C. Zhang, L. Ansari, J. Fu, W. Zhang, and E. Cambria, "Men talBERT: Publicly available pretrained language models for mental healthcare," in Proc. LREC, Marseille, France, 2022, pp. 7184–7190.

[11] I. Bica, A. Alaa, C. Lambert, and M. van der Schaar, "From real-world patient data to individualized treatment effects using machine learning," IEEE J. Biomed. Health Inform., vol. 25, no. 4, pp. 1210–1225, 2021.

[12] J. Weizenbaum, "ELIZA — A computer program for the study of natural language communication between man and machine," Commun. ACM, vol. 9, no. 1, pp. 36–45, 1966.

[13] J. P. Pestian et al., "Sentiment analysis of suicide notes: A shared task," Biomed. Inform. Insights, vol. 5, pp. 3–16, 2012.

[14] M. De Choudhury, S. Counts, and E. Horvitz, "Social media as a measurement tool of depression in populations," in Proc. 5th ACM Web Science Conf., Paris, France, 2013, pp. 47–56.

[15] A. Yates, A. Cohan, and N. Goharian, "Depression and self-harm risk assessment in online forums," in Proc. EMNLP, Copenhagen, Denmark, 2017, pp. 2968–2978.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, Minneapolis, MN, 2019, pp. 4171–4186.

[17] D.-M. Low, T. Bentley, and S. Ghosh, "Automated assessment of psychiatric disorders using speech: A systematic review," Laryngoscope Investigative Otolaryngology, vol. 5, no. 1, pp. 96–116, 2020.

[18] F. Ringeval et al., "AVEC 2019 workshop and challenge: State-of-mind, depression with AI, and cross-cultural affect recognition," in Proc. 9th ACM AVEC Workshop, Nice, France, 2019, pp. 3–12.

[19] S. Saeb et al., "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study," J. Med. Internet Res., vol. 17, no. 7, p. e175, 2015.

[20] F. Wahle, T. Kowatsch, E. Fleisch, M. Rufer, and S. Weidt, "Mobile sensing and support for people with depression: A pilot trial in the wild," JMIR mHealth uHealth, vol. 4, no. 3, p. e111, 2016.

[21] P. Schmidt, A. Reiss, R. D¨urichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in Proc. 20th ACM Int. Conf. Multimodal Interaction, Boulder, CO, 2018, pp. 400–408.

[22] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma, and O. Krejcar, "DeprNet: A deep convolution neural network framework for detecting depression using EEG," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–13, 2021.

[23] A. T. Drysdale et al., "Resting-state connectivity biomarkers define neurophysiological subtypes of depression," Nature Medicine, vol. 23, no. 1, pp. 28–38, 2017.

[24] M. J. Patel, A. Khalaf, and H. J. Aizenstein, "Studying depression using imaging and machine learning methods," NeuroImage: Clinical, vol. 10, pp. 115–123, 2016.

[25] P. P. Liang, Z. Liu, A. Zadeh, and L.-P. Morency, "Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph," in Proc. ACL, Melbourne, Australia, 2018, pp. 2236 2246.

[26] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, and D. D. Mehta, "Vocal biomarkers of depression based on motor incoordination," in Proc. ACM AVEC Workshop, Amsterdam, Netherlands, 2016, pp. 41 48.

[27] Y.-H. H. Tsai, S. Bai, P. P. Liang, J. Z. Kolter, L.-P. Morency, and R. Salakhutdinov, "Multimodal transformer for unaligned multimodal language sequences," in Proc. ACL, Florence, Italy, 2019, pp. 6558 6569.

[28] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, “Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot),” JMIR Mental Health, vol. 4, no. 2, p. e19, 2017.
[29] B. Inkster, S. Sarda, and V. Subramanian, “An empathy-driven, conver sational AI agent (Wysa): Real-world data evaluation mixed-methods study,” JMIR mHealth uHealth, vol. 6, no. 11, p. e12106, 2018.
[30] I.-Y. Chen, P. Szolovits, and M. Ghassemi, “Can AI help reduce disparities in general medical and mental health care?,” AMA Journal of Ethics, vol. 21, no. 2, pp. 167–179, 2019.
[31] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, “Dissecting racial bias in an algorithm used to manage the health of populations,” Science, vol. 366, no. 6464, pp. 447–453, 2019.
[32] L. Floridi et al., “An ethical framework for a good AI society: Oppor tunities, risks, principles, and recommendations,” Minds Mach., vol. 29, no. 4, pp. 689–707, 2019.