ICCK Transactions on Emerging Topics in Artificial Intelligence
http://dx.doi.org/10.62762/TETAI.2025.416797

REVIEW ARTICLE

Check for updates

Advances in Artificial Intelligence-Based Depression Diagnosis: A Systematic Review

Liangguo Wang[1] **and Jiaqian Wu**[2,*]

[1] Xinjiang Hetian College, Xinjiang 848000, China
[2] Beijing Anding Hospital, Capital Medical University, Beijing 100088, China

Abstract

Depression affects approximately 280 million people worldwide, yet remains substantially underdiagnosed due to the limitations of clinical assessment tools. This systematic review synthesizes 38 studies (2013–2025) on AI-based depression detection across voice, facial expression, physiological signals, and social media modalities, following PRISMA 2020 guidelines. Multimodal fusion consistently outperforms unimodal approaches, with reported F1 gains of 5–15% on benchmark datasets; however, no reviewed system has undergone prospective clinical validation, and reported accuracy figures should be interpreted as upper bounds on clinical utility. Critical open challenges include cross-cultural generalizability, longitudinal monitoring, interpretability, and ethical governance. This review reframes the field's central question from *can AI detect depression?* to *under what conditions and with what governance can it be safely deployed?*

Keywords: depression, artificial intelligence, multimodal features, deep learning, intelligent detection.

Submitted: 22 May 2025
Accepted: 18 July 2025
Published: 27 August 2025

Vol. 2, **No.** 3, 2025.
10.62762/TETAI.2025.416797

***Corresponding author:**
Jiaqian Wu
1310908555@qq.com

1 Introduction

Depression affects approximately 280 million people worldwide and ranks among the leading contributors to years lived with disability [3]. The COVID-19 pandemic drove a further 25% increase in prevalence during 2020 [4], while in China the lifetime prevalence reaches 6.8% [5, 6]. Despite this burden, depression remains substantially underdiagnosed. Standard instruments such as the PHQ-9 and Hamilton Depression Rating Scale are vulnerable to inter-rater variability, recall bias, and symptom concealment, and scale poorly to primary care and low-resource settings [7, 8].

AI offers an objective, scalable complement to clinical assessment. Machine learning algorithms have matched clinician-level accuracy on structured clinical data [1, 2], while deep learning now enables end-to-end analysis of speech acoustics, facial dynamics, EEG signals, and social media language [9]. Multimodal systems integrating these streams consistently outperform unimodal approaches, reflecting depression's multidimensional phenomenology [10, 11]. Yet critical obstacles remain: datasets are small and demographically narrow; models generalize poorly across recording

Citation
Wang, L., & Wu, J. (2025). Advances in Artificial Intelligence-Based Depression Diagnosis: A Systematic Review. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 2(3), 148–156.

conditions; deep learning architectures lack clinical interpretability; and privacy, bias, and regulatory questions remain unresolved.

This systematic review synthesizes AI-based depression detection across four modalities—voice, facial expression, physiological signals, and social media—alongside fusion strategies and benchmark datasets. Our contributions are fourfold: a modality- organized taxonomy of detection methods; a cross-benchmark performance comparison revealing generalization gaps; a critical analysis of methodological limitations; and an evidence-based research agenda for clinically deployable systems.

Search Strategy

Searches were conducted in PubMed, IEEE Xplore, and Google Scholar using: (*depression* OR *major depressive disorder* OR *MDD*) AND (*artificial intelligence* OR *machine learning* OR *deep learning*) AND (*detection* OR *diagnosis* OR *classification*), supplemented by modality-specific terms. Eligibility required studies to report quantitative performance on a defined dataset, focus on automated depression detection, and be published as peer-reviewed full-text articles in English (January 2013–March 2025). Of 2,847 records identified, 38 met inclusion criteria after deduplication and screening. This review follows PRISMA 2020 guidelines [43].

2 Multimodal Feature Analysis of Intelligent Depression Detection

Depression manifests across multiple behavioral and physiological dimensions simultaneously, making single-modality approaches inherently limited in capturing its full clinical picture. Multimodal systems that integrate complementary data streams have consistently demonstrated superior diagnostic performance over unimodal counterparts [9, 18], motivating the field's shift toward fusion-based architectures.

2.1 Voice Feature Analysis

Speech is among the most extensively studied biomarkers of depression. Prosodic features—particularly reductions in fundamental frequency (F0), decreased voice energy, and slowed speech rate—are consistently reported across studies [19]. Beyond acoustics, linguistic markers extracted via NLP have also proven informative: analysis of online text reveals depression- associated language patterns including increased use of first-person singular pronouns, negative affect words, and reduced social engagement vocabulary [10]. However, both acoustic and linguistic features exhibit substantial inter-individual variability and sensitivity to confounds such as age, gender, and recording conditions, limiting cross-corpus generalizability.

2.2 Facial Expression Analysis

Facial dynamics provide objective, non-invasive indicators of depressive states. Depressed individuals characteristically display reduced action unit activation, delayed expression transitions, and prolonged neutral affect [12]. Deep learning models—particularly spatiotemporal CNNs and attention-based architectures—have demonstrated strong performance on clinical interview datasets such as DAIC-WOZ and AVEC, with F1 scores exceeding 0.80 on binary classification tasks [39]. Nevertheless, performance degrades substantially in naturalistic settings due to illumination variability, pose changes, and the absence of controlled elicitation protocols.

2.3 Physiological Signal Analysis

Physiological signals offer direct windows into the neurobiological substrates of depression. EEG-based approaches have identified depression-associated patterns including alpha asymmetry and reduced theta coherence, achievable through both traditional feature engineering and end-to-end deep learning [14]. Wearable-derived digital biomarkers—including heart rate variability, actigraphy, and galvanic skin response—enable continuous passive monitoring in ecological settings, offering advantages over laboratory-bound measurements [15]. A key limitation is dataset scarcity: clinically annotated physiological recordings remain rare due to the cost and ethical complexity of collection.

2.4 Social Media Behavior Analysis

Social media platforms provide large-scale, longitudinal behavioral data reflecting real-world depressive symptomatology. Linguistic, temporal, and social interaction features extracted from platforms such as Twitter and Reddit have demonstrated predictive validity against clinical assessments [11, 17]. Population-level monitoring studies have further shown that aggregated social media signals can track depression prevalence trends at the community level [16, 29]. Critical challenges include the non-representative nature of social media users, platform policy constraints on data access, and the ethical implications of passive mental health surveillance.

2.5 Multimodal Feature Fusion

Fusion strategies can be broadly categorized as feature-level (early), decision-level (late), or hybrid. Feature-level fusion concatenates representations before classification, preserving cross-modal interactions but increasing dimensionality. Decision-level fusion aggregates modality-specific predictions, offering robustness to missing modalities [18]. More recently, attention-based fusion mechanisms—including cross-modal transformers—have enabled dynamic weighting of modality contributions, yielding state-of-the-art performance on multimodal benchmarks [21]. Despite these advances, the optimal fusion strategy remains dataset-dependent, and modality alignment across heterogeneous data streams poses ongoing technical challenges.

2.6 Benchmark Datasets

Progress in AI-based depression detection is tightly coupled to dataset availability. Table 1 summarizes the primary benchmarks used in the field. DAIC-WOZ [40] and the AVEC challenge series [12] remain the dominant multimodal benchmarks, providing synchronized audio- visual-text recordings with standardized PHQ-8 annotations. However, most datasets are small (fewer than 300 subjects), demographically homogeneous, and collected under laboratory conditions—factors that severely limit the ecological validity and generalizability of models trained on them. The emergence of naturalistic datasets such as D-VLOG [31] and social media corpora represents a meaningful step toward more representative evaluation, though annotation quality and label reliability remain concerns.

As summarized in Table 1, more recently, the D-VLOG dataset [31] has emerged as a valuable resource, containing 961 video blogs (approximately 160 hours) from 816 subjects in naturalistic settings, offering a more ecologically valid alternative to laboratory-collected data. The Pittsburgh Depression Dataset [32] includes 49 subjects with 130 audio-visual samples recorded during clinical interviews, with expert annotations of depression severity.

For physiological signal analysis, several datasets are available: MODMA [41] provides 53-channel EEG recordings from clinically diagnosed patients and matched controls; DEAP [33] contains EEG and peripheral physiological signals from 32 participants during emotion elicitation experiments; although originally designed for general emotion recognition, its physiological signal modalities have been adopted in depression-related affective computing research as a reference benchmark; and SEED [34] offers EEG data collected for emotion recognition research, investigating critical frequency bands and channels through deep neural networks; its methodology has been referenced in depression-related EEG studies for signal processing approaches.

Text-based depression research is supported by datasets such as the eRisk CLEF collection [35], which contains early risk detection data from social media platforms; the Reddit Depression Corpus [42], which includes user posts from depression-related subreddits; and various Twitter depression datasets [36] that enable large-scale analysis of linguistic patterns.

Recent studies have further enriched the pool of depression-relevant resources. Wang et al. [37] proposed a three-dimensional convolutional network with attention mechanism for depression speech recognition, demonstrating that acoustic features extracted from multi-channel speech signals can effectively capture depressive symptom severity. IEMOCAP [38] provides multimodal recordings of dyadic interactions with valence and arousal annotations; its audio-visual emotional features have been utilized in transfer learning approaches for depression detection, given the overlap between sadness-related affective states and depressive presentations. He et al. [39] proposed a CNN-based framework with attention mechanism for automatic depression recognition from facial videos, demonstrating that spatiotemporal attention effectively captures subtle facial dynamics associated with depressive states.

These diverse datasets collectively address different aspects of depression manifestation while presenting unique challenges in data quality, annotation consistency, and ethical considerations. Researchers must carefully select appropriate datasets based on their specific research questions and methodological approaches.

3 Algorithm Research on Intelligent Depression Detection

Algorithmic approaches to depression detection have evolved from interpretable statistical models to end-to-end deep learning architectures, with each generation offering improved representational capacity at the cost of increased data requirements

Table 1. Commonly used depression prediction datasets.

Dataset Name	Data Type	Key Features
DAIC-WOZ	Audio/Video/Text	189 participants in clinical interviews with PHQ-8 depression scores
AVEC Challenge	Multimodal	Annual competition datasets with synchronized audio-visual recordings for affective computing research
D-VLOG	Video/Audio	961 vlogs (160+ hours) from 816 subjects in naturalistic settings
Pittsburgh Depression	Audio/Video	49 subjects with 130 samples from clinical interviews
MODMA	EEG signals	53-channel EEG recordings from clinically diagnosed patients and matched controls
DEAP	EEG/Physiological	EEG and peripheral physiological signals from 32 participants during emotion elicitation; adopted as benchmark in affective computing research including depression studies
SEED	EEG	EEG data for emotion recognition research; methodology referenced in depression-related EEG signal processing studies
eRisk CLEF	Text	Early risk detection data from social media platforms
Reddit Depression Corpus	Text	User posts from depression-related subreddits with temporal information
Depression Twitter	Social media text	Large-scale collection of tweets with depression indicators
IEMOCAP	Audio/Video/Motion	Interactive Emotional Dyadic Motion Capture Database with valence/arousal annotations; audio-visual features utilized in transfer learning for depression detection

and reduced transparency [20]. This section traces this trajectory and critically evaluates the trade-offs involved.

3.1 Traditional Machine Learning Methods

Classical ML methods—including SVMs, ANNs, random forests, and logistic regression—dominated early depression detection research and retain practical relevance in low-resource clinical settings due to their interpretability and modest data requirements. SVMs with radial basis kernels have shown consistent efficacy on high-dimensional EEG features, achieving competitive accuracy through systematic hyperparameter optimization [1]. ANN-based approaches have demonstrated utility in community screening contexts where clinical infrastructure is limited [28]. A systematic meta-analysis further confirmed that ML algorithms can meaningfully predict treatment outcomes in depression, supporting their role in clinical decision support [27]. The primary limitation of classical methods is their reliance on hand-crafted features, which constrains their ability to capture the temporal and contextual complexity of depression-related signals.

3.2 Deep Learning Methods

Deep learning has substantially advanced depression detection by enabling automatic feature extraction from raw, unstructured data. NLP models applied to social media text have revealed depression-associated linguistic markers—including elevated use of negative affect words and reduced future-oriented language—that are imperceptible to rule-based systems [10]. Longitudinal social media analysis has further demonstrated that language shifts precede clinical symptom onset, suggesting potential for early intervention [13].

Multimodal deep learning architectures that jointly process audio, visual, and physiological streams have consistently outperformed unimodal baselines, with reported gains of 5–15% in F1 score on DAIC-WOZ and AVEC benchmarks [18]. Attention mechanisms have been particularly impactful: cross-modal attention enables dynamic weighting of modality contributions, while self-attention captures long-range temporal dependencies within single modalities. Critically, attention weight visualization provides a degree of post-hoc interpretability, allowing clinicians to identify which facial regions or speech segments drive model predictions [21]—a property essential for clinical trust and regulatory acceptance.

3.3 Algorithm Optimization

Translating high-performing models into clinical tools requires optimization beyond accuracy. For EEG-based systems, temporal convolutional networks and recurrent architectures have proven effective at capturing depression-related spectral dynamics, while lightweight model compression techniques enable deployment on resource-constrained wearable devices [22]. Regularization strategies including dropout, batch normalization, and domain adversarial training have improved cross-dataset generalization—a persistent bottleneck where models trained on one benchmark typically show 10–20% performance degradation when evaluated on another [23]. Data augmentation and transfer learning from large-scale affective computing corpora represent promising but underexplored directions for addressing the chronic data scarcity in clinical depression research. Table 2 summarizes the performance of representative methods reviewed, illustrating both the progress achieved and the cross-dataset comparability limitations discussed above.

4 Existing Problems and Challenges

Despite significant advancements in AI-based depression detection, several critical challenges remain across multiple domains including data management, technological development, clinical integration, and ethical considerations [20, 25]. Addressing these challenges is essential for translating research innovations into practical clinical applications that can meaningfully impact mental healthcare delivery.

4.1 Data Quality and Standardization Issues

High-quality, standardized data represents the cornerstone of effective intelligent detection systems. Several fundamental challenges in medical data acquisition and utilization directly impact model performance and clinical validity [25]. These include inconsistent data collection protocols, variable annotation standards, and limited representation of diverse patient populations.

To overcome these limitations, researchers advocate for establishing unified acquisition standards and rigorous quality control processes. This approach necessitates the formation of specialized annotation teams working with comprehensive guidelines and protocols. Implementing multi-rater annotation procedures with formal consistency assessments significantly enhances the reliability of training data, which directly translates to improved model performance in clinical applications.

4.2 Technical Reliability and Robustness

The reliability and robustness of depression detection technologies remain significant concerns for real-world deployment. In their comprehensive review, Shatte et al. [20] highlighted three primary technical challenges: limited model generalization across diverse populations and clinical contexts, heightened sensitivity to data noise and environmental interference, and insufficient stability in multimodal feature fusion frameworks. These limitations can lead to inconsistent performance in naturalistic settings, potentially undermining clinical trust and adoption.

4.3 Clinical Practice and Application Issues

Bridging the gap between technological innovation and clinical implementation presents multifaceted challenges. Kellogg and Sadeh-Sharvit [30] emphasized that successful clinical integration depends on addressing model interpretability, seamless workflow integration, and real-time performance capabilities. Clinicians require systems that not only provide accurate assessments but also offer transparent reasoning that aligns with established clinical frameworks. Well-designed AI-assisted diagnosis systems have demonstrated potential to enhance diagnostic efficiency and accuracy while providing reliable decision support, but their implementation requires careful attention to clinical workflows and practitioner needs.

4.4 Ethical Privacy and Security Issues

The deployment of intelligent detection systems raises significant ethical considerations regarding patient privacy and data security. Larson et al. [24] proposed a comprehensive framework for protecting patient privacy in medical data utilization, emphasizing that technological innovation must proceed in parallel with robust ethical safeguards. This framework highlights the importance of informed consent, data minimization principles, and transparent algorithmic processes. Establishing sound ethical codes and privacy protection mechanisms not only safeguards patient rights but also significantly influences social acceptance and long-term technological adoption.

4.5 Cost-Benefit and Implementation Considerations

The widespread adoption of intelligent detection technologies necessitates careful evaluation of economic factors and implementation requirements. Huckvale et al. [26] advocated for comprehensive assessment frameworks that consider system

Table 2. Performance comparison of representative AI-based depression detection methods.

Study	Modality	Method	Dataset	Metric	Performance
Yang et al. [1]	EEG	SVM-RBF	PRED+CT	Accuracy	96.7%
Sau & Bhakta [28]	Clinical	ANN	Custom	Accuracy	97.2%
Cai et al. [14]	EEG	KNN	Custom	Accuracy	79.3%
De Choudhury et al. [11]	Social media	SVM	Twitter	F1	70%
Tadesse et al. [17]	Text	LIWC+LDA+Bigram	Reddit	Accuracy	91.0%
Ringeval et al. [12]	Audio/Video	Multimodal	AVEC 2019	F1	52%
He et al. [39]	Video	CNN+Attention	DAIC-WOZ	F1	78%
Yang et al. [18]	Multimodal	Decision Fusion	DAIC-WOZ	F1	86%*

*Development set result; test set performance not separately reported.

development and maintenance costs, infrastructure requirements, training needs, and measurable clinical outcomes. Successful implementation strategies must balance technological capabilities with resource constraints, particularly in diverse healthcare settings. This requires establishing unified data standards, developing robust and interpretable algorithms, conducting rigorous clinical validation studies, implementing appropriate ethical safeguards, and creating evidence-based evaluation systems to demonstrate clinical and economic value [20, 25].

5 Discussion

5.1 Modality-Dataset Interaction

Multimodal fusion consistently outperforms unimodal baselines, yet individual modality contributions vary substantially across datasets. Speech models dominate on structured interview corpora such as DAIC-WOZ but degrade under naturalistic recording conditions [19]; facial models show the inverse pattern, with robustness to acoustic variability but sensitivity to demographic confounds [12]. This interaction suggests that benchmark rankings reflect dataset-specific biases as much as genuine algorithmic superiority, and that performance figures should not be assumed transferable across clinical contexts.

5.2 Performance versus Interpretability

Deep learning achieves the highest benchmark scores but produces explanations that are not clinically grounded—attention weights do not map onto DSM-5 criteria or established symptom dimensions [30]. Classical ML offers transparency but cannot capture temporal complexity. This trade-off cannot be fully resolved technically; deployment decisions must therefore be context-driven, with interpretable models preferred for individual diagnosis and deep learning reserved for population screening.

5.3 The Validation Gap

Every high-performing system reviewed was evaluated retrospectively on controlled benchmarks. No study identified completed prospective clinical validation, leaving unknown the failure modes—distributional shift, workflow friction, edge-case behavior—that emerge in real deployment [20]. Reported metrics should be interpreted as upper bounds on, not estimates of, clinical utility.

5.4 Advances Since Prior Reviews

Relative to Shatte et al. [20], three advances are evident: attention-based multimodal fusion has emerged as the dominant paradigm; naturalistic datasets have challenged laboratory-derived findings; and the field's central question has shifted from *can AI detect depression?* to *under what conditions and with what governance can it be safely deployed?*

6 Future Development Trends

The preceding analysis reveals several structural gaps that define the field's near-term research agenda. Rather than enumerating general directions, this section identifies four specific open questions whose resolution is prerequisite for clinical deployment.

6.1 Cross-Cultural and Demographic Generalizability

The vast majority of reviewed datasets originate from North American and East Asian populations under controlled laboratory conditions. Whether depression-associated acoustic, facial, and linguistic markers are culturally universal or context-dependent remains empirically unresolved. Future work should prioritize multi-site, cross-cultural data collection protocols and evaluate model performance explicitly across age, gender, ethnicity, and language subgroups. Domain adaptation and federated learning frameworks offer technically feasible pathways to cross-institutional generalization without requiring centralized data sharing [25].

6.2 Longitudinal and Ecological Validity

Existing systems are almost exclusively trained and evaluated on single-session, laboratory-collected data, which poorly approximates the episodic and fluctuating course of clinical depression. Passive sensing paradigms—integrating smartphone behavioral data, wearable physiological monitoring, and ambient acoustic capture—enable continuous, ecologically valid assessment at low burden to patients [15]. The critical research challenge is developing models that detect clinically meaningful symptom trajectories rather than cross-sectional severity scores, requiring longitudinal datasets and time-series modeling architectures that remain largely absent from current literature.

6.3 Interpretability and Clinical Integration

Regulatory approval and clinician adoption of AI diagnostic tools require that model outputs be explainable in clinically meaningful terms. Current attention-based interpretability approaches provide post-hoc feature attribution but do not yet map onto established clinical constructs such as DSM-5 diagnostic criteria or treatment response predictors [30]. Future systems should embed clinical knowledge through theory-informed architectures, symptom-level supervision signals, and structured output formats aligned with existing clinical workflows [20]. Prospective clinical validation studies—moving beyond retrospective benchmark evaluation—are essential to establish real-world utility.

6.4 Ethical Governance and Privacy-Preserving Deployment

The deployment of depression detection systems raises acute ethical concerns that technical solutions alone cannot resolve. Passive behavioral monitoring risks perpetuating algorithmic bias against underrepresented populations; training data reflecting historical diagnostic disparities will reproduce rather than correct those disparities [25]. Privacy-preserving techniques including federated learning, differential privacy, and on-device inference provide partial technical mitigations, but must be accompanied by regulatory frameworks specifying data governance, consent procedures, and audit requirements. Multidisciplinary collaboration among AI researchers, clinicians, ethicists, and policymakers is essential to develop governance structures commensurate with the sensitivity of mental health data [24].

7 Conclusion

This review has systematically examined AI-based depression detection across voice, facial expression, physiological signal, and social media modalities. Three conclusions emerge from this synthesis.

Multimodal fusion consistently outperforms unimodal approaches, yet current architectures remain brittle to missing modalities and cross-dataset distribution shift. Models achieving high accuracy on controlled benchmarks such as DAIC-WOZ have not been validated in prospective clinical settings, and the absence of standardized evaluation protocols renders cross-study comparison largely uninformative. Ethical and regulatory dimensions—including algorithmic bias, privacy governance, and equitable access—have received disproportionately little attention relative to their centrality for real-world deployment.

Closing these gaps requires richer ecological datasets, architectures that balance accuracy with interpretability, and governance frameworks that earn clinician and patient trust. These remain the field's most consequential open challenges.

Data Availability Statement

Not applicable.

Funding

This work was supported by the Key Project of Natural Science Foundation of Xinjiang Hetian College under Grant 2025ZR002.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Yang, C. Y., & Chen, Y. Z. (2024). Support vector machine classification of patients with depression based on resting-state electroencephalography. *Asian Biomedicine: Research, Reviews and News*, *18*(5), 212-223. [CrossRef]

[2] Su, C., Xu, Z., Pathak, J., & Wang, F. (2020). Deep learning in mental health outcome research: a scoping review. *Translational psychiatry, 10*(1), 116. [CrossRef]

[3] World Health Organization. (2022). *World mental health report: Transforming mental health for all*. WHO Press.

[4] Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., ... & Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet, 398*(10312), 1700-1712. [CrossRef]

[5] Wang, J., Zhao, Z., Yang, J., Wang, L., Zhang, M., & Zhou, M. (2024). The Association Between Depression and All-Cause, Cause-Specific Mortality in the Chinese Population—China, 2010–2022. *China CDC Weekly, 6*(40), 1022. [CrossRef]

[6] The Lancet Psychiatry. (2024). Global burden of disease 2021: Mental health messages. *The Lancet Psychiatry, 11*(8), 573. [CrossRef]

[7] Chinese Medical Association, Chinese Medical Association Publishing House, & Chinese Medical Association General Practice Branch. (2021). Guidelines for the diagnosis and treatment of depression in primary care (2021). *Chinese Journal of General Practitioners, 20*(12), 1101–1116. [CrossRef]

[8] Torous, J., Myrick, K. J., Rauseo-Ricupero, N., & Firth, J. (2020). Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. *JMIR mental health, 7*(3), e18848. [CrossRef]

[9] Sadeghi, M., Richer, R., Egger, B., Schindler-Gmelch, L., Rupp, L. H., Rahimi, F., ... & Eskofier, B. M. (2024). Harnessing multimodal approaches for depression detection using large language models and facial expressions. *npj Mental Health Research, 3*(1), 66. [CrossRef]

[10] Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. S. (2020). Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. *Journal of medical Internet research, 22*(10), e22635. [CrossRef]

[11] De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Proceedings of the international AAAI conference on web and social media* (Vol. 7, No. 1, pp. 128-137). [CrossRef]

[12] Ringeval, F., Schuller, B., Valstar, M., Cummins, N., Cowie, R., Tavabi, L., ... & Pantic, M. (2019, October). AVEC 2019 workshop and challenge: state-of-mind, detecting depression with AI, and cross-cultural affect recognition. In *Proceedings of the 9th International on Audio/visual Emotion Challenge and Workshop* (pp. 3-12). [CrossRef]

[13] Guntuku, S. C., Sherman, G., Stokes, D. C., Agarwal, A. K., Seltzer, E., Merchant, R. M., & Ungar, L. H. (2020). Tracking mental health and symptom mentions on Twitter during COVID-19. *Journal of general internal medicine, 35*(9), 2798-2800. [CrossRef]

[14] Cai, H., Han, J., Chen, Y., Sha, X., Wang, Z., Hu, B., ... & Gutknecht, J. (2018). A pervasive approach to EEG-based depression detection. *complexity, 2018*(1), 5238028. [CrossRef]

[15] Jacobson, N. C., Weingarden, H., & Wilhelm, S. (2019). Digital biomarkers of mood disorders and symptom change. *NPJ digital medicine, 2*(1), 3. [CrossRef]

[16] Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences, 18*, 43-49. [CrossRef]

[17] Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. *Ieee Access, 7*, 44883-44893. [CrossRef]

[18] Yang, L., Jiang, D., He, L., Pei, E., Oveneke, M. C., & Sahli, H. (2016, October). Decision tree based depression classification from audio video and language information. In *Proceedings of the 6th international workshop on audio/visual emotion challenge* (pp. 89-96). [CrossRef]

[19] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech communication, 71*, 10-49. [CrossRef]

[20] Shatte, A. B., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine, 49*(9), 1426-1448. [CrossRef]

[21] He, L., Niu, M., Tiwari, P., Marttinen, P., Su, R., Jiang, J., ... & Dang, W. (2022). Deep learning for depression recognition with audiovisual cues: A review. *Information Fusion, 80*, 56-86. [CrossRef]

[22] Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering, 16*(5), 051001. [CrossRef]

[23] Huang, Z., Epps, J., Joachim, D., Stasak, B., Williamson, J. R., & Quatieri, T. F. (2020, January). Domain adaptation for enhancing speech-based depression detection in natural environmental conditions using dilated CNNs. In *Interspeech 2020* (Vol. 2020, pp. 4561-4565). International Speech Communication Association. [CrossRef]

[24] Larson, D. B., Magnus, D. C., Lungren, M. P., Shah, N. H., & Langlotz, C. P. (2020). Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. *Radiology, 295*(3), 675-682. [CrossRef]

[25] Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: an overview. *Current psychiatry reports, 21*(11), 116. [CrossRef]

[26] Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *NPJ digital medicine*, *2*(1), 88. [CrossRef]

[27] Lee, Y., Ragguett, R. M., Mansur, R. B., Boutilier, J. J., Rosenblat, J. D., Trevizol, A., ... & McIntyre, R. S. (2018). Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. *Journal of affective disorders*, *241*, 519-532. [CrossRef]

[28] Sau, A., & Bhakta, I. (2017). Artificial neural network (ANN) model to predict depression among geriatric population at a slum in Kolkata, India. *Journal of clinical and diagnostic research: JCDR*, *11*(5), VC01. [CrossRef]

[29] De Choudhury, M., Counts, S., & Horvitz, E. (2013, May). Social media as a measurement tool of depression in populations. In *Proceedings of the 5th annual ACM web science conference* (pp. 47-56). [CrossRef]

[30] Kellogg, K. C., & Sadeh-Sharvit, S. (2022). Pragmatic AI-augmentation in mental healthcare: key technologies, potential benefits, and real-world challenges and solutions for frontline clinicians. *Frontiers in psychiatry*, *13*, 990370. [CrossRef]

[31] Yoon, J., Kang, C., Kim, S., & Han, J. (2022, June). D-vlog: Multimodal vlog dataset for depression detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 11, pp. 12226-12234). [CrossRef]

[32] Girard, J. M., Cohn, J. F., Mahoor, M. H., Mavadati, S. M., Hammal, Z., & Rosenwald, D. P. (2014). Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses. *Image and vision computing*, *32*(10), 641-647. [CrossRef]

[33] Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., ... & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18-31. [CrossRef]

[34] Zheng, W. L., & Lu, B. L. (2015). Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, *7*(3), 162-175. [CrossRef]

[35] Losada, D. E., & Crestani, F. (2016, August). A test collection for research on depression and language use. In *International conference of the cross-language evaluation forum for European languages* (pp. 28-39). Cham: Springer International Publishing. [CrossRef]

[36] Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality* (pp. 1-10).

[37] Wang, H., Liu, Y., Zhen, X., & Tu, X. (2021). Depression speech recognition with a three-dimensional convolutional network. *Frontiers in human neuroscience*, *15*, 713823. [CrossRef]

[38] Busso, C., Bulut, M., Lee, C. C., Kazemzadeh, A., Mower, E., Kim, S., ... & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, *42*(4), 335-359. [CrossRef]

[39] He, L., Chan, J. C. W., & Wang, Z. (2021). Automatic depression recognition using CNN with attention mechanism from videos. *Neurocomputing*, *422*, 165-175. [CrossRef]

[40] Gratch, J., Artstein, R., Lucas, G. M., Stratou, G., Scherer, S., Nazarian, A., ... & Morency, L. P. (2014, May). The distress analysis interview corpus of human and computer interviews. *In LREC* (Vol. 14, pp. 3123-3128).

[41] Cai, H., Yuan, Z., Gao, Y., Sun, S., Li, N., Tian, F., ... & Hu, B. (2022). A multi-modal open dataset for mental-disorder analysis. *Scientific Data*, *9*(1), 178. [CrossRef]

[42] Pirina, I., & Çöltekin, Ç. (2018, October). Identifying depression on reddit: The effect of training data. In *Proceedings of the 2018 EMNLP workshop SMM4H: the 3rd social media mining for health applications workshop & shared task* (pp. 9-12). [CrossRef]

[43] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. [CrossRef]

Liangguo Wang received the PhD degree in Computer Science and Technology from Beijing Institute of Technology, Beijing 100081, China, in 2018. His research interests include multimodal analysis and artificial intelligence applications.(Email: wlgbit@163.com)

Jiaqian Wu received the B.S. degree in Clinical Medicine from Anhui Medical University, Hefei 230032, China, in 2016. (Email:1310908555@qq.com)