Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 6 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i06.61567

SPECIAL FOCUS PAPER

Explainable Deep Learning Models for Individualized Mental Health Risk Assessment Using Wearable and Smartphone Sensing

Krisna Veni Balakrishnan[1], Geetika Parmar[2] (✉)

[1]Rushford Business School, Lucerne, Switzerland

[2]Dr. Vishwanath Karad MIT World Peace University, Pune, India

geetika.parmar@mitwpu.edu.in

ABSTRACT

Background: Mental health disorders represent a major global public health concern, creating a critical need for early detection and personalized intervention strategies. Advances in wearable devices and smartphone sensing enable continuous and unobtrusive monitoring of behavioral and physiological patterns, while deep learning offers powerful tools for analyzing such high-dimensional data. However, the limited interpretability of many deep learning models restricts their clinical adoption. **Methods:** The study presents a systematic literature review of explainable deep learning models for individualized mental health risk assessment using wearable and smartphone sensing data. A comprehensive search was conducted across IEEE Xplore, PubMed, Google Scholar, and the Directory of Open Access Journals (DOAJ). Studies were screened using predefined inclusion and exclusion criteria, resulting in the selection of 25 relevant articles for qualitative synthesis. **Results:** The reviewed studies demonstrate a clear shift toward multimodal sensing and advanced deep learning architectures, including attention-based and temporal models, to capture complex behavioral dynamics. Despite strong predictive performance, explainability techniques were inconsistently applied across studies. Challenges related to data quality, validation practices, generalizability, and real-world deployment were frequently identified. **Conclusion:** The findings highlight the importance of integrating explainability into deep learning-based mental health assessment systems to enhance trust, clinical relevance, and regulatory compliance. This review provides a structured synthesis of current approaches and outlines key research directions for developing transparent, reliable, and clinically meaningful mental health assessment frameworks.

KEYWORDS

mental health risk assessment, wearable sensing, smartphone sensing, deep learning, explainable artificial intelligence

Balakrishnan, K. V., Parmar, G. (2026). Explainable Deep Learning Models for Individualized Mental Health Risk Assessment Using Wearable and Smartphone Sensing. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(6), pp. 124–138. https://doi.org/10.3991/ijoe.v22i06.61567

Article submitted 2026-03-16. Revision uploaded 2026-04-07. Final acceptance 2026-04-07.

1 INTRODUCTION

Mental illnesses represent an important global health concern that affects hundreds of millions of people, resulting in substantial impairment, low quality of life, and socioeconomic cost. Estimates by the World Health Organization indicate that depression and other mental illnesses are some of the top causes of years lived with disability and that there is a dire need to come up with adequate preventive measures, monitoring, and early intervention strategies [1]. Psychological methods of mental health assessment are mostly based on self-reported questionnaires and infrequent clinical assessments, and they are prone to recall bias, subjectivity, and problems with time resolution. As a result of new developments in mobile and sensing technologies, human behavior may now be monitored in real-time and without being obtrusive in real-life contexts. Nowadays, the idea of digital phenotyping has become a potentially successful paradigm of capturing moment-to-moment behavioral and psychological patterns based on the data gathered through smartphones and wearable devices [2]. Such technologies can offer deep movements of sensor data on mobility, physical activity, sleep, social interaction, and device use and new prospects of a personalized mental health risk evaluation and personalized care [3]. These streams of data enable longitudinal and real-time study of the behavioral patterns and enable observing mental health changes in a more naturalistic setting and not just during a single clinical appointment. Its level of importance is quite high in identifying nuanced risk indicators, in helping to generate individualized action, and in predictive analytics of individualized mental health examinations.

An increasing amount of literature has investigated how wearable and smartphone-based sensing data can be used to identify and forecast mental health states. The initial research revealed that passive data collected by mobile phone sensors, including location logs and activity intensity, have a strong correlation with the severity of depressive symptoms [4]. Sub-research also indicated that passive sensing over smartphones can be applied to anticipate moment-to-moment changes of depressed mood in individuals with clinical symptoms of significance [5]. Systematic reviews have validated the practicability and potential of smartphone-based passive sensing strategies for monitoring the state of health and well-being, as well as mental health outcomes [6].

Simultaneously, machine learning and deep learning methods have been used with growing opportunities to analyze high-dimensional and multimodal sensing data to determine mental health risk. Surveys of machine learning applications in mental health have already pointed to the utility of sophisticated models in enhancing the quality of prediction as compared to the conventional statistical techniques [7]. The concept of data-driven models is also being applied in the new area of computational psychiatry to learn more about mental disorders and aid in clinical decision-making [8]. The most recent works have proven the effective implementation of machine learning models trained on mobile phone usage metadata to screen depression, which once again confirms the importance of mobile sensing in mental health assessment [9]. In the reviewed literature, explainability emerged as a fundamental requirement for deploying deep learning models in high-stakes application domains such as mental health assessment. Foundational work emphasized the need for a rigorous and well-defined scientific framework for interpretable machine learning to ensure reliability and trust in such contexts [10]. To support safe and reliable use in healthcare, interpretable machine learning must be grounded in well-defined and rigorous scientific principles. Several studies highlight the importance of an adequate scientific basis of interpretable machine learning, especially in high-stakes healthcare settings. This is acutely needed, especially in the area of mental health assessment, where the decisions arrived at may have significant clinical and ethical ramifications and where transparency and trust are needed to replicate the same in real life.

Although the investigation of wearable sensing, deep learning, and explainability has been done independently in most studies, few systematic reviews will be done that will synthesize explainable deep learning models of personalized mental health risk assessment using wearable and smartphone-generated data. This gap has motivated the present systematic literature review, which aims to explore the existing practices, identify key trends and gaps, and also indicate where future research should head in this highly emerging field. Particularly, several aspects that are focused on in this paper include the nature of the sensing modalities employed, the deep learning architecture employed, and the explainability employed to strengthen interpretability and trust. Moreover, it speaks about the methodological problems that are related to the quality of data and model validation and clinical applicability. The purpose of this review is to inform the scientists and the practitioners and enable the development of clear, dependable, and clinically significant systems of mental health assessment through the provision of a critical and systematic synthesis of the literature.

1.1 Objectives of the study

- To systematically review and categorize explainable deep learning models used for individualized mental health risk assessment based on wearable and smartphone sensing data.
- To identify commonly used sensing modalities, deep learning architectures, and explainability techniques reported in the existing literature.
- To analyze current limitations, challenges, and research gaps to inform future development of transparent and clinically applicable mental health assessment systems.

2 MATERIALS AND METHODS

2.1 Data sources and search strategy

The literature search was performed in online databases such as IEEE Xplore, PubMed, Google Scholar, and the Directory of Open Access Journals (DOAJ) that contain a massive repository of peer-reviewed journals and conference proceedings in the domains of artificial intelligence, biomedical engineering, mobile health, and mental health. The choice of search terms was a difficult task, as the range of terms to describe mental health issues, sensing technologies, and explainable artificial intelligence methods is wide.

The original query, which was mental health AND deep learning, brought up a small set of relevant results, many of which have not deployed sensing technologies or addressed the aspects of explainability. To achieve more in-depth and applicable results, the search strategy was narrowed down to incorporate a combination of terms associated with mental health disorders, wearable and smartphone-based sensing, deep learning models, and explainability. The last search query was based on the following terms: mental health AND wearable AND smartphone AND deep learning AND explainable. It included the research papers that used sensor-based data gathered by using mobile phones and implemented explainable deep learning algorithms to assess the mental health risk. A search was performed in the titles, and abstracts and by keywords, and only the studies published in English and available as full-text articles were included.

2.2 Selection criteria for studies

The inclusion and exclusion criteria were formulated based on the area of study that the studies listed in the results section cover, as summarized in Table 1. The review was restricted to the studies employing wearable or smartphone sensing data to determine mental health and using deep learning models with consideration of explainability, validation, and clinical relevance. Papers that did not include sensing-based data, deep learning procedures, and interpretability were not included.

Table 1. Selection criteria for studies

Inclusion Criteria	Exclusion Criteria
Publications written in English	Editorials, commentaries, or opinion-only articles
Any year of publication	Studies lacking wearable or smartphone-based sensing data
Journal articles, conference papers, and book chapters	Studies not employing deep learning models
Focus on mental health risk assessment or monitoring	Works focusing only on general well-being without risk assessment
Use of wearable devices and/or smartphone sensing	Studies using sensing data without mental health relevance
Application of deep learning techniques	Traditional machine learning without deep learning architectures
Consideration of explainability, interpretability, validation, or clinical relevance	Absence of explainability or interpretability considerations
Relevance to real-world, longitudinal, or applied settings	Purely theoretical AI studies without a healthcare application

Source: Compiled by authors.

2.3 Study selection process

Figure 1 illustrates the procedure for selecting the papers. Those records that did not match the objectives of this review were filtered out. Records were based on the titles and abstracts of the articles with the help of the predefined selection criteria. The complete articles of the rest of the studies were then evaluated thoroughly in order to ascertain their applicability to explainable deep learning-based mental health risk assessment using wearable and smartphone sensing data. At this step, studies in which sensing modalities were not adequately covered, deep learning approaches, or explainability were not covered were filtered out. Finally, a conclusive number of studies were identified that qualitatively fit all the inclusion criteria and were qualified to be included in this systematic literature review's qualitative synthesis.

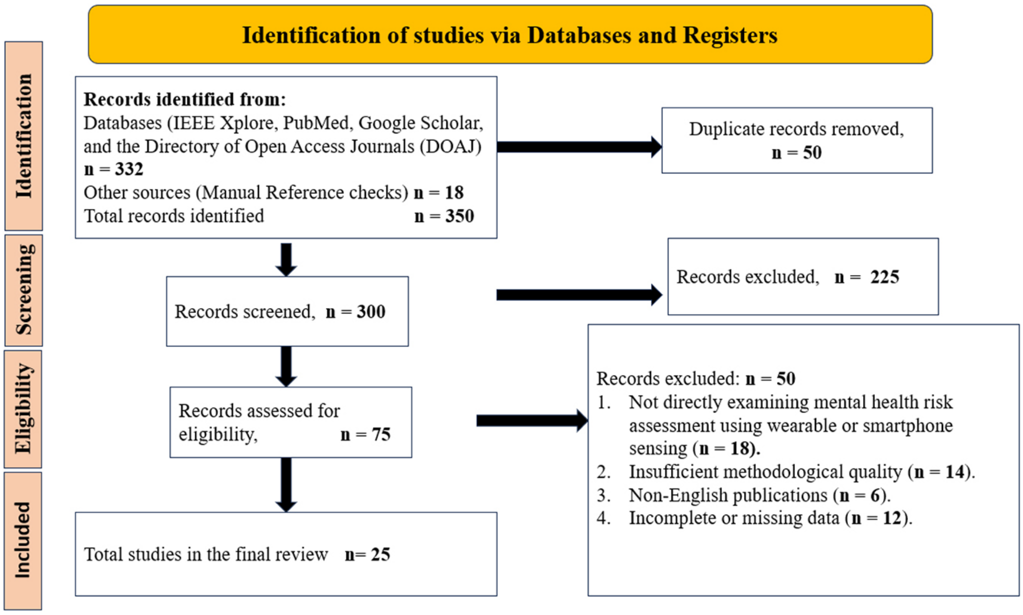

Fig. 1. PRISMA flow diagram illustrating the study identification, screening, eligibility, and inclusion process for the systematic review

Source: Compiled by authors.

3 RESULTS

3.1 Explainability foundations in mental health modeling

Early studies on model-agnostic explanation techniques demonstrated that providing interpretable explanations for classifier predictions can substantially improve user understanding and trust in model outputs [11]. Later developments added single set frameworks of attributing feature significance in complex models, allowing a similar perspective of feature significance across learning structures [12]. A study of the medical field further pushed the point that explainable AI systems need to be domain-specific in both clinical relevance, usability, and transparency to be deemed practically relevant [13]. Clinically, the research has emphasized the need to have context-sensitive and process-oriented explanations in accordance with clinician decision-making patterns to adopt in the real health care facilities [14]. Additionally, larger scopes of artificial intelligence in medicine have determined interpretability as a major determinant of clinical impact, as well as workflow integration and accountability issues [15]. Meanwhile, the idea of interpretability has been discussed critically to emphasize how the concept has not yet achieved a universal definition, and it is necessary to conduct a systematic assessment and synthesis of explainable methods [16]. The insights give a fundamental conceptual basis to the analysis of the explainable deep learning models.

3.2 Overview of selected studies

The literature demonstrates the evolutionary movement of traditional machine learning to modern deep learning and explainable models in the field of mental assessment. Some of the initial investigations showed that mobile phone usage metadata could be used to accurately screen and detect depression, indicating that passive behavioral information can be used for prediction outside of a clinical environment [17]. Additional investigations aimed at longitudinal follow-up have shown that continuous sensing makes it possible to identify developing mental health risks, in contrast to a single-point measure [18]. On the whole, these tendencies indicate the movement toward active, individual assessment and personalized intervention strategies. Table 2 demonstrates the summary of a representative of the studies incorporated in the study of this systematic literature review, summarizing the presence of the targeted mental health conditions, data sources, sample sizes, and study designs.

Table 2. Summary of included studies

Thematic Category	Scope of Studies	Reference IDs (each used once)	Number of Studies
Explainability foundations	Core interpretability/XAI concepts for high-stakes ML	[11]–[16]	6
Smartphone-based mental health sensing	Smartphone sensing + passive data for mental health assessment	[17], [18], [25], [28]	4
Wearable-based mental health sensing	Wearable sensor data for mental health/stress monitoring	[19], [22], [26]	3
Multimodal sensing approaches	Combined wearable + smartphone sensing	[20], [24]	2
Deep learning surveys and reviews	Reviews/surveys/human-centered perspectives in sensing + DL	[21], [23], [27], [29]	4
Clinical validation and evaluation studies	Validation, generalizability, metrics, and reporting guidelines	[31]–[35]	5
Related clinical AI reference	Closely related clinical DL study (supporting evidence)	[30]	1
TOTAL (unique included studies)	Final studies are synthesized in Results	[11]–[35]	25

Source: Compiled by authors.

3.3 Sensing modalities and data sources

In the studies reviewed, sensing modalities became more complex and integrated streams of data than single-source data. The first methods were mainly based on sensor data related to activities to identify depressive conditions, and it was proven that motion patterns can be used as significant predictors of mental health only [19]. More modern studies have implemented multimodal sensing approaches integrating heterogeneous sensor measurements and highly developed encoding methods to more effectively represent the complexity of mental well-being in its daily aspects of life [20]. Passive sensing systematic reviews have also emphasized the fact that multimodal data integration is superior in robustness and contextual awareness to unimodal sensing [21]. These findings indicate that multi-dimensional sensing plans are becoming more popular to embrace the complexity of mental health. Table 3 is the summary of sensing modalities and types of data used in the reviewed studies, differentiating wearable-based sensors and smartphone-based sensors. It shows the continuum of behavioral and physiological indicators of mental health assessment and displays the relative frequency of each modality of sensing.

Table 3. Sensing modalities and data types used

Sensor Type	Device Type	Data Category	Example Features	Representative Studies (IEEE Ref. IDs)
Accelerometer	Smartphone/ Wearable	Activity	Steps, movement intensity	[19], [25]
GPS	Smartphone	Mobility	Location variance, travel distance	[17], [18]
Heart rate sensor	Wearable	Physiological	Heart rate, heart rate variability (HRV)	[22], [26]
Sleep sensor	Wearable	Behavioral	Sleep duration, sleep efficiency	[20], [24]
App usage logs	Smartphone	Behavioral	App frequency, screen time	[17], [28]
Communication logs	Smartphone	Social	Call and SMS frequency	[18]

Source: Compiled by authors.

3.4 Deep learning models for mental health assessment

It is observed in the reviewed literature that a shift is approaching deep learning models that can absorb temporal and nonlinear relationships between sensing data. Attention-based architectures provide an improvement to feature fusion of wearable sensors by prioritizing informative signals to assess mental well-being [22]. Deep learning surveys of stress detection methods assert that these kinds of models perform better than conventional ones when they are used to analyze complex physiological and behavioral data [23]. Studies also suggest that deep learning has since emerged as the leading methodology in the field of mental health monitoring with wearable and smartphone data, owing to its capacity to acquire representations directly out of raw data [24]. Although these models can scale and adapt to assessment, interpretability, cost of computation, and implementations are still a challenge. There is a need to balance the complexity and transparency of models to apply them in clinical practices. The deep learning architectures and explainability methods utilized in the reviewed studies are summarized in Table 4. It indicates the way various modeling and interpretability strategies are implemented in mental health applications to strike a balance between predictive performance and model transparency.

Table 4. Deep learning models and explainability techniques

Deep Learning Model	Explainability Technique	Application Area	Key Outcome	Representative Studies (IEEE Ref. IDs)
CNN	Feature importance analysis	Depression detection	Improved spatial feature learning	[17]
LSTM	Attention weights	Mood prediction	Captured temporal dependencies	[18]
Attention-based model	Attention visualization	Mental well-being assessment	Enhanced interpretability	[22]
Hybrid CNN–LSTM	SHAP	Stress detection	Balanced accuracy and transparency	[26]
Deep autoencoder	Latent feature analysis	Mental health screening	Reduced dimensionality	[28]

Source: Compiled by authors.

3.5 Longitudinal monitoring and explainability

Some of the studies had gone further in terms of model performance to consider real-world practicability and disclosures. The practicality of passive sensing methods was supported by field-based pilot studies that showed that mobile sensing systems could be implemented in normal environments to support depressed individuals [25]. More recent literature proposed explainable deep learning models to overcome issues of transparency and trust, especially in tasks of detecting stress using wearable sensors [26]. Reviews of artificial intelligence in mental health also highlighted the importance of explainability to clinical acceptability, adherence to ethics, and meaningful human-AI interaction [27]. A combination of these findings highlights the increasing awareness of the importance of interpretability along with predictive accuracy in mental health applications. Practically, predictive models may be utilized to assist clinicians in comprehending the contributors to risk forecasts. This knowledge aids in making informed decisions and in seamless integration into the current clinical practice.

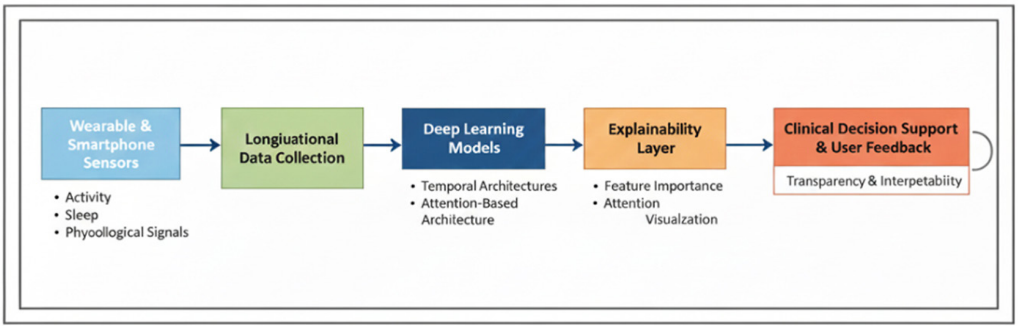

Fig. 2. Role of explainability across the mental health assessment pipeline

Source: Compiled by authors (Synthesized based on concepts from wearable and smartphone sensing, deep learning architectures, and explainable AI frameworks reported in the reviewed literature).

Figure 2 shows the end-to-end process of longitudinal monitoring of mental health with wearable and smartphone sensing alongside explainable deep learning. The collected behavioral and physiological data on mobile and wearable sensors are interpreted in the form of continuous longitudinal data collection. The clinical decision support and user feedback are the final step in the workflow, which focuses on transparency, interpretability, and real-life implementation.

3.6 Human-centered perspectives and emerging gaps

The issues of human-centered design have been increasingly involved in the research of mental health sensing systems. Explainable behavior modeling based on smartphones has been suggested to improve user knowledge and confidence in automated mental health detection models [28]. Further conversations about human-centered artificial intelligence place greater significance on the need to align the explanations of models with user requirements, expectations of privacy, and real-world limitations, especially in mobile health settings [29]. Although deep learning models have delivered very good predictive accuracy in related fields in the medical world, like the diagnosis of neurological diseases, it is difficult to translate these models into clinically actionable, transparent, and interpretable mental health systems [30]. The literature on the topic reveals that there are still many gaps regarding explainability, generalizability, and the possibility of a smooth clinical integration. The communication and contextualization of explanations affect user engagement and acceptance in a strong manner. More intuitive user-clinician and intelligent system interactions can be enabled through human-centered approaches. Feedback mechanisms can also enhance customization and its flexibility. Nonetheless, there is no guarantee that designing explanations with accurate and interpretable capabilities is an easy task. These are some of the issues that should be addressed to ensure the successful implementation of explainable mental health technologies in everyday clinical practice.

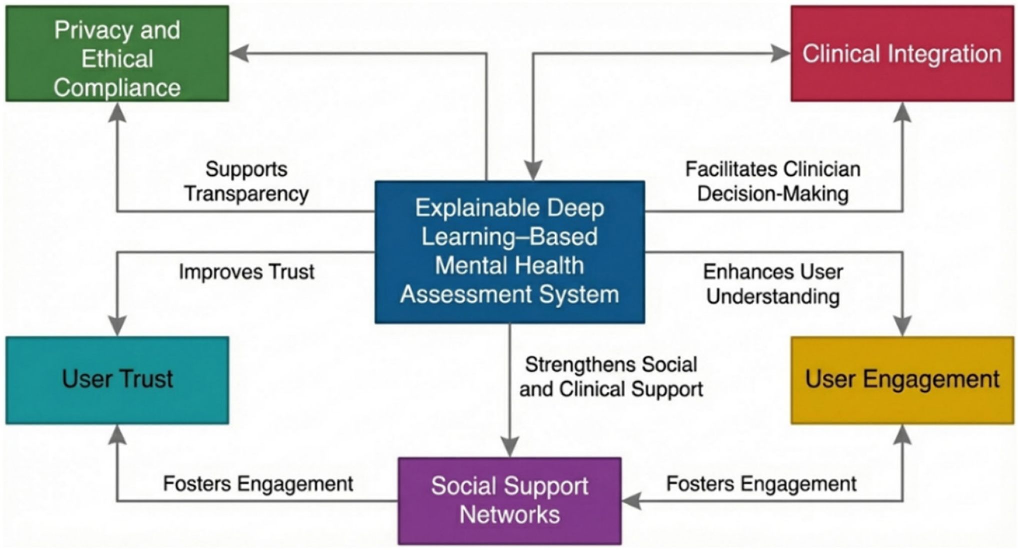

Fig. 3. Human-centered factors influencing explainable mental health systems

Source: Compiled by authors (Synthesized based on concepts from wearable and smartphone sensing, deep learning architectures, and explainable AI frameworks reported in the reviewed literature)

Figure 3 shows a human-oriented, explainable deep learning-based mental health assessment system. It emphasizes the importance of explainability in bridging the gap between the technical outputs and the clinical and user requirements. On the whole, the figure illustrates that explainable systems make it possible to conduct ethical, trustworthy, and clinically integrated mental health assessments.

3.7 Evaluation metrics and validation strategies

In the studies reviewed, various metrics of evaluation and validation strategies were employed in evaluating the performance and reliability of deep learning models in mental health assessment. The majority of studies included standard measures of classification, such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve, which indicate that predictive performance was of high priority. However, there was marked variation in procedures of validation, particularly in how models have been tested through dissimilar populations, sensing circumstances, and deployment situations. Issues regarding the clinical validation and the applicability in the real world were often pointed out, and it was noted that systematic and context-specific validation is necessary within AI systems in healthcare [31]. There were widespread concerns with the generalizability of models, particularly when these models were trained on small-scale or homogeneous data. Mathematical experiments have to demonstrate that methodological biases, including data leakage, unreasonable validation splits, and population bias, may result in overstated performance estimates, which decrease real-life reliability [32]. Also, due to the pervasive nature of class imbalance in mental health data, researchers have focused on more reasonable metrics of evaluation since accuracy, in these cases, can be deceiving [33]. Longitudinal and time-series sensing data make validation more challenging as they inherit the problem of temporal dependencies and missing data, and evaluate potential data generation and augmentation methods to make data more robust [34]. Lastly, multidisciplinary guidelines also focus on open reporting of the evaluation protocols to enable reproducibility and clinical confidence [35].

4 DISCUSSION

According to the results, explainable deep learning models, together with wearable and smartphone sensing data, represent a promising avenue to machine-individualized mental health risk assessment. In the papers under consideration, the learning toward more sophisticated approaches that use multimodal data and time series is evident, instead of the focus on the feasibility of the studies. Yet, with all these technical advances, most of the systems are weak in converting predictions into clinically meaningful activities. This implies that future studies must focus not only on how to increase prediction accuracy but also on how to produce useful insights and a decision support system to make sensing-based mental health applications more useful in real-world settings.

Furthermore, the results indicate that model performance is not enough to identify real values of the work without the evaluation of the usability and interpretation. Explainability integration will help the stakeholders to know more about the behavioral drivers of risk predictions. This kind of knowledge is especially crucial for building trust between clinicians and users. Moreover, the noisy outputs can assist with the individual intervention planning as opposed to generic risk typification. All these points bring up the necessity of the future mental health assessment systems being evaluated in an outcome-oriented manner.

Normative and regulatory considerations turned out to be an important factor that determines the practical implementation of AI-based mental health assessment systems. The ongoing gathering and observation of delicate behavioral and physiological information brings up significant issues associated with privacy, consent, responsibility, and disclosure. The global recommendations regarding artificial intelligence in health highlight that the use of such systems must be explainable and under human control to maintain ethical and credible use of these systems [36]. The ethics also stipulate that transparency is not to be kept as an abstract concept but rather operationalized in the system design and assessment procedures [37]. The regulatory views on medical software based on artificial intelligence and machine learning also support the significance of validation, risk management, and lifecycle monitoring and highlight the importance of regulatory-conscious development of explainable mental health technologies [38]. The results reveal that explainability can be used as a balance between technical innovation and regulatory acceptance. Explicit explanation mechanisms may promote system lifecycle auditability and accountability. Furthermore, regulatory principles, when incorporated at an initial stage of development, can potentially lower obstacles to clinical adoption.

This highlights the need to have interdisciplinary cooperation among the developers, clinicians, and policy stakeholders.

The social support networks are also seen as a key supplement to the technology-based mental health assessment. Digital mental health tools have a greater effect with the help of a human support service, including clinician involvement, caregiver feedback, or peer-based support, especially with diverse and resource-restrained situations [39]. The results of this review indicate that wearable and smartphone sensing systems can enhance social support through the developed systems that allow timely interventions and informed communication. Social context and support structures have, however, not been well integrated into explainable deep learning models. The solution to this gap can lead to improved trust, user engagement, and the overall success of mental health assessment systems, promoting a shift from passive monitoring to action-oriented and supportive care models [40]. The findings also show that explainability may be instrumental in the process of communication in social support networks. Explainable outputs can be used to offer clinicians and caregivers a better perspective on fluctuations in mental health status. It can facilitate faster and more suitable responses to support [41–43]. Also, explanations that are provided to the users can enable individuals to have more active roles in their mental health management. The introduction of social context in model design holds good potential for future research.

5 CONCLUSION

The above papers discussed explainable deep learning models that can assess individual mental health risks based on wearable and smartphone sensing data. The review outlines increasing opportunities of mobile and wearable technologies to assist in continuous and real-world monitoring of mental health because it records behavioral and physiological indicators of daily life activities. The results show that there is a progression in the early feasibility tests to more sophisticated deep learning methods, which can capture the dynamics over time and combine multimodal sensing data to provide more personalized and proactive mental health measurements. The findings show that deep learning models, especially those that include attention patterns and multimodal feature integration, are now the most commonly used tools of analysis in the field because of their ability to process high-dimensional and complex data. Nevertheless, the lack of integration of explainability, regardless of better predictive performance, is still a major obstacle to clinical usage. The review supports the fact that interpretability, transparency, and usability are essential in establishing trust between the clinicians, the user, and other stakeholders, and usability in achieving meaningful implementation in the actual mental health care setting. Besides, the findings confirm the assumption that aligning technical progress with the level of regulation and ethics is essential, and it is of particular importance in the case of mental health information, which is particularly delicate. The social support networks were another critical but less studied element that explained why explainable AI with the support systems based on people could be more helpful regarding engagement and long-term success. Overall, the review provides a structured overview of the current practices, shows the critical gaps in the methodology and practice, and the future directions of the research. In order to transform technological innovation into ethical, clinically significant mental health care, explainable, robust, and socially integrated mental health assessment systems will have to be developed.

6 REFERENCES

[1] World Health Organization, "World mental health report: Transforming mental health for all," Mental Health, Brain Health and Substance Use (MSD), p. 296, 2022. [Online]. Available: https://www.who.int/publications/i/item/9789240049338 [Accessed: Mar. 5, 2026].

[2] T. R. Insel, "Digital phenotyping: Technology for a new science of behavior," *JAMA*, vol. 318, no. 13, pp. 1215–1216, 2017. https://doi.org/10.1001/jama.2017.11295

[3] J. P. Onnela and S. L. Rauch, "Harnessing smartphone-based digital phenotyping to enhance behavioural and mental health," *Neuropsychopharmacology*, vol. 41, no. 7, pp. 1691–1696, 2016. https://doi.org/10.1038/npp.2016.7

[4] S. Saeb *et al.*, "Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study," *Journal of Medical Internet Research*, vol. 17, no. 7, p. e4273, 2015. https://doi.org/10.2196/jmir.4273

[5] N. C. Jacobson and Y. J. Chung, "Passive sensing of moment-to-moment depressed mood among undergraduates with clinical levels of depression using smartphones," *Sensors*, vol. 20, no. 12, p. 3572, 2020. https://doi.org/10.3390/s20123572

[6] V. P. Cornet and R. J. Holden, "Systematic review of smartphone-based passive sensing for health and wellbeing," *Journal of Biomedical Informatics*, vol. 77, pp. 120–132, 2018. https://doi.org/10.1016/j.jbi.2017.12.008

[7] 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. https://doi.org/10.1017/S0033291719000151

[8] Q. J. M. Huys, T. V. Maia, and M. P. Paulus, "Computational psychiatry: From mechanistic insights to the development of new treatments," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 1, no. 5, pp. 382–385, 2016. [Online]. Available: https://www.biologicalpsychiatrycnni.org/article/S2451-9022(16)30082-9/abstract. [Accessed: Mar. 6, 2026].

[9] R. Razavi, A. Gharipour, and M. Gharipour, "Depression screening using mobile phone usage metadata: A machine learning approach," *Journal of the American Medical Informatics Association*, vol. 27, no. 4, pp. 522–530, 2020. https://doi.org/10.1093/jamia/ocz221

[10] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017. https://doi.org/10.48550/arXiv.1702.08608

[11] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144. https://doi.org/10.1145/2939672.2939778

[12] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf [Accessed: Mar. 7, 2026].

[13] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable AI systems for the medical domain?" *arXiv preprint arXiv:1712.09923*, 2017. https://doi.org/10.48550/arXiv.1712.09923

[14] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, "What clinicians want: Contextualising explainable machine learning for clinical end use," in *Proceedings of the 4th Machine Learning for Healthcare Conference, PMLR 106*, 2019, pp. 359–380. [Online]. Available: https://proceedings.mlr.press/v106/tonekaboni19a/tonekaboni19a.pdf [Accessed: Mar. 5, 2026].

[15] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key challenges for delivering clinical impact with artificial intelligence," *BMC Medicine*, vol. 17, no. 1, p. 195, 2019. https://doi.org/10.1186/s12916-019-1426-2

[16] Z. C. Lipton, "The mythos of model interpretability," *Queue*, vol. 16, no. 3, pp. 31–57, 2018. [Online]. Available: https://spawn-queue.acm.org/doi/pdf/10.1145/3236386.3241340 [Accessed: Mar. 5, 2026].

[17] R. A. Rahman *et al.*, "Application of machine learning methods in mental health detection: A systematic review," *IEEE Access*, vol. 8, pp. 183952–183964, 2020. https://doi.org/10.1109/ACCESS.2020.3029154

[18] P. Chikersal *et al.*, "Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing," *ACM Transactions on Computer-Human Interaction*, vol. 28, no. 1, pp. 1–41, 2021. https://doi.org/10.1145/3422821

[19] A. K. M. Masum, M. F. I. Khan, F. Anjum, and S. Alam, "Depression detection through activity recognition: Deep learning models using synthesized sensor data," *Journal of Basic Science and Engineering*, vol. 21, no. 1, pp. 571–590, 2024. [Online]. Available: https://www.researchgate.net/publication/380459231_DEPRESSION_DETECTION_THROUGH_ACTIVITY_RECOGNITION_DEEP_LEARNING_MODELS_USING_SYNTHESIZED_SENSOR_DATA [Accessed: Mar. 5, 2026].

[20] K. Woodward, E. Kanjo, and A. Tsanas, "Combining deep learning with signal-image encoding for multi-modal mental wellbeing classification," *ACM Transactions on Computer-Human Interaction*, vol. 5, no. 1, pp. 1–23, 2024. https://doi.org/10.1145/3631618

[21] L. S. Khoo, M. K. Lim, C. Y. Chong, and R. McNaney, "Machine learning for multimodal mental health detection: A systematic review of passive sensing approaches," *Sensors*, vol. 24, no. 2, p. 348, 2024. https://doi.org/10.3390/s24020348

[22] J. Jin *et al.*, "Attention-block deep learning based feature fusion in wearable social sensors for mental wellbeing evaluation," *IEEE Access*, vol. 8, pp. 89258–89268, 2020. https://doi.org/10.1109/ACCESS.2020.2994124

[23] M. Kyrou, I. Kompatsiaris, and P. C. Petrantonakis, "Deep learning approaches for stress detection: A survey," *IEEE Transactions on Affective Computing*, vol. 16, no. 2, pp. 499–517, 2024. https://doi.org/10.1109/TAFFC.2024.3455371

[24] S. Shen *et al.*, "Passive sensing for mental health monitoring using machine learning with wearables and smartphones: A scoping review," *Journal of Medical Internet Research*, vol. 27, p. e77066, 2025. https://doi.org/10.2196/77066

[25] F. Wahle *et al.*, "Mobile sensing and support for people with depression: A pilot trial in the wild," *JMIR mHealth and uHealth*, vol. 4, no. 3, p. e5960, 2016. https://doi.org/10.2196/mhealth.5960

[26] M. K. Moser, M. Ehrhart, and B. Resch, "An explainable deep learning approach for stress detection in wearable sensor measurements," *Sensors*, vol. 24, no. 16, p. 5085, 2024. https://doi.org/10.3390/s24165085

[27] M. Ali *et al.*, "Artificial intelligence for mental health: A narrative review of applications, challenges, and future directions," *Digital Health*, vol. 11, 2025. https://doi.org/10.1177/20552076251395548

[28] H. Lee, T. Park, and U. Lee, "Smartphone-based human behaviour task modelling for explainable mental health detection," in *Proceedings Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2024, pp. 894–896. https://doi.org/10.1145/3675094.3679003

[29] T. Dang, D. Spathis, A. Ghosh, and C. Mascolo, "Human-centred artificial intelligence for mobile health sensing: Challenges and opportunities," *Royal Society Open Science*, vol. 10, no. 11, p. 230806, 2023. https://doi.org/10.1098/rsos.230806

[30] Y. Ding *et al.*, "A deep learning model to predict a diagnosis of Alzheimer's disease using 18F-FDG PET," *Radiology*, vol. 290, no. 2, pp. 456–464, 2019. https://doi.org/10.1148/radiol.2018180958

[31] S. Nabhani-Gebara *et al.*, "Clinical validation in AI for healthcare: An experiential learning," in *Trustworthy AI in Cancer Imaging Research*, 2025, pp. 267–282. https://doi.org/10.1007/978-3-031-89963-8_12

[32] F. Maleki *et al.*, "Generalizability of machine learning models: Quantitative evaluation of three methodological pitfalls," *Radiology: Artificial Intelligence*, vol. 5, no. 1, p. e220028, 2022. https://doi.org/10.1148/ryai.220028

[33] M. Owusu-Adjei, J. B. Hayfron-Acquah, T. Frimpong, and G. Abdul-Salaam, "Imbalanced class distribution and performance evaluation metrics in healthcare systems," *PLoS Digital Health*, vol. 2, no. 11, p. e0000290, 2023. https://doi.org/10.1371/journal.pdig.0000290

[34] B. M. Maweu, R. Shamsuddin, S. Dakshit, and B. Prabhakaran, "Generating healthcare time-series data for improving diagnostic accuracy of deep neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–5, 2021. https://doi.org/10.1109/TIM.2021.3077049

[35] W. Luo *et al.*, "Guidelines for developing and reporting machine learning predictive models in biomedical research," *Journal of Medical Internet Research*, vol. 18, no. 12, p. e323, 2016. https://doi.org/10.2196/jmir.5870

[36] World Health Organization, "Ethics and governance of artificial intelligence for health: Executive summary," World Health Organization, Health Ethics & Governance (HEG), 2021. [Online]. Available: https://www.who.int/publications/i/item/9789240037403 [Accessed: Mar. 5, 2026].

[37] N. Balasubramaniam, M. Kauppinen, K. Hiekkanen, and S. Kujala, "Transparency and explainability of AI systems: Ethical guidelines in practice," in *International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ)*, 2022, pp. 3–18. [Online]. Available: https://doi.org/10.1007/978-3-030-98464-9_1 [Accessed: Mar. 7, 2026].

[38] Food and Drug Administration, "Artificial intelligence and machine learning in software as a medical device," Food and Drug Administration, 2021. [Online]. Available: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device [Accessed: Mar. 8, 2026].

[39] J. A. Naslund *et al.*, "Digital technology for treating and preventing mental disorders in low-income and middle-income countries," *The Lancet Psychiatry*, vol. 4, no. 6, pp. 486–500, 2017. https://doi.org/10.1016/S2215-0366(17)30096-2

[40] D. A. Adler *et al.*, "Beyond detection: Towards actionable sensing research in clinical mental healthcare," *Proceedings of the ACM Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 4, pp. 1–33, 2024. https://doi.org/10.1145/3699755

[41] A. Kumar and P. Banerjee, "Leading with intelligence: How AI is reshaping creative innovation in organizations," *Prabandhan: Indian Journal of Management*, vol. 18, no. 9, pp. 54–62, 2025. https://doi.org/10.17010/pijom/2025/v18i9/174843

[42] H. Sikandar, A. F. Abbas, N. Khan, and M. I. Qureshi, "Digital technologies in healthcare: A systematic review and bibliometric analysis," *International Journal of Online and Biomedical Engineering*, vol. 18, no. 8, pp. 34–48, 2022. https://doi.org/10.3991/ijoe.v18i08.31961

[43] A. U. Adoghe, E. Noma-Osaghae, and R. I. Yabkwa, "Photonic crystal and its application as a biosensor for the early detection of cancerous cells," *International Journal of Online and Biomedical Engineering*, vol. 16, no. 3, pp. 86–94, 2020. https://doi.org/10.3991/ijoe.v16i03.12523

7 AUTHORS

Krisna Veni Balakrishnan is a Doctorate Candidate in Business Administration (Healthcare Management). She is a healthcare professional with over 17 years of academic teaching experience and 11 years of expertise in clinical trial management. Currently pursuing a Doctorate in Business Administration (Healthcare Management) at Rushford Business School, she serves as the Head of the Clinical Trial Unit in Malaysia's private healthcare sector. She is certified in Good Clinical Practice by the Ministry of Health Malaysia, she led clinical research initiatives with a strong focus on compliance, ethical standards, and innovation (E-mail: dba1055@rushford.eu).

Dr. Geetika Parmar holds a Ph.D. in Business Analytics and currently serves as an Assistant Professor in the Department of Computer Science and Applications at Dr. Vishwanath Karad MIT World Peace University, Pune, India. Her current research interests include Industry 4.0 with publications on big data analytics, HR analytics, sustainable development, Cloud Computing, Artificial Intelligence and Machine Learning in journals of good repute. Her expertise lies in integrating cutting-edge analytics with practical pedagogy, and she strongly advocates continuous learning for academic and personal growth (E-mail: geetika.parmar@mitwpu.edu.in).