Global Journal of Modeling and Intelligent Computing (GJMIC)

Vol. 3, No. 2, July, 2023 ISSN: 2767-1917

A SYSTEMATIC REVIEW OF MULTIMODAL MACHINE LEARNING APPLICATIONS IN MENTAL HEALTH

Bhushan Kumar Kashyap, Atal Bihari Vajpayee Vishwavidyalaya, India (bkk.csa@gmail.com)
Richa Handa, Atal Bihari Vajpayee Vishwavidyalaya, India (proffhanda@gmail.com)
H. S. Hota, Atal Bihari Vajpayee Vishwavidyalaya, (proffhota@gmail.com)

ABSTRACT

Mental health research is increasingly complex, requiring the integration of diverse data sources to gain comprehensive insights. This paper presents a systematic review of Multimodal Machine Learning (MML) applications in the field of mental health. By examining studies that leverage multiple data modalities such as text, audio, video, physiological signals and analysing that how their combined approach enhance the management of mental health issues by its diagnosis, treatment, and prevention. The findings of this review contribute to the advancement of mental health research by providing a comprehensive overview of the current state-of-the-art and identifying promising directions for future research in this rapidly evolving area.

Keywords: Multimodal, Machine Learning, Mental Health, data modalities.

1. INTRODUCTION

Mental health disorders, such as depression, anxiety, and stress, are among the most pressing global health concerns, affecting over 970 million individuals worldwide (World Health Organization [WHO], 2021). The economic and social burden of these conditions is immense, with untreated mental health issues often leading to reduced productivity, strained healthcare systems, and diminished quality of life.

Artificial intelligence (AI) aims to mimic human cognitive functions. Artificial Intelligence (AI), and machine learning in particular, has radically altered our interactions with the world, fostering rapid advancements in various domains. It is bringing a paradigm shift to healthcare, powered by increasing availability of (Nojavanasghari et al., 2016) healthcare data and rapid progress of analytics techniques(Jiang et al., 2017). However, the adoption of machine learning approaches in healthcare has been slower than in other fields despite the increasing pressure on healthcare systems and the urgent demand for high quality, personalised care(Kirch & Petelle, 2017; Topol, 2019). Despite the growing awareness, mental health detection remains challenging due to reliance on subjective assessments and single-modal diagnostic tools, which are prone to biases and inconsistencies (Garcia-Ceja et al., 2018; Priya et al., 2020). Single modality cannot capture the full context of the scenario like in textual data don't have visual and auditory cues, it leads to incomplete comprehension. Multimodal machine learning (ML) systems have emerged as a promising approach to address these limitations by integrating data from multiple sources, such as text, audio, video, and physiological signals, to provide a more comprehensive and reliable mental health assessment(Nojavanasghari et al., 2016). Each modality provides distinct perspectives regarding an individual's psychological status, facilitating a comprehensive comprehension of intricate mental health disorders. For instance, textual data from social media posts or self-reported questionnaires can reveal cognitive and emotional patterns. Audio features, such as pitch and tone, provide vocal markers of anxiety and stress (Czyzewski et al., 2017). Similarly, visual data, including facial expressions and body gestures, capture non-verbal cues critical for diagnosing emotional states (Monisha et al., 2022)Physiological signals, such as electroencephalograms (EEG), heart rate variability, and galvanic skin response, add an objective dimension to the assessment by directly measuring biological responses associated with mental health disorders(Dura & Wosiak, 2021; Pisanski et al., 2018).

Text data like medical history, questionnaire, survey plays an important role in mental health detection same like this tabular like pathological reports are also useful in mental health detection. Audio data allows to analyze some acoustic features like pitch, speech rate, tone and intonation these are helpful to know mental health conditions include depression, anxiety, bipolar disorder etc. Video data possesses considerable utility in the identification of mental health conditions, as it facilitates the observation and analysis of non-verbal indicators such as facial expressions, bodily movements, and speech patterns by researchers and clinicians. An image is also useful in mental health detection, most medical images are stored in 2D and 3D slices. From facial expression mental status can be detected. In terms of medical science, an image is a representation of internal side of the body that is

captured or created using various technologies like X-ray imaging, Magnetic resonance imaging (MRI), Computed tomography (CT) these are also called radiology. Physiological data is a collection of measurement that can used to monitor persons mental and physical health like Electroencephalogram (EEG), Pupil diameter, Body temperature, Electromyogram (EMG), Breathing rate, Electrocardiogram (ECG), Skin conductance, Pulse rate, Skin color, and Perspiration.

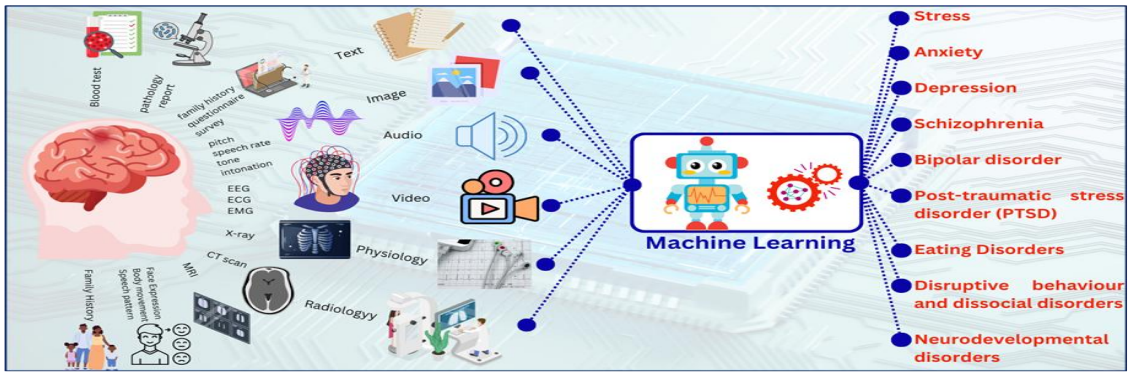

Figure 1: Multimodal Machine Learning for Mental Health Assessment and Disorder Detection

The integration of these modalities into a unified framework requires advanced machine learning techniques capable of processing and fusing heterogeneous data. Classical algorithms, such as support vector machines (SVMs) and random forests, have been widely employed for feature classification (Priya et al., 2020). However, the advent of deep learning has revolutionized multimodal analysis by enabling the automatic extraction of complex patterns through architectures like convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and transformer models. Fusion strategies, categorized as early (feature-level), intermediate (representation-level), and late (decision-level), further enhance the robustness of multimodal predictions by allowing seamless integration of diverse data sources

Despite the promise of multimodal systems, their development and deployment face several challenges. One of the primary concerns is data privacy, particularly given the sensitivity of mental health information. Ethical considerations surrounding informed consent and data ownership are critical to ensuring user trust and regulatory compliance. Additionally, the synchronization of data from multiple modalities in real time poses significant technical hurdles, particularly in dynamic and resource-constrained settings(Garcia-Ceja et al., 2018). Furthermore, the “black-box” nature of many ML models limits their interpretability, making it difficult for clinicians to trust and adopt these systems in practice (Shrotri et al., 2022).

The motivation for developing a multimodal machine learning system for mental health detection is rooted in addressing these challenges while leveraging the strengths of multimodal analysis. By combining diverse modalities, such a system can emulate the comprehensive diagnostic methods used by mental health professionals, thereby improving diagnostic accuracy and enabling early intervention(Rizzo et al., 2016; Zhang et al., 2020). Moreover, advancements in explainable AI (XAI) offer opportunities to enhance the transparency and usability of these systems in clinical settings (Thalpage, 2023).

2. DETAILS OF SOURCE OF RESEARCH ARTICLE

In this literature review research articles and data taken from standard databases, journals, proceeding of national and international conference, websites etc. distribution of source is represented in the first pie chart (Figure 2). The sources of research articles used in this review. It reveals that the majority of articles were sourced from IEEE and ScienceDirect, each contributing 21.4% of the total. This is followed by PubMed at 16.7% and Elsevier at 9.5%. Other sources, including ACM Digital Library, arXiv, Nature, and others, contributed smaller percentages ranging from 7.1% to 2.4%. This diverse range of sources underscores the interdisciplinary nature of the research on this topic, drawing from both medical and technical domains.

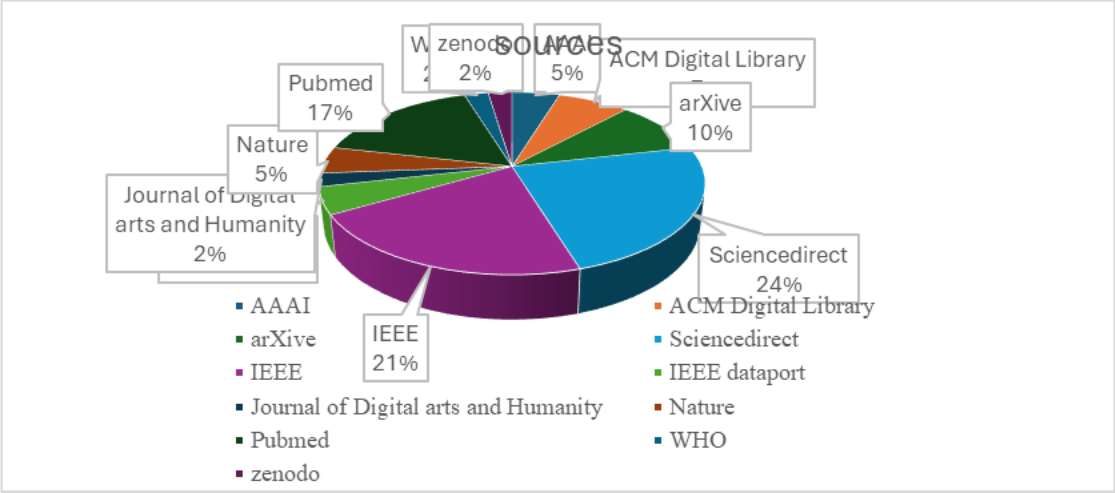

Figure 2: Distribution of source of literatures.

The bar chart (Figure 3) highlights the year-wise distribution of research articles related to "multimodal-based machine learning systems for mental health detection." The data shows an increasing trend in the number of articles published over the years, with a notable peak in 2022, where more than 10 articles were recorded. After 2022, a slight decline in publications is observed for 2023 and 2024. However, the "not specified" category also accounts for a significant number of articles, indicating that some publication years were not documented

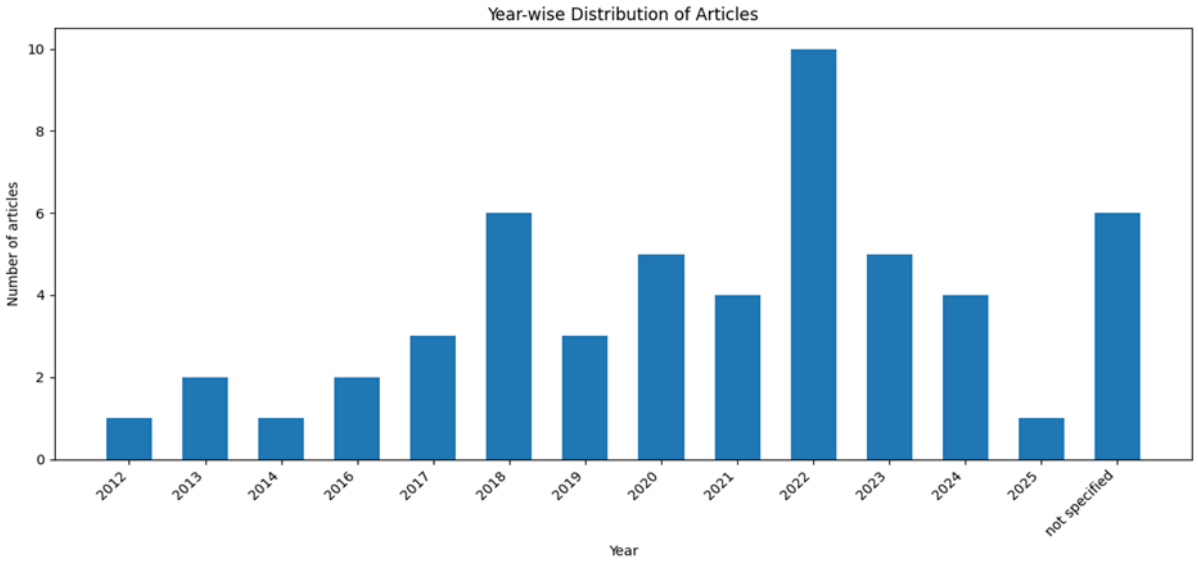

Figure 3: Year wise distribution of published research articles.

3. LITERATURE REVIEW

In this section, we present an overview of the available multimodal mental health datasets that are currently available (refer table 1). The mental health disorders examined encompass depression, post-traumatic stress disorder (PTSD), bipolar disorder, stress responses, and emotional recognition (refer to Table 1). Nevertheless, specific mental health conditions, such as schizophrenia. Furthermore, we presented data regarding gender distribution and the geographic locations from which the data was sourced, in order to assist in reducing bias and fostering equity in research endeavors. It is also crucial to acknowledge that the datasets incorporated in this analysis are subject to access limitations due to concerns regarding privacy and ethical implications(Al Sahili et al., n.d.). It is essential that any research undertaken utilizing these datasets is conducted in accordance with rigorous ethical standards, with the aim of enhancing mental health outcomes for all individuals. The datasets encompass a diverse array of mental health disorders, with a significant emphasis on depression and stress, which represent the most frequently addressed conditions. Specifically, seven datasets pertain to depression, three datasets concentrate on post-traumatic stress disorder (PTSD), ten datasets examine stress, two datasets investigate

bipolar disorder, one dataset pertains to behavioural disorders, and six datasets focus on emotion recognition (Figure 4).

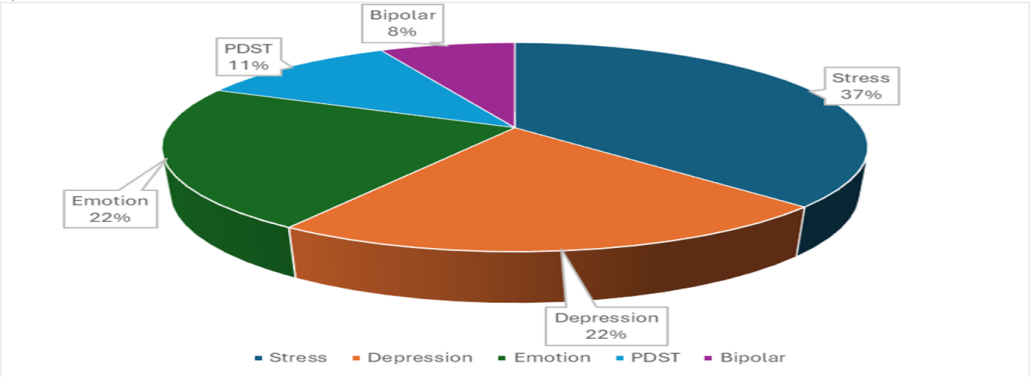

Figure 4: Distribution of various mental health condition in this review.

This predominance of datasets concerning depression and stress signifies its substantial impact on global mental health. Regarding data modalities, the majority of datasets incorporate both video and audio data, with 16 datasets dedicated to each modality. Textual data is represented in 11 datasets, while physiological data appears in 13 datasets (Figure 5).

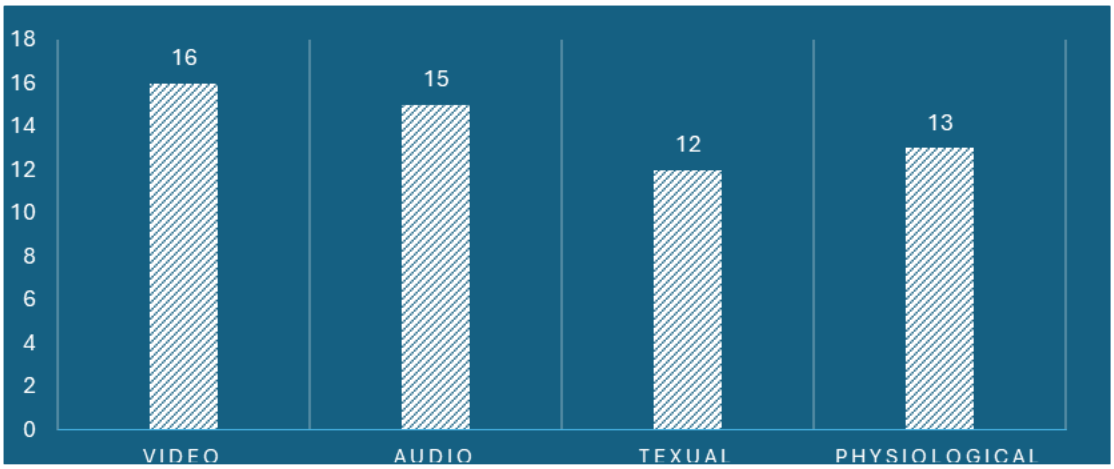

Figure 5: Modality wise data distribution in dataset study

The amalgamation of these diverse modalities enhances the robustness and precision of machine learning models utilized in mental health research. Nonetheless, the integration of physiological data with other modalities remains comparatively infrequent.

The amalgamation of diverse modalities is prevalent, thereby enhancing the robustness of machine learning models. However, the utilization of physiological data is comparatively infrequent notwithstanding its prospective advantages. In the analysis of multimodal integrations, six datasets encompass video, audio, and text; an additional six datasets comprise video and audio. Four datasets are exclusively dedicated to physiological data, whereas two datasets each amalgamate text and physiology, video and physiology, audio and text and physiology, along with video and text and physiology. Furthermore, one dataset each incorporates audio and physiology, video, audio, and physiology, and video, text, and physiology. Additionally, the datasets exhibit substantial variability in both size and composition, ranging from extensive datasets such as E-DAIC to more focused datasets like the Chinese Multimodal Depression Corpus. The examination of these 24 datasets underscores the pivotal significance of multimodal data in propelling advancements in mental health research through machine learning methodologies. Despite considerable advancements, there persists a necessity for more extensive and heterogeneous datasets to fully actualize the potential of these technological innovations. By furnishing a thorough comprehension of the intricate nature of mental health disorders, these datasets possess the capacity to catalyze significant progress within the domain, ultimately enhancing outcomes for individuals grappling with mental health challenges. To augment the utility of forthcoming datasets, it is imperative to incorporate a broader spectrum of demographic information, thereby ensuring that models can generalize effectively across diverse populations. Moreover,

guaranteeing that datasets are both extensive and varied could bolster the generalizability of models. Furthermore, broadening the inclusion of physiological data may yield additional insights, potentially resulting in more precise mental health evaluations.

Table 1: Details of multimodal mental health datasets

Reference	Dataset	Disorder	Number of modality	video	audio	text	physiology	Number of physiology	gender	country
(Rizzo et al., 2016)	DAIC-WOZ	depression PTSD	3	√	√	√	-	-	-	-
(Ringeval et al., 2018)	E-DAIC dataset	depression PTSD	3	√	√	√	-	-	-	-
(Valstar et al., 2013)	AVEC 2013	depression	2	√	√				-	-
(Cai et al., 2022)	A multi-modal open dataset for mental disorder analysis	depression	2			√	√	1	-	china
(Koldijk et al., 2014)	SWELL	stress	5	-	-	-	√	5	32% Female	Netherland
(Yoon et al., 2022)	D -vlog	depression	2	√	√				-	-
(Zou, 2022)	Chinese Multimodal Depression Corpus	depression	2	√	√	-	-	-	-	china
(Ringeval & Multimedia, 2018)	Bipolar Disorder and Cross - Cultural Affect Recognition	bipolar/ behavior	3	√	√	-	√	1	-	Turkey+ Germany+ Hunagry+ France
(Kaya et al., n.d.)	Turkish Bipolar Disorder Corpus	bipolar	2	√	√				31% Female	Turkey
(Sawadogo et al., 2023)	PTSD in the Wil	PTSD	3	√	√	√	-	-	-	-
(Kutt et al., 2022)	BIRAFFE 2	Emotion	3			√	√	2	33% Female	Poland
(Koelstra et al., 2012)	Deap	Emotion	3	√	-	-	√	2	-	-
(Jaiswal et al., 2020)	MuSE	stress	5	√		√	√	3	32% F	US
(Panda et al., 2013)	MIREX	Emotion	3	√	√	√	-	-	-	-
(Poria et al., n.d.)	MELD	Emotion	3	√	√	√	-	-	-	-

(Miranda et al., 2022)	WEMAC	Emotion	5		√	√	√	3	100 %	Spain
(Nojavanasghari et al., 2016)	EmoReact	Emotion	2	-	-	-	-	-	-	-
(Schmidt et al., 2018)	WESAD	stress and affect	3	-	-	-	√	3	-	-
(S. Hosseini et al., 2022)	A multimodal sensor dataset for continuous stress detection of nurses in a hospital	stress	6	-	-	-	√	6	-	-
(Stappen et. al.,2021)	MuSe-Stress	stress	2	√	√	-	-	-	69.9 % Female	-
(Pisanski et al., 2018)	Multimodal stress detection	stress	6		√	√	√	4	58.7 % Female	
(M. Hosseini et al., n.d.)	Empathic School	stress	4	√	-	-	√	3	-	Finland and US
(Lin et al., 2020)	Automatic Detection of Self-Adaptors for Psychological Distress	stress	3	√	√	√	-	-	-	-
	CLAS	stress	5			√	√	4	-	-
(S. Hosseini et al., 2022)	MuSe2022	stress	2	√	√	-	-	-	-	-
(Meziati et al., 2021)	UBFC-PHYS	stress	2	-	-	-	√	2	-	-

Recent advancements in machine learning (ML) have significantly enhanced mental health detection through the integration of multimodal data, including text, audio, visual, physiological signals, and structured data. Textual data, such as clinical notes, social media posts, and conversational transcripts, are analyzed using natural language processing (NLP) techniques to identify indicators of depression, anxiety, and stress)(Dixit et al., 2023). Audio data, particularly speech, provides insights into an individual's emotional state through acoustic features like pitch, prosody, and voice quality, with studies highlighting these features as biomarkers for mental health conditions(Ringeval et al., 2018; Schuller, 2021). Visual data, including facial expressions, eye gaze, and body movements, are critical for understanding emotional states, with advancements in computer vision enabling the analysis of microexpressions and gestures for detecting depression and anxiety(Sun et al., 2020). Physiological signals, such as heart rate variability (HRV), electrodermal activity (EDA), and electroencephalography (EEG), provide objective measures of mental health, often linked to stress and anxiety levels. These signals, captured through wearable devices or medical sensors, have been widely used for mental state classification(Bruin et al., 2024). Structured data, including demographic information, clinical questionnaires, and medical histories, often serve as complementary features, improving the contextual understanding and predictive accuracy of ML models(Su et al., 2020). Combining these modalities has led to more robust systems, with multimodal approaches outperforming unimodal systems by leveraging the strengths of each modality to provide a comprehensive view

of an individual's mental state. However, challenges such as data privacy concerns, limited availability of diverse datasets, and the complexity of multimodal data integration persist, emphasizing the need for future research to develop ethical frameworks, standardized datasets, and improved fusion techniques to enhance the effectiveness of ML applications in mental health.

Feature extraction is a critical step in machine learning applications for mental health detection, as it transforms raw multimodal data into meaningful representations that facilitate accurate classification and prediction. In text-based analyses, techniques such as term frequency-inverse document frequency (TF-IDF) and word embeddings like Word2Vec, GloVe, and BERT are employed to capture semantic nuances indicative of mental health conditions(Alfarizi et al., 2022). For audio data, acoustic features such as pitch, energy, and spectral properties, extracted using tools like Praat or OpenSMILE, are linked to speech patterns associated with disorders like depression and anxiety(Ringeval et al., 2018; Valstar et al., 2013). Visual data analysis involves extracting facial landmarks, action units, and micro-expressions using techniques like OpenFace or convolutional neural networks (CNNs)(Pham & Won, 2019) to identify emotional states. Physiological signals such as electroencephalography (EEG) and electrocardiography (ECG) are used to extract time-domain and frequency-domain features, such as power spectral density and heart rate variability, to assess mental health status(Alhussein et al., 2019). Structured data, such as demographic and clinical information, is processed to extract relevant features that complement other modalities(Goncalves & Busso, 2022). Deep learning approaches, including CNNs and recurrent neural networks (RNNs), have enabled automatic feature extraction from raw data, learning hierarchical representations in multimodal contexts (Yang et al., 2021). Advanced methods like feature fusion integrate multimodal features, enhancing the robustness of mental health detection systems through early fusion, late fusion, and hybrid techniques (Majumder et al., 2018). Despite these advancements, challenges such as feature selection, dimensionality reduction, and interpretability persist, requiring further research to optimize feature extraction processes for multimodal mental health.

Machine learning algorithms used to improve diagnostic accuracy and provide a holistic view of mental health conditions. Traditional machine learning models, such as Support Vector Machines (SVMs) and Random Forests, have been applied to classify mental health states based on features derived from single and multiple modalities (Al Sahili et al., n.d.; Ringeval et al., 2018). However, deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have revolutionized this field by enabling automated feature extraction and capturing temporal dependencies in complex datasets(Pi, 2024; Thomson & Emery, 2024). Transformers, including BERT and multimodal extensions, have further advanced textual and multimodal mental health detection by leveraging attention mechanisms to integrate multiple data streams(Makhmudov et al., 2024; Verma et al., 2023). For instance, integrating textual data with acoustic and visual features improves detection accuracy in disorders like depression and anxiety (Poria et al., n.d.). Advanced multimodal fusion strategies—such as early, late, and hybrid fusion—allow models to combine complementary information from various modalities, enhancing performance (Nojavanasghari et al., 2016). Despite these advancements, challenges remain, including the lack of large annotated datasets, heterogeneity in data sources, and interpretability of deep models. Ongoing research focuses on addressing these limitations by developing scalable, explainable, and robust systems to detect and monitor mental health conditions effectively (Chuang et al., 2023)

Multimodal mental health detection systems leverage fusion techniques to integrate diverse data modalities Early fusion combines raw data from multiple modalities at the input level, enabling models to learn joint representations; however, it struggles with heterogeneous data types and temporal synchronization (Baltrušaitis et al., 2019; Poria et al., 2017). Late fusion integrates unimodal predictions at the decision level, offering flexibility but potentially missing inter-modal correlations (Xie et al., 2022). Hybrid fusion, which combines early and late fusion strategies, captures both low- and high-level interactions among modalities, improving performance in mental health detection (Qi et al., 2025). Advanced methods, such as tensor fusion networks and attention mechanisms, dynamically weigh the contribution of each modality, addressing issues of modality relevance, noise, and missing data(Majumder et al., 2018). Graph-based approaches have also been introduced to model complex modality interdependencies. Despite these advancements, challenges like missing modalities, limited annotated multimodal datasets, and the interpretability of fusion-based models persist (Koelstra et al., 2012) Ongoing research explores strategies such as hierarchical fusion, robust handling of missing data, and incorporating self-supervised learning to improve scalability and robustness in mental health detection systems (Baltrusaitis et al., 2018; Poria et al., 2020).

4. RESEARCH GAP

Multimodal machine learning systems for mental health detection face several critical research gaps. One major limitation is the insufficient integration of imaging and non-imaging data, such as MRI, CT scans, blood tests, and genetic markers, which restricts comprehensive assessments. Additionally, there is a lack of longitudinal studies exploring long-term changes in mental health conditions using multimodal medical datasets. Advanced fusion techniques capable of handling and combining high-dimensional data from sources like MRI, CT, and EEG are underdeveloped, further impeding progress. Moreover, the integration of diverse medical imaging and signal data, such as EEG, MRI, and CT scans, remains underexplored, hindering holistic mental health assessments. Another significant gap lies in the underutilization of multimodal data for AI explainability, as limited efforts have been made to interpret AI models combining text sentiment, vocal tone, and facial emotions. Text-based systems also neglect non-linguistic features like punctuation, emojis, or typing patterns, which could provide valuable insights. Furthermore, current systems lack emotional granularity, often classifying emotions broadly (e.g., happy, sad) while failing to capture nuanced emotions such as guilt or boredom. Addressing these gaps could lead to more robust, comprehensive, and interpretable multimodal systems for mental health detection.

5. TENTATIVE OBJECTIVES

The primary objective of this research is to collect data across various modalities, such as text, images, audio, video, physiological signals, and radiological data, from both primary and secondary sources, depending on their availability, accessibility, and relevance to the study. This will be followed by the implementation of optimized multimodal fusion techniques to effectively integrate information from these diverse data sources. Finally, the goal is to develop a machine learning system that leverages this multimodal data to detect and assess mental health conditions, offering a more comprehensive and accurate understanding of mental health states.

6. CONCLUSION

In conclusion, the development of multimodal machine learning systems marks a significant advancement in mental health diagnostics, addressing the limitations of traditional unimodal approaches. By integrating modalities such as text, audio, video, physiological signals, and imaging data, these systems emulate the comprehensive diagnostic methods used by clinicians, enabling more accurate and detailed mental health assessments. Technological innovations, particularly in fusion strategies like early, late, and hybrid fusion, have further enhanced the ability to process and integrate diverse data modalities, improving the precision and reliability of mental health detection systems.

However, challenges remain, including the need for diverse and extensive datasets, better synchronization of multimodal data, and improved interpretability of machine learning models. Overcoming these hurdles will be critical to enabling the widespread adoption of these systems in clinical and non-clinical settings. Ethical and privacy considerations must also take center stage, with robust frameworks to ensure informed consent, safeguard sensitive information, and build user trust.

Looking ahead, multimodal systems hold immense potential to transform mental healthcare globally by enabling early diagnosis, personalized interventions, and real-time monitoring. These advancements promise not only to improve individual outcomes but also to reduce the overall burden of mental health disorders on healthcare systems worldwide, paving the way for a more inclusive and effective approach to mental healthcare.

REFERENCES

Al Sahili, Z., Patras, I., & Purver, M. (n.d.). Multimodal Machine Learning in Mental Health: A Survey of Data, Algorithms, and Challenges.

Alfarizi, M. I., Syafaah, L., & Lestandy, M. (2022). Emotional Text Classification Using TF-IDF (Term Frequency-Inverse Document Frequency) And LSTM (Long Short-Term Memory). JUITA : Jurnal Informatika, 10(2), 225. https://doi.org/10.30595/juita.v10i2.13262

Alhussein, M., Muhammad, G., & Hossain, M. S. (2019). EEG Pathology detection based on deep learning. IEEE Access, 7, 27781–27788. https://doi.org/10.1109/ACCESS.2019.2901672

Bruin, W. B., Oltedal, L., Bartsch, H., Abbott, C., Argyelan, M., Barbour, T., Camprodon, J., Chowdhury, S., Espinoza, R., Mulders, P., Narr, K., Oudega, M., Rhebergen, D., Ten Doesschate, F., Tendolkar, I., Van Eijndhoven, P., Van Exel, E., Van Verseveld, M., Wade, B., … Van Wingen, G. (2024). Development and validation of a multimodal neuroimaging biomarker for electroconvulsive therapy outcome in depression:

A multicenter machine learning analysis. Psychological Medicine, 54(3), 495–506. https://doi.org/10.1017/S0033291723002040

Cai, H., Yuan, Z., Gao, Y., Sun, S., Li, N., Tian, F., Xiao, H., Li, J., Yang, Z., Li, X., Zhao, Q., Liu, Z., Yao, Z., Yang, M., Peng, H., Zhu, J., Zhang, X., Gao, G., Zheng, F., … Hu, B. (2022). A multi-modal open dataset for mental-disorder analysis. Scientific Data, 9(1). https://doi.org/10.1038/s41597-022-01211-x

Chuang, C. Y., Lin, Y. T., Liu, C. C., Lee, L. E., Chang, H. Y., Liu, A. S., Hung, S. H., & Fu, L. C. (2023). Multimodal Assessment of Schizophrenia Symptom Severity From Linguistic, Acoustic and Visual Cues. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 3469–3479. https://doi.org/10.1109/TNSRE.2023.3307597

Czyzewski, A., Kostek, B., Kurowski, A., Szczuko, P., Lech, M., Odya, P., & Kwiatkowska, A. (2017). Multimodal approach for polysensory stimulation and diagnosis of subjects with severe communication disorders. Procedia Computer Science, 121, 238–243. https://doi.org/10.1016/j.procs.2017.11.033

Dixit, K. K., Pundir, S., Shrivastava, A., Kumar, C. P., Srivastava, A. P., & Singh, P. (2023). Analyzing Textual Data for Mental Health Assessment: Natural Language Processing for Depression and Anxiety. 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 1796–1802. https://doi.org/10.1109/UPCON59197.2023.10434291

Dura, A., & Wosiak, A. (2021). EEG channel selection strategy for deep learning in emotion recognition. Procedia Computer Science, 192, 2789–2796. https://doi.org/10.1016/j.procs.2021.09.049

Garcia-Ceja, E., Riegler, M., Nordgreen, T., Jakobsen, P., Oedegaard, K. J., & Tørresen, J. (2018). Mental health monitoring with multimodal sensing and machine learning: A survey. In Pervasive and Mobile Computing (Vol. 51, pp. 1–26). Elsevier B.V. https://doi.org/10.1016/j.pmcj.2018.09.003

Goncalves, L., & Busso, C. (2022). Robust Audiovisual Emotion Recognition: Aligning Modalities, Capturing Temporal Information, and Handling Missing Features. IEEE Transactions on Affective Computing, 13(4), 2156–2170. https://doi.org/10.1109/TAFFC.2022.3216993

Hosseini, M., Sohrab, F., Gottumukkala, R., & Katragadda, S. (n.d.). EmpathicSchool: A multimodal dataset for real-time facial expressions and physiological data analysis under different stress conditions. https://doi.org/10.48550/arXiv.2209.13542

Hosseini, S., Gottumukkala, R., Katragadda, S., Bhupatiraju, R. T., Ashkar, Z., Borst, C. W., & Cochran, K. (2022). A multimodal sensor dataset for continuous stress detection of nurses in a hospital. Scientific Data, 9(1). https://doi.org/10.1038/s41597-022-01361-y

Jaiswal, M., Bara, C.-P., Luo, Y., Burzo, M., Mihalcea, R., & Provost, E. M. (2020). MuSE: a Multimodal Dataset of Stressed Emotion.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. In Stroke and Vascular Neurology (Vol. 2, Issue 4, pp. 230–243). BMJ Publishing Group. https://doi.org/10.1136/svn-2017-000101

Kaya, H., Çiftçi, E., Salah, A. A., Güleç, H., Çiftçi, E., Kaya, H., Güleç, H., & Salah, A. A. (n.d.). The Turkish Audio-Visual Bipolar Disorder Corpus.

Kirch, D. G., & Petelle, K. (2017). Addressing the Physician Shortagee: the peril of ignoring demography. JAMA, 317(19), 1947. https://doi.org/10.1001/jama.2017.2714

Koelstra, S., Muhl, C., Soleymani, M., Jong-Seok Lee, Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2012). DEAP: A Database for Emotion Analysis ;Using Physiological Signals. IEEE Transactions on Affective Computing, 3(1), 18–31. https://doi.org/10.1109/T-AFFC.2011.15

Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M. A., & Kraaij, W. (2014). The Swell knowledge work dataset for stress and user modeling research. ICMI 2014 - Proceedings of the 2014 International Conference on Multimodal Interaction, 291–298. https://doi.org/10.1145/2663204.2663257

Kutt, K., Drążyk, D., Żuchowska, L., Szelążek, M., Bobek, S., & Nalepa, G. J. (2022). BIRAFFE2, a multimodal dataset for emotion-based personalization in rich affective game environments. Scientific Data, 9(1), 274. https://doi.org/10.1038/s41597-022-01402-6

Lin, W., Orton, I., Liu, M., & Mahmoud, M. (2020). Automatic Detection of Self-Adaptors for Psychological Distress. Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, 371–378. https://doi.org/10.1109/FG47880.2020.00032

Majumder, N., Hazarika, D., Gelbukh, A., Cambria, E., & Poria, S. (2018). Multimodal sentiment analysis using hierarchical fusion with context modeling. Knowledge-Based Systems, 161, 124–133. https://doi.org/10.1016/j.knosys.2018.07.041

Makhmudov, F., Kultimuratov, A., & Cho, Y.-I. (2024). Enhancing Multimodal Emotion Recognition through Attention Mechanisms in BERT and CNN Architectures. Applied Sciences, 14(10), 4199. https://doi.org/10.3390/app14104199

Meziati, R., Benezeth, Y., De Oliveira, P., Chappé, J., & Yang, F. (2021, March 3). UBFC-Phys [Data set]. IEEE Dataport. https://doi.org/10.21227/5da0-7344

Miranda, J. A., Rituerto-González, E., Gutiérrez-Martín, L., Luis-Mingueza, C., Canabal, M. F., Bárcenas, A. R., Lanza-Gutiérrez, J. M., Peláez-Moreno, C., & López-Ongil, C. (2022). WEMAC: Women and Emotion Multi-modal Affective Computing dataset. http://arxiv.org/abs/2203.00456

Monisha, G. S., Yogashree, G. S., Baghyalaksmi, R., & Haritha, P. (2022). Enhanced Automatic Recognition of Human Emotions Using Machine Learning Techniques. Procedia Computer Science, 218, 375–382. https://doi.org/10.1016/j.procs.2023.01.020

Nojavanasghari, B., Baltrušaitis, T., Hughes, C. E., & Morency, L. P. (2016). Emo react: A multimodal approach and dataset for recognizing emotional responses in children. ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction, 137–144. https://doi.org/10.1145/2993148.2993168

Panda, R., Malheiro, R., Rocha, B., Oliveira, A. P., Panda, R., Malheiro, R., Rocha, B., Oliveira, A., & Paiva, R. P. (2013). Multi-Modal Music Emotion Recognition: A New Dataset, Methodology and Comparative Analysis. https://www.researchgate.net/publication/257409136

Pham, T. T. D., & Won, C. S. (2019). Facial Action Units for Training Convolutional Neural Networks. IEEE Access, 7, 77816–77824. https://doi.org/10.1109/ACCESS.2019.2921241

Pi, W. (2024). An Introduction to Recurrent Neural Networks (RNNs). https://doi.org/10.59350/gjsvh-zbk81

Pisanski, K., Kobylarek, A., Jakubowska, L., Nowak, J., Walter, A., Błaszczyński, K., Kasprzyk, M., Łysenko, K., Sukiennik, I., Piątek, K., Frackowiak, T., & Sorokowski, P. (2018). Multimodal stress detection: Testing for covariation in vocal, hormonal and physiological responses to Trier Social Stress Test. Hormones and Behavior, 106, 52–61. https://doi.org/10.1016/j.yhbeh.2018.08.014

Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., & Mihalcea, R. (n.d.). MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations.

Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. Procedia Computer Science, 167, 1258–1267. https://doi.org/10.1016/j.procs.2020.03.442

Qi, X., Wen, Y., Zhang, P., & Huang, H. (2025). MFGCN: Multimodal fusion graph convolutional network for speech emotion recognition. Neurocomputing, 611. https://doi.org/10.1016/j.neucom.2024.128646

Ringeval, F., Schuller, B., Valstar, M., Cowie, R., Kaya, H., Schmitt, M., Amiriparian, S., Cummins, N., Lalanne, D., Michaud, A., Ciftçi, E., Güleç, H., Salah, A. A., & Pantic, M. (2018). AVEC 2018 Workshop and Challenge. Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop, 3–13. https://doi.org/10.1145/3266302.3266316

Ringeval, Fabien., & Multimedia, A. S. I. G. on. (2018). Proceedings of the 2018 on AudioVisual Emotion Challenge and Workshop. ACM.

Rizzo, A., Scherer, S., & Gratch, J. (2016). Detection and computational analysis of psychological signals using a virtual human interviewing agent. www.almaden.

Sawadogo, M. A. L., Pala, F., Singh, G., Selmi, I., Puteaux, P., & Othmani, A. (2023). PTSD in the wild: a video database for studying post-traumatic stress disorder recognition in unconstrained environments. Multimedia Tools and Applications, 83(14), 42861–42883. https://doi.org/10.1007/s11042-023-17203-x

Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., & Van Laerhoven, K. (2018). Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. Proceedings of the 20th ACM International Conference on Multimodal Interaction, 400–408. https://doi.org/10.1145/3242969.3242985

Schuller, Bjorn. (2021). MuSe '21 : proceedings of the 2nd Multimodal Sentiment Analysis Challenge : October 24, 2021, Virtual Event, China. The Association for Computing Machinery.

Shrotri, A. A., Narodytska, N., Ignatiev, A., Meel, K. S., Marques-Silva, J., & Vardi, M. Y. (2022). Constraint-Driven Explanations for Black Box ML Models. https://gitlab.com/Shrotri/clime

Stappen, L., Baird, A., Christ, L., Meßner, E.-M., & Schuller, B. (2021). MuSe-Stress: Multimodal Emotional Stress (MuSe2021) (Version V1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4659752

Su, C., Xu, Z., Pathak, J., & Wang, F. (2020). Deep learning in mental health outcome research: a scoping review. In Translational Psychiatry (Vol. 10, Issue 1). Springer Nature. https://doi.org/10.1038/s41398-020-0780-3

Sun, Z., Huang, Z., Duan, F., & Liu, Y. (2020). A Novel Multimodal Approach for Hybrid Brain-Computer Interface. IEEE Access, 8, 89909–89918. https://doi.org/10.1109/ACCESS.2020.2994226

Thalpage, N. (2023). Unlocking the Black Box: Explainable Artificial Intelligence (XAI) for Trust and Transparency in AI Systems. Journal of Digital Art & Humanities, 4(1), 31–36. https://doi.org/10.33847/2712-8148.4.1_4

Thomson, R. E., & Emery, W. J. (2024). Neural Networks, Convolutional Neural Networks and Deep Learning. In Data Analysis Methods in Physical Oceanography (pp. 755–774). Elsevier. https://doi.org/10.1016/B978-0-323-91723-0.00015-5

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., Schnieder, S., Cowie, R., & Pantic, M. (2013). AVEC 2013 - The continuous Audio/Visual Emotion and depression recognition challenge. AVEC

2013 - Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, 3–10. https://doi.org/10.1145/2512530.2512533

Verma, S., Vishal, Joshi, R. C., Dutta, M. K., Jezek, S., & Burget, R. (2023). AI-Enhanced Mental Health Diagnosis: Leveraging Transformers for Early Detection of Depression Tendency in Textual Data. 2023 15th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 56–61. https://doi.org/10.1109/ICUMT61075.2023.10333301

Xie, W., Wang, C., Lin, Z., Luo, X., Chen, W., Xu, M., Liang, L., Liu, X., Wang, Y., Luo, H., & Cheng, M. (2022). Multimodal fusion diagnosis of depression and anxiety based on CNN-LSTM model. Computerized Medical Imaging and Graphics, 102. https://doi.org/10.1016/j.compmedimag.2022.102128

Yoon, J., Kang, C., Kim, S., & Han, J. (2022). D-vlog: Multimodal Vlog Dataset for Depression Detection. https://sites.google.com/view/jeewoo-yoon/dataset

Zhang, Z., Lin, W., Liu, M., & Mahmoud, M. (2020). Multimodal Deep Learning Framework for Mental Disorder Recognition. Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020, 344–350. https://doi.org/10.1109/FG47880.2020.00033

World Health Organization. (2021). Mental health: Strengthening our response. Retrieved from https://www.who.int

Zou, B. (2022, May 9). Chinese multimodal depression corpus. IEEE Dataport. https://doi.org/10.21227/jng1-rn06