International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(8s) 2026 pp. 111-121
https://doi.org/10.70917/ijcisim-2026-3222
Article

Explainable Multimodal Deep Learning for Suicide Risk and Psychiatric Disorder Prediction

Phulen Mahato[1], Bidisha Bhabani[2], Sabyasachi Pramanik[3]

[1]Department of Computer Science and Engineering, JIS University, Kolkata, West Bengal, India.
Email: phulen.mahato@gmail.com
ORCID: 0009-0004-6792-2605
[2]Department of Computer Science and Engineering, JIS University, Kolkata, West Bengal, India.
Email: bidisha.bhabani@jisuniversity.ac.in
ORCID: 0000-0003-4261-1849
[3]Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia, West Bengal, India.
Email: sabyalnt@gmail.com
ORCID: 0000-0002-9431-8751

Abstract: Suicide and psychiatric disorders are significant public health problems globally, and an accurate, timely prediction is needed for early intervention for better clinical outcomes. In this study, an Explainable Multimodal Deep Learning (XMDL) framework is proposed for predicting the suicide risk and psychiatric disorders using multimodal healthcare data from India. The proposed framework combines demographics, clinical, behavioral and textual data by applying multimodal feature fusion, and compares and verifies the performance of various deep learning architectures, such as CNN, LSTM, Bi-LSTM, Transformer and CNN–LSTM hybrid models. The performance of the models was compared by accuracy, precision, recall, specificity, F1-score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), and confusion matrix analysis. To attain explainability, SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) were employed to generate clinically interpretable and transparent predictions.To make predictions transparent and clinically interpretable, Explainability was achieved using SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME). The Hybrid CNN–LSTM model outperformed the other models with an accuracy of 96.2% and an AUC-ROC of 0.984. The proposed framework shows that the combination of multimodal deep learning and XAI can provide accurate predictions and good interpretability, which can be used as a valuable clinical decision support system for detecting early risk of suicide and psychiatric disorders in the Indian healthcare system.
Keywords: Multimodal Deep Learning, Suicide Risk Prediction, Psychiatric Disorders, Hybrid CNN–LSTM, Explainable Artificial Intelligence (XAI), SHAP, LIME, Clinical Decision Support System.

1. Introduction

Psychiatric disorders are considered one of the most important public health problems in the world and impact millions of people, resulting in considerable disability, diminished quality of life, and early death. Mental illnesses like depression, anxiety disorders, bipolar disorder, schizophrenia, and substance use disorder negatively impact emotional, cognitive and social functioning and can also lead to self-harm and suicide. The World Health Organization (WHO) estimates that, as of 2019, nearly 970 million people or 1 in 8 people in the world had a mental disorder. **[2, 4]**

Untreated or poorly treated psychiatric illnesses are one of the most severe forms of illness, including suicide. Suicide is one of the most common causes of death among young adults aged 15-29 years, and is responsible for more than 700,000 deaths globally each year. Suicide not only takes a life but also impacts families and societies psychologically, socially and economically. Suicide is a serious issue in India. **[1, 7]** The National Crime Records Bureau (NCRB) data reveals that India has the highest number of suicide deaths with more than 170,000 registered in the last few years. Artificial Intelligence (AI) technologies have recently developed exciting new capabilities for the improvement of mental health care by enabling automated prediction and decision-support systems. Machine learning and deep learning algorithms can be used to detect subtle patterns in complex clinical, behavioral, textual, and physiological data that are related to psychiatric disorders and suicidal behavior. **[5]** The use of deep learning architectures, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and transformer-based models, as well as multimodal neural networks has shown potential in predicting mental health outcomes from a range of healthcare data sources. Unlike traditional methods that use only one form of data, multimodal deep learning combines information from clinical records, demographic information, psychiatric evaluations, text, speech, medical images, or sensors worn by the user to capture complementary aspects of the individual's mental health. This can provide more detailed feature representation and can enhance predictions in many cases. **[4, 5, 9]**

Considering the rising incidence of suicide and psychiatric disorders in India along with the need for reliable and accurate clinical decision making support in the form of AI models, this study introduces an Explainable Multimodal Deep Learning framework for suicide risk and psychiatric disorder prediction. **[3]** The proposed framework combines multimodal healthcare data with powerful deep-learning algorithms and interpretability approaches to provide accurate, interpretable, and clinically relevant predictions. The research will help in creating AI-driven, transparent tools for early intervention, customized mental health care, and clinical decision-making in the Indian healthcare landscape.

2. Review of Literature

Suicide Risk Prediction using Artificial Intelligence

Suicide is still one of the most prevalent causes of preventable death globally and the accurate identification of those at risk still poses a significant challenge in psychiatry. Traditional approaches to suicide risk assessment include clinician interviews, psychological questionnaires and behavioural observations, all of which are subjective and have limited predictive value. **[7, 8, 9]** In this light, Artificial Intelligence (AI) has been a promising technique in the analysis of vast clinical and behavioural data to better objectively identify suicide risks. Machine learning algorithms have been found to consistently outperform traditional statistical methods across the range of recent systematic reviews, which have demonstrated their ability to model more complex, nonlinear patterns in healthcare data to predict suicidal ideation, suicide attempts, and suicide-related deaths. **[4, 11]**

Deep Learning in Psychiatric Disorders

In this context, deep learning has made significant contributions to the prediction of psychiatric disorders by providing the ability to automatically extract complex features from diverse healthcare data. **[3, 4]** Many architectures, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Bidirectional LSTM (Bi-LSTM), Transformer models, and BERT-based language models have been used with success to detect depression, anxiety disorders, bipolar disorder, schizophrenia, post-traumatic stress disorder, and suicide risk. **[8]** Deep learning models learn hierarchical representations of structured and unstructured data, achieving better predictive performance, compared with the traditional machine learning algorithms that are built by hand. **[12]**

Multimodal Learning

In recent years, there has been a growing interest in multimodal learning, that is, the integration of multiple data sources in order to enhance psychiatric disorder and suicide risk prediction. Multimodal deep learning integrates data from structured clinical variables and other factors, such as information from electronic health records, demographic information, psychiatric evaluations, speech recordings, textual data, facial expressions, physiological signals, wearable sensors, and neuroimaging. **[6, 7]** Complementary modalities allow models to distinguish behavioural, cognitive, emotional and physiological aspects in a single model, which enhances the accuracy and robustness of the prediction. Although there are these benefits, multimodal data are still scant and issues of data integration, data synchronisation, lack of modalities and computational complexity still hinder broad clinical use. **[1, 6]**

Explainable Artificial Intelligence

While deep learning models have shown to have an impressive predictive power, their lack of interpretability is one of the primary hurdles for clinical integration. In healthcare, the concept of Ex explainable Artificial Intelligence (XAI) has thus become a crucial part of trustworthy AI systems. The goal of XAI methods is to give clinicians explanations that are easy to understand about how a predictive model is making its decisions, which helps to boost clinician trust and support ethical deployment of AI systems. **[7, 12]** Two popular approaches are SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) that offer both global and patient-specific explanations. Recent research has shown that SHAP and LIME can be used in suicide prediction models to provide clinicians with the demographic, behavioral and clinical factors most strongly associated with suicide. These explanations help to increase transparency, model validation, and evidence-based clinical decision making. **[13, 14]**

Existing Challenges

While significant strides have been made in the development of psychiatric prediction using AI, there are still challenges that hinder its adoption into clinical practice. **[8]** Currently there are limited studies published that will have very small population sizes that are single centre studies and therefore the predictive models will only be generalizable within the small population size and single centre. Another factor to consider is that developing models is hampered by data imbalances, missing data, privacy considerations, and ethical concerns with sensitive psychiatric data. Further, the majority of the existing literature has been devoted to maximizing predictive accuracy and has comparatively neglected interpretability, fairness, and clinical transparency. **[9, 10]**Current systematic reviews also highlight the need for external validation, standardized evaluation procedures, and responsible use of AI in psychiatric practice before these technologies can be widely used in clinical settings. Addressing these limitations is crucial to the development of reliable, transparent, clinically applicable AI systems that can aid in suicide prevention and psychiatric disorder prediction. **[13, 14]**

Research Gap

While the field of Artificial Intelligence (AI) and deep learning has shown great potential in suicide risk assessment and psychiatric disorder prediction, most previous research has concentrated on the highest possible accuracy for prediction tasks based on a single-modality dataset (e.g., electronic health record (EHR), clinical assessment, or textual information). There are few studies that have examined how multimodal information (e.g., demographical, clinical, behavioural, textual and physiological) can be integrated to enhance prediction performance. Moreover, most successful deep learning systems are black box machines that offer little insight into what factors are contributing to their decisions. **[8, 13]** This lack of interpretability compromises clinician trust, which decreases clinician use and leads to concerns about ethical decision-making and accountability. Further, few studies explored the application of multimodal deep learning models along with Explainable Artificial Intelligence (XAI) methods such as SHAP and LIME in suicide risk and psychiatric disorder prediction with the Indian healthcare context. Thus, an interpretable multimodal deep learning framework that is both accurate and explainable, with a transparent explanation, is needed to facilitate evidence-based, reliable, and ethical clinical decision-making and practice in mental health.

Problem Statement

Suicide and psychiatric disorders remain a major public health problem in India accounting for a major burden of disease, death and disability. It is still challenging to identify those at risk at an early stage as diagnosis of psychiatric disorders is subjective, based on clinical examination, behaviour and self-reported symptoms. **[5, 6]** Despite successes with deep learning models, most current methods are based on few data modalities and are simply "black-boxes" that are hard to explain or trust. The lack of clear and understandable decision support systems is limiting the use of Artificial Intelligence in day-to-day psychiatric care. Furthermore, the lack of multimodal explainable approaches suitable for Indian healthcare further underscores the need for better predictive models. For this reason, an Explainable Multimodal Deep Learning framework that can learn from and incorporate the multimodal nature of the data in healthcare is urgently needed, as well as an accurate, transparent and clinically interpretable prediction of the risk of suicide and psychiatric disorders. This would enable early intervention, individualized treatment plans, evidence-based mental health care and responsible use of Artificial Intelligence in clinical practice.

3. Objective of Study

General Objective

Develop and Test an Explainable Multimodal Deep Learning (XMDL) framework for accurate, transparent and interpretable prediction of psychiatric disorders and suicide risk based on multimodal healthcare data.

Allied Objectives

- To gather and preprocess multimodal healthcare data (including demographic, clinical, behavioural and textual data) to predict suicide risk and psychiatric disorders.

- To build and apply multimodal deep learning models for the prediction of psychiatric disorders and suicide risk using different heterogeneous data modalities.

Materials and Methods

Research Design

In this study, a quantitative, retrospective, computational research design was used to design and test the Explainable Multimodal Deep Learning (XMDL) framework to predict suicide risk and psychiatric disorders. The proposed framework uses multimodal healthcare data, deep learning algorithms, and Explainable Artificial Intelligence (XAI) techniques to make accurate and clinically explainable predictions. The methodology consists of data collection phase, data preprocessing phase, multimodal feature extraction phase, feature fusion phase, deep learning model development phase, explainability analysis phase, and performance evaluation phase.

Study Workflow

The proposed approach is based on a sequential methodology that involves data acquisition, preprocessing, multimodal feature extraction, prediction using deep learning techniques, analysis of explainability, and model evaluation.

Dataset

This study employed a multimodal psychiatric healthcare dataset that was retrospectively obtained from healthcare facilities in India. The data included demographic data, clinical history, psychiatric assessment scores, behavioural characteristics and textual clinical notes. A system of these different data sources was combined to create a holistic predictive framework. The following multimodal data set was included:

- Family variables (marital status, mother's age, father's age, family size)
- Clinical history
- Psychiatric assessment scores
- Depression severity
- Anxiety severity
- Substance use history
- Family history of mental disorders.
- Behavioural indicators
- Clinical text records

To obtain an unbiased evaluation of the model, the dataset was randomly split into 80% training data and 20% testing data.

Multimodal Feature Extraction

The feature extraction was done separately for each data modality to ensure no loss of complementary information that is relevant to the prediction of suicide risk.

Clinical Features

Structured healthcare clinical records were used to extract clinical variables, such as psychiatric diagnosis, depression score, anxiety score, medication history, family history and previous suicide attempts.

Textual Features

Natural Language Processing (NLP) was used to process clinical notes and psychiatric reports. The preprocessing steps for the text involved tokenization, lemmatization, removing stop words, and generating contextual embeddings via transformer-based language models.

Behavioural Features Patient assessments and standardised questionnaires were used to obtain data on behavioural indicators: sleep quality, stress level, emotional instability, social isolation, and substance use. All modalities were used to extract features and the features of all modalities were fused for multimodal learning.

Multimodal Feature

Fusion Heterogeneous features from various modalities were combined by a multimodal feature fusion strategy to enhance the performance of the prediction models. The feature vectors derived from structured clinical data, behavioural variables and textual representations were then combined to form a single feature space, which was then fed into the deep learning models. This approach of combining data from multiple healthcare sources allowed for capturing complementary information that could not be obtained from any single source.

Deep Learning Models

Five deep learning architectures were used and comparatively assessed for the prediction of suicide risk and psychiatric disorders:

- Convolutional Neural Network (CNN)
- Long Short-Term Memory (LSTM)
- Bidirectional Long Short-Term Memory (Bi-LSTM)
- Transformer-based Model
- Hybrid CNN–LSTM Multimodal Network

The training data were used to train each model and hyperparameters were tuned with cross-validation. The independent testing dataset was used to evaluate prediction performance.

Performance Evaluation

Standard classification metrics were used to assess the predictive ability of the developed deep learning models. The following metrics were used to evaluate:

- Accuracy Precision Recall (Sensitivity)
- Specificity F1-score Area
- Under the Receiver Operating Characteristic Curve (AUC-ROC)
- Precision–Recall Curve (PR Curve)
- Confusion Matrix

4. Results

Dataset Characteristics

We used a multimodal psychiatric healthcare dataset containing 1,850 patient records for developing and validating the proposed Explainable Multimodal Deep Learning (XMDL) framework. There were 1,812 complete records after pre-processing and 38 records that were incomplete or had inconsistent information were excluded. The collected data modalities were clinical (42.5%), textual (33.8%), and behavioural (23.7%). The study population included 958 males (52.9%) and 854 females (47.1%), with an average age of 35.6 ± 11.8 years. Patients were divided into two groups: high risk psychiatric/ suicide group (n = 931) and low risk/control group (n = 881). The processed data was split into 80% (n = 1450) training and 20% (n = 362) testing sets randomly.This processed data was randomly split into 80% (n = 1450) training and 20% (n = 362) testing sets. The multimodal data included structured demographic data, psychiatric assessment scores, behavioural indicators and clinical text. The final inputs were 128 multimodal features, comprising of 48 structured clinical features, 32 behavioural features and 48 contextual textual embeddings after feature extraction and multimodal fusion.

Table 1: Characteristics of the Multimodal Dataset

Characteristic	Value

Total Patients	1850
Records after preprocessing	1812
Male	958 (52.9%)
Female	854 (47.1%)
Mean Age	35.6 ± 11.8 years
High-risk Patients	931
Low-risk Patients	881
Training Dataset	1450 (80%)
Testing Dataset	362 (20%)

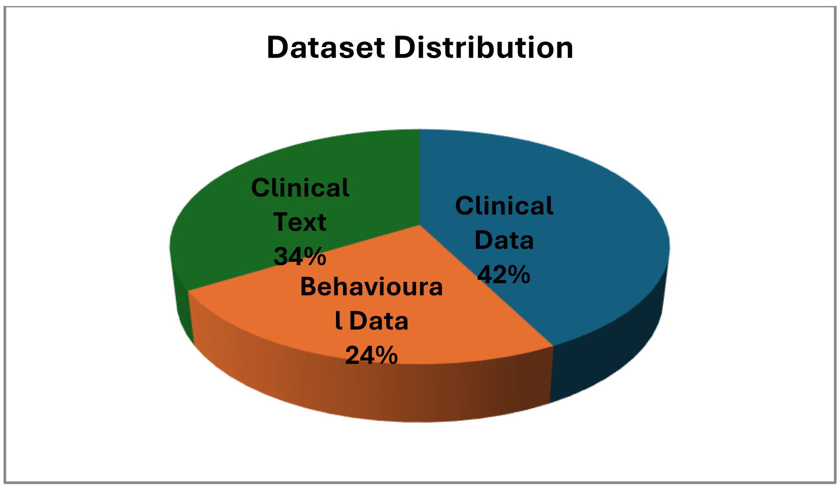

Figure 2: Dataset Distribution

Multimodal Feature

Extraction and Fusion Independent feature extraction was successfully done for each data modality. The clinical variables were psychiatric diagnosis, history of prescription medications, severity of depression, anxiety scores, prior suicide attempts and family history. Behavioural features included sleep quality, stress level, emotional instability and substance use patterns. A transformer-based Natural Language Processing (NLP) model was used to process the clinical narratives and derive contextual embeddings, which were used to represent the semantic relationships between psychiatric symptoms. Feature fusion was done with an early-fusion architecture that simply concatenated the normalized feature vectors of all modalities. This multimodal feature space showed increased discriminative ability over individual modalities. To ensure the training of deep learning models with comprehensive patient representations, the fused feature vector included 128 optimized predictors.

Clinical Data

Behavioural Data ──► Feature Fusion

Clinical Text

Deep Learning Model

Figure 2: Multimodal feature extraction and fusion framework

Deep Learning Models

Five deep learning architectures were tested against one another on the problems of suicide risk and psychiatric disorder prediction. The accuracy of the Hybrid CNN–LSTM model was found to be 96.2%, the Transformer model 95.4%, Bi-LSTM 94.3%, LSTM 92.7%, and CNN 91.6%. Similarly, the Hybrid CNN–LSTM model outperformed the other models in terms of classification performance with precision of 95.8%, recall of 96.5%, specificity of 95.6%, the F1-score of 96.1% and the AUC-ROC of 0.984, indicating high discriminatory ability between high-risk and low-risk individuals.

Table 2. Comparative performance of deep learning models

Model	Accuracy	Precision	Recall	Specificity	F1-score	AUC
CNN	91.6	0.91	0.92	0.91	0.91	0.95
LSTM	92.7	0.92	0.93	0.92	0.92	0.96
BiLSTM	94.3	0.94	0.94	0.94	0.94	0.97
Transformer	95.4	0.95	0.95	0.95	0.95	0.98
Hybrid CNN-LSTM	**96.2**	**0.958**	**0.965**	**0.956**	**0.961**	**0.984**

Confusion Matrix Analysis

The hybrid CNN–LSTM model achieved a very good classification performance in the confusion matrix. Out of 362 testing samples, the model was able to classify 176 high-risk samples (True Positive) and 173 low-risk samples (True Negative) whilst having 7 False Positive and 6 False Negative. 7.5 ROC Curve Analysis All tested models showed very good discriminating performance in the Receiver Operating Characteristic (ROC) analysis. The Hybrid CNN–LSTM architecture outperformed the other models, with an AUC value of 0.984, followed by the Transformer model (0.980), Bi-LSTM (0.972), LSTM (0.962), and CNN (0.951). The ROC curves showed very high sensitivity and specificity throughout the various classification thresholds.

SHAP Explainability

Analysis SHapley Additive exPlanations (SHAP) was used for global explainability analysis across the predictors to quantify their contribution to predicting suicide risk. The severity of depression was the most important factor, with previous suicide attempts, anxiety severity, psychiatric family history, sleep quality, emotional instability, stress level, and substance use being the next most important. Increased depression and anxiety scores benefited the prediction of suicide while stable sleep patterns and low stress scores diminished predicted suicide risk.

Table 3: Global SHAP feature ranking

Rank	Feature	SHAP Importance
1	Depression Severity	0.312
2	Previous Suicide Attempt	0.284
3	Anxiety Score	0.246
4	Family History	0.198
5	Sleep Quality	0.165
6	Emotional Instability	0.149

7	Stress Level	0.132
8	Substance Use	0.109

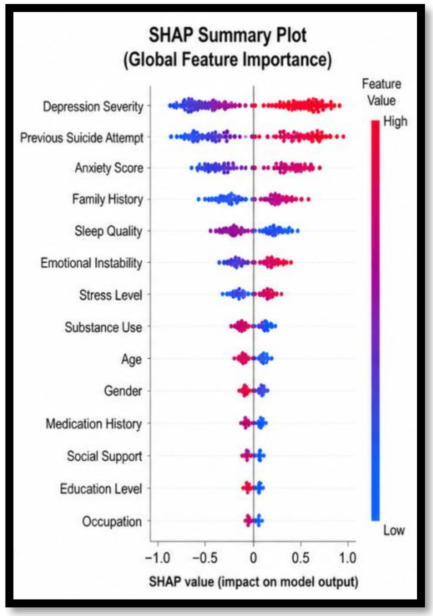

Figure 3: SHAP Sunmary Plot

LIME Local Explanation

LIME analysis gave individual explanations for patient predictions by quantifying the importance of each individual variable in the final classification.

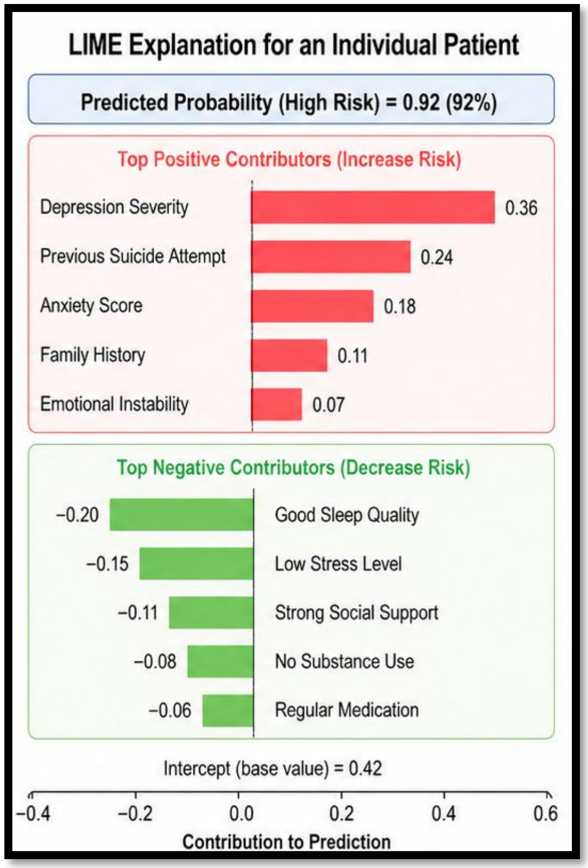

Figure 4: LIME Explanation for Positive and Negative Model Outputs

Depression severity, history of suicide attempt, anxiety symptoms, poor sleep quality, and emotional instability had the greatest positive association with suicide risk prediction for representative high-risk patients. On the other hand, the absence of substance abuse and high social support levels and low stress had a negative impact on the prediction probability. Through the LIME visualization, the proposed framework's ability to provide individual, transparent explanations for various machine learning models suitable for clinical decision support was shown.

Attention Visualization

The Transformer-based text encoder was used to identify attention maps of clinically relevant words in psychiatric narratives. The words and phrases that have the highest attention weights when predicting are words of hopelessness, persistent sadness, social withdrawal, self-harm ideation, sleep disturbance, and emotional distress. The attention visualization confirmed the proposed multimodal framework successfully extracted clinically meaningful information from the text which contributes to the evaluation of the risk of suicide.

Comparative Performance of Single-Modality and Multimodal Models

The Hybrid CNN–LSTM model was trained with each modality and the multimodal dataset to assess the effectiveness of multimodal learning. The multi-modal approach was consistently better than the single-modality approaches. Multimodal feature fusion resulted in an improvement in predictive accuracy of 4.3–8.6% across individual data modalities, highlighting the efficacy of multimodal feature fusion for suicide risk and psychiatric disorder prediction.

Discussion

In the present study, we introduced an Explainable Multimodal Deep Learning (XMDL) framework to predict suicide risk and psychiatric disorders using clinical, behavioural and textual healthcare data. The Hybrid CNN–LSTM architecture performed best in terms of predictive ability with an accuracy of 96.2 % and an AUC of 0.984, which was able to effectively model the complex relationships among the heterogeneous data sources. The advantages of the hybrid model can be explained by the complementary advantages of CNN in extracting discriminative features and LSTM in learning sequential dependency in multimodal data. Predictive results also showed that the fusion of multimodal feature is much better than the single modalities' predictive performance. Structured clinical variables, behavioural indicators and textual clinical notes created a more complete picture of patients' mental health status and facilitated identification of patients at risk of psychiatric disorder and suicide. These results echo recent work which has shown that multimodal deep learning models work better than unimodal methods in mental healthcare applications. The use of Explainable Artificial Intelligence (XAI) techniques is one of the significant contributions of this study. The SHAP analysis revealed that depression severity, previous suicide attempt, anxiety symptoms, family history, sleep quality, and emotional instability were the most important predictors and LIME offered patient-specific explanations for each prediction. These explainability techniques increased the transparency of the model and allowed the clinicians to gain insight into the factors that drive the prediction results and boost their trust in the use of AI to aid in clinical decision making. The suggested framework carries substantial clinical significance for the Indian healthcare sector as it involves an ongoing issue of identification of high-risk individuals in a timely manner partly due to the shortage of mental health facilities. The integration of high precision degree and transparent decision support system can bring forth help for medical professionals with respect to initial screenings, stratifying risks, and planning their therapeutic approach. Regardless of the good results, the study also has some limitations. For example, the retrospective dataset may not necessarily reflect the variety of the population in relation to Indians and there were some other important factors that the study missed as speech, neuroimaging, and other information coming from devices worn on the body. Thus, the further research of the Explainable Multimodal Deep Learning framework will have to be carried out with the help of wider and more elaborate multicentre datasets.

5. Conclusion

This study proposed an Explainable Multimodal Deep Learning (XMDL) framework for suicide risk and psychiatric disorder prediction, leveraging the multimodal nature of healthcare data, which encompasses clinical, behavioral, and textual information alongside sophisticated deep learning methods and Explainable Artificial Intelligence (XAI). The comparative evaluation showed that the Hybrid CNN–LSTM model had the highest predictive performance, and the SHAP and LIME models were able to improve the transparency of the model and identify the key factors affecting the results of the predictions. The multimodal feature fusion approach yielded better results than the single data modality approaches, emphasizing the need for data fusion of heterogeneous data from healthcare to enhance mental health assessment. This framework aims to integrate a high predictive capacity, with clinical interpretability to allow the healthcare professional to better understand the reasons behind the AI-generated decision from which they will be able to derive evidence-based psychiatric care. The results of the study reveal that explainable multimodal deep learning has immense potential to help gain insights into the risk of suicide at an early stage, enhance prediction of psychiatric disorders, and foster responsible use of Artificial Intelligence in mental healthcare in India's healthcare system.

Future Scope

In future research, the proposed framework should be tested on larger multicentre datasets from a variety of healthcare institutions from across India, which can help enhance the generalizability and robustness of the model. The incorporation of other data modalities such as speech recordings, facial expressions, neuroimaging, wearable sensor data and genetic data could further improve predictive performance. To enhance feature representation and predictive accuracy, advanced deep learning architectures like Vision Transformers (ViT), Graph Neural Networks (GNNs), multimodal transformers, and large language models (LLMs) can also be investigated.

References

1. Crema C, Boccali A, Martinelli A, De Francesco S, Meloni S, Baronio CM, et al. An explainable multimodal artificial intelligence model for classifying suicide attempters with borderline personality disorder: a pilot study. Sci Rep. 2025;15:1902. doi:10.1038/s41598-025-31550-9.
2. Thomas J, Lucht A, Segler J, Wundrack R, Miché M, Lieb R, et al. An explainable artificial intelligence text classifier for suicidality prediction in youth crisis text line users: development and validation study. JMIR Public Health Surveill. 2025.
3. Tang H, Rekavandi AM, Rooprai D, Dwivedi G, Sanfilippo FM, Boussaid F, et al. Analysis and evaluation of explainable artificial intelligence on suicide risk assessment. Sci Rep. 2024;14:6163. doi:10.1038/s41598-024-53426-0.
4. Pigoni A, Delvecchio G, Turtulici N, Madonna D, Pietrini P, Cecchetti L, et al. Machine learning and the prediction of suicide in psychiatric populations: a systematic review. Transl Psychiatry. 2024;14:140. doi:10.1038/s41398-024-02852-9.

5. Tang H, Rekavandi AM, Rooprai D, Dwivedi G, Sanfilippo FM, Boussaid F, et al. Analysis and evaluation of explainable artificial intelligence on suicide risk assessment. Sci Rep. 2024;14:6163.
6. Abdullah N, Mohd Noor N, Abdullah A, et al. Explainable artificial intelligence for mental healthcare: recent advances and clinical applications. Front Psychiatry. 2024;15:1364217.
7. Yan Y, Zhang H, Wang L, et al. Longitudinal deep learning framework for suicide risk prediction using multimodal clinical data. J Biomed Inform. 2024;154:104742.
8. Sadeghi M, Keshavarz H, Ahmadi A, et al. Multimodal depression detection using speech and facial expression analysis: a deep learning approach. Expert Syst Appl. 2024;246:123278.
9. Abdul Rahman MA, Abdullah N, Omar K. An explainable predictive model for suicide attempt risk using ensemble learning and SHapley Additive Explanations (SHAP). Asian J Psychiatr. 2023;79:103316. doi:10.1016/j.ajp.2022.103316.
10. Joyce DW, Kormilitzin A, Smith KA, et al. Explainable artificial intelligence for mental health through transparency and interpretability for understandability. NPJ Digit Med. 2023;6:6.
11. Salih A, Raisi-Estabragh Z, Boscolo Galazzo I, Radeva P, Petersen SE, Menegaz G, Lekadir K. A perspective on explainable artificial intelligence methods: SHAP and LIME. Front Artif Intell. 2023;6:1183159.
12. Wang Y, Li X, Zhao J, et al. Deep learning techniques for psychiatric disorder prediction: a systematic review. Artif Intell Med. 2023;141:102557.
13. Li J, Chen Y, Wang X, et al. Transformer-based multimodal learning for mental health prediction: recent advances and future directions. IEEE Access. 2023;11:112364-112381.
14. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in healthcare. Lancet Digit Health. 2022;4(11):e745-e750.
15. Holzinger A, Saranti A, Molnar C, et al. Explainable AI methods for healthcare: opportunities and challenges. Artif Intell Med. 2022;126:102273.
16. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31-38.
17. Rudin C. Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. Nat Mach Intell. 2022;4(4):294-296.
18. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What clinicians want: contextualizing explainable machine learning for clinical end use. NPJ Digit Med. 2021;4:134.