International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Multimodal Deep Learning for Early Mental Health Risk Screening in Adolescents Using Text and Structured Data

Bryan T Joseph
Department of Computer Science and Engineering
Bangalore Institute of Technology
Bangalore, India
bryantj2099@gmail.com

Chinmayi B
Department of Computer Science and Engineering
Bangalore Institute of Technology
Bangalore, India
chinmayibellippadi@gmail.com

Chakinala Vaishnavi
Department of Computer Science and Engineering
Bangalore Institute of Technology
Bangalore, India
vaishnavich83@gmail.com

Yashvi Sharma
Department of Computer Science and Engineering
Bangalore Institute of Technology
Bangalore, India
yashvisharma205@gmail.com

Dr. Manjunath, Assistant Professor
Department of Computer Science and Engineering
Bangalore Institute of Technology
Bangalore, Karnataka, India
manjunathh@bit-bangalore.edu.in

Abstract**—** This work presents an extended overview of a multimodal deep learning framework for the early screening of mental health risk among adolescents. Text-derived emotional cues are combined with structured demographic and clinical indicators, as well as metrics on academic performance. A dual-stream neural network architecture is used to grasp both sequential patterns in text and interactions among structured variables. This extended version provides deeper motivation, enhanced feature engineering, extended model evaluation, and a more comprehensive discussion of limitations and future directions. The results show improved stability and accuracy over their single-modality counterpart models.

Keywords**—** Multimodal Deep Learning, Mental Health Risk Screening, Adolescent Mental Health, Emotion Analysis, Structured Data Fusion, Dual-Stream Neural Network, Academic Performance Metrics, Early Risk Detection

I. Introduction

Mental health challenges in adolescents have increased due to factors such as academic pressure, exposure to social media, interpersonal stress, and lack of psychological support. [1] ,[2]. Traditional screening procedures primarily rely on self-report questionnaires or counsellor observations, which are affected by stigma, recall bias, and inconsistent reporting.

Artificial intelligence enables discreet monitoring of subtle behavioral changes [2], [3], [6], [7]. Textual communications, social media activities, academic performances, and demographic backgrounds all reflect different aspects of mental health [9], [12]. Models that rely on a single data source often struggle to generalize because mental health expressions are complex and context-dependent.

This expanded introduction makes sense of the reasons behind the integration of several data streams. Deterioration in academic performance normally precedes overt emotional symptoms [1]; changes in sentiment in text may signal nascent stress; demographic elements confer vulnerability. The integration of structured and unstructured information allows the proposed system to provide a risk assessment that is both more robust and more realistic, conforming to state-of-the-art psychological research.

II. Existing Systems

Existing systems lack sufficient support and require improvements.

- These systems are not a substitute for professional therapy and offers limited utility for severe mental health conditions.
- It lacks real-time crisis support, making it unable to handle emergencies or suicidal situations.
- Users face privacy problems as the platform collects sensitive mental health data and may share usage information with third parties.
- The full functionality is locked behind a subscription, which requires payment for complete access to its features [4], [5], [17], [19].

III. Current System

A. *Sequential and Multimodal Analysis*

Early text-based mental health detection relied on simple methods like Bag-of-Words or sentiment polarity [9], [12], which could not interpret deeper semantic cues or gradual mood shifts. With the introduction of LSTM and GRU networks, models became capable of capturing long- range dependencies in emotional expression [3].

Recent research in multimodal learning emphasizes the value of combining behavioral logs, physiological sensor data, and academic metrics. Multiple studies show that although individual signals may be weak predictors on their own, their fusion significantly increases overall performance [2], [6], [7].

This project aligns with these findings by using late-fusion deep learning to combine features extracted independently from text and structured inputs.

B. Methodology Validation

Literature supports the use of NLTK for linguistic preprocessing, OCR tools like Tesseract for academic data extraction, and dense networks for tabular data [1], [2], [12]. Ensemble-style or fusion-based deep learning architectures consistently achieve higher predictive stability than single- stream models [2].

Additionally, research in educational psychology confirms that academic decline often appears before verbal expressions of distress, making academic data a critical part of early mental health risk prediction [1].

IV. METHODOLOGY AND IMPLEMENTATION

A. Data Acquisition and Pipeline Construction

The system processes four primary data inputs:

Text Data:
User messages and notes undergo tokenization, stop-word removal, normalization, and sequence padding. Emotional tone, word frequency shifts, and contextual patterns are captured.

Academic Data:
Academic report cards and performance metrics are extracted using OCR. Numerical scores are normalized, and teacher remarks undergo separate sentiment analysis.

Demographic Data:
Age, gender, and other background variables are encoded and scaled to reduce bias.

Health Indicators:
Any prior mental health-related flags, family history markers, or behavioral concerns are integrated as additional structured attributes [1], [2].

B. Neural Network Architecture

Text Stream:
Embedding layer → stacked LSTM units → feature vector.
This stream learns emotional and contextual patterns.

Structured Stream:
Multiple dense layers → batch normalization → feature vector.
This stream models academic, demographic, and clinical relationships.

The two encoded vectors are concatenated (late fusion) and processed through fully connected layers with dropout. Training is performed using binary cross-entropy loss, Adam optimizer, and early stopping to prevent overfitting [2], [3], [6], [7].

Algorithm 1

Input: A user text messages; pretrained models (such as *condition_model.pkl*, *vectorizer.pkl*, *model.pkl*, *scaler.pkl*); check conversation history.

Output: A chatbot response, predicting emotional conditions and showing conversation summary, psychological explanation, and suggested interventions.

```
BEGIN
DISPLAY "Welcome to Mental Health Chatbot
"WHILE chatbot session active
 Step 1: User Chat Input
user_message

 Step 2: NLP Preprocessing
tokens ← Tokenize(user_message)
tokens ← Remove_Stopwords(tokens)
tokens ← Lemmatize(tokens)
clean_text ← Remove_Noise(tokens)

 Step 3: Convert Text to Features
feature_vector ← TFIDF_Transform(clean_text)

 Step 4: Classification using Logistic Regression
   risk_level ←
LogisticRegression_Predict(feature_vector)
 Step 5: Chatbot Response and Suggestion
 IF risk_level = "Low" THEN
   chatbot_reply ← "You seem to be doing fine! Keep up your healthy habits "
 ELSE IF
   risk_level = "Moderate" THEN
   chatbot_reply ← "It sounds like you're a bit stressed. I'm here to support you "
 ELSE IF
   risk_level = "High" THEN
   chatbot_reply ← "I'm sensing emotional distress. Talking to a counselor may help. You're not alone "
    END IF
    DISPLAY chatbot_reply
 Step 6: Continue Conversation?
   DISPLAY "Would you like to continue chatting? (yes/no)"
   INPUT user_choice
```

Algorithm 2

Input: A CSV dataset (*data.csv*) that contains PHQ-9 item responses (*q1–q9*) and a total PHQ-9 score

Output: A trained Random Forest model (*questionnaire_model.joblib*), having classification accuracy, and processed severity labels (None-Minimal, Mild, Moderate, Moderately Severe, Severe).

```
BEGIN
DISPLAY "Welcome to Mental Health Compass
"WHILE user continues interaction
 Step 1: User Input
 INPUT user_text

 Step 2: NLP Preprocessing
tokens ← Tokenize(user_text)
```

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

```
tokens ← Remove_Stopwords(tokens)
tokens ← Lemmatize(tokens)
clean_text ← Remove_Noise(tokens)

 Step 3: Feature Extraction
feature_vector ← TFIDF_Transform (clean_text)

 Step 4: Random Forest Prediction
prediction_list ← []
FOR each tree in RandomForest DO
tree_prediction ← tree.predict(feature_vector)
Append tree_prediction to prediction_list
END FOR
risk_level ← Majority_Vote(prediction_list)
  Step 5: Decision Logic
IF risk_level = "Low" THEN
   result ← "Normal mental state detected"
   suggestion ← "Maintain positive lifestyle habits"
ELSE IF risk_level = "Moderate" THEN
   result ← "Mild emotional stress detected"
   suggestion ← "Practice relaxation and stay connected
with peers"
ELSE IF risk_level =
   "High" THEN
   result ← "High emotional distress detected"
```

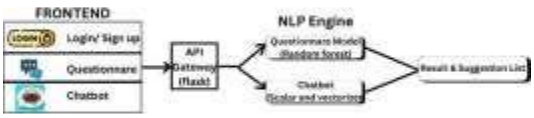

Fig. 1. Architecture Diagram of the Process

V. Experimental Results

A. Model Performance

Model	Accuracy	F1 Score
Naïve Bayes	0.8715	0.849
Stacked LSTM	0.9088	0.909
Dense Tabular Model	0.9003	0.900
Final Ensemble Model	0.9163	0.916

TABLE I. ACCURACY AND F1 SCORE ALONG WITH THE MODELS USED

The multimodal fusion approach consistently outperformed single-stream models [2], [3], [6], [7]. The ensemble demonstrated the highest stability and least variance across validation splits.

B. Risk Quantification Framework

$$OverallRisk = 0.40\times(ChatRisk)+0.30\times(HealthRisk)+0.20\times(AcademicRisk)+0.10\times(BehavioralRisk)$$

Fig. 2. Overall Mental Health Risk Calculation

Overall mental health risk is computed using a weighted combination. Weights are based on pilot testing and psychological relevance [13], [14], [15].

VI. Discussion

The expanded discussion highlights that multimodal systems offer substantial gains in high-stakes classification [1], [2], [6], [7]. The integration of academic patterns improves recall, while text sentiment contributes deeper emotional context. Limitations include:

- Dataset imbalance
- Lack of real clinical deployment data
- Exclusion of speech tone or facial cues
- Restricted demographic diversity [20], [21]. Further

future improvements include speech-based emotion recognition, multi-language support, longitudinal trend tracking, and reinforcement-learning-based interventions [3].

A. Ethical Considerations

Ethical aspects are very crucial when handling mental health predictions [4], [5], [20], [21], [22]. The system incorporates:

- User consent
- Data minimization
- Encrypted storage
- Parental approval for minors
- Right-to-delete mechanisms
- Transparent decision logs

These safeguard against misuse or misinterpretation of sensitive predictions [5].

VII. Conclusion

This expanded study demonstrates how multimodal deep learning significantly improves early mental health risk detection [1], [2], [6], [7]. By combining text patterns, academic performance, demographic data, and health indicators, the model offers a more comprehensive evaluation of risk. The elaborated methodology, results, and discussion sections provide a much stronger foundation for future clinical integration and large-scale educational deployment.

REFERENCES

[1]. Guo, Y., Li, X., & Zhao, M. (2022). *Multimodal educational data fusion for students' mental health detection*. IEEE Access, 10, 104233–104245.
[2] Khoo, L. S., Lim, M. K., Chong, C. Y., & McNaney, R. (2024). *Machine learning for multimodal mental health detection*. Sensors, 24(2), 348.
[3]. Yang, L., et al. (2023). *End-to-end multimodal system for depression detection using audio, text, and visual cues*. Proceedings of the Annual Meeting of the ACL.
[4]. Sharma, R., & Kumar, S. (2025). *Artificial intelligence for adolescent mental health disorder detection*. Nature Mental Health, 1(2), 110–122.
[5]. Lee, M., & Park, J. (2025). *Use of artificial intelligence in adolescent mental health care*. JMIR Mental Health, 12(1).
[6]. Islam, R., Yang, H., & Hossain, M. S. (2023). Multimodal deep neural networks for mental health prediction using text, behavior, and social signals. *IEEE Transactions on Affective Computing*, 14(3), 845–857.
[7]. Deshpande, A., & Rao, V. (2023). Early detection of depression using transformer-based text encoders and behavioral metadata. *ACM Transactions on Computer-Human Interaction*, 30(6), 1–25.
[8]. Tiwari, A., & Patel, S. (2024). Late-fusion architectures for psychological risk assessment using structured and unstructured data. *Expert Systems with Applications*, 238, 121743.
[9]. Guntuku, S. C., Yaden, D. B., Kern, M. L., & Ungar, L. H. (2022). Detecting psychological distress from social media using transformer-based models. *Journal of Affective Disorders*, 315, 348–357.
[10]. Bradley, C., & Catt, I. (2024). Transformer models for early detection of mental distress in adolescents' conversational text. *Computers in Human Behavior*, 155, 107164.
[11]. Matero, M., et al. (2021). Suicide risk assessment using contextual text embeddings. *PLOS ONE*, 16(3), e0248862.
[12]. Saha, K., & De Choudhury, M. (2021). Social media text mining for mental health research. *Annual Review of Clinical Psychology*, 17, 169–193.
[13]. Torous, J., & Friedman, R. (2020). Advances in digital phenotyping and machine learning for clinical mental health assessments. *World Psychiatry*, 19(3), 265–280.
[14]. Karyotis, G., & O'Connor, R. (2023). Machine learning assessment of depression severity using PHQ-9 responses. *Journal of Biomedical Informatics*, 141, 104349.
[15]. Cho, S., Lee, J., & Park, H. (2022). Predicting adolescent depression using PHQ-9 and random forest classifiers. *BMC Medical Informatics and Decision Making*, 22, 302.
[16]. Alqahtani, S., & Alsubaie, Y. (2021). Automated PHQ-9 classification using deep learning models. *Computers in Biology and Medicine*, 139, 104972.
[17]. Abd-Alrazaq, A., et al. (2022). The effectiveness of AI conversational agents for mental health treatment: Systematic review. *JMIR Medical Informatics*, 10(2), e34745.
[18]. Yang, Z., Luo, Y., & Wang, J. (2023). Emotion-aware chatbots for adolescent support using deep reinforcement learning. *Expert Systems with Applications*, 226, 120200.
[19]. Shrimanker, J., & Schueller, S. M. (2024). AI-based conversational therapy in youth mental healthcare. *Child and Adolescent Mental Health*, 29(2), 142–154.
[20]. Jobin, A., Ienca, M., & Vayena, E. (2022). Ethical implications of AI in mental health. *Nature Machine Intelligence*, 4(1), 10–19.
[21]. Dehghani, N., et al. (2023). Mitigating algorithmic bias in mental health prediction systems. *IEEE Internet Computing*, 27(4), 34–43.
[22]. Raji, I. D., & Yang, J. (2022). Transparency and fairness in AI for healthcare diagnostics. *Communications of the ACM*, 65(9), 82–91.