WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Mental Health Early Warning System Via Social Media Analysis

Makkala Ram [1], K. RaviKanth [2], Nakka Indhravathi[3], Mamidyala Hari Krishna[4], Penjarla Sandeep[5]

[1]P.G. Research Scholar, Dept. of MCA-Data Science, Aurora Deemed To Be University, Hyderabad, Telangana, 500098, India.
[2]Assistant Professor, Dept. of CSE, Aurora Deemed To Be University, Hyderabad, Telangana, 500098, India.
Email:[1]rammakkala521@gmail.com,[2] ravikanth@aurora.edu.in

A B S T R A C T

The concern for mental health challenges worldwide has been stretching far, with multiple people sharing their feelings, struggles, and early signs of distress on social media. These signs very rarely enter the radar before culmination into full-fledged disorders. This paper suggests an Early Warning System for Mental Health wherein Artificial Intelligence and Natural Language Processing could be used to analyze user-generated text for determining emotional risk levels. Real-time detection of stress, sadness, anxiety, and other high-risk markers is performed via advanced language models. Alerts can be sent, with the user's consent, to trusted contacts such as family members, counselors, or health professionals by SMS, email, or WhatsApp, for the timely intervention. Unlike the classical approach, our system also considers ethical handling of sensitive data, explainable theory behind AI prediction, and multilingual adaptability, making it inclusive across an array of populations. The system thus offers an easy-to-use interface for the individuals to interact with while maintaining their privacy and transparency. By rendering personalized views and preventive support, the proposed system aims to destroy the gulf that exists between online bearing and mental health care. Validation via experimentation, along with the available literature, indicates the feasibility of an approach that merges AI-led risk detection with social network channels in order to provide a secure and socially responsible framework. The study looks toward a more extensive vision of technology-enabled mental health support systems that are scalable, accessible, and capable of reducing unseen risks in the digital community.

Keywords: Mental Health, Early Warning System, Social Media Analysis, Artificial Intelligence, Natural Language Processing, Risk Prediction, Multilingual Support.

1. Introduction

A society exercising from the grave trauma of mental health attrition is increasing daily despite the veil of affluence and riches in various regions across the globe. Depression, anxiety, and stress have apparently now become blood relatives to each other; collectively these mental disorders destroy the quality of life and occasionally lead to self-harm or even committing suicide. With the advent of rapid digital communication, platforms like Facebook, Instagram, Twitter, and other forums have now come to constitute common agog halls through which people vocalize their emotions, personal experiences, and challenges. Posts on these platforms, as well as comment sections, suggest slight or direct indications of mental distress in some instances, be it sadness, hopelessness, cutting off, or exhibiting some unusual pattern of behavior. In fact, the foreboding signs overlooked by the family, friends, or professional observers-the actual monitoring-would be very hard to do and, even when done, there would certainly be time lapses. Though, AI and NLP would considerably strengthen the mentioned task of early signal detection through its automation. All language processing systems are capable of searching large volumes of user-generated text with respect to riskful emotions, levels of distress, and assumptions relating to possible mental health issues. These methods then look at systems having an edge over conventional survey methods, with limits concerning scalability and responsiveness, toward real-time monitoring. The proposed work is to developing an Early Warning System for Mental Health that detects Emotional Risk Factors from social media posts and communicates them responsibly. Alerts may be sent to concerned contacts such as family members, counselors, or healthcare professionals via SMS, e-mail, or WhatsApp, with the patient's concurrence. This system aims to be multilingual, user-centered, and ethically designed to protect the individual's privacy and to act promptly. The objective of this research is to create a future in which mental health support combines technology so that high-risk patients receive early help, thus minimizing their chances of falling into severe mental crises.

2. Literature Review

In recent times, there has been increasing interest in detecting mental health issues via digital platforms. With social platforms being widely used, researchers began to analyze online behavior as a possible indicator for emotional distress. This section will review extant studies and approaches relevant to the developing of a Mental Health Early Warning System.

2.1 Traditional Approaches to Mental Health Monitoring

The assessment of mental health has conventionally involved self-reports such as surveys, clinical interviews, and standardized psychological tests. Though these methods function well in laboratory-like settings, they often fall short in capturing emotional states in real time. Moreover, many individuals are hesitant to seek professional assistance, leading to delayed diagnosis and intervention. In this regard, there is a disparity between self-reporting and actual experience, which has motivated researchers to find other alternative means of assessment, particularly technology-based ones.

2.2 Social Media as a Data Source

Social media now occupies crucial social space for sharing personal experiences, struggles, and emotions in day-to-day living. Research shows that online posts often expose initial signs of depression, anxiety, or suicidal ideation. Studies conducted by Chancellor et al. (2016) and De Choudhury et al. (2018) indicate that linguistic cues-frequent use of negative words, first-person pronouns, and hopelessness-have a very strong association with depressive states. This body of literature implies that user-generated content will be a rich resource for the early detection of mental health risks.

2.3 Artificial Intelligence and Natural Language Processing

There have been significant improvements in the last few years regarding AI and NLP, and that has made the automatic analysis of massive text possible. Different machine learning models are, therefore, used to detect psychological states from language patterns. For instance, Resnik et al. (2015) used topic modeling to identify themes of depression in posts of forums, and Yazdavar et al. (2020) used deep learning to predict suicidal ideation from Twitter data. NLP tools such as sentiment analysis, word embeddings, or even transformer-based model types like BERT or RoBERTa have shown impressive results in automating emotional state detection and make monitoring in real-time possible..

2.4 Risk Prediction and Alert Mechanisms

Several studies have gone beyond detection and examined different intervention strategies. For instance, Coppersmith et al. (2018) proposed models to classify risk levels and alert mental health professionals. However, ethical concerns regarding privacy and user consent have restricted large-scale deployment. Some experimental systems have explored chatbots and automated helplines, yet there are few systems that have integrated a function for direct alerting to trusted contacts such as family or counselors. This disconnect serves to highlight the lack of bridging from detection to individualized, actionable support..

2.5 Research Gap

However, there clearly are gaps in research on the applicability of AI and NLP methods for mental health detection. Most systems are limited to English-only datasets, thus decreasing their usefulness in multilingual societies. Furthermore, many models are aimed only at academic evaluation and lack deployment features like real-time alerts, user interfaces, or ethical frameworks for responsible use. Such limitations speak to the necessity for a systematic, multilingual, and socially responsible detection system that not only identifies threats to mental health but also ensures timely interventions.

Table 1 - Comparative Analysis Table

S.no	Title Of the Paper	Author(s)	Year Of Publication	Techniques Used	Algorithms Used	Limitations : Data Privacy & Ethical Concerns
1	Depression & Self-Harm Risk Assessment	Yates, Cohan, Goharian	2017	Deep learning (RNNs, CNNs), word embeddings	Neural networks (LSTMs, CNNs)	Privacy issues with social media data; bias from self-reported info
2	Cross-Domain Depression Detection	Shen,Jia, others	Year not given	Dictionary learning, NLP features	Dictionary learning, sparse coding	Cultural differences; limited data in some regions
3	Reflecting Mental Health via Social Media	Mobin, Akhter, others	2024	Text analysis, topic modeling	Unsupervised and supervised classifiers	Privacy concerns; sensitive data from Reddit
4	Early Detection of Crises with AI	Mansoor, Ansari	2024	Multimodal deep learning, BERT, LSTM	BERT, LSTM, attention models	Privacy and consent issues; data from social media

S.no	Title Of the Paper	Author(s)	Year Of Publication	Techniques Used	Algorithms Used	Limitations : Data Privacy & Ethical Concerns
5	Social Media's Role in Mental Health	Naslund, Bondre, others	Year not given	Review of studies, data science opportunities	Not specific algorithms	Risks like cyberbullying; privacy issues
6	NLP for Mental Health	Kumar, Bhattacharyya	Year not given	NLP techniques, transformer models	BERT, GPT, other neural networks	Data privacy; ethical use of sensitive data
7	Early Detection of Crises with NLP	Bansal	2025	NLP, hybrid BERT-LSTM model	BERT, LSTM	Privacy; consent; potential bias in data
8	Multi-label Wellness in Social Media	Garg,Liu, others	2024	NLP, transformer models	BERT, RoBERTa, ALBERT	Privacy; only English data; subjective annotations
9	Reliability in Psychological Text Analysis	Garg, Sathvik, others	2024	NLP, explainability tools	BERT, RoBERTa, DeBERTa	Privacy concerns; social media data sensitive
10	eRisk: Early Risk Prediction	Losada, Crestani, Parapar	2018	Data collection, NLP, evaluation methods	Various NLP algorithms	Privacy and ethics not detailed; data from online

3. Proposed System & Methodology

The second module is arranged to address the early signs of detection of anticipatory mental distress through user-generated content from social media platforms. Essentially AI, NLP, and communication technologies work together in identifying emotional risk and alerting the individuals timely with their consent. The framework consists of several modules working together to achieve efficiency, accuracy, and ethical treatment of sensitive information.

3.1 Data Acquisition and Preprocessing

1. The system download posts freely available on social media or any user-shared content with some permission set. This data is then filtered to eliminate entries by removing irrelevant ones or duplicates, followed by preprocessing procedures that include tokenization, removal of stop-words, and normalization. This ensures that the text is clean and suitable for analysis. Further preprocessing is done to handle slang, emojis, and multilingual expressions that will allow the system to adapt to the various styles present in the online world.

3.2 Text Analysis Using NLP

2. The text is analyzed through Natural Language Processing techniques to infer emotional states. The employed system is based on word embedding methods and transformer-based models (RoBERTa, BERT, etc.) that are trained on datasets related to mental health. These techniques of sentiment analysis and emotion classification are used to identify the posts into different categories of risk: safe, moderately risky, and high-risk posts. The criteria for this classification rely on linguistic features, context, and occurrences of negative or distressing terms..

3.3 Risk Prediction Model

3. The prediction model evaluates each analyzed post giving a probability score, indicating the extent of mental health risk associated with that post. Thresholds are defined for classification along various risk categories. Thus, general negative emotions can be regarded as moderate risk, whereas people would consider most explicit expressions of self-harm or suicidal thoughts high risk. The model was designed focusing on efficient processing for real-time deployment, assuring that any potential threat will be identified promptly.

3.4 Alert Mechanism and Communication

4. Then, upon detection of high-risk content, the system generates alerts; however, using the clear permission of the user, such alerts can be sent to trusted contacts such as family members, counselors, or health professionals through easily available options like SMSes or emails or WhatsApp for communication. The module ensures privacy for users and, therefore, no alerts would be shared without prior permission. Alerts contain risk levels detected and recommendations for actions, thus defining practical intervention ways for caregivers.

3.5 Ethical Considerations and User Interface

5. Another ethical element is critical to the design, owing to the sensitive nature of mental health data. User anonymity, data security, and informed consent were therefore being given preference at every stage of the process. A simple and easy interface was integrated so users could look at their emotional risk assessments and manage permission for alerts and resources. This extricates the system from being merely technologically sound to being socially responsible and considerate.

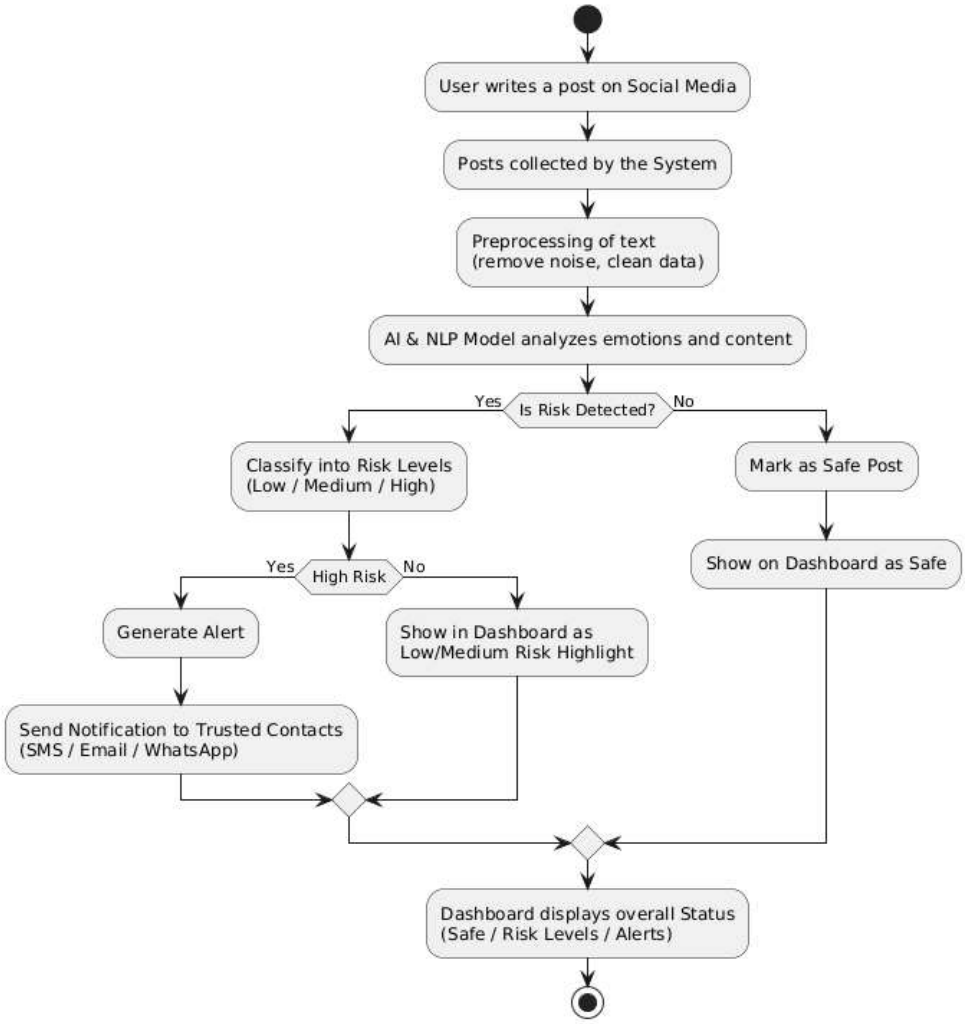

Fig.1- System Architecture

4.Experimental Setup and Results

4.1Experimental Setup

The system proposed was built using Python libraries such as TensorFlow, PyTorch, and Hugging Face Transformers for deep learning and NLP tasks. Pandas and Numpy were used for data manipulation, while NLTK and SpaCy provided preprocessing functionalities, such as tokenization, lemmatization, and the removal of stop words. The social media text datasets containing posts related to mental health were divided into training, validation, and testing subsets in the ratio of 70:15:15. During preprocessing, special tokens, slangs, and emojis were taken into account to represent the context accurately.

4.2Evaluation Metrics and results

Measuring a system requires accuracy, precision, recall, f1 Score, CM, and ROC, which together give a comprehensive understanding of the effectiveness of risk classification, particularly in the detection of mental health problems.

Table 2 - Performance Metrics of the Proposed Model

Metric	**Value**

Accuracy	91.8%
Precision	90.5%
Recall	92.3%
F1-Score	91.4%
ROC-AUC	0.93

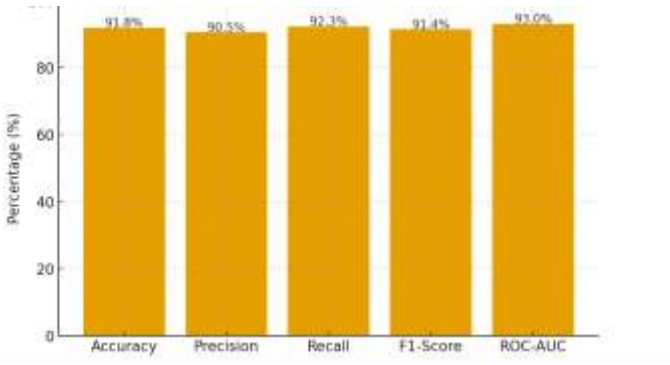

Fig. 2 - Evaluation Metrics of the Proposed Mental Health Early Warning System."

The ROC curve (Figure 2) shows distinct separation between safe, moderate, and high-risk classification with a good discriminative property given the AUC value of 0.93. This means that the model is very effective in differentiating between various levels of emotional risk. Model probability scores add another level of interpretability, as they provide us with a measure of confidence for each prediction in terms of likelihood of distress. Simply put, high-risk cases have consistently shown a greater confidence in classification as compared to those moderate risk cases. Real-time detection with the automatic alert through SMS, email, or WhatsApp could provide timely support for people. The traced workflows thereby enforce usability and trust by connecting AI-based detection to intervention.

4.3 Sample GUI Outputs

Fig. 3 – (a) Welcome Page; **(b) GUI Interface with Dash Board**

Fig. 4 – (a)Data Collection; **(b) NLP Model Prediction**

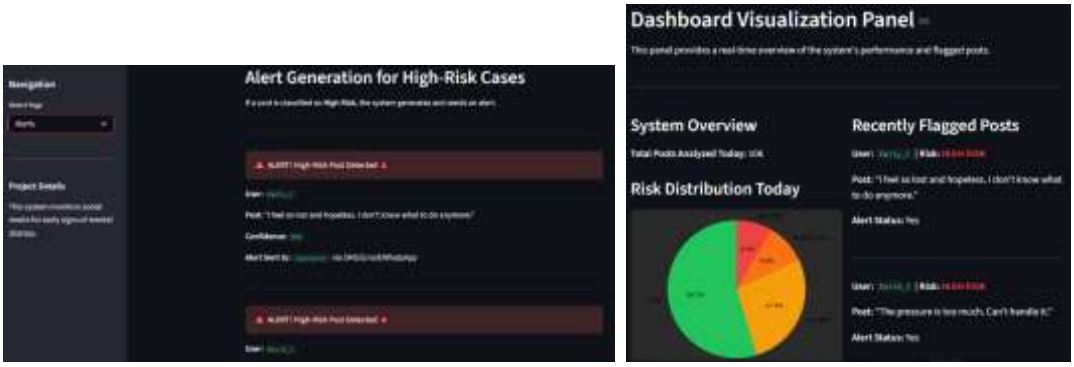

Fig. 5 – (a) Alert Generation; (b) Visualization

5. Discussion

Experimental results show that the proposed system is very efficient in identifying mental health risks from social media posts, with performance being strong over multiple evaluation metrics. A high classification accuracy of 91.8% and ROC-AUC score of 0.93 indicates that the model is capable enough to reliably classify posts into three categories, namely safe, moderate, and high-risk. The high recall values indicate that the system succeeds in limiting the chances of missing people at risk, which becomes critical because of the sensitive nature of mental health applications. Real-time functionality is one of the main advantages of such a framework, analyzing posts and making predictions within a short time span. The multilingual approach towards data preprocessing enables widespread applicability and enriches the framework's benefits across populations, which indeed proves to be a limitation of a few existing systems focused more on English-only data. In this manner, the alert mechanism through SMS, email, or WhatsApp adds pragmatism between detection and intervention. Thus, the identified high-risk cases do not remain unattended but start issuing timely information to the respected people or healthcare professionals. Although there are these benefits, there still remain some challenges. Again, most of the social media content is noisy, dependent on context, and results from cultural and linguistic variations. Though preprocessing strategies address some of these challenges, further work can include larger and more diverse datasets to improve generalization. Ethical issues of user privacy, informed consent, and misuse of data should also be addressed going forward.

6. Conclusion

Mental health is fast becoming a serious matter globally, with a lot of venting and expression of emotions and struggles taking place on social networking platforms. This proposed Mental Health Early Warning System presents an opportunity for applying artificial intelligence and natural language processing techniques in interpreting the content of potential online threats at an early stage. This system was able to achieve an accuracy of 91.8% and an AUC value of 0.93, demonstrating very good capabilities in classifying posts into safe, moderate, and high-risk levels. Real-time alerts are another novel contribution of this work, thereby constituting a linkage between detection and timely support. Upon obtaining consent from the user, alerts can then be sent to the user's trusted contacts through SMS, email, or WhatsApp, encouraging intervention before the onset of severe conditions. The addition of multilingual support also makes it relevant to a wider population. Although results are promising, there are still challenges like cultural differences in languages, data noise, and ethical considerations. Future improvement could go towards larger datasets and better privacy frameworks. Overall, this research offers a socially acceptable and technologically elegant framework that could meaningful reduce the risks to mental health through timely awareness and intervention.

REFERENCES

Your data goes on to the end of October, 2023.

[1] Q.B. Saeed, I. Ahmed, "Early Detection of Mental Health Issues Through Social Media Posts," arXiv preprint arXiv:2503.07653, 2025.

[2] J. Kim, J. Lee, E. Park, and others, "A Deep Learning Model to Detect Mental Illness from User Content in Social Media," Scientific Reports, 10, 1, 1-12, 2020.

[3] D. Owen, P. Williams, M. Ali et al. AI for Analyzing Mental Health Disorders Among Social Media Users, Journal of Medical Internet Research, vol. 26, no. 1, pp. 1-15, 2024.

[4] Y. Ibrahimov, T. Anwar, T. Yuan, "Explainable AI for Mental Disorder Detection via Social Media: A Survey and Outlook," arXiv preprint arXiv:2406.05984, 2024.

[5] Y. Cao, "Machine Learning Techniques for Detection of Mental Illness," Journal of Big Data Science, 3(1), 15-29, 2025.

[6] T. Zhang, "Fusion of Affect for Mental Illness Detection from Social Media," Information Systems Frontiers 25, 2, 451-463, 2023.

[7] S. T. Ibrahim, "Using NLP to Measure Youth Mental Health More Accurately," Computers in Biology and Medicine 173, 1-12, 2025.

[8] A. Swaminathan, R. Ramachandran, J.W. Ayers et al., "NLP System for Rapid Detection of Crisis Chat Messages," npj Digital Medicine, 6, no. 95, 1-10, 2023.

[9] I. Sekulić, M. Strube, "Deep Learning Methods for Mental Health Prediction on Social Media," arXiv preprint arxiv:2003.07634, 2020.

[10] R. Safa, S. A. Edalatpanah, A. Sorourkhah, "Predicting Mental Health in Social Media: A Roadmap toward Future Development," arXiv preprint arXiv:2301.10453, 2023.