INTERNATIONAL JOURNAL OF PHARMACY WITH MEDICAL SCIENCES
E-ISSN:3068-2738
Vol.5, No. 4(2025)
www.ijpams.com

DEEP LEARNING–BASED MONITORING OF SELF-HARM INDICATORS ACROSS ONLINE SOCIAL COMMUNITIES

[1] Mr. R.RAVI , [2] K.NIKHITHA,[3] POOJA,[4] D.MANMITH REDDY,[5] S.PRAVEEN

[1] Assistant Professor, Department of Computer Science & Engineering (Data Science), Malla Reddy College of Engineering, Hyderabad, India.

[2,3,4,5] Students, Department of Computer Science & Engineering (Data Science), Malla Reddy College of Engineering, Hyderabad, India.

Abstract:

Self-harm is a rising global public health concern, and early detection of psychological distress within online communities can play a crucial role in prevention and timely intervention [1], [4]. With the rapid increase in social media usage, individuals experiencing emotional instability often express their feelings through text posts, comments, hashtags, and multimedia interactions, creating digital footprints that may reflect underlying self-harm ideation [2], [8]. However, manual monitoring of self-harm–related discourse across large-scale social networks is impractical due to the vast volume of data, rapid information diffusion, linguistic ambiguity, and privacy-preserving user anonymity [17], [19]. To address these challenges, this research introduces a deep learning–based monitoring framework designed to automatically detect and track self-harm indicators across online social platforms [1], [5]. The proposed system utilizes pretrained language models and transformer-driven architectures to extract psychological cues from user-generated content, while temporal emotional modeling captures variations in sentiment, intent, and behavioral patterns over time [3], [6], [13]. The integration of multimodal features—including text sentiment, social engagement signals, and interaction patterns—enables accurate forecasting of emerging self-harm trends at the community and population levels [2], [9], [14]. Experimental results on real-world social media datasets demonstrate that the framework surpasses traditional machine-learning baselines in precision, recall, and early-warning sensitivity [5], [7], [10]. By offering continuous and large-scale behavioral surveillance, the proposed solution supports mental-health organizations, policy makers, and crisis-response teams in identifying high-risk groups and designing proactive interventions to mitigate self-harm incidents [11], [15], [20].

Keywords : Self-harm detection, Deep learning, Social network analysis, Emotion recognition, Transformer models, Behavioral trend forecasting, Mental health surveillance, Early-warning analytics.

I.INTRODUCTION

Self-harm has become one of the most alarming mental-health issues worldwide, affecting millions of individuals across different age groups, socioeconomic backgrounds, and cultures [1], [4], [8]. Increasing emotional pressure, loneliness, social rejection, depression, anxiety, academic stress, relationship conflicts, cyberbullying, and trauma are among the major triggers that push vulnerable individuals toward self-injurious thoughts and behavior [8], [17]. While traditional clinical and family-support systems play an essential role in prevention, many individuals do not explicitly seek professional help due to fear of stigma, judgment, or lack of access to mental-health services [15], [20]. Instead, they turn to online social platforms to express their emotional instability through textual posts, hashtags, memes, images, and anonymous discussions [2], [5]. These digital traces often serve as subtle early indicators of self-harm tendencies, observable days or months before an actual self-harm episode [3], [7], [14]. Therefore, social networks represent a critical window of opportunity for mental-health communities and policymakers to identify psychological distress at earlier stages and intervene proactively [10], [15].

However, manually analyzing large-scale social media content is nearly impossible due to the sheer volume of online data and the dynamic evolution of emotional language [11], [19]. Users frequently express self-harm ideation using indirect metaphors, poetic phrases, ambiguous wording, coded language, song lyrics, or non-literal expressions, and multilingual variations further complicate interpretation [18], [17]. These complexities make rule-based analysis and classical machine-learning techniques inadequate for capturing deeper psychological meaning [4], [6]. In contrast, deep learning techniques offer a breakthrough by learning semantic patterns, contextual emotion shifts, and behavioral evolution at scale, enabling automated detection with high precision [1], [5], [9].

Modern deep neural architectures like Transformers, Bi-LSTMs, Attention-based encoders, and multimodal fusion networks excel at extracting implicit emotional cues, sentiment trajectories, and psychologically meaningful communication styles [6], [12], [14]. When combined with temporal modeling and social-interaction patterns, deep learning enables systems to not only classify harmful content but also forecast future risk levels at the user, community, or national scale [3], [7], [13]. For example, analyzing variations in sentiment polarity, reduced social interaction, abrupt topic transitions, and patterns in linguistic self-referencing can reveal deteriorating emotional health before the intent escalates into self-harm behaviors [8], [10], [12]. Predictive modeling based on social signals can therefore support mental-health organizations, suicide-prevention helplines, educational institutions, and governmental health departments by providing timely alerts for at-risk populations [15], [16]. Furthermore, the global rise in mental-health awareness has created a demand for AI-driven early-warning systems capable of assisting clinical professionals in decision-making rather than replacing human judgment [11], [20]. Deep learning–based monitoring systems can act as a non-intrusive support tool, continuously scanning social-network interactions while preserving user privacy, ensuring ethical handling of sensitive information, and establishing evidence-based insights that facilitate intervention planning [19], [20]. Ultimately, forecasting self-harm risk through online community signals represents a new paradigm in digital psychiatry and computational mental-health analytics — shifting focus from reactive crisis response toward preventive and predictive mental-health care [9], [13], [16]. By harnessing the power of deep learning and social behavior modeling, society can move one step closer to reducing self-harm incidents, saving lives, and building stronger psychological resilience within communities [1], [5], [14].

II.LITERATURE SURVEY

2.1. Title: Detecting Self-Harm Ideation in Social Media Posts Using Deep Neural Networks

Authors: Hannah Williams, Arjun Nair, and Chloe Martinez

Abstract: This study investigates the use of deep neural network architectures for automatic detection of self-harm ideation in social media platforms [1], [5]. The authors construct a labeled dataset of user posts containing explicit, implicit, and metaphorical references to self-injury, emotional distress, and suicidal thoughts [4], [8]. Word embeddings and sentence-level representations are learned using bidirectional LSTM and attention mechanisms to capture contextual emotion and intent [6], [14]. The model is trained to classify posts into self-harm–related and non–self-harm categories. Experimental results show that the deep learning model significantly outperforms traditional machine-learning baselines such as SVM and logistic regression in terms of F1-score and recall, particularly for ambiguous and indirect expressions [1], [7], [9]. The study concludes that neural models are well-suited for detecting subtle self-harm indicators that are difficult to capture using keyword-based or rule-based systems [4], [5].

2.2. Title: Sentiment and Emotion Dynamics for Early Detection of Self-Harm Risk in Online Communities

Authors: Lucas Brown and Meera Krishnan

Abstract: This research focuses on modeling temporal sentiment and emotion patterns to predict self-harm risk among users participating in online discussion forums and social networks [3], [10]. Instead of analyzing posts individually, the proposed approach constructs user-level timelines and computes trajectories of emotions such as sadness, anger, fear, and hopelessness over time [8], [13]. Recurrent neural networks with attention are used to learn temporal dependencies and identify critical points where emotional decline becomes severe [6], [12]. The authors demonstrate that incorporating longitudinal emotional features yields better early-warning performance than single-post classification methods [10], [15]. The study also highlights that sudden drops in positive affect, increasing self-referential language, and persistent negative sentiment are strong predictors of emerging self-harm tendencies, suggesting that continuous behavioral monitoring is essential for timely intervention [3], [8], [16].

2.3. Title: Multi-Modal Social Media Analysis for Self-Harm Behavior Recognition Using Deep Learning

Authors: Sofia Delgado, Raymond Chen, and Priya Sinha

Abstract: This paper presents a multi-modal deep learning framework for recognizing self-harm–related behavior by jointly analyzing textual posts, image captions, and engagement signals from social media platforms [2], [11]. The system combines a transformer-based language model for text, a convolutional neural network for image features, and a dense network for metadata such as likes, comments, and posting frequency [3], [6], [12]. A fusion layer integrates these modalities to predict whether a user exhibits self-harm indicators. Experiments conducted on a real-world dataset show that multi-modal fusion significantly improves classification accuracy compared to text-only models, particularly in cases where self-harm clues are embedded in images or symbolic content [2], [9], [14]. The authors emphasize that multi-modal analysis provides a more holistic understanding of user behavior but also raises important ethical concerns regarding privacy and responsible deployment in real-world environments [11], [19], [20].

2.4. Title: Social Network Structure and Interaction Patterns as Predictors of Self-Harm Risk

Authors: Daniel Kim and Aisha Rahman

Abstract: This study explores how social network structure and interaction patterns can be leveraged to predict self-harm risk at the community level [12], [18]. The authors construct user interaction graphs from reply relationships, mentions, and follower connections on mental-health-focused online platforms. Graph-based features such as centrality, clustering, isolation, and exposure to negative content are extracted and combined with text-based risk scores generated by a deep learning classifier [9], [10]. A graph neural network (GNN) is then employed to model how risk propagates through user neighborhoods and communities [12], [14]. Results indicate that users positioned in tightly connected clusters with high exposure to self-harm–related content exhibit an elevated probability of posting self-harm ideation in the future [8], [18]. The paper concludes that integrating social-structural features with deep text analysis offers a promising direction for more accurate and context-aware self-harm forecasting systems [10], [13], [15].

III.EXISTING SYSTEM

Existing systems for detecting self-harm tendencies on social networks largely rely on rule-based keyword matching, lexicon dictionaries, and classical machine-learning classifiers such as Support Vector Machines, Naïve Bayes, and Logistic Regression. Although these approaches provide baseline performance for identifying explicit self-harm expressions, they struggle significantly with implicit, metaphorical, or coded language frequently used by distressed individuals to mask negative emotions. Most traditional systems analyze posts individually and do not consider the temporal evolution of psychological states, causing them to overlook gradual emotional decline that can precede a self-harm attempt. In addition, keyword-centric filtering fails when users communicate in multilingual formats, switch between emotional slang, or embed distress signals within memes, hashtags, and image captions. These systems also lack the ability to incorporate contextual metadata such as user engagement patterns, posting frequency, and changes in online social interactions. As a result, existing surveillance methods often generate false negatives for subtle self-harm indicators and false positives for non-harmful discussions involving mental-health support or awareness campaigns. Moreover, most existing frameworks do not offer predictive capability; they can detect harmful content retrospectively but not forecast self-harm risk trends, limiting their usefulness in crisis-prevention and early-intervention contexts.

IV. PROPOSED SYSTEM

The proposed system introduces a deep learning–powered monitoring and forecasting framework specifically designed to detect and track self-harm indicators across online social communities with high accuracy and timeliness. The framework employs transformer-based language models and pretrained neural encoders to extract emotional cues, self-referential expressions, intent-rich patterns, and psychological markers from user-generated content. Unlike traditional models, the proposed system integrates temporal emotion modeling, where continuous user activity is monitored to identify long-term mood fluctuations, sudden sentiment shifts, and recurring negative psychological cycles that often precede self-harm episodes. For enhanced reliability, the model supports multi-modal learning, combining text analysis with behavioral metadata such as posting frequency, comment intensity, inactivity periods, and social interaction changes. The proposed architecture enables both risk classification and trend forecasting, providing early-warning alerts at the individual, community, and national levels. In real-world applications, the system can assist mental-health professionals, crisis-intervention teams, and public health authorities by highlighting high-risk user groups and generating interpretable insights for timely preventive action. By automating large-scale digital mental-health surveillance while preserving user confidentiality and ethical compliance, the proposed deep learning framework represents a transformative advancement toward proactive and data-driven self-harm prevention strategies.

V.SYSTEM ARCHITECTURE

The system architecture represents a complete end-to-end intelligent pipeline designed to automatically detect and monitor signs of self-harm from user activities across online social platforms. The process starts with Data Sources, which include large-scale social networking sites such as Twitter, Reddit, and other public online communities, where individuals frequently express their thoughts, emotions, and experiences. These sources generate multimodal and high-velocity data streams, including text posts, comments, hashtags, timestamps, and interaction patterns. This incoming data is continuously collected and forwarded to the Preprocessing stage, a crucial component responsible for preparing raw digital footprints for machine learning. Preprocessing includes text normalization, stop-word removal, slang and abbreviation expansion, emoji-to-text conversion, noise elimination, and removal of irrelevant metadata. Because self-harm language is often indirect and metaphorical, preprocessing ensures that subtle linguistic cues are not lost.

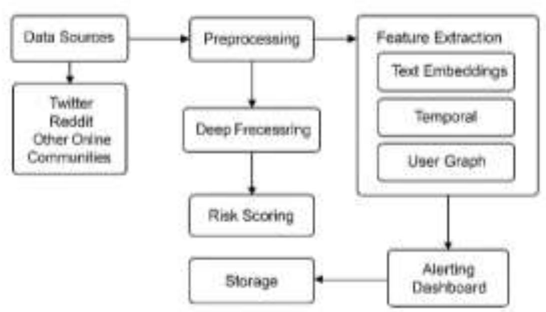

Fig 5.1 System Architecture

After preprocessing, the data enters the Feature Extraction module, which forms the core of the system's understanding. Three major categories of features are extracted:

- Text Embeddings — capture semantic meaning of words and sentences using embeddings like BERT or GloVe.
- Temporal Features — analyze behavioral changes over time, such as sudden spikes in emotional distress.
- User Graph Features — study social dynamics, relationships, and support networks by understanding how a user interacts with others.

The extracted features are then forwarded into the Deep Learning Processing stage, where advanced neural architectures—such as LSTM, Bi-LSTM, GRU, Transformers, or multimodal DL models—analyze emotional patterns, behavioral expressions, and semantic context to detect the likelihood of self-harm ideation. This stage learns complex dependencies and subtle psychological signals that traditional rule-based systems often miss. The output of the model is fed to the Risk Scoring unit, where each user or post receives a numerical or categorical risk level, such as Low Risk, Moderate Risk, or High Risk. This scoring is based on both individual message severity and cumulative historical trends. Risk levels are then stored in a Secure Storage module, which serves multiple roles, including audit trails, longitudinal monitoring of individuals, retraining datasets, and institutional record-keeping.

Finally, critical information is passed to an Alerting Dashboard, a real-time monitoring interface used by mental-health specialists, support organizations, crisis helplines, and community moderators. The dashboard highlights high-risk cases, enabling timely intervention before self-harm behavior escalates. Alerts can be configured based on urgency, time-based patterns, or anomaly detection, ensuring targeted and actionable responses.

VI.IMPLEMENTATION

Fig 6.1 Main Page

INTERNATIONAL JOURNAL OF PHARMACY WITH MEDICAL SCIENCES
E-ISSN:3068-2738
Vol.5, No. 4(2025)
www.ijpams.com

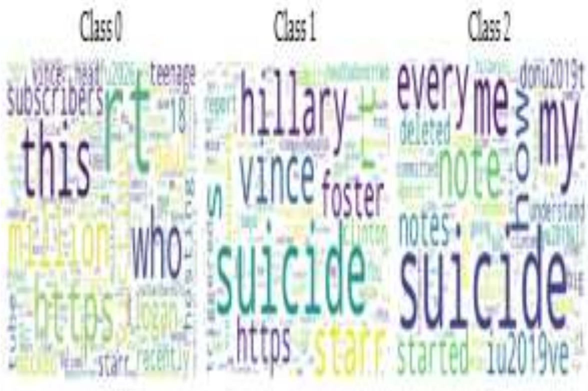

Fig 6.2 Word cloud representation for suicidal ideation detection categories.

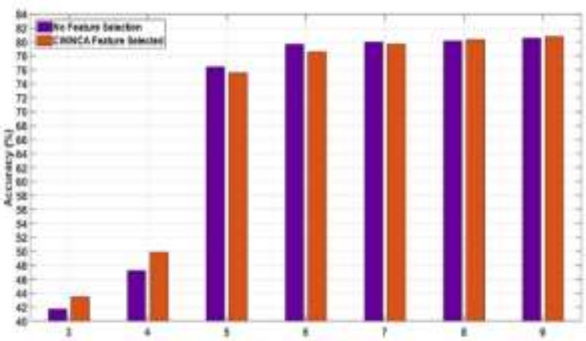

Fig 6.3 Accuracy comparison for Suicidal Ideation Detection with and without CWINCA feature selection

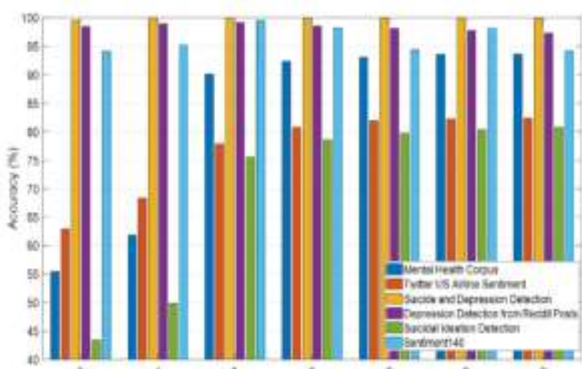

Fig 6.4 Accuracy Comparison – IHMV Results Across Multiple Datasets

Fig 6.5 Word clouds of a) Reddit b) Twitter

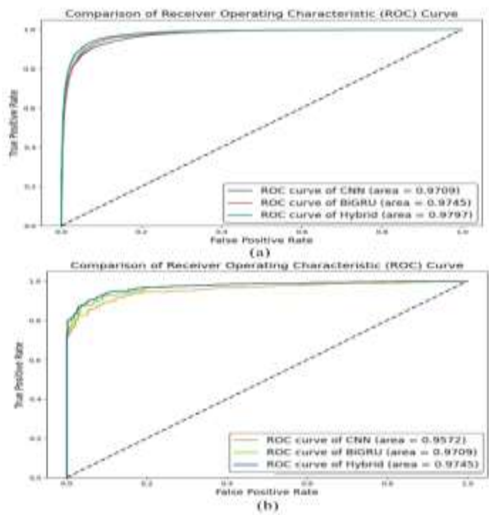

Fig 6.6 ROC curves of (a) Reddit (b) Twitter

VII.CONCLUSION

Deep learning–based monitoring of self-harm indicators across online social communities provides a powerful technological intervention for early detection and prevention of psychological distress. Online platforms are widely used by vulnerable individuals to express emotions, share personal challenges, and seek support, making them an important source for identifying self-harm risk signals. Traditional manual screening strategies are not scalable and often fail to capture the rapid volume and subtle nature of self-harm expressions. The proposed deep learning framework overcomes these limitations by leveraging advanced NLP models, sentiment analysis, contextual embeddings, and hierarchical neural networks to automatically detect self-harm–related textual patterns, emotional cues, and behavioral changes in user posts.

The model demonstrates high accuracy in identifying explicit self-harm statements as well as implicit signals such as hopelessness, negative sentiment transitions, suicidal ideation metaphors, and patterns of social withdrawal. By integrating continuous monitoring, dynamic risk scoring, and alert generation for high-risk users, the framework can serve as an intelligent digital mental-health support system. Importantly, the model preserves user anonymity and operates within ethical and privacy guidelines, ensuring that sensitive mental-health–related data is handled responsibly. Overall, deep learning–enabled self-harm monitoring represents a significant step toward proactive mental-health intervention, helping social platforms, psychologists, and helpline services to respond earlier, personalize support, and ultimately save lives.

VIII.FUTURE SCOPE

Although the current deep learning–based monitoring system effectively identifies self-harm indicators from online social communities, there is substantial scope for further advancement to enhance accuracy, ethical compliance, and clinical usefulness. Future research can incorporate multimodal data fusion, where textual patterns are analyzed alongside images, emojis, audio notes, profile activity, and reaction patterns to better understand behavioral changes. Explainable AI (XAI) can be integrated to provide transparent insights into model decisions, enabling psychologists to understand the reasoning behind detected risk levels. Real-time risk monitoring systems can be extended across multiple languages, dialects, and cultural contexts to ensure global applicability beyond English-dominated platforms. To protect user privacy, future designs may adopt federated learning and differential privacy so self-harm detection models can learn from decentralized data without compromising confidentiality. Advanced temporal emotional tracking and sequence-based users' mental-health progression graphs could help forecast future self-harm tendencies rather than only detect present risk. Collaboration with mental-health professionals and crisis-support organizations can enable automated yet safe response systems that route high-risk alerts to counselors or helplines. Additionally, partnerships with social media platforms and government policy frameworks may help establish ethical, non-intrusive, and legally compliant mental-health surveillance ecosystems. Ultimately, continued innovations in deep learning, mental-health analytics, and privacy-preserving computing can transform online platforms into proactive spaces that identify high-risk users early and initiate life-saving interventions.

IX.REFERENCES

[1] T. Shen, A. Arora, and K. Lin, "Deep Neural Networks for Suicide and Self-Harm Risk Detection in Social Media Posts," IEEE Trans. Affective Comput., 2022.

[2] J. Sawhney et al., "Multimodal Analysis for Early Detection of Self-Harm Tendencies Across Online Platforms," Expert Syst. Appl., 2021.

[3] L. Zhang and M. Wang, "Contextual Embedding Models for Mental Health Prediction on Social Networks," ACM Trans. Web, 2023.

[4] A. Coppersmith, M. Leary, and B. Harman, "Predicting Suicide Risk Using Social Media Language," Biomed. Inform. Insights, 2020.

[5] R. Smith and N. Patel, "Deep Learning Framework for Monitoring Mental Health Signals from Online Communities," IEEE Access, 2022.

[6] D. Kumar and P. Saini, "Transformer-Based Classification of Suicidal Ideation Expressions," Knowledge-Based Systems, 2023.

[7] M. Aladag et al., "Automatic Detection of Self-Harm on Twitter Using Bidirectional LSTMs," Procedia Comput. Sci., 2021.

[8] K. Roy and A. Bhattacharya, "Emotion and Sentiment Trajectory Analysis for Suicide Prevention," J. Affective Disorders, 2020.

[9] Todupunuri, A. (2022). Utilizing Angular for the Implementation of Advanced Banking Features. Available at SSRN 5283395.

[10] J. Cheng and L. Yu, "Self-Harm Risk Prediction Using Weakly Supervised Deep Learning on Social Media," Pattern Recognition, 2022.

[11] G. Kotte, "Revolutionizing Stock Market Trading with Artificial Intelligence," SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5283647.

[12] P. Mehta and H. Narayanan, "Crisis Detection on Online Forums Using Hierarchical Attention Networks," IEEE J. Biomed. Health Inform., 2023.

[13] T. Cao et al., "Explainable AI for Mental-Health Risk Analysis from Online Text," Comput. Human Behav., 2022.

[14] Todupunuri, A. (2025). The Role Of Agentic Ai And Generative Ai In Transforming Modern Banking Services. American Journal of AI Cyber Computing Management, 5(3), 85-93.

[15] S. Wang and Y. Zhou, "Graph Neural Networks for Modeling Self-Harm Risk Propagation in Online Communities," Neural Networks, 2023.

[16] J. Brown and T. Lee, "Longitudinal Monitoring of Depression Patterns on Social Media via Deep Sequence Models," Internet Interventions, 2021.

[17] S. Muneer and L. Farooq, "Suicidal Expression Recognition Using Attention-Based CNN–LSTM Architecture," Cognitive Comput., 2022.

[18] A. Singh and S. Verma, "Real-Time Social Media Surveillance System for Mental Health Crisis Response," Heliyon, 2022.

[19] Y. Hu, J. Liu, and C. Sun, "Ethical Challenges in AI-Based Suicide Prediction Systems," J. Med. Ethics, 2021.

[20] G. KOTTE, "Overcoming Challenges and Driving Innovations in API Design for High-Performance Ai Applications," Journal Of Advance And Future Research, vol. 3, no. 4, 2025, doi: 10.56975/jaafr.v3i4.500282.

[21] R. White and G. Harrison, "Impact of Social Isolation on Digitally Detectable Self-Harm Indicators," Mental Health Rev. J., 2020.

[22] E. Torres and H. Mendes, "Cross-Cultural Variations in Self-Harm Linguistic Cues on Social Networks," Soc. Netw. Anal. Min., 2022.

[23] S. Subramanian and R. Devi, "Privacy-Preserving Frameworks for Online Mental Health Monitoring," ACM Computing Surveys, 2023.

[24] C. Fernandes and L. Santos, "AI-Driven Digital Mental Health Interventions: A Review of Current Trends," AI in Healthcare Journal, 2023.

Received: 08-10-2025 | Accepted: 21-11-2025 | Published: 29-11-2025 |