

Prediction of Suicidal Ideation Using Deep Learning Models

Vipin Kumar Sahu¹ and Dr. Brijendra Gupta²

Department of Information Technology^{1,2}

Siddhant College of Engineering Pune, Maharashtra, India

Abstract: *In recent years, the incidence of fatalities resulting from suicide has risen. Suicide is increasingly a leading cause of mortality globally. This growth has resulted in a perilous scenario, that jeopardises human life. Numerous investigations have been conducted to ascertain the causes of such suicides and their prevention. Literature indicates that early diagnosis of suicidal ideation can aid in saving lives. The concept of early detection has prompted several scholars to conduct studies in this area. Numerous research studies have employed machine learning and deep learning models to forecast suicidal ideation. This paper examines the current research conducted on the identification of suicidal ideation utilising machine learning and deep learning methodologies.*

Keywords: Machine Learning, Artificial Intelligence, Natural Language Processing

I. INTRODUCTION

Suicide remains one of the world's most severe public health challenges, taking more than 700,000 lives every year and leaving millions of others struggling with suicidal thoughts or tendencies (WHO, 2023). Identifying suicidal ideation—continuous reflections or fears related to self-harm—at an early stage is vital to ensuring timely care and prevention. Traditional assessment techniques such as mental health evaluations and clinical interviews, while important, often suffer from subjectivity, limited reach, and delayed response times. In contrast, the growing use of machine learning (ML) and deep learning (DL) methods has demonstrated remarkable potential in automatically recognizing people at risk by studying their behaviour, written communication, and activity on social media. In [1], several ML algorithms were analyzed for detecting suicidal ideation, with a focus on supervised models like Support Vector Machines (SVM), decision trees, and ensemble approaches that capture both linguistic and behavioural features. Expanding upon that work, [2] compared natural language processing (NLP), ML, and DL approaches using social media data. Their findings revealed that deep learning networks such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) performed considerably better than traditional systems in understanding emotional and contextual expressions. Likewise, [3] integrated hybrid DL architectures into the analysis of social media content and user metadata, achieving superior accuracy and recall in identifying suicidal intent. Recent literature underscores the growing importance of creating AI systems that are not only accurate but also interpretable and ethically responsible. For instance, [4] reviewed emerging ML-based approaches and highlighted the significance of explainable AI along with responsible data handling. Moreover, [5] presented a new framework that merges ML with genetic algorithms for optimised feature selection, resulting in more reliable and robust models. Collectively, these studies illustrate a strong movement toward combining deep learning, optimization techniques, and explainable AI to enhance suicide risk prediction. However, persistent challenges—such as skewed datasets, privacy issues, and limited generalisation—still demand further research to develop trustworthy, transparent, and inclusive AI systems for mental health prediction. In recent years, the incidence of fatalities resulting from suicide has risen. Suicide is increasingly a leading cause of mortality globally. This growth has resulted in a perilous scenario, jeopardizing human life. Numerous investigations have been conducted to ascertain the causes of such suicides and their prevention. Literature indicates that early diagnosis of suicidal ideation can aid in saving lives. The concept of early detection has prompted several scholars to conduct studies in this area. Numerous research studies have employed machine learning and deep learning models to forecast suicidal ideation. This paper examines the current research conducted on the identification of suicidal ideation



utilizing machine learning and deep learning methodologies. In recent years, the number of suicide-related deaths has risen significantly, making suicide one of the leading causes of mortality worldwide. This alarming trend poses a serious threat to human life and well-being. Numerous studies have sought to understand the underlying causes of suicide and explore effective prevention strategies. Research in this area emphasizes that early detection of suicidal thoughts can play a crucial role in saving lives. Motivated by this, many researchers have focused on developing models capable of identifying suicidal ideation at an early stage. Accordingly, this paper reviews existing studies that employ machine learning (ML) and deep learning (DL) techniques for the detection and prediction of suicidal thoughts.

II. LITERATURE SURVEY

A. Suicidal Ideation Detection (SID) Techniques

Suicidal Ideation Detection (SID) employs computational models to identify individuals at risk by analysing textual, behavioural, and multimodal signals. Recently, both machine learning (ML) and deep learning (DL) approaches have become prominent in this domain, offering scalable and automated alternatives to conventional clinical assessments.

Classical Machine Learning Approaches

Traditional ML techniques remain widely used for SID, particularly in analysing textual and behavioural data with handcrafted features. Support Vector Machines (SVM), Random Forests (RF), Logistic Regression (LR), and Decision Trees (DT) are some of the most common algorithms [1][2][6]. Feature engineering typically relies on linguistic markers, psycholinguistic cues, sentiment scores, and n-grams extracted from social media posts or clinical records. For example, Ji et al. [1] showed that SVM and ensemble classifiers can tell the difference between suicidal and non-suicidal text, and Haque et al. [2] showed that classical ML models give clear results at a lower cost. However, these methods often face challenges in capturing complex contextual and semantic nuances in large datasets.

Deep Learning Approaches

Deep learning models automatically learn hierarchical features, enabling more context-aware analysis than classical ML. Popular architectures include Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks for sequential text modelling [3][8]; Convolutional Neural Networks (CNNs) for extracting local patterns in text or multimodal data [12]; and transformer-based models like BERT and RoBERTa, which capture long-range contextual dependencies [2][13]. For example, Aldhyani et al. [3] applied CNN and LSTM models to Reddit posts, achieving higher recall than traditional ML methods. Verma et al. [13] also looked into hybrid sequential-transformer models to improve detection. While deep learning excels with large-scale, unstructured data, it often requires significant computational resources and may lack transparency.

Hybrid and Multimodal Techniques

Hybrid and multimodal SID techniques combine multiple approaches or diverse data sources to boost detection performance. Hybrid models integrate ML and DL, such as applying SVM or RF on features extracted from neural networks [5][14]. Multimodal models incorporate text, audio, images, and behavioural indicators to capture richer signals of suicidal ideation [12][16]. Explainable AI (XAI) tools, such as attention mechanisms in transformers, help identify which features or phrases most influence predictions, improving interpretability [4][5]. For instance, Basyouni et al. [5] combined ML classifiers with genetic algorithms for feature optimization, while Li et al. [7] used NLP and ML on clinical interviews for cross-domain detection. These approaches represent the forefront of SID research, prioritising accuracy, interpretability, and integration of diverse data types.

B. Literature survey

Ji, Shaoxiong, et al. (2020) [1] review various machine learning approaches for detecting suicidal ideation (SI) in text and social media data. It categorises techniques into classical ML, deep learning, and hybrid methods, discussing their effectiveness in early SI identification. The study highlights challenges such as data scarcity, privacy issues, and feature extraction limitations. It also emphasises the importance of integrating multimodal data to improve detection accuracy. The paper serves as a foundational reference for SI detection research.

Haque, Rezaul, et al. (2022) [2] the authors perform a comparative analysis of NLP, machine learning, and deep learning techniques for suicidal ideation detection. The study evaluates performance metrics such as precision, recall,



and F1-score across different approaches. It identifies deep learning models as more robust for capturing complex textual patterns, while classical ML techniques are faster and computationally efficient. The paper also highlights challenges like imbalanced datasets and noisy social media text.

Aldhyani, Theyazn HH, et al. (2022) [3] the study focuses on detecting suicidal ideation on social media using both machine learning and deep learning models. It explores feature extraction methods, including text embeddings and sentiment analysis, to improve prediction accuracy. The authors compare the performance of several classifiers, showing that ensemble deep learning models outperform traditional ML models. Ethical concerns and data privacy issues are also discussed.

Abdulsalam, Asma, and Areej Alhothali (2024) [4] review recent machine learning techniques for suicidal ideation detection on social media platforms. It categorises approaches into supervised, unsupervised, and hybrid models, discussing their relative advantages. The study highlights challenges such as limited labelled data, annotation bias, and the need for context-aware models. The authors suggest that integrating multiple modalities and real-time monitoring could improve early intervention strategies.

Basyouni, Abdallah, et al. (2024) [5] the authors propose a framework combining machine learning and genetic algorithms to detect suicidal ideation on social media. Feature selection using genetic algorithms improves model performance and reduces computational complexity. Various classifiers are tested, showing significant improvements in prediction accuracy. The paper emphasises the potential of optimization techniques in enhancing SI detection.

Yeskuatov, Eldar, et al. (2022) [6] – the review examines Reddit as a primary source for suicidal ideation detection using ML and NLP techniques. It highlights the importance of domain-specific preprocessing and semantic analysis for accurate detection. The study notes challenges in noisy, unstructured data and the need for ethical considerations. Recommendations include using ensemble methods and transformer-based models for improved performance.

Li, Tim MH, et al. (2023) [7] conducted a study on the use of NLP and machine learning to detect suicidal ideation in clinical interview transcripts. It demonstrates that automated methods can effectively supplement clinical assessment, especially in identifying subtle linguistic cues. The study compares multiple ML models and shows that deep learning-based NLP methods achieve higher detection accuracy. The paper also addresses clinical implications and potential integration into healthcare systems.

Renjith, Shini, et al. (2022) [8] the authors propose an ensemble deep learning method for detecting suicidal ideation in social media posts. Their approach combines multiple neural network architectures to capture different textual features. Experimental results show significant improvements over individual models in terms of accuracy and F1-score. The paper also highlights challenges related to imbalanced datasets and context understanding.

Macalli, Melissa, et al. (2021) [9] focus on predicting suicidal thoughts and behaviour among college students using machine learning. Various models are trained on survey and behavioral data to identify high-risk individuals. Results suggest that ML techniques can effectively support early intervention strategies. The paper also discusses ethical considerations and the need for longitudinal data to improve predictive reliability.

In [10] it reviews ML techniques for suicidal ideation detection on social media, highlighting challenges like limited labelled data and annotation bias. The study recommends hybrid and multimodal approaches for improved detection accuracy and real-time monitoring.

Table 1: Comparison table for the literature review

Title	Methodology	Algorithm Used	Limitations
[11] Sawhney, Ramit, et al. "Towards suicide ideation detection through online conversational context." (2022)	Contextual analysis of online conversations	NLP, LSTM, Bi-LSTM	Limited to conversational datasets; may not generalize to other social media platforms
[12] Chatterjee, Moumita, et al. "Suicide ideation detection from online social media: A multi-modal feature based technique." (2022)	Multi-modal feature extraction from text, images, and metadata	Random Forest, SVM, CNN	Complexity in fusing multi-modal features; computationally intensive



[13] Verma, Aniket, et al. "Suicide ideation detection: a comparative study of sequential and transformer hybrid algorithms." (2022)	Sequential and transformer hybrid model comparison	LSTM + Transformer, BERT-based models	Requires large labeled datasets; high training time
[14] Shah, Faisal Muhammad, et al. "A hybridized feature extraction approach to suicidal ideation detection from social media post." (2020)	Feature engineering + hybrid classification	SVM, Naive Bayes, Logistic Regression	Limited feature representation; may miss contextual nuances
[15] Liu, Dexi, et al. "Suicidal ideation cause extraction from social texts." (2020)	Cause identification in text	Rule-based NLP, ML classifiers	Dependence on manually defined rules; may not capture implicit causes
[16] Sardari, Sara, et al. "Audio based depression detection using convolutional autoencoder." (2022)	Audio signal analysis	Convolutional Autoencoder (CAE), CNN	Limited to speech data; background noise affects performance
[17] Bhadra, Sweta, and Chandan Jyoti Kumar. "Enhancing the efficacy of depression detection system using optimal feature selection from EHR." (2024)	EHR data-based feature selection	Random Forest, Gradient Boosting	Data privacy issues; model performance depends on quality of EHR data
[18] Ansari, Luna, et al. "Ensemble hybrid learning methods for automated depression detection." (2022)	Ensemble hybrid learning	Random Forest + XGBoost + SVM	Computational complexity; needs large datasets for optimal performance
[19] Marriwala, Nikhil, and Deepti Chaudhary. "A hybrid model for depression detection using deep learning." (2023)	Deep learning hybrid model	CNN + LSTM	High computational requirements; risk of overfitting on small datasets
[20] Khan, Shakir, and Salihah Alqahtani. "Hybrid machine learning models to detect signs of depression." (2024)	Hybrid ML for depression detection	SVM + Random Forest + ANN	Performance depends on feature quality; may not capture temporal patterns

C. Gap Analysis

- Datasets are highly imbalanced; fewer SI cases lead to biased models and poor real-world performance [1,2,8].
- Proposed a multimodal (text, image, metadata) model, but integrating audio or behavioural cues remains a challenge [12].
- Highlight conversational context importance, but most models ignore temporal links between messages [11,13].
- Few systems operate in real time; most models are tested offline only [4,6].
- Language and cultural bias persist, with datasets mainly in English and Western-focused [1,10].
- Deep learning achieves high accuracy but lacks explainability for clinical use [8,19].
- Models often fail on new domains; cross-domain adaptation is limited [7,9]
- Explored audio-based depression detection; non-verbal cues are still underused in SI studies [16].
- Analysed SI causes, but most models detect suicidal content only without causal reasoning [15].
- Lack of standardised datasets and metrics makes comparison difficult [2,13].



- Hybrid/ensemble models were explored, but combining rule-based and ML approaches needs improvement [5,18].
- Integration of clinical and social media data is rare but promising [7,17].
- Need proactive systems to alert professionals or offer support beyond detection [1,4]

III. CHALLENGES IN SUICIDAL IDEATION DETECTION

Despite significant progress in applying machine and deep learning methods to suicidal ideation (SI) detection, several critical challenges persist that hinder real-world implementation and scalability

Table II: Key Challenges in Suicidal Ideation Detection

Sr. No.	Challenges	References
1	Data imbalance leading to biased models and poor real-world generalization	[1], [2], [8]
2	Limited integration of multimodal data (text, image, audio, behavioral cues)	[12], [6], [16]
3	Lack of contextual and temporal understanding in posts and conversations	[11], [13], [8]
4	Models designed for offline evaluation rather than real-time monitoring	[4], [6], [11]
5	Language and cultural bias — dominance of English and Western datasets	[1], [10], [4]
6	Lack of model transparency and explainability in deep learning systems	[8], [19], [5]
7	Poor cross-domain generalization across platforms or demographics	[9], [7], [13]
8	Underutilization of non-verbal cues such as tone, pauses, or physiological data	[16], [11]
9	Limited exploration of causal inference to identify root causes of SI	[15], [1]
10	Privacy, consent, and ethical issues in collecting user-generated data	[1], [2], [4]
11	Absence of standardized datasets and evaluation benchmarks	[2], [13], [12]
12	Inefficient hybrid and ensemble model integration	[5], [18], [19]
13	Static modeling of suicidal ideation ignoring behavioral changes over time	[1], [8], [11]
14	Lack of integration between clinical (EHR) and social media data	[7], [17], [9]
15	Focus on detection rather than proactive intervention or real-time alerting	[1], [4], [10]

IV. APPLICATIONS OF MACHINE LEARNING TECHNIQUES ON DIFFERENT DATASETS

Machine learning (ML) has been extensively applied to diverse datasets to detect suicidal ideation (SI), depression, and related mental health risks. These datasets vary in nature, including social media texts, clinical interviews, electronic health records (EHRs), and multimodal datasets integrating audio, image, and behavioral cues.

Table III: Overview of Machine Learning Applications on Different Datasets

Authors	Dataset	ML Technique / Algorithm
Ji et al. [1]	Social media (Reddit, Twitter)	SVM, RF, CNN, LSTM
Haque et al. [2]	Social media text	NLP, ML, DL
Aldhyani et al. [3]	Twitter, Reddit	CNN, LSTM, RF
Abdulsalam & Alhothali [4]	Social media	ML & DL review
Basyouni et al. [5]	Reddit & Twitter	ML + Genetic Algorithms
Yeskuatov et al. [6]	Reddit	ML, NLP
Li et al. [7]	Clinical interviews	NLP + ML (SVM, RF)
Renjith et al. [8]	Social media posts	Ensemble DL (CNN + LSTM)
Macalli et al. [9]	College surveys	SVM, RF, ANN
Abdulsalam & Alhothali [10]	Social media	ML & DL review
Sawhney et al. [11]	Conversational online data	LSTM, Contextual ML
Chatterjee et al. [12]	Text + Image + Metadata	Multimodal ML
Verma et al. [13]	Social media posts	Sequential + Transformer hybrid
Shah et al. [14]	Social media posts	Hybrid feature extraction + ML



Liu et al. [15]	Social text	NLP + ML
Sardari et al. [16]	Audio datasets	Convolutional Autoencoder
Bhadra & Kumar [17]	HER	Optimal feature selection + ML
Ansari et al. [18]	EHR + social data	Ensemble Hybrid Learning
Marriwala & Chaudhary [19]	Clinical + sensor data	Deep Learning Hybrid Models
Khan & Alqahtani [20]	Multimodal clinical datasets	Hybrid ML models

V. TRENDS IN SUICIDAL IDEATION DETECTION

- Text-based data from social media posts, comments, or clinical transcripts heavily influences most existing research.
- A few studies have tried to improve detection accuracy by combining different types of information—such as text, images, audio, and metadata—but these multimodal systems are still rare.
- Alongside this, the focus is shifting from static, offline analysis toward dynamic frameworks capable of real-time SI detection on live platforms.
- Building explainable and interpretable models that healthcare professionals can trust is becoming increasingly important.
- Efforts are also underway to create shared benchmark datasets and standardised evaluation metrics that allow fair comparisons and reproducibility across different studies.
- Finally, growing awareness around ethics has led to a stronger focus on privacy, informed consent, and the responsible handling of sensitive mental health information, ensuring that detection technologies remain both effective and respectful of user rights.

VI. FUTURE RESEARCH DIRECTIONS IN SUICIDAL IDEATION DETECTION

- Combine textual (images and videos), visual (audio), and behavioural data to enhance detection accuracy and minimise false positives.
- Develop real-time monitoring systems capable of continuously analysing social media activity to identify risks early and enable timely interventions.
- Broaden datasets and models to include multiple languages and cultural contexts, ensuring the global relevance and fairness of SI detection systems.
- Prioritise interpretable and transparent models to strengthen clinical reliability and ethical accountability.
- Improve model performance by integrating rule-based approaches with machine learning and deep learning methods for more robust predictions.
- Merge social media insights with clinical data sources such as electronic health records (EHRs), interviews, and psychological surveys to form a comprehensive view of mental health.
- Design detection systems that not only identify suicidal ideation but also notify professionals or provide immediate access to mental health resources.
- Incorporate conversational flow, behavioural cues, and temporal patterns to better understand how suicidal thoughts evolve over time.
- Establish widely accepted datasets, evaluation metrics, and benchmarking standards to ensure consistency and reproducibility across studies.
- Ensure the secure, transparent, and ethical handling of sensitive mental health data in compliance with privacy and data protection regulations.
- Tailor models to individual user behaviour, demographic factors, and historical data for more accurate and context-aware predictions.
- Deploy lightweight SI detection models on mobile applications or edge devices to increase accessibility and real-time user engagement.



- Move beyond detection to explore the underlying causes, triggers, and risk factors associated with suicidal ideation.
- Develop systems that can update and adapt to changing online behaviours, linguistic trends, and emerging data patterns over time.
- Combine AI-based detection mechanisms with therapeutic tools and mental health support systems to provide holistic care and continuous emotional support.

VII. CONCLUSION

Suicidal ideation poses a serious threat to human life, and significant efforts are being made to identify and prevent such thoughts to save lives. This paper reviews the various approaches the research community has developed for detecting suicidal ideation. Most of these studies rely on machine learning and deep learning techniques for feature extraction, using different data sources and evaluating classifier performance. Evidence indicates that online social networking platforms serve as a significant repository of data for the identification of psychological health concerns, including anxiety, depression, and suicidal inclinations. Conventional methods, such as clinical interviews and self-report questionnaires, cannot entirely supplant these online sources but can function as supplementary data. Future research is likely to benefit from integrating data from both social media and traditional sources, providing a more holistic and effective approach to mental health monitoring and prevention.

REFERENCES

- [1]. Ji, Shaoxiong, et al. "Suicidal ideation detection: A review of machine learning methods and applications." *IEEE Transactions on Computational Social Systems* 8.1 (2020): 214-226.
- [2]. Haque, Rezaul, et al. "A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning." *Technologies* 10.3 (2022): 57.
- [3]. Aldhyani, Theyazn HH, et al. "Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models." *International journal of environmental research and public health* 19.19 (2022): 12635.
- [4]. Abdulsalam, Asma, and Areej Alhothali. "Suicidal ideation detection on social media: A review of machine learning methods." *Social Network Analysis and Mining* 14.1 (2024): 188.
- [5]. Basyouni, Abdallah, et al. "A suicidal ideation detection framework on social media using machine learning and genetic algorithms." *IEEE Access* 12 (2024): 124816-124833.
- [6]. Yeskuatov, Eldar, Sook-Ling Chua, and Lee Kien Foo. "Leveraging reddit for suicidal ideation detection: A review of machine learning and natural language processing techniques." *International journal of environmental research and public health* 19.16 (2022): 10347.
- [7]. Li, Tim MH, et al. "Detection of suicidal ideation in clinical interviews for depression using natural language processing and machine learning: cross-sectional study." *JMIR medical informatics* 11.1 (2023): e50221.
- [8]. Renjith, Shini, et al. "An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms." *Journal of King Saud University-Computer and Information Sciences* 34.10 (2022): 9564-9575.
- [9]. Macalli, Melissa, et al. "A machine learning approach for predicting suicidal thoughts and behaviours among college students." *Scientific reports* 11.1 (2021): 11363.
- [10]. Abdulsalam, Asma, and Areej Alhothali. "Suicidal ideation detection on social media: A review of machine learning methods." *Social Network Analysis and Mining* 14.1 (2024): 188.
- [11]. Sawhney, Ramit, et al. "Towards suicide ideation detection through online conversational context." *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*. 2022.
- [12]. Chatterjee, Moumita, et al. "Suicide ideation detection from online social media: A multi-modal feature based technique." *International Journal of Information Management Data Insights* 2.2 (2022): 100103.



- [13]. Verma, Aniket, et al. "Suicide ideation detection: a comparative study of sequential and transformer hybrid algorithms." the International Conference on data Science and emerging technologies. Singapore: Springer Nature Singapore, 2022.
- [14]. Shah, Faisal Muhammad, et al. "A hybridized feature extraction approach to suicidal ideation detection from social media post." 2020 IEEE Region 10 Symposium (TENSYP). IEEE, 2020.
- [15]. Liu, Dexi, et al. "Suicidal ideation cause extraction from social texts." Ieee Access 8 (2020): 169333-169351.
- [16]. Sardari, Sara, et al. "Audio based depression detection using convolutional autoencoder." Expert Systems with Applications 189 (2022): 116076.
- [17]. Bhadra, Sweta, and Chandan Jyoti Kumar. "Enhancing the efficacy of depression detection system using optimal feature selection from EHR." Computer methods in biomechanics and biomedical engineering 27.2 (2024): 222-236.
- [18]. Ansari, Luna, et al. "Ensemble hybrid learning methods for automated depression detection." IEEE transactions on computational social systems 10.1 (2022): 211-219.
- [19]. Marriwala, Nikhil, and Deepti Chaudhary. "A hybrid model for depression detection using deep learning." Measurement: Sensors 25 (2023): 100587.
- [20]. Khan, Shakir, and Salihah Alqahtani. "Hybrid machine learning models to detect signs of depression." Multimedia Tools and Applications 83.13 (2024): 38819-38837.

