Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/71/20241627

Suicidal ideation detection on social media using machine learning: A review

Zhuo Sheng

School of Computer Science, Nanjing University of Posts and Telecommunications, 9 Wenyuan Road, Nanjing, Jiangsu, 210023, China

b23040326@njupt.edu.cn

Abstract. Suicide remains a critical global health issue, with over 800,000 deaths annually and profound impacts on individuals and their communities. The increasing role of social media platforms in expressing personal struggles has highlighted the potential of these platforms for early detection of suicidal tendencies. Advances in machine learning and natural language processing (NLP) have revolutionized this detection process, allowing for more accurate identification of suicidal ideation in social media posts. This review examines the integration of machine learning algorithms in suicide detection, including the use of multi-layered neural networks for analyzing textual, audio, and visual data. It explores various machine learning models and their performance in predicting suicide attempts, with a focus on their accuracy, precision, and recall. The review also discusses applications such as real-time monitoring systems and emergency response mechanisms, emphasizing how these technologies can enhance early intervention and support. By surveying current research and identifying challenges and future directions, this review aims to provide a comprehensive understanding of how machine learning can significantly contribute to suicide prevention efforts through effective social media monitoring.

Keywords: Suicidal Ideation Detection, Social Media, Machine Learning.

1. Introduction

According to the World Health Organization (WHO) [1], approximately 16 million suicide attempts occur annually, with over 800,000 people succumbing to suicide, suggesting a truly alarming rate of about one death every 40 seconds. Suicide was identified as the fourth leading cause of death in the 2019 report by WHO, attesting to its tragic pervasiveness. Moreover, the repercussions of an individual's suicide extend far beyond the person, profoundly affecting families, friends, colleagues, and beyond. Research indicates that for every suicide death, over 20 individuals could be contemplating similar actions [2, 3].

The advent of social networks has significantly changed how people, particularly the younger generation, communicate their feelings. Platforms like Facebook, Instagram, and Twitter have become venues for sharing personal struggles and seeking support, often revealing textual cues indicative of suicidal tendencies. It is estimated that more than 20% of suicide attempters and approximately 50% of those who complete suicide leave behind notes [4]. This underscores the critical importance of early

DOI: 10.54254/2755-2721/71/20241627

detection of suicidal intent in social media postings, as it can lead to timely interventions that save lives [5-8].

Recent advances in machine learning, a subset of artificial intelligence, have revolutionized natural language processing. By employing multi-layered neural networks, machine learning can identify intricate patterns within textual data, enabling more accurate predictions. This technology is pivotal in the early identification of suicidal tendencies in social media posts, potentially leading to life-saving interventions. The machine learning for suicide attempt detection has been studied by many researchers. Bernert et al. [9] identified and summarized original reports employing use of an AI/ML framework to predict suicidal behaviors as an outcome of risk according to systematic review. Linthicum et al. [10] introduced machine learning and its potential application to open questions in suicide research. Sanchez et al. [11] aims to identify the machine learning techniques used to predict suicide risk based on information posted on social networks. Ji et al. [12] summarized the limitations of current work and provided an outlook of further research directions. Kusuma et al. [2] evaluated how well machine learning models can predict long-term outcomes related to suicide, such as thoughts of suicide, attempts, and deaths. They also looked at different types of outcomes, data, and models to see how these factors might affect the models' performance. D'Hotman and Loh [13] presented a qualitative narrative review of research focusing on two categories of suicide prediction tools: medical suicide prediction and social suicide prediction. McHugh et al. [14] summarized recent studies on predicting suicide, including those using machine learning, to understand the value of these new approaches. Their research highlighted the scalability and robustness of machine learning models in handling diverse datasets and achieving state-of-the-art performance in suicide ideation detection [8, 15-19].

This review aims to explore the intersection of machine learning and the detection of suicidal intentions. By surveying the current research landscape, identifying challenges, and highlighting potential future directions, this review seeks to provide a comprehensive understanding of how machine learning can contribute to more effective suicide prevention efforts through social media monitoring.

2. Literature Statistics

Figure 1 shows the number of papers searched using "suicide detection" and "machine learning" per year on Google Scholar. The relevant literature has increased steadily from 2010 to 2023. The number of papers has risen from about 1,530 in 2010 to a peak of 8,400 papers in 2023. This indicates a growing interest in detecting suicidal ideation on social media using machine learning.

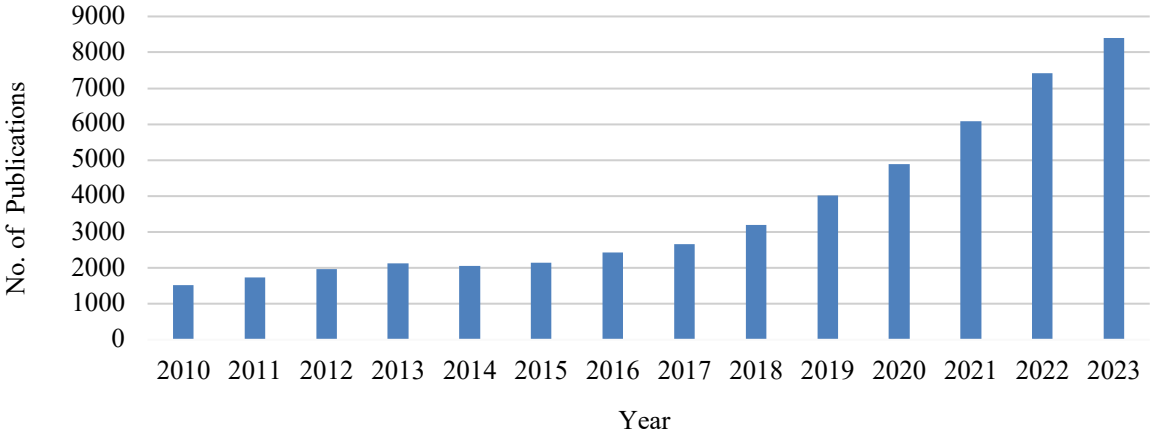

Figure 1. The number of papers searched using "suicide detection" and "machine learning" per year

DOI: 10.54254/2755-2721/71/20241627

3. Natural Language Processing for Suicide Ideation Detection on Social Media

Currently, diagnosing suicidal ideation is mainly done by psychiatrists, which can be both expensive and time-consuming. Patients usually interact with psychiatrists either face-to-face or online, allowing the professionals to assess for suicidal tendencies [20]. Additionally, some agencies distribute psychological questionnaires that psychiatrists analyze to make their final judgments. However, these traditional methods often prove to be ineffective. Moreover, individuals with psychological issues might avoid seeing psychiatrists, causing delays in detecting suicidal ideation [21].

Nowadays, people frequently share their feelings on social media platforms like Facebook and Twitter, creating an opportunity to assess suicidal risk based on their posts. With the rapid advancements in Natural Language Processing (NLP) techniques, we can now identify patterns in language used on social media with greater accuracy.

Many researchers have made significant strides in this area. For instance, Desmet et al. [22] developed a method for analyzing suicide notes using binary Support Vector Machine (SVM) classifiers to detect suicidal ideation. Huang et al. [23] created a psychological lexicon based on a Chinese sentiment dictionary (Hownet) and used the SVM approach to classify posts, developing a real-time suicide ideation detection system for Sino Weibo. O'Dea et al. [24] demonstrated the feasibility of distinguishing levels of concern in suicide-related posts by using both human coding and automatic machine learning classifiers, such as SVM and Logistic Regression, on TF-IDF features. These studies suggest that deep learning can be effectively applied to detect suicidal ideation on social media.

4. Machine Learning Models for Suicide Attempt Detection

Machine learning models for detecting suicide attempts utilize diverse data sources and sophisticated algorithms to identify individuals at risk. These models integrate clinical data, including electronic health records (EHR), patient history, and medication records, with behavioral data from social media activity, text messages, and phone calls. They also incorporate psychological assessments and demographic information to form a comprehensive dataset. Key features extracted for analysis include sentiment analysis, keyword extraction from text data, temporal patterns in social media use or hospital visits, and clinical indicators such as previous mental health diagnoses or substance abuse history. This multi-faceted approach enhances the accuracy and reliability of identifying at-risk individuals, facilitating timely interventions and support.

Table 1. model performance for suicide attempt detection

Study	Model	Accuracy	Precision	Recall	F1 Score
Aldhyani et al. [7]	CNN-BiLSTM	95%	-	-	-
	XGBoost	91.50%	-	-	-
Aladag et al. [25]	Logistic Regression	80%-92%	-	-	-
Haque et al. [26]	Random Forest (RF)	93%	-	-	0.92
	BiLSTM	93.60%	-	-	0.93
Jain et al. [27]	Logistic Regression	77.29%	-	-	0.77
	Naive Bayes	74.35%	-	-	0.74
	Support Vector Machine (SVM)	77.12%	-	-	0.77
	Random Forest (RF)	77.30%	-	-	0.77
Sawhney et al. [28]	Recurrent Neural Network (RNN)	73.70%	0.72	0.817	0.753
	Long Short-Term Memory (LSTM)	78.90%	0.745	0.874	0.796
	CNN-LSTM	81.20%	0.787	0.872	0.827

Commonly used algorithms include logistic regression, random forests, support vector machines (SVM), and deep learning techniques such as recurrent neural networks (RNN) and convolutional neural

DOI: 10.54254/2755-2721/71/20241627

networks (CNN). These models aim to provide timely and accurate predictions, facilitating early intervention and support for at-risk individuals [8, 29, 30].

Table 1 provides an overview of the performance of various machine learning and deep learning models for detecting suicidal ideation, using metrics such as accuracy, precision, recall, and F1 score. The studies compared include models like CNN-BiLSTM, XGBoost, Logistic Regression, Random Forest, BiLSTM, Naive Bayes, Support Vector Machine (SVM), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and CNN-LSTM. The results show that deep learning models, particularly CNN-BiLSTM and CNN-LSTM, generally outperform traditional machine learning models, with CNN-BiLSTM reaching the highest accuracy of 95%.

5. Applications

Machine learning techniques offer promising applications for detecting suicidal ideation on social media platforms by analyzing textual, audio, and visual content. By leveraging natural language processing (NLP) models, such as transformers, deep learning can identify patterns and linguistic cues indicative of suicidal thoughts. These models can be trained on large datasets containing labeled examples of distressing and non-distressing content to enhance their accuracy. Additionally, integrating sentiment analysis and emotion recognition can further refine detection capabilities. Advanced methods like multimodal machine learning, which combines text with images and audio, can provide a more comprehensive understanding of the user's emotional state, potentially leading to more effective interventions and support mechanisms. Some instances are listed as follows:

1. Real-time monitoring system: A real-time monitoring system on social media platforms enhance the detection of suicidal ideation by continuously analyzing user posts and interactions. This system would use machine learning models to scan for patterns and signals associated with suicidal thoughts, providing a proactive approach to identifying at-risk individuals. When the system encounters uncertainty in its assessments, it would promptly alert mental health professionals for further evaluation and intervention. This integration of automated analysis with expert human judgment seeks to offer timely and precise support, improving outcomes for those in need [31, 32].

2. Emergency response mechanisms: When the real-time monitoring system identifies a potential suicide attempt, it will promptly notify local police, emergency services, and the user's listed emergency contact. By utilizing the user's IP address to determine their location, the system enables rapid dispatch of law enforcement to intervene if the user is at imminent risk (Castillo-Sánchez et al. [10]). In addition to immediate intervention, the system ensures that psychological support and guidance are provided, aiming to offer comprehensive care and assistance during this critical period. This approach integrates technology with emergency response to address urgent situations effectively.

6. Challenges and Limitations

Several challenges remain in detecting suicidal ideation on social media using machine learning:

1. Data privacy and ethical concerns: The collection and analysis of social media data for detecting suicidal ideation engender privacy and ethical issues. It is necessary to establish strict ethical guidelines that prioritize user consent and robust data protection measures. These guidelines should address the nuances of data usage, ensuring transparency about how data is collected, analyzed, and stored [10].

2. Language and cultural differences: Variations in language and cultural contexts significantly influence how suicidal expressions are conveyed, presenting challenges for the creation of universally applicable detection models. To address these variations, it is crucial to tailor models specifically to accommodate such differences. This customization demands an intricate understanding of the linguistic and cultural peculiarities inherent to each demographic. Incorporating region-specific slang, idioms, and cultural nuances into the training data is part of this process. Furthermore, collaboration with local experts and the use of culturally relevant datasets can increase the model's precision across diverse populations, ensuring broader applicability and more dependable detection.

DOI: 10.54254/2755-2721/71/20241627

7. Summary

Suicide is a major global health issue with significant impacts on individuals and their communities. Social media platforms have emerged as crucial sources for detecting suicidal ideation due to the growing trend of individuals sharing personal struggles online. Advances in machine learning and natural language processing (NLP) have enhanced the ability to identify suicidal tendencies in social media posts with greater accuracy. This review explores how machine learning models, such as CNN-BiLSTM and CNN-LSTM, outperform traditional methods in predicting suicide attempts. It also discusses applications like real-time monitoring systems and emergency response mechanisms, which combine automated analysis with expert intervention to improve early detection and support.

References

[1] World Health Organization. (2019). Suicide in the world: Global health estimates. World Health Organization.

[2] Kusuma, K., Larsen, M., Quiroz, J. C., Gillies, M., Burnett, A., Qian, J., & Torok, M. (2022). The performance of machine learning models in predicting suicidal ideation, attempts, and deaths: A meta-analysis and systematic review. Journal of Psychiatric Research, 155, 579-588.

[3] Heckler, W. F., de Carvalho, J. V., & Barbosa, J. L. (2022). Machine learning for suicidal ideation identification: A systematic literature review. Computers in Human Behavior, 128, 107095.

[4] DeJong, T. M., Overholser, J. C., & Stockmeier, C. A. (2010). Apples to oranges?: A direct comparison between suicide attempters and suicide completers. Journal of Affective Disorders, 124(1-2), 90-97.

[5] Nordin, N., Zainol, Z., Noor, M. H., & Chan, L. F. (2022). Suicidal behaviour prediction models using machine learning techniques: A systematic review. Artificial Intelligence in Medicine, 132, 102395.

[6] Rabani, S. T., Khan, Q. R., & Khanday, A. M. (2020). Detection of suicidal ideation on Twitter using machine learning & ensemble approaches. Baghdad Science Journal, 17(4), 1328-1328.

[7] Aldhyani, T. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. A. (2022). 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), 12635.

[8] Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of suicide ideation in social media forums using deep learning. Algorithms, 13(1), 7.

[9] Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., Shah, N. H., & Abnousi, F. (2020). Artificial intelligence and suicide prevention: A systematic review of machine learning investigations. International Journal of Environmental Research and Public Health, 17(16), 5929.

[10] Linthicum, K. P., Schafer, K. M., & Ribeiro, J. D. (2019). Machine learning in suicide science: Applications and ethics. Behavioral Sciences & the Law, 37(3), 214-222.

[11] Castillo-Sánchez, G., Marques, G., Dorronzoro, E., Rivera-Romero, O., Franco-Martín, M., & De la Torre-Díez, I. (2020). Suicide risk assessment using machine learning and social networks: A scoping review. Journal of Medical Systems, 44(12), 205.

[12] Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. IEEE Transactions on Computational Social Systems, 8(1), 214-226.

[13] D'Hotman, D., & Loh, E. (2020). AI enabled suicide prediction tools: A qualitative narrative review. BMJ Health & Care Informatics, 27(3).

[14] McHugh, C. M., & Large, M. M. (2020). Can machine-learning methods really help predict suicide?. Current Opinion in Psychiatry, 33(4), 369-374.

[15] Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. In Fundamentals of Artificial Intelligence (pp. 603-649). Springer.

[16] Jones, K. S. (1994). Natural language processing: A historical review. In Current Issues in Computational Linguistics: In Honour of Don Walker (pp. 3-16).

DOI: 10.54254/2755-2721/71/20241627

[17] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. Multimedia Tools and Applications, 82(3), 3713-3744.
[18] Kalyanathaya, K. P., Akila, D., & Rajesh, P. (2019). Advances in natural language processing–a survey of current research trends, development tools and industry applications. International Journal of Recent Technology and Engineering, 7(5C), 199-202.
[19] Torous, J., Larsen, M. E., Depp, C., Cosco, T. D., Barnett, I., Nock, M. K., & Firth, J. (2018). Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: A review of current progress and next steps. Current Psychiatry Reports, 20(7), 1-6.
[20] Hawton, K. (1987). Assessment of suicide risk. The British Journal of Psychiatry, 150(2), 145-153.
[21] Bertolote, J. M., Fleischmann, A., De Leo, D., & Wasserman, D. (2004). Psychiatric diagnoses and suicide: Revisiting the evidence. Crisis, 25(4), 147-155.
[22] Desmet, B., & Hoste, V. (2013). Emotion detection in suicide notes. Expert Systems with Applications, 40(16), 6351-6358.
[23] Huang, X., Zhang, L., Chiu, D., Liu, T., Li, X., & Zhu, T. (2014). Detecting suicidal ideation in Chinese microblogs with psychological lexicons. In 2014 IEEE 11th International Conference on Ubiquitous Intelligence and Computing and 2014 IEEE 11th International Conference on Autonomic and Trusted Computing and 2014 IEEE 14th International Conference on Scalable Computing and Communications and Its Associated Workshops (pp. 844-849). IEEE.
[24] O'Dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., & Christensen, H. (2015). Detecting suicidality on Twitter. Internet Interventions, 2(2), 183-188.
[25] Aladağ, A. E., Muderrisoglu, S., Akbas, N. B., Zahmacioglu, O., & Bingol, H. O. (2018). Detecting suicidal ideation on forums: Proof-of-concept study. Journal of Medical Internet Research, 20(6), e9840.
[26] Haque, R., Islam, N., Islam, M., & Ahsan, M. M. (2022). A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning. Technologies, 10(3), 57.
[27] Jain, P., Srinivas, K. R., & Vichare, A. (2022). Depression and suicide analysis using machine learning and NLP. In Journal of Physics: Conference Series (Vol. 2161, No. 1, p. 012034). IOP Publishing.
[28] Sawhney, R., Manchanda, P., Mathur, P., Shah, R., & Singh, R. (2018). Exploring and learning suicidal ideation connotations on social media with deep learning. In Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 167-175).
[29] Mbarek, A., Jamoussi, S., Charfi, A., & Hamadou, A. B. (2019). Suicidal profiles detection in Twitter. In Proceedings of WEBIST (pp. 289-296).
[30] Kodati, D., & Tene, R. (2023). Identifying suicidal emotions on social media through transformer-based deep learning. Applied Intelligence, 53(10), 11885-11917.
[31] Ryu, S., Lee, H., Lee, D. K., Kim, S. W., & Kim, C. E. (2019). Detection of suicide attempters among suicide ideators using machine learning. Psychiatry Investigation, 16(8), 588.
[32] Ophir, Y., Tikochinski, R., Asterhan, C. S., Sisso, I., & Reichart, R. (2020). Deep neural networks detect suicide risk from textual Facebook posts. Scientific Reports, 10(1), 16685.