Suicidal Ideation Detection from social media: A Detailed Review of Machine Learning and Deep Learning

K.Senil Seby [1], Dr.M.Elamparithi[2], Dr.V.Anuratha[2]

[1] Research Scholar, Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Bharathiar University, Coimbatore, Tamilnadu, India

[2] Associate Professor, Department of Computer Science, Kamalam College of Arts and Science, Anthiyur, Bharathiar University, Coimbatore, Tamilnadu, India

Abstract: Social networks are crucial tools for learning about people's attitudes towards various issues since they allow them to express their thoughts to friends and family. Recent years have seen considerable challenges in natural language processing (NLP) and psychology regarding the detection of suicidal ideation via online social network analysis. The complex early signs of suicide ideations can be identified with the proper use of social media information, which can afterwards save many lives. Even though numerous strategies have been used over the past few decades to identify suicidal thoughts, machine learning (ML) and deep learning (DL) methods offer more insightful results. So, in this study, we examine several cutting-edge ML and DL approaches for detecting suicidal ideation. We also observe a few significant issues in the entire corpus of literature that could be explored in further research. Finally, we hope this study will shed light on critical issues in identifying suicidal thoughts on social media for readers and ML and DL researchers.

Keywords: Suicidal ideation Detection, Social Media, Machine Learning, Deep Learning, Word Embedding.

1. INTRODUCTION

One of the major social problems is suicide. According to the World Health Organization (WHO), over 700 million people worldwide commit suicide yearly. Many more people, particularly those in their 20s and 30s, also make suicide attempts. The second most common cause of death for those between the ages of 10 and 34 is suicide. The prevalence of suicide is rising quickly due to mental health issues such as anxiety and depression [1]. Social media today has drawn billions of individuals to engage in interpersonal communication. Today, many people regularly engage in such media to socialize with one another and exchange thoughts, observations, and personal experiences [2]. As a result, the link between social media and suicide has drawn attention as a public health issue. For instance, a study found that eight out of ten individuals express suicidal ideation thoughts online.

Consequently, the research focus has switched to using social media to analyze mental health and identify at-risk human beings [3]. For various reasons, many people prefer to write blogs about their feelings or plans to commit suicide because they find it difficult to discuss their personal stories and express their emotions in person. Sadly, people frequently disregard or overlook these suicide posts. The use of this information can facilitate widespread suicidality screening. The term "suicidal ideation" refers to suicidal intentions, whether they are explicit or ambiguous, that are more prevalent than suicide deaths and that might foretell future suicide attempts. ML has become a potential method for identifying at-risk persons and preventing suicide [4]. Most earlier studies examined the relationship between suicidal thoughts and lexical or posting characteristics of the contents (i.e., frequency of posting on social media platforms). However, ML models like support vector machines (SVM), k-nearest neighbour (KNN), random forests (RF), logistic regression (LR), Naïve Bayes (NB), etc., categorize tweets by the severity of the distress, the sort of concern expressed, or the style of suicide communication. However, the ML models do not work well when the dataset is significant, and they also need human interventions to perform the feature learning process in suicidal ideation detection. They are not impervious to errors and inaccuracies, especially in dynamic and complicated situations. Factors like noisy or imperfect data, biased training sets, or unforeseen events may cause erroneous predictions. Recent initiatives have tried to categorize tweets using DL models, which perform

exceptionally well. Word embedding and DL continue to succeed, as seen by their superior performance in predicting suicidal thoughts compared to more conventional ML systems. This motivates us to present a survey of recent ML and DL-based models for suicide ideation detection in social media platforms.

The remainder of this survey is divided into sections in the following sequence. Section 2 covers the details of the background information of suicidal ideation. Section 3 reviews methods used to detect suicidal ideation. Section 4 and 5 presents the discussion and conclusion with future directions.

2. BACKGROUND INFORMATION

Suicidal ideation is a growing problem in society that knows no boundaries. Recent observations show an apparent increase in the severity of this issue. People worldwide are using social media as a solid platform to share their feelings and views. Many social networking sites encourage suicidal behaviour or thoughts, especially in people who are suffering from a mental health issue. In general, there are four stages to the suicidal ideation detection process: data collection, data preprocessing, word embedding, and classification. The architecture diagram of suicidal ideation from social media posts is shown in Figure 1.
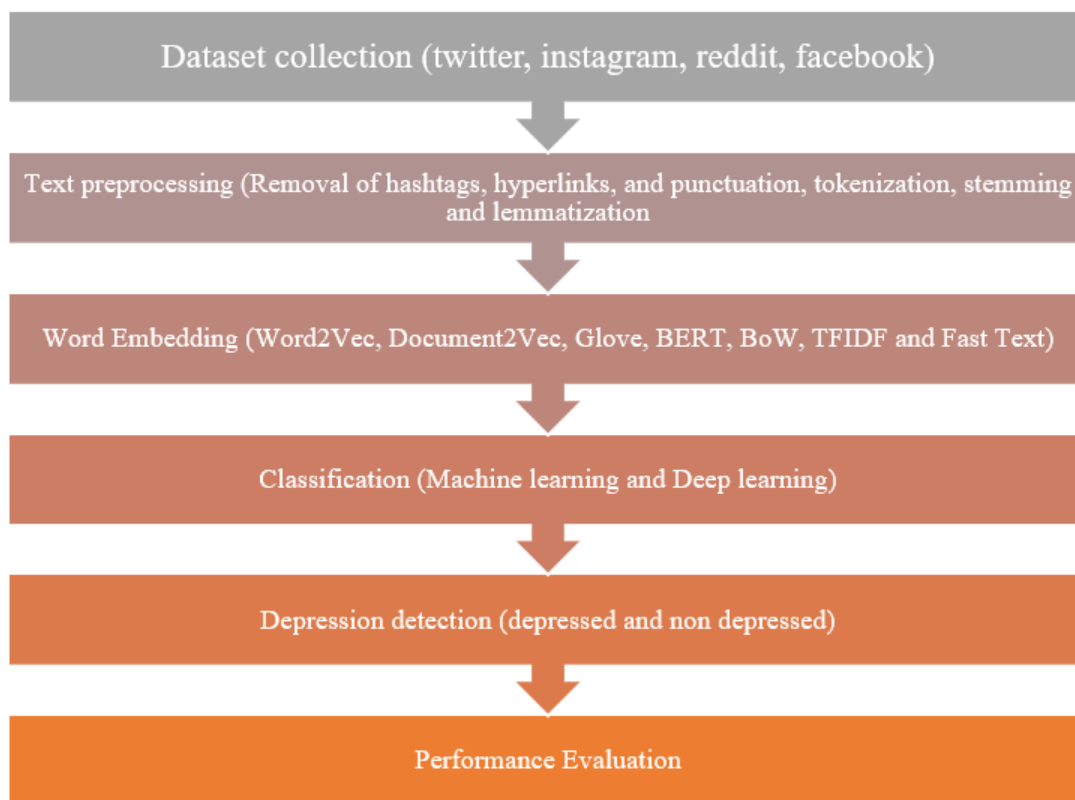
Figure 1: Suicidal Ideation Detection System

2.1 Data Collection

The first step in the analysis process is data collecting because we are required to train our classifiers with the collected social media data. Many researchers can learn people's thoughts and opinions via social media sites like Twitter, Reddit, Facebook, Instagram, and others. To find posts related to suicide, several information sets, social media platforms, and data sources were examined. One of the critical issues is preparing social network data to determine whether or not it contains depression or suicide-related content. To address this problem, most researchers employ NCapture, an effective tool for data collection. Its goal is to make it possible to organize, deconstruct, and find knowledge in unstructured data sources like social media, interviews, open-ended survey replies, articles, and web content. Additionally, it provides a location to organize and manage information so that knowledge may be discovered more effectively.

2.2 Data Preprocessing

The acquired dataset will subsequently move into the data preprocessing stage. Data preprocessing is a procedure that transforms the form of acquired data into an understandable form. The research involved data preprocessing, which included data cleaning such as removal of usernames, punctuation, hyperlinks, commas, hashtags, and whitespace, tokenization, lemmatization and stemming [5]. Inflected words are reduced to their root form through stemming. The process of lemmatization returns inflected words to their primary form.

2.3 Word Embedding

After preprocessing, the word embedding stage is crucial for predicting suicidal ideation. Word vectorization, also known as word embedding, transforms or maps the text or words to real-valued vectors [6]. The network can learn the meaning of each word in the people's thoughts. There are numerous embedding techniques, such as TF-IDF, Word2Vec, Fast Text, Global vector (GloVe), N-gram feature, word sentiment polarity score representation, Twitter-specific feature representation, temporal feature vector, non-temporal feature vector, etc., available for scoring each word. Even though several embedding methods and technologies are used, each has its drawbacks. Most authors attempt to address these shortcomings by combining two embedding techniques, which improved model results and demonstrated the potential for extracting features from tweets while improving classification performance.

2.4 Classification

Classification is the last step for detecting suicidal thoughts. The classification system aims to classify social media messages as suicidal or non-suicidal. ML and DL models are crucial techniques in suicide research for automatically identifying suicidal ideation and preventing suicide [7, 8]. A powerful way to identify suicidal thoughts from social media posts is ML. NB, LR, SVM, RF, KNN, and DT are popular ML approaches. Features are defined and extracted in this ML model manually or using feature selection techniques. With outstanding outcomes, DL models can learn traits and classify suicidal ideation. The neural network's hidden layers are modified by DL using a multilayer technique. Deep Neural Network, Long Short-Term Memory (LSTM), and Convolution Neural Network (CNN) are the most often used DL models, producing superior outcomes.

3. REVIEW ON METHODS USED FOR SUICIDAL IDEATION DETECTION

Due to the recent increase in suicide rates, some researchers have become more concerned with suicidal identification. The communication of so many different topics is overwhelming, complicating the causes of suicide. There has been a significant amount of research on suicide ideation on social media texts. Table 1 summarizes the recent research works published for suicidal ideation detection using ML and DL with several word embedding approaches.

Table 1: Summary of methods for suicidal ideation detection

Author & Ref No.	Methodology used	Dataset used	Outcomes	Benefits	Limitations
Moumita Chatterjee *et al.* [9]	ML classifiers such as LR, RF, SVM, and extreme gradient boosting (XGB)	Twitter dataset	Accuracy=0.87%, Precision=0.83%, recall=0.82%, f1-score=0.841%, and Cohen kappa=0.78%	It attained excellent prediction performance by carefully selecting characteristics.	Manual feature extraction decreases the efficacy of the system and accuracy.
Pratyaksh Jain *et al.* [10]	ML models such as LR, SVM, RF, and NB classifiers	Reddit dataset	It attained 77.29 % accuracy for LR, 74.35 % for NB, 77.12% for SVM, and	The system had many benefits, including improved accuracy,	The results may have needed to have been correctly interpreted by

			77.298% for RF.	efficiency, and decision-making.	the ML. The algorithm set served as the foundation for interpretation; if faulty, the results would be incorrectly understood.
Samer Muthana Sarsam *et al.* [11]	Yet Another Two-Stage Idea (YATSI)	Twitter dataset	Accuracy=86.97%	It could improve accuracy and aid in developing clinical decision support systems.	The algorithm was complicated and slow for real-time forecasting.
Shini Renjith *et al.* [12]	CNN and LSTM	Reddit Suicidality Dataset	Accuracy= 90.3% and F1-score=92.6%	This could provide proper medical assistance and help to those in need.	The random selection of hyperparameters in LSTM and CNN impacted the model results and made the computational processes more difficult.
Syed Tanzeel Rabani *et al.* [13]	ML models such as SVM, RF, and XGB	Twitter and Reddit	Best overall accuracy=96.33%	Improved efficiency, accuracy and decision-making.	The imbalanced dataset has bias and variance issues that were inherent.
Tianlin Zhang *et al.* [14]	GloVe with Bidirectional LSTM (BiLSTM)	Subreddit dataset	Precision =95.0%, Recall =94.9%, and F-measure =94.9%	It was useful for categorizing suicide notes, which may contribute to creating social media suicide prevention tools.	However, in this dataset, the suicide notes gathered were insufficient for training a stable and reliable supervised learning-based model due to the number of data.

Shaoxiong Ji *et al.* [15] presented attention relation networks (RN) for suicidal ideation and mental disorder detection. Initially, post representation of the tweets was carried out by extracting risk-related state indicators and LSTM text encoder. Then, the RN was utilized to learn the important relation scores from the post-representation. The system was tested on the Reddit dataset and attained a 0.6478 % f-score and 0.6474% accuracy. **Jingfang Liu *et al.* [16]** developed a k-means clustering model for detecting suicidal ideation in social media. Initially, the system used the 'Text Mind' software to count the frequency of different word categories of posts. Then, the suicidal-related keywords were clustered using k-means and word2vec models. The system attained an f-score of 79.20% and an accuracy of 80.61% for the Weibo dataset.

Zepeng Li *et al.* [17] developed a suicidal ideation detection model based on a deep hierarchical ensemble model (DHE). The data was collected from the Weibo social media posts and was preprocessed by removing emoticons, URLs, and meaningless keywords. Then, DHE was utilized to detect suicidal thoughts from the preprocessed data. The system achieved maximum f-score and accuracy of 89.6% and 93.11%. **Damien Lekkas *et al.* [18]** suggested an ensemble of ML frameworks for predicting acute suicidal ideation on Instagram. The Instagram texts were first gathered, and the features were extracted based on the four conditions: following ratio, engagement, average comments to average likes ratio, and the sum of moderate comments and likes per follower. Then, each of these extracted features was standardized to produce data with a mean and standard deviation of 0 and 1, respectively. The classification of suicidal ideation was finally done using an ensemble of ML models. The results demonstrated that the system outperformed existing techniques by attaining 0.702% accuracy and 0.769% sensitivity.

Mohamed Ali Ben Hassine *et al.* [19] recommended a classifier based on association rules (CBA) for detecting suicidal ideation on social media. The system collected the data from Twitter and performed preprocessing in it to improve data quality. Then, the preprocessed tweets were changed into a vector representation using the Empath tool, and finally, suicidal ideation was done using the CBA method. The system attained an accuracy of 0.71%, lower than the models we surveyed in the literature. **Woojin Jung *et al.* [20]** presented an RF and gradient boosting machine (GBM) for suicidal ideation detection. The system collected the data from Twitter, and preprocessing operations, such as the removal of stop words and hyperlinks, were performed. Next, the features such as text, n-gram, time, non-time, domain-specific lexicon, and topic polarity lexicon were extracted and fed into RF and GBM for classification. The system attained an f-score of 0.846 on the tested dataset.

4. DISCUSSION

Suicidal ideation problems can be detected early through social media posts, and suitable treatments can help people with potential mental diseases. The literature review has investigated numerous depression detection and prevention methodologies, which were implemented using ML and DL methods, by gathering and analyzing data from suicidal ideation-related social media platforms like Twitter, Reddit, Weibo, Instagram, etc., focusing on suicidal issues. The ML algorithm is often used in research to identify suicidal ideation at an early stage. ML has shown promise in identifying at-risk persons and assisting in preventing suicide. LR, SVM, RF, and NB are the most often used ML algorithms. These are quite popular because they produce significant accuracy with less computational resources. However, suppose the dimension of the collected data is too small. In that case, it is impossible to divide the training set and test set as in a traditional validation without suffering a considerable loss in modelling. Due to their extremely high accuracy at lower levels of engineering and processing, DL architectures provide several advantages for identifying suicidal ideation. The above literature mostly used the CNN and LSTM models. Both do not require manual feature extraction, as a backpropagation algorithm automatically extracts the distinct characteristics and boosts the overall performance of the suicidal ideation detection system.

5. CONCLUSION

It has become essential for our society to lower the suicide rate. Our contribution is intended to aid the community in the early detection of suicidal ideation. The first contribution made by this study is to the system's overall architecture for detecting suicidal ideation. Next, it provided the literature on the prediction of suicidal thoughts by suggesting ML and DL methodologies and presenting it in a tabular form for academics to draw from and reference. The study then highlights the present status and inadequacies of existing research in the literature through a critical evaluation of research on the topic. By choosing features based on a better comprehension of the sequential relationship between phrases, improved DL approaches could significantly increase accuracy.

REFERENCES

[1] Kumar, Ashok, Tina Esther Trueman, and A. K. Abinesh. "Suicidal risk identification in social media." *Procedia Computer Science* 189 (2021): 368-373.

[2] Elgeldawi, Enas, Awny Sayed, Ahmed R. Galal, and Alaa M. Zaki. "Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis." In *Informatics*, vol. 8, no. 4, p. 79. MDPI, 2021.

[3] Sawhney, Ramit, Shivam Agarwal, Atula Tejaswi Neerkaje, Nikolaos Aletras, Preslav Nakov, and Lucie Flek. "Towards suicide ideation detection through online conversational context." In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pp. 1716-1727. 2022.

[4] Gupta, Sumit, Dipnarayan Das, Moumita Chatterjee, and Sayani Naskar. "Machine learning-based social media analysis for suicide risk assessment." In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 2*, pp. 385-393. Springer Singapore, 2021.

[5] Amanat, Amna, Muhammad Rizwan, Abdul Rehman Javed, Maha Abdelhaq, Raed Alsaqour, Sharnil Pandya, and Mueen Uddin. "Deep learning for depression detection from textual data." *Electronics* 11, no. 5 (2022): 676.

[6] Nandwani, Pansy, and Rupali Verma. "A review on sentiment analysis and emotion detection from text." *Social Network Analysis and Mining* 11, no. 1 (2021): 81.

[7] Aldhyani, Theyazn HH, Saleh Nagi Alsubari, Ali Saleh Alshebami, Hasan Alkahtani, and Zeyad AT Ahmed. "Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models." *International journal of environmental research and public health* 19, no. 19 (2022): 12635.

[8] Haque, Rezaul, Naimul Islam, Maidul Islam, and Md Manjurul Ahsan. "A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning." *Technologies* 10, no. 3 (2022): 57.

[9] Chatterjee, Moumita, Piyush Kumar, Poulomi Samanta, and Dhrubasish Sarkar. "Suicide ideation detection from online social media: A multi-modal feature based technique." *International Journal of Information Management Data Insights* 2, no. 2 (2022): 100103.

[10] Jain, Pratyaksh, Karthik Ram Srinivas, and Abhishek Vichare. "Depression and suicide analysis using machine learning and NLP." In *Journal of Physics: Conference Series*, vol. 2161, no. 1, p. 012034. IOP Publishing, 2022.

[11] Sarsam, Samer Muthana, Hosam Al-Samarraie, Ahmed Ibrahim Alzahrani, Waleed Alnumay, and Andrew Paul Smith. "A lexicon-based approach to detecting suicide-related messages on Twitter." *Biomedical Signal Processing and Control* 65 (2021): 102355.

[12] Renjith, Shini, Annie Abraham, Surya B. Jyothi, Lekshmi Chandran, and Jincy Thomson. "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, no. 10 (2022): 9564-9575.

[13] Rabani, Syed Tanzeel, Akib Mohi Ud Din Khanday, Qamar Rayees Khan, Umar Ayoub Hajam, Ali Shariq Imran, and Zenun Kastrati. "Detecting suicidality on social media: Machine learning at rescue." *Egyptian Informatics Journal* 24, no. 2 (2023): 291-302.

[14] Zhang, Tianlin, Annika M. Schoene, and Sophia Ananiadou. "Automatic identification of suicide notes with a transformer-based deep learning model." *Internet interventions* 25 (2021): 100422.

[15] Ji, Shaoxiong, Xue Li, Zi Huang, and Erik Cambria. "Suicidal ideation and mental disorder detection with

attentive relation networks." *Neural Computing and Applications* 34, no. 13 (2022): 10309-10319.

[16] Liu, Jingfang, Mengshi Shi, and Huihong Jiang. "Detecting suicidal ideation in social media: An ensemble method based on feature fusion." *International journal of environmental research and public health* 19, no. 13 (2022): 8197.

[17] Li, Zepeng, Jiawei Zhou, Zhengyi An, Wenchuan Cheng, and Bin Hu. "Deep hierarchical ensemble model for suicide detection on imbalanced social media data." *Entropy* 24, no. 4 (2022): 442.

[18] Lekkas, Damien, Robert J. Klein, and Nicholas C. Jacobson. "Predicting acute suicidal ideation on Instagram using ensemble machine learning models." *Internet interventions* 25 (2021): 100424.

[19] Ben Hassine, Mohamed Ali, Safa Abdellatif, and Sadok Ben Yahia. "A novel imbalanced data classification approach for suicidal ideation detection on social media." *Computing* 104, no. 4 (2022): 741-765.

[20] Jung, W., Kim, D., Nam, S. and Zhu, Y., 2023. Suicidality detection on social media using metadata and text feature extraction and machine learning. *Archives of suicide research*, *27*(1), pp.13-28.