International Journal of Computer Sciences and Engineering
Vol.11, Special Issue.1, pp.29-35, November 2023
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Research Paper

Detecting Depression and Suicidal Ideation from Texts using Machine Learning & Deep Learning Techniques

Sumit Gupta[1*], Arya Manikya Sinha[2], Debjyoti Prodhan[3], Nirnay Ghosh[4], Souvik Modak[5]

[1,2,3,4,5]Dept. of Computer Science and Engineering, University Institute of Technology, The University of Burdwan, Golapbag (North), Burdwan-713104, West Bengal, India

Corresponding Author: sumitsayshi@gmail.com

Abstract: Suicide has emerged as a pressing societal health concern in contemporary times. Suicidal intent refers to an individual's contemplation of taking one's own life, and such tragedies have far-reaching impacts on families, communities, and nations. The global standardized rate of suicides per population suggests that in 2022, there were approximately 903,450 completed suicides, alongside a staggering 18,069,000 cases of individuals having suicidal thoughts but not acting upon them. These distressing figures highlight the widespread nature of this issue, affecting people of all ages, nationalities, races, beliefs, socioeconomic backgrounds, and genders. Additionally, it is pertinent to acknowledge that depression, a prevalent mental disorder, can significantly hinder daily functioning and potentially contribute to developing suicidal thoughts. This work focuses on detecting suicidal intent by employing Machine Learning classifiers such as Support Vector Machines, Naive Bayes, Logistic Regression, and Random Forest. Furthermore, this research extends its analysis by incorporating Deep Learning classifiers such as Convolutional Neural Networks, Long Short-Term Memory, Bidirectional Long Short-Term Memory, and Bidirectional Encoder Representations from Transformers on six selected datasets. The primary aim is to identify signs of depression as a means to gauge the likelihood of suicidal thoughts. In addition to the classification algorithms, various features are extracted to provide insights into an individual's emotional state and mindset. By combining these techniques, the study aims to improve the understanding and prediction of suicidal tendencies.

Keywords: Depression, Suicide, Depression Detection, Suicidal Ideation, Text Classification, Machine Learning, Deep Learning.

1. Introduction

Suicide has been around for as long as human society, ranking among the top 13 causes of death in all ages worldwide, and continues to challenge our collective wisdom. Suicide is an unsocial act that has an overwhelming impact on relations and families [1]. As per the World Health Organization (WHO), suicide is a primary cause of death among individuals between 15-29 years old worldwide. Depression and suicide are two of the most critical mental health problems prevalent in the current era. It is estimated that depression affects over 264 million people globally, while suicide is responsible for over 800,000 deaths annually, leading to increased suicidal ideation [2, 3]. Suicidal ideation is one of the most critical symptoms of depression and is often difficult to detect [4]. The use of Artificial Intelligence, specifically Machine Learning (ML) and Deep Learning (DL) techniques can help to identify and predict depression and suicidal ideation in individuals.

This work focuses on using text analytics to extract meaningful insights from accumulated texts, enabling the identification of various emotions and intentions. The goal is to detect signs of suicidal tendencies by analyzing depression levels in textual inputs. The process involves collecting six datasets from past research, preprocessing the data, and splitting it into training and testing sets. To classify suicidal texts, four machine learning (ML) and four deep learning (DL) classifiers are utilized, including Support Vector Machine (SVM), Naive Bayes (NB), Logistic Regression (LR), Random Forest (RF), Bidirectional Encoder Representations from Transformers (BERT), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Convolutional Neural Network (CNN). The findings underscore the potential of ML and DL classification techniques in identifying individuals with depression based on text analysis.

The rest of the paper is organized as follows: Section 2 discusses few related works on various techniques proposed by authors in detecting suicidal tendency. In section 3, the working principle of the proposed system has been explained in a step-by-step manner, starting from the data collection phase and data preprocessing phase to the text classification phase. Section 4 presents the tools and datasets used for this research work and the results so obtained are analyzed to compare the performances of different classifiers. The contribution of this work is summarized in section 5 followed by some insights into what can be some lines of improvements that can be undertaken in future.

2. Related Work

A good amount of work has been done in the area of Depression Detection and Suicidal Ideation from textual data using ML and DL classification approaches. This section discusses few prominent existing works in this domain.

Authors in their research works [7, 8, 9, 12, 15, 20] have used Naive Bayes' classifier due to its simplicity and strong independence assumptions, making it suitable for multiclass classification. On the other hand, classification employing Support Vector Machines (SVMs) [5, 6, 7, 9, 15, 16, 19, 22, 23] are preferred for binary classification as they are supervised learning models known for their effective classification and regression analysis. Researchers [7, 12, 15, 19, 20, 22] also use Logistic Regression (LR) for text classification as LR classifier describes data aptly and explains the relationship between dependent and independent variables. Random Forest (RF) [5, 7, 9, 12, 19, 20] is known to be a versatile Machine Learning algorithm that combines multiple decision trees for classification and regression.

Gupta et al. [9] have used the "Reddit CSSRS Suicide Dataset" [39], which is a multiclass dataset labeled with five emotions, thereby obtaining precision scores of 0.714, 0.696, and 0.667 using NB, RF and SVM respectively. In paper [6], the authors have considered the "Life Corpus" [42] dataset with SVM and Bag of Words (BoW) for binary classification, achieving a macro F1 score of 0.78-0.81. In the research works [15, 12], tweets are gathered and labeled into suicidal and non-suicidal classes. Various ML techniques and Ensemble Methods have been applied. Paper [19] has reported F1 scores of 0.81, 0.80, and 0.77 using SVM, LR, and RF, respectively. In the work [12], RF classifier achieves an accuracy of 0.996, while LR and NB techniques yield poor result. Rabani et al. [15] suggest that the performance of RF is best with an accuracy of 98.5%, while NB has the lowest accuracy at 81.7%. It is observed that Naive Bayes is more suitable for multiclass classification, while SVM, RF, and LR are more effective for binary classification. But all these ML classifiers are found to perform better on smaller datasets. Figure 1 depicts the four commonly used ML classifiers viz. SVM, NB, RF and LR that have been employed to detect depression and presence of suicidal tendency, thereby showcasing that SVM is the more popular among the lot.

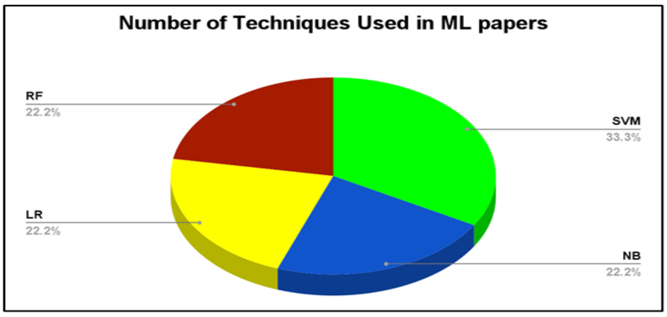

Figure 1. Pie chart depicting the commonly used ML techniques as per ML-based existing works on Suicidal Ideation.

To handle larger datasets, some research works also focus on Deep Learning (DL) approaches. Convolutional Neural Networks (CNN) have been utilized in [24, 8, 19, 26-38]. CNNs are capable of understanding relationships between variables and have been widely used for data description. Long-Short Term Memory (LSTM) Networks [24, 26, 27, 33, 36-38], a type of Neural Network, address long-term dependencies by employing specialized gates to carefully manage data, whereas Bidirectional LSTM (Bi-LSTM) used in [8, 19, 25, 29, 30, 34, 35, 37] is a DL algorithm that analyzes data for classification and data analysis by considering dependencies in both directions. Bi-LSTM used in BERT, a model designed to understand the meaning of ambiguous language, is employed in papers [29, 33, 35, 36, 38], either alone or in hybrid forms.

Tadesse et al. [24] have used CNN-LSTM and this hybrid model achieves an accuracy of 93.8% is achieved, with LSTM slightly outperforming CNN individually. In paper [34], CNN-Bi-LSTM achieves 95% accuracy using LIWC textual features. Papers [19, 28] report CNN accuracies of 74% and 82.14%, respectively. Ghosh & Bhattacharyya [26] use the "CEASE" [43] dataset with multiple classes and achieves only 59.54% accuracy with CNN. However, Kim et al. [32] demonstrate CNN's strong performance by achieving an accuracy range of 75.13% to 96.96%. In [35], Bi-LSTM and BERT outperform CNN on a large dataset with well-selected embedding. Ghosh et al. [36] employ the "CEASE-v2.0" dataset for depression detection and achieve an improved accuracy of 76.86% using a VAD-BERT multitask framework. In paper [38], utilizing the "Suicide and Depression Detection (SDD) dataset" [41] with 232,074 rows, BERT and ELECTRA pre-trained models perform the best, achieving accuracies of 97.57% and 97.92%, respectively. Overall, CNN is the most frequently used DL classifier and performs decently, while LSTM, Bi-LSTM, and BERT or hybrid BERT models excels when applied to larger datasets. Figure 2 depicts the four commonly used DL classifiers in order viz. CNN, Bi-LSTM, LSTM and BERT that have been employed to detect depression and presence of suicidal tendency.

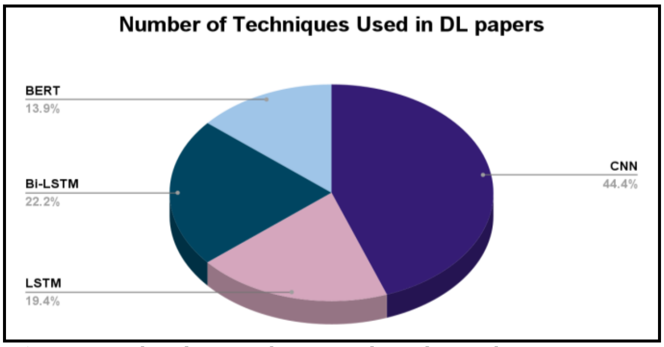

Figure 2. Pie chart depicting the commonly used DL techniques as per DL-based existing works on Suicidal Ideation.

3. Proposed Methodology

This research work is divided into two main categories to address the text classification problem of suicidal ideation. The methodology encompasses two distinct approaches: one

utilizing ML classifiers, and the other employing DL classifiers. Figure 3 illustrates the text classification process for suicidal ideation detection using ML and DL approaches.
Detecting suicidal ideation is a serious matter, and it requires a comprehensive approach that goes beyond the steps. Nonetheless, an overview of how these steps can be incorporated into the process is elaborated. Here's a high-level outline of the steps:

Step 1. *Data Preprocessing:* Collect relevant data related to suicidal ideation, such as text or social media posts, interviews, or medical records. For this work, six datasets are collected from various sources. Next clean the data by removing irrelevant information, special characters, and potentially sensitive information. Normalize the text data by converting it to lowercase, removing stop words, and applying lemmatization techniques.

Step 2. *Feature Extraction:* Extract meaningful features from the preprocessed texts. Some common approaches include Bag-of-Words (BoW); which represents the text as a frequency count of individual words or n-grams, TF-IDF, which weighs the importance of words based on their frequency in a document and the entire corpus, Sentiment Analysis; which analyzes the emotional tone of the text using sentiment lexicons or ML models.

Step 3. *Encoding:* Convert the extracted features into numerical representations that Deep Learning algorithms can process. If needed, perform additional encoding steps like one-hot encoding or label encoding for categorical variables.

Step 4. *Data Splitting:* Split the encoded data into training and testing sets.

Step 5. *Model Selection and Training:* Apply various Machine Learning models on the training data. The models such as SVM, LR, RF and NB are traditional ML algorithms, while CNN, LSTM, Bi-LSTM, and BERT are DL models commonly used for text classification. Train each model using the training data and evaluate their performance using the validation set. Tune hyperparameters for each model to optimize their performance.

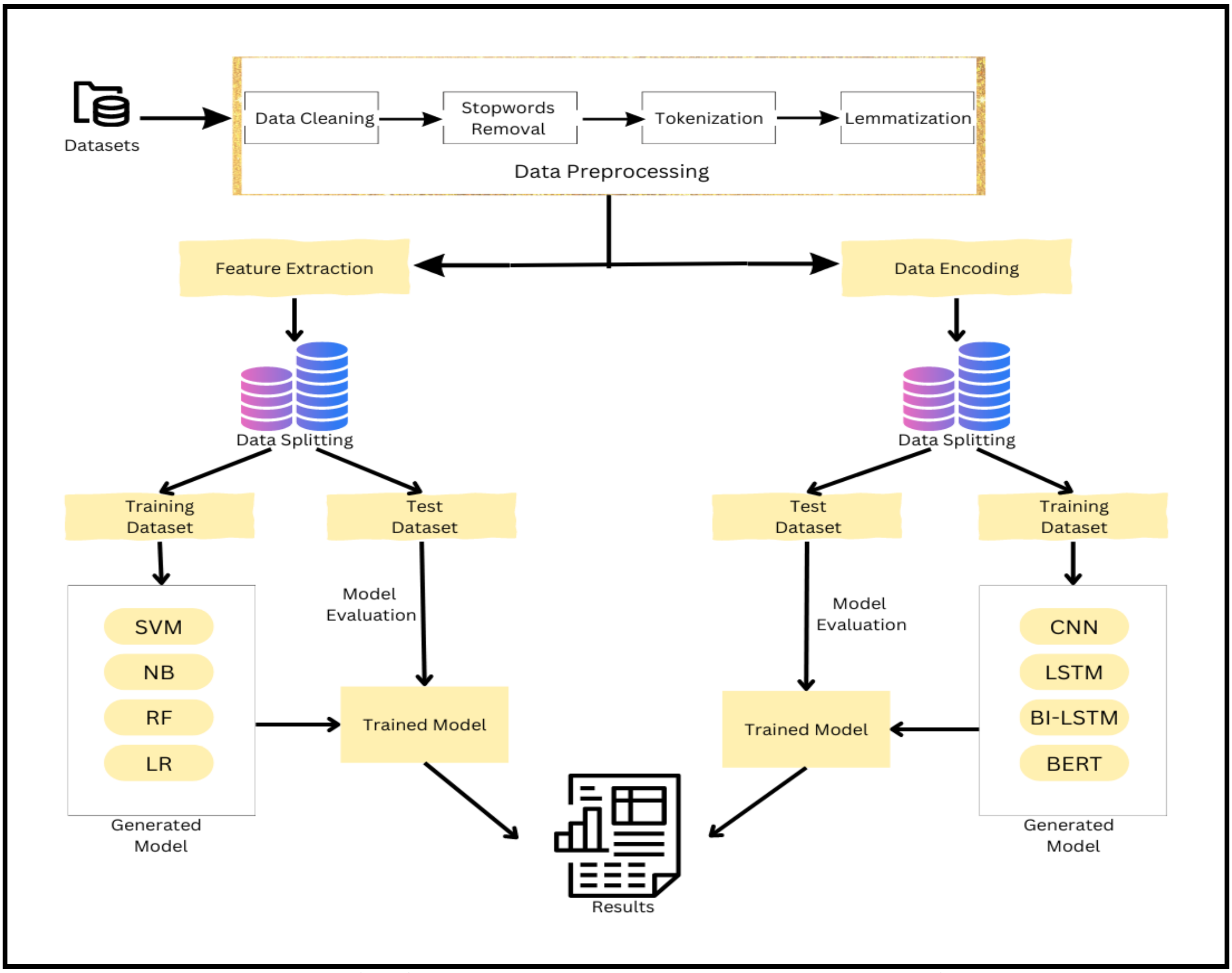

Figure 3. Workflow of the Proposed Suicidal Ideation System Employing ML and DL Classifiers.

Step 6. ***Evaluation:*** Assess the performance of each model based on evaluation metrics such as accuracy, precision, recall, and F1-score. Select the model with the best performance on the validation set.

Step 7. ***Testing and Deployment:*** Assess the final model's performance on the testing set to obtain a more accurate estimate of its generalization ability.

4. Implementation and Results

Implementing ML and DL models for suicidal ideation detection typically involves writing code in a programming language like Python and utilizing popular libraries such as scikit-learn, Keras, or TensorFlow. In this work, "Google colab" is used as the coding interface or platform. To conduct the experiment, a train-test ratio of 80:20 has been adopted.

Datasets: The proposed system has been trained and tested on the following six datasets:

I) The *Reddit C-SSRS Suicide Dataset (Reddit Dataset)* [43] contains 500 Reddit posts with three columns: "User," "Post," and "Label." The usernames in the "User" column are removed. The "Label" column originally had five labels: "Ideation," "Supportive," "Attempt," "Behavior," and "Indicator," which has been simplified by replacing "Ideation," "Attempt," "Behavior," and "Indicator" with label 1, and "Supportive" with label 0. 400 posts have been used for training and 100 posts for testing.

II) The *SDCNL Dataset* [44] comprises 1894 rows and 14 columns, including "selftext" and "is_suicide". These two columns have been selected for this work as they contain the text messages and corresponding labels (0 and 1) respectively.

III) The *Life_Corpus Dataset* [46] is small, with 273 rows and 2 columns: "text" and "cls." The "text" column contains text messages, and the "cls" column labels them as "No risk" or "Risk." "No risk" labels have been converted to 0 and "Risk" labels to 1.

IV) The *CEASE Dataset* [47] comprises 15 text files containing messages of various emotions. These files are merged into a single text file named "Total_dataset.txt." Sentiment intensity analysis using the VADER lexicon is conducted on each message, labeling depressive messages as 1 and non-depressive messages as 0. The resulting dataset has 2393 rows.

V) The *SWMH Dataset* [48], collected from mental health-related subreddits, is divided into three parts: "train.csv," "test.csv," and "val.csv." For this experiment, the "train.csv" with 34,823 rows for training and "test.csv" with 10,883 rows for testing are used. Each CSV file contains two columns: "text" containing the messages and "label" consisting of five different labels: 'self.Anxiety,' 'self.bipolar,' 'self.depression,' 'self.SuicideWatch,' and 'self.offmychest.'

VI) The *Suicide and Depression Detection (SDD) Dataset* [45] is a large dataset comprising posts from the "SuicideWatch" and "Depression" subreddits on Reddit. It consists of 232,074 rows and 3 columns: "Unnamed: 0," "text," and "class." For the experiment, the "Unnamed: 0" column has been removed as it is not necessary for this work. The "text" column contains the text messages, while the "class" column contains the corresponding labels viz. "non-suicide" and "suicide." The labels are converted to numeric values, with "non-suicide" labeled as 0 and "suicide" as 1. Due to the dataset's size, only the first 30,000 rows are used for the experiment.

Results: Table 1 shows the accuracy of the proposed Suicidal Ideation system on all selected datasets using ML & DL classifiers.

It can be observed from Table 1 that SVM outperforms other classifiers in the SDD and CEASE datasets with 94% and 79% accuracy value respectively. Bi-LSTM proves more effective with Reddit and SWMH datasets, with its LSTM counterpart just close behind in terms of performance. Logistic Regression performs a bit better than other classification approaches for the SDCNL dataset whereas in case of the Life-Corpus, the CNN model proves more accurate.

Table 2 compares the performance of the proposed system with existing works.

Table 1. Accuracy Scores Obtained for Different Datasets using Machine Learning and Deep Learning Classifiers

Datasets	**Text Classifiers**							
	Machine Learning				**Deep Learning**			
	SVM	**NB**	**RF**	**LR**	**BERT**	**LSTM**	**Bi-LSTM**	**CNN**
Reddit [39]	77	77	77	77	78	**82**	**82**	81
SDCNL [40]	70	69	68	**71**	55	67	62	66
SDD [41]	**94**	92	83	93	80	88	87	86
Life_Corpus [42]	80	78	80	80	85	85	85	**87**
CEASE [43]	**79**	75	67	77	68	76	77	77
SWMH [44]	54	52	53	54	60	71	**73**	68

Table 2. Performance Comparison of Proposed Work with other existing Suicidal Ideation systems

Datasets	Authors	Technique / Features	Reported Results
Reddit [39]	Gupta et al. [9]	NB, Semantic Features	71
	Proposed work	LSTM	**82**
SDCNL [40]	*Proposed work*	LR, TF-IDF(1), TF-IDF(2)	71
	Haque et al. [29]	guse-dense with UMAP-KMeans	**98.18**
SDD [41]	*Proposed work*	SVM, TF-IDF(1), TF-IDF(2)	94
	Wen et al. [38]	Electra	**97**
Life_Corpus [42]	Caicedo et al. [6]	Rasa	0.49 ± 0.02 (Macro f1)
	Parraga-Alava et al. [18]	Hierarchical clustering. Average	0.79 (F1)
	Proposed work	CNN	87(Accuracy)
	Caicedo et al. [14]	POS, SYNSETS, lemma, word RandomCommittee	**0.958 (F1)**
CEASE [43]	Ghosh et al. [26]	CNN	59.54
	Proposed work	CNN	**77**
SWMH [44]	Ji et al. [37]	RN	64
	Proposed work	Bi-LSTM	**73**

Table 2 showcases that the proposed system manages to perform better when compared to other existing works for three datasets viz. Reddit, CEASE and SWMH when LSTM, CNN and Bi-LSTM models are used. The performance for SDCNL dataset is very poor with only 71% accuracy score obtained compared to the existing best score of 98.18% [29]. Though the proposed system manages to perform fairly for the SDD dataset with 94% accuracy score with TFIDF (unigram and bigram) features and SVM classification technique but it performs miserably in case of the Life_Corpus [42] with only 87% accuracy in spite of using the CNN model.

5. Conclusion and Future Scope

The main focus of this work is to identify suicidal intent through depression detection from texts using Machine Learning and Deep Learning techniques. Based on the summary findings and gap analysis done on various previous related works, six depression and/or suicide-related datasets have been collected and features such as BoW (unigram and bigram), TF-IDF (unigram and bigram) and sentiment scores have been extracted. These individual features and their combination have been used as input while training and testing the proposed model using four different Machine Learning classifiers, namely SVM, NB, RF and LR. Furthermore, text classification models using four Deep Learning classifiers viz. LSTM, Bi-LSTM, CNN, and BERT have also been built. The performance of the proposed system is found to be better when compared to other related systems for three datasets viz. Reddit, CEASE and SWMH when LSTM, CNN and Bi-LSTM models are used with accuracy scores of 82%, 77% and 73% respectively.

As a future improvement, implementation of word embedding techniques such as “Word2vec”, “GloVe embedding”, “FastText” & “ELMo embedding” can be initiated. The proposed model’s performance remains to be tested on other datasets such as CEASE v2.0. Plans are also ongoing to build a GUI based application for common user to detect the presence of depression in a given text. Finally, to enhance the challenge, handwritten notes can be collected and this image dataset can be analyzed to identify the presence of depression and suicidal tendency amidst the writer of the text document.

Conflict of Interest
Authors declare that they do not have any conflict of interest.

Funding Source
None.

Authors’ Contributions
Author-1 conceived the study, mentored the proposed approach, reviewed and edited the manuscript and approved the final version of the manuscript. Experimentation task has been performed by Author-2 and Author-3, alongwith the preparation of the first draft of the manuscript. Author-4 and Author-5 have done the literature study and have performed the gap analysis on existing works.

Acknowledgments
The authors extend their thanksgiving and regards to their parents, teachers and friends for being the constant pillar of strength, encouragement and support throughout their journey.

References

[1]. S. Kumar, A.K. Verma, S. Bhattacharya, S. Rathore, “Trends in rates and methods of suicide in India,” *Egyptian Journal of Forensic Sciences*, **3(3), 75-80, 2013**.

[2]. S. Bachmann, “Epidemiology of suicide and the psychiatric perspective,” *International journal of environmental research and public health*, **15(7), 1425, 2018**.
[3]. L. Brådvik, “Suicide risk and mental disorders,” *International journal of environmental research and public health*, **15(9), 2028, 2018**.
[4]. M.W. Gijzen, S.P. Rasing, D.H. Creemers, F. Smit, R.C. Engels, D.D. Beurs, “Suicide ideation as a symptom of adolescent depression. A network analysis,” *Journal of Affective Disorders*, **278, 68-77, 2021**.
[5]. A. Mbarek, S. Jamoussi, A.B. Hamadou, “An across online social networks profile building approach: Application to suicidal ideation detection,” *Future Generation Computer Systems*, **133**, **171-183**, **2022**.
[6]. R.W.A. Caicedo, J.M. Soriano, H.A.M. Sasieta, “Bootstrapping semi-supervised annotation method for potential suicidal messages,” *Internet Interventions*, **100519**, **2022**.
[7]. R. Haque, N. Islam, M. Islam, M.M. Ahsan, “A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning,” *Technologies*, **10(3)**, **57**, **2022**.
[8]. T. Zhang, A.M. Schoene, S. Ananiadou, “Automatic identification of suicide notes with a transformer-based deep learning model,” *Internet interventions*, **25**, **100422**, **2021**.
[9]. S. Gupta, D. Das, M. Chatterjee, S. Naskar, “Machine Learning-Based Social Media Analysis for Suicide Risk Assessment,” *In Emerging Technologies in Data Mining and Information Security*, Springer, Singapore, pp. **385-393**, **2021**.
[10]. A.M. Schoene, A. Turner, G.R. De Mel, N. Dethlefs, “Hierarchical multiscale recurrent neural networks for detecting suicide notes,” *IEEE Transactions on Affective Computing*, **2021**.
[11]. Z. Xu, Y. Xu, F. Cheung, M. Cheng, D. Lung, Y.W. Law, B. Chiang, Q. Zhang, P.S.F. Yip, “Detecting suicide risk using knowledge-aware natural language processing and counseling service data,” *Social Science & Medicine*, **283**, **114176**, **2021**.
[12]. E.R. Kumar, K.V.S.N. Rama Rao, S.R. Nayak, R. Chandra, “Suicidal ideation prediction in twitter data using machine learning techniques,” *Journal of Interdisciplinary Mathematics*, **23(1)**, **117-125**, **2020**.
[13]. S. Parrott, B.C. Britt, J.L. Hayes, D.L. Albright, “Social media and suicide: a validation of terms to help identify suicide-related social media posts,” *Journal of Evidence-Based Social Work*, **17(5)**, **624-634**, **2020**.
[14]. R.W.A. Caicedo, J.M.G. Soriano, H.A.M. Sasieta, “Assessment of supervised classifiers for the task of detecting messages with suicidal ideation,” *Heliyon*, **6(8)**, **e04412**, **2020**.
[15]. S.T. Rabani, Q.R. Khan, A.M.U.D. Khanday, “Detection of suicidal ideation on Twitter using machine learning & ensemble approaches,” *Baghdad Science Journal*, **17(4)**, **1328-1328**, **2020**.
[16]. S.B. Hassan, S.B. Hassan, U. Zakia, “Recognizing Suicidal Intent in Depressed Population using NLP: A Pilot Study,” *In 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, pp. **0121-0128**, **2020**.
[17]. A. Alambo, M. Gaur, U. Lokala, U. Kursuncu, K. Thirunarayan, A. Gyrard, A. Sheth, R.S. Welton, J. Pathak, “Question answering for suicide risk assessment using reddit,” *In 2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, IEEE, pp. **468-473**, **2019**.
[18]. J. Parraga-Alava, R.A. Caicedo, J.M. Gómez, M. Inostroza-Ponta, “An unsupervised learning approach for automatically to categorize potential suicide messages in social media,” *In 2019 38th International Conference of the Chilean Computer Science Society (SCCC)*, IEEE, pp. **1-8**, **2019**.
[19]. J. Du, Y. Zhang, J. Luo, Y. Jia, Q. Wei, C. Tao, H. Xu, “Extracting psychiatric stressors for suicide from social media using deep learning,” *BMC medical informatics and decision making*, **18(2)**, **77-87**, **2018**.
[20]. M.J. Vioules, B. Moulahi, J. Azé, S. Bringay, “Detection of suicide-related posts in Twitter data streams,” *IBM Journal of Research and Development*, **62(1)**, **7-1**, **2018**.
[21]. S.R. Braithwaite, C. Giraud-Carrier, J. West, M.D. Barnes, C.L. Hanson, “Validating machine learning algorithms for Twitter data against established measures of suicidality,” *JMIR mental health*, **3(2)**, **e4822**, **2016**.
[22]. B. O'dea, S. Wan, P.J. Batterham, A.L. Calear, C. Paris, H. Christensen, “Detecting suicidality on Twitter,” *Internet Interventions*, **2(2)**, **183-188**, **2015**.
[23]. A. Nikfarjam, E. Emadzadeh, G. Gonzalez, “A hybrid system for emotion extraction from suicide notes,” *Biomedical informatics insights*, **5**, **BII-S8981**, **2012**.
[24]. M.M. Tadesse, H. Lin, B. Xu, L. Yang, “Detection of suicide ideation in social media forums using deep learning,” *Algorithms*, **13(1)**, **7**, **2019**.
[25]. F.M. Shah, F. Ahmed, S.K.S. Joy, S. Ahmed, S. Sadek, R. Shil, M.H. Kabir, “Early depression detection from social network using deep learning techniques,” *In 2020 IEEE Region 10 Symposium (TENSYMP)*, IEEE, pp. **823-826**, **2020**.
[26]. S. Ghosh, A. Ekbal, P, Bhattacharyya, “Cease, a corpus of emotion annotated suicide notes in English,” *In Proceedings of the 12th Language Resources and Evaluation Conference*, pp. **1618-1626**, **2020**.
[27]. M. Morales, P. Dey, T. Theisen, D. Belitz, N. Chernova, “An investigation of deep learning systems for suicide risk assessment,” *In Proceedings of the sixth workshop on computational linguistics and clinical psychology*, pp. **177-181**, **2019**.
[28]. S. Boukil, F. El Adnani, L. Cherrat, A.E. El Moutaouakkil, M. Ezziyyani, “Deep learning algorithm for suicide sentiment prediction,” *In Advanced Intelligent Systems for Sustainable Development (AI2SD’2018)* Vol **4**: Advanced Intelligent Systems Applied to Health, Springer International Publishing, pp. **261-272**, **2019**.
[29]. A. Haque, V. Reddi, T. Giallanza, “Deep learning for suicide and depression identification with unsupervised label correction,” *In Artificial Neural Networks and Machine Learning–ICANN 2021: 30th International Conference on Artificial Neural Networks*, Bratislava, Slovakia, Vol **30,** Springer International Publishing, pp. **436-447**, **2021**.
[30]. C.S. Wu, C.J. Kuo, C.H. Su, S.H. Wang, H.J. Dai, “Using text mining to extract depressive symptoms and to validate the diagnosis of major depressive disorder from electronic health records,” *Journal of affective disorders*, **260**, **617-623**, **2020**.
[31]. K.S. Choi, S. Kim, B.H. Kim, H.J. Jeon, J.H. Kim, J.H. Jang, B. Jeong, “Deep graph neural network-based prediction of acute suicidal ideation in young adults,” *Scientific reports*, **11(1)**, **1-11**, **2021**.
[32]. J. Kim, J. Lee, E. Park, J. Han, “A deep learning model for detecting mental illness from user content on social media,” *Scientific reports*, **10(1)**, **1-6**, **2020**.
[33]. N.A. Baghdadi, A., Malki, H.M. Balaha, Y. AbdulAzeem, M. Badawy, M. Elhosseini, “An optimized deep learning approach for suicide detection through Arabic tweets,” *PeerJ Computer Science*, **8**, **e1070**, **2022**.
[34]. T.H. Aldhyani, S.N. Alsubari, A.S. Alshebami, H. Alkahtani, Z.A. 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(19)**, **12635**, **2022**.
[35]. S. Long, R. Cabral, J. Poon, S.C. Han, “A Quantitative and Qualitative Analysis of Suicide Ideation Detection using Deep Learning,” arXiv preprint arXiv:2206.08673, **2022**.
[36]. S. Ghosh, A. Ekbal, P. Bhattacharyya, “VAD-assisted multitask transformer framework for emotion recognition and intensity prediction on suicide notes,” *Information Processing & Management*, **60(2)**, **103234**, **2023**.
[37]. S. Ji, X. Li, Z. Huang, E. Cambria, “Suicidal ideation and mental disorder detection with attentive relation networks,” *Neural Computing and Applications*, **34(13)**, **10309-10319**, **2022**.
[38]. A.S.S. Wen, G.J. Yi, L.Z. Hui, L. Xiao, Q.Y. Zhen, “Suicidal Text Detection,” Retrived from: https://github.com/gohjiayi/suicidal-text-detection, **2021**.
[39]. M. Gaur, A. Alambo, J.P. Sain, U. Kurscuncu, K. Thirunarayan, R.

Kavuluru, A. Sheth, R. Welton, J. Pathak, "Reddit C-SSRS suicide dataset," *Zenodo*, https://doi.org/10.5281/zenodo.2667859, **2019**.
[40]. https://github.com/ayaanzhaque/SDCNL
[41]. https://www.kaggle.com/datasets/nikhileswarkomati/suicide-watch
[42]. https://github.com/PlataformaLifeUA.
[43]. https://www.iitp.ac.in/~ai-nlp-ml/resources.html#CEASE
[44]. https://zenodo.org/record/6476179#.ZGuJF3ZX6SU

AUTHORS PROFILE

Mr S Gupta is currently working as Assistant Professor in the Department of Computer Science & Engineering, University Institute of Technology, The University of Burdwan, West Bengal, India. He has an Academic Experience of 11 years and a Research Experience of more than 8 years. He has published more than 30 articles, book chapters, research papers in reputed international and national conferences including IEEE, Springer, IETE, CRC Press etc. His main research work focuses on Natural Language Processing, Machine Learning, Deep Learning and Information Security.

Mr A M Sinha is currently pursuing Bachelor of Engineering in the Department of Computer Science & Engineering, University Institute of Technology, The University of Burdwan, West Bengal, India. His main research work focuses on Natural Language Processing and Cyber Security.

Mr D Prodhan is currently pursuing Bachelor of Engineering in the Department of Computer Science & Engineering, University Institute of Technology, The University of Burdwan, West Bengal, India. His main research work focuses on Natural Language Processing and Data Mining.

Mr N Ghosh is currently pursuing Bachelor of Engineering in the Department of Computer Science & Engineering, University Institute of Technology, The University of Burdwan, West Bengal, India. His main research work focuses on Java-based Application Development and Text Mining.

Mr S Modak is currently pursuing Bachelor of Engineering in the Department of Computer Science & Engineering, University Institute of Technology, The University of Burdwan, West Bengal, India. His main research work focuses on Mobile Application Development and Text Analytics.

