Detection of Suicidal Ideation From Social Media Posts Using BERT And LSTM Models

Ramandeep Singh Bawa[1], Anshul Khurana[2]
[1]Dept of CSE
[2]Assistant Professor, Dept of CSE
[1, 2] SRIT, Jabalpur, MP

Abstract- *Twitter and Facebook are the most popular and commonly used social media platforms among people of all ages when textually expressing their state of mind or emotions. Some of these emotions and thoughts are from people who are suicidal. Suicide is a serious public health problem in countries around the world. To detect suicidal thoughts or opinions from users' tweets, we can examine and classify textual content on Twitter. In this work, we train two deep learning models, LSTM and transformers including BERT, to classify text as either suicidal or no-suicidal. Our results show that LSTM performed best with an accuracy of 92% on Reddit Dataset, while the BERT model achieved an accuracy of 51%. We also tested the effectiveness of some machine learning a model achieving good accuracy with MLP 65%, accuracy with Extra Tree Classifier is58%, accuracy using LightGBM is 61% and accuracy with RF is 59%. With these results, applying deep learning models to social media can help to timely identify an individual likely to commit suicide based on their text and aid the relatives in providing the necessary help.*

Keywords- Suicidal behaviour, Social Media, BERT, LSTM, TF-Idf, Count-Vectorizer, Chi-square, feature Selection, Machine Learning.

I. INTRODUCTION

The widespread use of online platforms provides a unique opportunity for in-depth understanding of mental health issues1 and providing support to struggling individuals [1]. Among these platforms, social media sites have become an important way for users to express their emotions and struggles, including self-harm and suicidal thoughts. To meet the urgent need for timely intervention and psychological support, this study introduces a max voting ensemble classifier model designed to detect homicidal intent on personal social media websites [2], focusing on reddit. Early detection of suicidal ideation is a challenging task because it requires the integration of many factors, including psychological, social, and environmental changes [3]. The opportunity to capture and analyze big data on mental health [4], [5]. Machine learning and natural language processing (nlp) techniques are promising for detecting language patterns and signs of suicidal ideation in various text formats, such as social media, online chat, and electronic medical records [6] , [7], [8] , [9]. The data collection and annotation process is time-consuming and incurs large financial costs [10]. Obtaining more information about suicide can be difficult and limited, especially in suicide research, due to the many variables involved. The sensitive and stigmatized nature of suicide often creates problems with data collection. Individuals and organizations may be reluctant to share personal or confidential information about suicide for fear of negative consequences. Mental health, or how people experience and evaluate many things in life, is often used to measure their own happiness. Poor health is associated with higher levels of suicidal ideation, depression, and hopelessness in those who attempt suicide [11]. No motivation to continue, believing that there is nothing worth living for, and not wanting to live anymore [12]. . Suicidal behavior progresses from suicidal thoughts to suicide threats, suicide attempts, and suicide attempts. Fsome suicidal ideation tests are used as a tool to assess whether a person is at risk for self-harm and suicide and to facilitate timely intervention and treatment. Suicidality assessment is based on self-reports or is administered by trained clinicians and includes the personal thoughts and behaviors (ass) questionnaire, the Columbia-suicidality severity rating scale (c-ssr), the five-step assessment and classification of suicidality (safe-t), and the patient health questionnaire-9 (phq9) [13]. The following five suicidal ideation subscales of the Columbiasuicidality severity rating scale (c-ssrs) [15] have been used in many studies to assess suicidal ideation. Developing effective systems for detecting suicidal behavior using deep learning involves addressing several important challenges.

II. RELATED WORK

Suicidal ideation is a widely studied phenomenon in psychology. Understanding the underlying factors and risk factors associated with suicidal ideation and behavior is critical to developing effective prevention strategies and interventions. The impact of psychological factors such as depression, anxiety, hopelessness, and a sense of worthlessness on the development of suicidal ideation have been investigated in many studies. These studies have investigated the relationship between suicidal ideation and conditions such as depression [16], bipolar disorder [17], personality disorder [18], and drug addiction [19]. By examining the interaction of these factors, researchers aim to develop intervention plans that address the specific issues that individuals struggle with suicidal ideation [20], [21]. Environmental factors such as trauma history, social isolation, and access to death have also been identified as risk factors [22], [23], [24]. Psychological theories and frameworks such as suicide self-concept [25], suicide cognitive behavior [26], and the social model [27] provide the theoretical basis for understanding the interaction between personal vulnerability and the environment. Extensive research on suicidal ideation and related topics in psychology has contributed greatly to the understanding of these complex issues. By uncovering the causes, risk factors, and prevention factors associated with suicidal ideation, researchers aim to develop effective prevention strategies, improve mental health, and ultimately reduce the burden of suicide worldwide. The literature review encompasses a diverse range of methodologies and techniques employed for the detection of suicidal ideation across various online platforms. Transformer-based models like RoBERTa and BioBERT are evaluated in conjunction with dataset balancing strategies, primarily focusing on health-related tweets [28]. Detection of individuals with suicidal tendencies on platforms like online forums and Reddit is conducted using classification algorithms such as Logistic Regression, Random Forest, and Support Vector Machine (SVM) [29]. The embedding technique of term frequency-inverse document frequency (TF-IDF) is combined with algorithms like Machine, Logistic Regression, and AdaBoost for analyzing Reddit posts. A plethora of classification models, including Random Forest, Decision Tree, Naïve Bayes variants, Recurrent Neural Networks, and Artificial Neural Networks, is utilized for identifying suicidal tweets among non-suicidal ones[30]. A curated word-list is employed for detecting depression tendencies in individual tweets, utilizing Support Vector Machines and Naïve-Bayes classifiers for prediction [31]. Additional models such as Naïve Bayes, Logistic Regression, Random Forest, and Support Vector Machine Classifiers are also deployed for classification, using datasets like CLEF 2020[32]. Deep learning classifiers incorporating word embeddings, namely BiLSTM, LSTM, BiGRU, CLSTM, along with traditional models like Random Forest, Support Vector Classifier, and XGBoost, are utilized for tweet classification [33]. Pre-processing techniques involving Natural Language Processing (NLP) methods and sentiment analysis are adopted to clean and analyze tweets, with Bag-of-Words and TF-IDF features used in conjunction with SVM and Naïve Bayes [34] [35]. GloVe embeddings and deep neural networks are also employed for detecting suicidal texts, with Random Forest proving to be fast and accurate [36]. Twitter datasets are utilized with various machine learning algorithms including Naïve Bayes, Logistic Regression, J48, Random Forest, and Support Vector Machines, combined with NLP and lexicon-based approaches [37]. Other methods like K-NN, Decision Trees, and XGBoost are employed using data extracted from YouTube comments [38]. Comparisons are made between models like TF-IDF, N-gram, and LinearSVC for suicide prediction using tweets containing suicidal thoughts [39]. Finally, deep learning and machine learning-based classification approaches, such as LSTM-CNN combined models, are applied to Reddit posts to evaluate and compare different classification techniques [40].

In this study [41], we developed a Bangladesh suicide database named BanglaSPD and compared the performance of various machine learning and deep learning methods in predicting suicide by training and evaluating them on this information. The paper [42] reports the use of two deep learning algorithms, lstm and distils, for selfdestructing Bert content; the latter performs better. Finally, we leverage deep learning insights for the detection of suicidal content on social media and pre-deploy a telegram bot model that detects messages containing suicide hotspots and sends encouraging messages in response, ringing a warning bell to friends and relatives. When comparing machine learning classifiers and hybrid models, the neural network (nn) classifier outperformed the baseline with 94% precision, 94% recall, 94% f1 score, and 94% overall accuracy. Online social network (sn) data provides a data stream rich in terms of content and physical structure. It holds promise in predicting suicidal thoughts and behaviors. Combining sn data with machine learning algorithms provides a way forward. This work proposed the maximum voter turnout of a group of actors and applied it to the reddit dataset to determine suicidal thoughts. Preprocessing includes data cleaning, tokenization, and lemmatization. Tf-idf and word2vec word embedding techniques are also used. It uses various machine learning algorithms, including support vector machine (svm), logistic regression (lr), random forest (RF), multivariate naive bayes (mnb), Adaboost, and XGBoost. The results of machine learning classifiers (mlc) are combined with the maximum the results clearly show that the top voters in the voting group increased the accuracy rate by 91.39%, while the precision reached 87.5%. The application of ensemble techniques (ets) to Sn data has the potential to address the complexities and challenging models encountered in predicting suicidality in dynamic time.

III. METHODOLOGY

Steps of the proposed used in this work is shown below in figure 4.1. For this work we are using Reddit dataset. We have collected the text data from Reddit related to suicidal and non-suicidal posts.

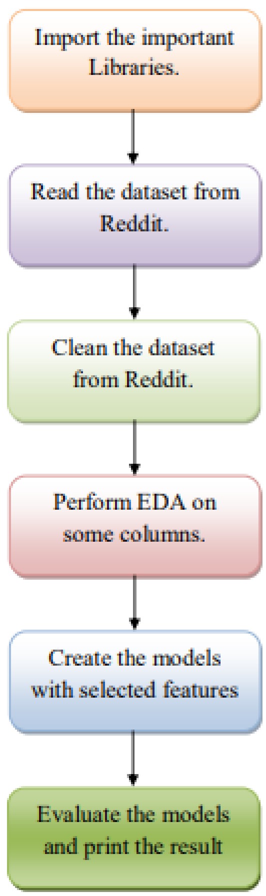


Figure 1: Proposed workflow

To build architecture for detecting suicidal behavior using an LSTM (Long Short-Term Memory) model, we need to ensure that the system can process and analyze sequential data, such as text, speech, or even time-series data. Below is a proposed step that outlines how to approach this task:

1. Data Collection and Preprocessing

Input Data Types: o Text: Social media posts, text messages, forum posts, or any form of textual communication.

Preprocessing Steps:
o Text Data:
Tokenization: Convert raw text into sequences of words or characters.

- Removal of stopwords, special characters, and normalization.
- Lemmatization: Reduce words to their base form.
- Vectorization: Convert text into numerical vectors using techniques like TF-IDF, word embeddings (Word2Vec, GloVe, or contextual embeddings like BERT).

2. Feature Engineering

Embedding layers: Used fine-tune contextual embeddings (e.g., BERT) depending on the model's performance needs. o Sequence padding: Make sure that the text data is of a consistent length for LSTM input.

3. Model Architecture

Input Layer: For text: An embedding layer that converts words into vectors.

LSTM Layer(s): Stack multiple LSTM layers (2-3 layers) to capture temporal dependencies in sequential data. Each LSTM layer learns patterns over times that are important for understanding behavioral changes and predicting suicidal ideation. The number of LSTM units (neurons) depends on the complexity of the data. Typically, between 64 and 512 units work well. Use Bidirectional LSTM to capture both past and future context in the data. Include Dropout layers between LSTM layers to reduce overfitting, especially for smaller datasets. Attention Mechanism (Optional):Introduce an attention layer to allow the model to focus on important words or timeframes in the sequence. This helps to better capture the nuances in emotional shifts, tone, or physiological stress signals. The attention mechanism can help weigh the importance of certain parts of the text (or time-series data), which is crucial in detecting specific suicidal signals.

Dense Layer(s): After the LSTM layers, you can include one or two fully connected (Dense) layers to learn higher-level representations of the sequential patterns. Typically, a layer with 128 or 64 units followed by an activation function such as ReLU. Output Layer: Binary Classification: Use a sigmoid activation function for binary classification (suicidal vs non-suicidal).

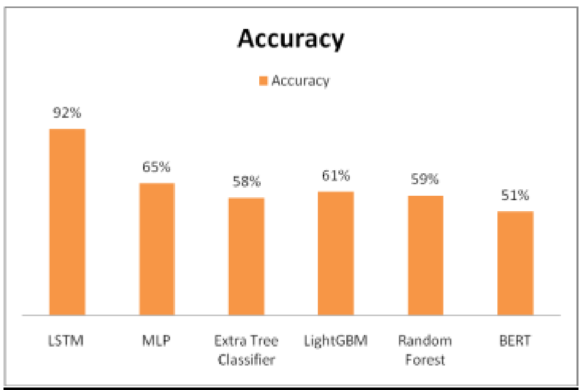
IV. PERFORMANCE EVALUATION

Implementing suicidal behavior detection using Long Short-Term Memory (LSTM) networks in Python involves several steps. LSTM is a type of recurrent neural network (RNN) that is wellsuited for time-series data and sequential data, which is particularly useful for text-based data like social media posts, conversations, or even text entries that may contain subtle patterns indicative of suicidal ideation.

Table below represents accuracy of various methods in the study:

Algorithm Used	Accuracy
LSTM	92%
MLP	65%
Extra Tree Classifier	58%
LightGBM	61%
Random Forest	59%
BERT	51%

Figure below represents graphical comparison of accuracy of different methods.



V. CONCLUSION

Suicidal behavior detection is a critical field in mental health, aiming to identify individuals at risk for suicide and provide timely interventions. Advances in technology, such as machine learning algorithms, natural language processing (NLP), and wearable devices, have significantly enhanced the ability to detect suicidal ideation and behaviors. These tools enable the analysis of a variety of data sources—ranging from social media posts and electronic health records to physiological indicators and speech patterns. In conclusion, while we are making significant strides in detecting suicidal behavior with technology, ongoing research and a careful focus on ethical practices, data privacy, and the integration of human expertise remain essential for these systems to be truly effective in preventing suicides and providing the necessary support to those at risk.

REFERENCES

[1] Almouzini S, Khemakhem M, Alageel A. Detecting Arabic Depressed Users from Twitter Data. Procedia Comput Sci 2019; 163:257– 65. https://doi.org/10.1016/j.procs.2019.12.107.

[2] Ashok Kumar J, Trueman TE, Abinesh AK. Suicidal risk identification in social media. Procedia CIRP 2021; 189:368–73.

[3] D. De Berardis, G. Martinotti, and M. Di Giannantonio, ''Understanding the complex phenomenon of suicide: From research to clinical practice,'' Frontiers Psychiatry, vol. 9, p. 61, 2018.

[4] E. R. Kumar and N. Venkatram, ''Predicting and analyzing suicidal risk behavior using rule-based approach in Twitter data,'' Soft Comput., pp. 1–9, 2023.

[5] A. Raza, F. Rustam, H. U. R. Siddiqui, I. D. L. T. Diez, B. Garcia-Zapirain, E. Lee, and I. Ashraf, ''Predicting genetic disorder and types of disorder using chain classifier approach,'' Genes, vol. 14, no. 1, p. 71, Dec. 2022.

[6] A. Abdulsalam and A. Alhothali, ''Suicidal ideation detection on social media: A review of machine learning methods,'' 2022, arXiv: 2201.10515.

[7] Z. Li, J. Zhou, Z. An, W. Cheng, and B. Hu, ''Deep hierarchical ensemble model for suicide detection on imbalanced social media data,'' Entropy, vol. 24, no. 4, p. 442, Mar. 2022.

[8] D. Kodati and R. Tene, ''Identifying suicidal emotions on social media through transformer-based deep learning,'' Appl. Intell., vol. 53, no. 10, pp. 11885–11917, 2023.

[9] M. M. S. Fareed, A. Raza, N. Zhao, A. Tariq, F. Younas, G. Ahmed, S. Ullah, S. F. Jillani, I. Abbas, and M. Aslam, ''Predicting divorce prospect using ensemble learning: Support vector machine, linear model, and neural network,'' Comput. Intell. Neurosci. vol. 2022, pp. 1– 15, Jul. 2022.

[10] Q. Wei, A. Franklin, T. Cohen, and H. Xu, ''Clinical text annotation— what factors are associated with the cost of time?'' in Proc. AMIA Annu. Symp. Bethesda, MD, USA: American Medical Informatics Association, Dec. 2018, pp. 1552–1560.

[11] Sisask, M., V¨ arnik, A., Kolves, ˜ K., Konstabel, K., & Wasserma, D. (2008). Subjective psychological well-being (WHO-5) in assessment of the severity of suicide attempt. Nordic Journal of Psychiatry. 62(6), 431–435. https://doi.org/10.1080/ 08039480801959273

[12] Find a Helpline. (2023). Help lines in Thailand for suicide. https://findahelpline.com/th /topics/suicidal-thoughts (accessed 1 July 2023).

[13] Allarakha, S., & Uttekar, P. S. (2021). What is a suicidal ideation scale?. https://www.medi

cinenet.com/what_is_a_suicidal_ideation_scale/article.htm (accessed 1 July 2023). American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5). Washington DC, USA: American Psychiatric Press.
[14] Beck, A. T., Steer, R. A., & Ranieri, W. F. (1988). Scale for suicide ideation: Psychometric properties of a self-report version. Journal of clinical psychology, 44(4), 499–505. https://doi.org/10.1002/1097-4679.
[15] Posner, K., Brent, D., Lucas, C., Gould, M., Stanley, B., Brown, G., Fisher, P., Zelazny, J., Burke, A., Oquendo, M., Mann, J. (2010). Columbia-suicide severity rating scale (CSSRS). https://cssrs.columbia.edu/wp-content/uploads/C-SSRS_Pediatric-SLC_11. 14.16.pdf (accessed 1 July 2023)
[16] N. H. Kalin, ''Insights into suicide and depression,'' Amer. J. Psychiatry, vol. 177, no. 10, pp. 877–880, Oct. 2020.
[17] L. da Silva Costa, Á. P. Alencar, P. N. Neto, M. do Socorro Vieira dos Santos, C. G. L. da Silva, S. de França Lacerda Pinheiro, R. T. Silveira, B. A. V. Bianco, R. F. F. P. Júnior, M. A. P. de Lima, A. O. A. Reis, and M. L. R. Neto, ''Risk factors for suicide in bipolar disorder: A systematic review,'' J. Affect. Disorders, vol. 170, pp. 237–254, Jan. 2015.
[18] J. Paris, ''Suicidality in borderline personality disorder,'' Medicina, vol. 55, no. 6, p. 223, May 2019.
[19] K. H. Lee, J. S. Jun, Y. J. Kim, S. Roh, S. S. Moon, N. Bukonda, and L. Hines, ''Mental health, substance abuse, and suicide among homeless adults,'' J. Evidence-Informed Social Work, vol. 14, no. 4, pp. 229–242, Jul. 2017
[20] C. Okolie, M. Dennis, E. S. Thomas, and A. John, ''A systematic review of interventions to prevent suicidal behaviors and reduce suicidal ideation in older people,'' Int. Psychogeriatrics, vol. 29, no. 11, pp. 1801–1824, Nov. 2017.
[21] E. M. Kleiman, B. J. Turner, S. Fedor, E. E. Beale, J. C. Huffman, and M. K. Nock, ''Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies,'' J. Abnormal Psychol., vol. 126, no. 6, pp. 726–738, 2017.
[22] N. Leigh-Hunt, D. Bagguley, K. Bash, V. Turner, S. Turnbull, N. Valtorta, and W. Caan, ''An overview of systematic reviews on the public health consequences of social isolation and loneliness,'' Public Health, vol. 152, pp. 157–171, Nov. 2017.
[23] J. Holt-Lunstad, T. B. Smith, M. Baker, T. Harris, and D. Stephenson, ''Loneliness and social isolation as risk factors for mortality: A Meta analytic review,'' Perspect. Psychol. Sci., vol. 10, no. 2, pp. 227–237, Mar. 2015.
[24] A. Allchin, V. Chaplin, and J. Horwitz, ''Limiting access to lethal means: Applying the social ecological model for firearm suicide prevention,'' Injury Prevention, vol. 25, no. 1, pp. i44–i48, Sep. 2019.
[25] K. A. V. Orden, T K. Witte, K. C. Cukrowicz, S. R. Braithwaite, E. A. Selby, and T. E. Joiner Jr., ''The interpersonal theory of suicide,'' Psychol. Rev., vol. 117, no. 2, pp. 575–600, 2010.
[26] A. Wenzel and A. T. Beck, ''A cognitive model of suicidal behavior: Theory and treatment,'' Appl. Preventive Psychol., vol. 12, no. 4, pp. 189–201, Oct. 2008.
[27] R. J. Cramer and N. D. Kapusta, ''A social-ecological framework of theory, assessment, and prevention of suicide,'' Frontiers Psychol., vol. 8, p. 1756, Oct. 2017
[28] Pimpalkhute V, Nakhate P, Diwan T. IIITN NLP at SMM4H 2021 Tasks: Transformer Models for Classification of Health-Related Tweets. Soc Media Min Heal SMM4H 2021 - Proc 6th Work Shar Tasks 2021:118–22.
[29] Aladag AE, Muderrisoglu S, Akbas NB, Zahmacioglu O, Bingol HO. Detecting suicidal ideation on forums: Proof-of-concept study. J Med Internet Res 2018;20:1–10. https://doi.org/10.2196/jmir.9840.
[30] Chadha A, Kaushik B. Performance Evaluation of Learning Models for Identification of Suicidal Thoughts. Comput J 2022; 65:139–54. https://doi.org/10.1093/comjnl/bxab060.
[31] Deshpande M, Rao V. Depression detection using emotion artificial intelligence. Proc Int Conf Intell Sustain Syst ICISS 2017 2018:858–62. https://doi.org/10.1109/ISS1.2017.8389299.
[32] Samanvitha S, Bindiya AR, Sudhanva S, Mahanand BS. Naive Bayes Classifier for depression detection using text data. 2021 5th Int Conf Electr Electron Commun Comput Technol Optim Tech ICEECCOT 2021 - Proc 2021:418–21. https://doi.org/10.1109/ICEECCOT52851.2021.9708014.
[33] Haque R, Islam N, Islam M, Ahsan MM. A Comparative Analysis on Suicidal Ideation Detection Using NLP, Machine, and Deep Learning. Technologies 2022;10. https://doi.org/10.3390/technologies10030057.
[34] Ruhi S, Diwan T. Estimating User's Social Behaviour by Analysing Online Tweet Pattern. HELIX 2018;8:4071–6.
[35] Pachouly SJ, Raut G, Bute K, Tambe R, Bhavsar S. Depression Detection on Social Media Network (Twitter) using Sentiment Analysis. Int Res J Eng Technol 2021;08:1834–9.
[36] Thongsi K, Booncherd N, Songmuang P. Time and performance comparison on suicide detection using various feature engineering and machine learning models.

15th Int Conf Knowl Smart Technol KST 2023 2023:1–4. https://doi.org/10.1109/KST57286.2023.10086874.

[37] Profile SEE. Machine Learning and Natural Language Processing for Detecting SuicideRelated Posts in Social Media 2021.

[38] Saifullah S, Fauziyah Y, Aribowo AS. Comparison of machine learning for sentiment analysis in detecting anxiety based on social media data. J Inform 2021;15:45. https://doi.org/10.26555/jifo.v15i1.a20111.

[39] Shanthi N, Muthuraja M, Sharmila C, Jagadeesh S, Karthick R, Bharanidharan M. Suicidal Ideation Prediction Using Machine Learning. 2023 Int Conf Comput Commun Informatics, ICCCI 2023 2023:1–4. https://doi.org/10.1109/ICCCI56745.2023.10128254.

[40] Tadesse MM, Lin H, Xu B, Yang L. Detection of suicide ideation in social media forums using deep learning. Algorithms 2020;13:1– 19. https://doi.org/10.3390/a13010007.

[41] S. Islam, M. S. Alam Forhad and H. Murad, "BanglaSAPM: A Deep Learning Model for Suicidal Attempt Prediction Using Social Media Content in Bangla," 2022 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2022, pp. 1122-1126, doi: 10.1109/ICCIT57492.2022.10055237.

[42] N. Sharma and P. Karwasra, "Suicidal Text Detection on Social Media for Suicide Prevention Using Deep Learning Models," TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON), Chiang Mai, Thailand, 2023, pp. 25-30, doi: 10.1109/TENCON58879.2023.10322499