

AI Against Suicide: Real – Time Risk Detection from Social Media Texts

V V N S S Raghunath¹, Tejinder Thind², Uminder Kaur³

¹Student, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India. Email: vnssrnath2002@gmail.com

²Assistant Professor, School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India. Email: tejinder.15312@lpu.co.in

³Assistant Professor, School of Computer Applications, Lovely Professional University, Phagwara, Punjab, India. Email: uminder.23377@lpu.co.in

Abstract: The rapid increase in the social media platforms has promoted the need to have smart systems that could identify suicidal thoughts in real-time. The study is aimed at automated suicide risk detection on the dataset Suicide Post Detection on Social Media Articles published by Google. Our machine learning and deep learning techniques varied and included Logistic Regression, Support Vector Machine (SVM), and ensemble techniques, including Stacking and Voting Classifiers. Deep learning architectures such as Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and CNN- LSTM hybrid, and finely tuned transformer-based models such as RoBERTa and ELECTRA were also used. The ELECTRA model performed far better compared to the rest with an accuracy of 95.52 and a global F1-score of 95.53. To simplify the interpretation of the results, explainable AI systems such as LIME and SHAP were applied in order to highlight specific characteristics of the language that influenced the selection of models. This practice will combine transparency and predictive intelligence, thus helping to identify high-risk posts effectively and useful assistance in the early intervention of suicide by analyzing text in real-time.

Index Terms - suicide detection, social media analysis, deep learning, transformer models, ELECTRA, RoBERTa, explainable AI, mental health monitoring.

How to cite this article: Nath VVNSSR, Thind T, Kaur U. AI Against Suicide: Real – Time Risk Detection from Social Media Texts. *Int J Drug Deliv Technol.* 2026;16(13s): 137-144. DOI: 10.25258/ijddt.16.13s.14

1. INTRODUCTION

Suicide has become one of the most urgent worldwide health emergencies that have caused millions of deaths each year and have left social and emotional consequences that affect families and communities long term. Social media is now a part of everyday reality in the digital age, and it offers people a platform to share their feelings, experiences, and psychological hardships. Centers of posts and comments on Twitter, Reddit, and Facebook usually have some slight signs of mental distress or suicide tendencies that, when identified in time, can be used to avert a terrible result. But the large quantity and the complexity of online data makes it impossible to monitor it manually, which is why automated detection systems operating on the principle of advanced computational procedures are in high demand to detect high-risk people [1][2].

AI and ML have transformed the possibility to process textual information and reveal emotional, linguistic, and semantic patterns that are correlated with suicidal thoughts. Recent research has established that NLP-based models can be successfully used to differentiate between suicidal and non-suicidal texts by using

context-sensitive word embeddings, affective features, and psychological indicators [3][4]. More complex architectures based on ensemble learning, deep learning, and genetic algorithms have been established in order to improve detection and interpretability [5]. In addition, explainable AI (XAI) has become a trend thanks to the possibility to give transparency to the results of predictions, which means that it is possible to make decisions based on AI models and trust them to clinicians and mental health professionals [6].

Continued studies also highlight the need to incorporate psychological findings and behavioral knowledge into the system to enhance its situational awareness of suicidal rhetoric [7]. Predictive abilities are further enhanced by the incorporation of multimodal data, which is textual, visual, and social interaction, and reflects complex emotional expression [8]. The use of real-time analytics systems that utilize big data has become more common to track social media feeds in real-time to provide early warnings to intervene in mental health [9]. Also, other new directions are investigating the integration of

environmental and lifestyle risk factors (e.g. diet) with textual data to improve suicide risk prediction [10].

The objective of the research is to design and evaluate an artificial intelligence system with the capability to detect suicidal ideation on social media. Text with explainable and ensemble-based methods. The proposed piece of work, combining the NLP methods with explainable MLs, will contribute to the process of early detection, assist in the intervention strategies that will ultimately prevent suicide in the digital communities.

2. RELATED WORK

The latest developments in AI and DL have made a major step forward regarding the possibility of detecting suicidal ideation and mental disorders with the help of the analyzed social media. Ezerceci and Dehkharghani [11] created a model of deep neural network (DNN) to detect mental disorders and suicidal thoughts using linguistic and behavioral features of user-generated posts on social media. Their research also focused on semantic representations in the differentiation of depressive, neutral and suicidal articulations. The authors showed that deep learning models could successfully represent emotional cues and were more accurate and understanding of the context than more traditional machine learning approaches. On the same vein, Kancharapu and Ayyagari [12] have suggested a GAN-based deep learning model with genetic optimization and fusion of word embedding, which performed better in detecting suicidal tendencies. The combination of generative adversarial networks allowed synthesizing data which helped to resolve the problem of class imbalance and improve the model robustness.

The article by Owen et al. [13] is a 25-year-long narrative review, which illustrates the development of AI-based analysis to detect mental health disorders in individuals using social media. They reviewed the advancements and issues related to implementing machine learning and large language models (LLMs) on sensitive psychological information. The authors highlighted ethical issues, explainability, and data transparency requirement to guarantee the responsible and reliable use of AI in suicide prevention. To further elaborate on the sophistication of the model, Yang, Zhou, and Wei came up with a framework of time perspective-enhanced multi-task learning to detect suicidal ideation [14]. This method incorporated temporal and contextual aspects of user posts to enhance the detection accuracy based on the change in user sentiment as time lapses which is essential in early intervention.

Gorai and Shaw [15] presented an ensemble CNN model based on BERT-encoded posts to identify suicide risk in social media posts. Their system utilized the contextual power of the BERT embeddings and the feature extraction ability of the convolutional networks, and they were able to perform on benchmark datasets with high quality. The ensemble model showed better performance than isolated CNN or transformer structures, which highlights the significance of hybrid structures in the process of suicidal ideation detection. In a different field, Rhased et al. [16] examined how to apply face image analysis to identify suicidal risk in deep neural networks. Their study involved visual emotion recognition with mental health indicators and showed the prospects of multimodal AI systems which combine textual and facial information to identify risks fully.

Explainable AI (XAI) has also been used to deal with the increasing demand of model interpretability and transparency. Ibrahimov, Anwar, and Yuan [17] gave an extensive overview of the use of XAI in enhancing the detection of mental disorders using social media data. They talked about the ways in which model decisions can be made to be understandable to clinicians as a way of connecting the gap between computational outputs and psychological tests. Their study highlighted that explainability does not only increase trust but can also make AI models ethically commit to mental health model applications. In line with this, Tejaswini, Sathya Babu and Sahoo [18] suggested a deep learning framework that combines NLP and deep networks to detect depression in social media text. Their method effectively represented the syntactic and semantic details and demonstrated better classification properties than traditional models at the same time consuming less computational resources.

Bokolo and Liu [19] conducted a comparative study of different models of machine learning and transformer based models in detecting depression and suicide in social media text. Their results showed that transformer models, especially BERT and RoBERTa were more accurate and sensitive to context than conventional ML classifiers. The research came to the conclusion that complex suicidal expressions would be identified through deep contextualized representations. Based on this observation, Lin et al. [20] developed a hybrid model (RoBERTa-CNN) that demonstrates the significance of data quality in suicide intention prediction. Their findings showed that combining high quality labeled data with hybrid neural architectures is very effective in improving the model performance and reliability.

3. MATERIALS AND METHODS

The proposed AI-based hybrid model is recommended to identify suicidal intent in social media texts. It is a hybrid of classical-based, deep learning, and transformer-based models. In order to predict a baseline and collective forecasting, there are such methods as Stacking and Voting Classifiers and some common machine learning algorithms as Logistic Regression and Support Vector Machine. Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and CNN+LSTM that are examples of deep learning models can extract out of written data emotional tone, the sequence of events and minor details of the context. Transformer-based models such as the fine-tuned version of RoBERTa and ELECTRA are added to achieve better contextual knowledge and select suicidal thoughts with accurate accuracy. The hybrid approach of the system features the accumulation of the statistical, semantic, and contextual learning advantages to achieve the best accuracy and sensitivity in the category of suicidal intents, which conforms to current developments in the suicide prevention models based on large language models [21], [23].

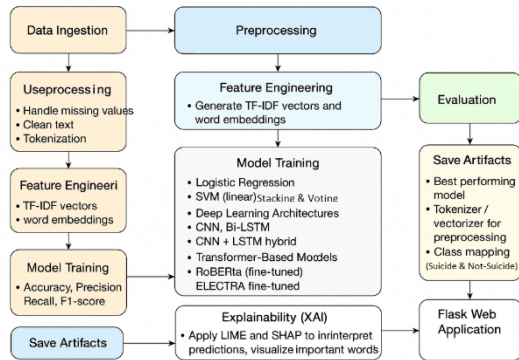


Fig.1 Proposed Architecture

Figure 1 describes the suicidal tweet identification system. It begins by Data Ingestion, which is followed by Preprocessing (dealing with missing values, cleaning etc.), and Feature Engineering (TF-IDF/embeddings). Data is prepared with these steps, which include classical ML, deep learning, and fine-tuned Transformer models (RoBERTa, ELECTRA). Evaluation, artifact saving and Explainability (XAI) are the last steps to be followed before deploying it in Flask Web Application.

i) Dataset Collection:

The Suicidal Tweet Detection Dataset is a formatted binary classification dataset to analyze a text of Natural Language Processing. It is mostly filled with a Tweet column, which has posts of social media and a Suicide column, in which 1 refers to a possible suicidal post and 0 refers to a non-suicidal post. It is the dataset,

which is usually in CSV format and collected in the form of a social media scraper, and the preprocessing of the data, including text cleaning, elimination of URLs, and class imbalance. The dataset will be helpful in providing the effective training and assessment of the AI models to recognize suicide-related content in the online communication and as a part of the early mental health risk detection and prevention process [24].

	Tweet	Suicide
0	making some lunch	Not Suicide post
1	@Alexia You want his money.	Not Suicide post
2	@dizzyhrvy that crap took me forever to put to...	Potential Suicide post
3	@jnaylor #kiwitweets Hey Jer! Since when did y...	Not Suicide post
4	Trying out "Delicious Library 2" wit...	Not Suicide post

Fig.2 Dataset Collection

ii) Pre-Processing:

Preprocessing is an important stage which makes the datasets clean, organized and analyzable. It entails handling the missing values, eliminating irrelevant data such as URLs and emojis, transforming text into lower case and tokenizing the text into useful pieces that can be further used to extract features.

a) **Data Processing:** The process of data processing is aimed at converting raw social media text into a readable machine format. Noise, stopwords and redundant symbols are removed to clean up the text. At this point the process of tokenization is applied to divide the text into tokens and subsequently converting these tokens into a number representation. Stratified splitting divides the processed data into training, validation, and testing datasets, so that the ratio of classes is balanced and to ensure that there is performance evaluation consistency across all model architectures.

b) **Feature Engineering:** The process of feature engineering consists of converting processed text into well-representative numerical features, which are understandable in semantic meaning and in linguistic patterns. It involves the creation of text vectors, sequence embedding, and contextual representations, which make the model better comprehend the word dependencies. Learning emotional tone and contextual nuances have their input in these engineered features, which can be better classified and interpreted as suicidal and non-suicidal expressions.

c) **Data Resampling:** The sampling of data is necessary to overcome the imbalance of classes in the data set. When either aspect is hegemonic, synthetic data generation or the application of weighted loss functions are applied in order to balance fairness in training. This measure will help avoid model biasing the majority

classes and will improve sensitivity to minority (suicidal) samples. The balance in the distribution of classes increases the system in terms of higher accuracy, generalization, and reliability when it comes to the real-world situation of suicide risk detection.

iii) Train & Test:

The dataset consists of two components training and testing. The training information is what allows the model to learn with the text being inputted how to detect patterns, relationships, and other significant linguistic features. The data of testing is just utilized to visualize how excellently the model learned works on fresh samples that belong to the actual world. It is not observed during training. This distinction is so that model learning remains unbiased and is not overfitted to known data. The procedure of division provides a balanced distribution of the classes in subsets, and both sets have an equal proportion of suicidal and non-suicidal samples, which allows valid and reliable performance validation.

iv) Algorithms:

Logistic regression is a statistical machine learning method which is utilized to address two-option issues. It makes an educated guess on the probability that a piece of text that you feed it on falls into a particular category, e.g. suicidal or non-suicidal. Here, [22] Logistic Regression is used to assess the textual patterns and the frequency of words obtained out of the textual posts in social media. It is used to offer a simple and yet an effective baseline model that learns a linear relation in the features and the output classes in order to detect suicidal intent. Its interpretability and efficiency make the algorithm an appropriate choice when it comes to establishing patterns in smaller data sets as well as assisting in the real-time classification task and contributing to the ability of the system to anticipate the future and its clarity in identifying mental health risk.

Text is sorted using a ML approach known as SVM with a linear kernel. It does this by identifying the optimal hyperplane that separates suicidal and non-suicidal texts most. In this generation, [25] SVM works with TF-IDF or embedding-based features to discover the latent structure in posts of social media. Its application is concerned with the high generalization performance, despite the limited data, because it can effectively work with sparse feature representations. The aim of SVM is to give a high accuracy of classification to detect a suicide by way of maximizing the accuracy and minimizing the misclassifications, and ensuring that the computational efficiency is maintained when doing real time prediction.

Ensemble model is a combination of more than two classifiers which make it predict more and be more stable. In this, the ensemble methods, Stacking and Voting Classifiers are methods that combine the output of the Logistic Regression, SVM, and other models to arrive at final predictions. In stacking, predictions are aggregated by using meta-learners and in Voting, predictions are aggregated by majority or weighted voting. Its aim is to reduce the amount of errors in individual models and use them to their advantage to achieve enhanced accuracy and reliability. Ensemble learning improves on generalization, variance, and more uniform results on suicide-related text classification than single-model methods.

One of the DL models used is CNN, which learns local text data dependencies with convolutional filters. It determines emotional and word-related patterns of suicide posts on the basis of n-grams. The use of CNN is related to the processing of word embeddings to extract the main contextual characteristics that lead to the suicide intent detection. It is intended to provide a higher classification precision with learning spatial relationships of text sequences, thus allowing the model to perceive a subtle tone of emotion. The parallel computation and hierarchical feature learning of CNN that is efficient makes it applicable to the real-time and scalable suicide risk analysis applications.

Bidirectional Long Short-Term Memory (Bi-LSTM) is a type of a RNN that reads text forward and backward to get the entire contextual meaning. Bi-LSTM is used in this system to analyze emotional dependencies in sentence to predict the intent of suicide. It is used in learning long-term contextual relationships between words to detect subtle manifestations of distress. This is to learn sequential emotional flow in text and achieve better and context-sensitive detection. Bi-LSTM enhances the capacity of the model to comprehend linguistic expressions and emotional patterns that require complexity in the social media post.

The CNN + LSTM hybrid model is a combination between CNN and LSTM with features and sequential learning power respectively. The CNN layers learn local patterns of phrases, whereas the LSTM layers learn long-term patterns in the text. Such a mix design is applied to strike a balance between the efficiency and contextual richness in text classification of suicidal information. It is intended to be more precise and accomplish this through combining spatial and temporal representation, which is also effective to identify the slightest linguistic hints. The hybrid architecture provides strong, contextualized predictions, which upgrades the system to identify

different intensity and intentions of emotions in the user generated social media posts.

RoBERTa (Robustly Optimized BERT Approach) is a transformer-based fine-tuned suicide text detector. It employs self-attention processes in providing a contextual analysis of word relationship and identification of subtle emotion pattern. RoBERTa, in this, works on an input text at the sentence level, making sense out of more than just surface words. Its application improves context interpretation in the process of classifying suicidal and non-suicidal posts. The aim is to understand the meaning in depth, which is enhanced with a transfer learning. The diverse linguistic expressions that RoBERTa can work with make it effective in providing an accurate classification in real time, which contributes to mental health risk prediction, which is delicate.

ELECTRA is a transformer model that is pre-trained on a replace-token detection task, enabling the model to be computationally efficient and powerful in contexts. Within the framework of this system, fine-tuned ELECTRA is used to analyze the text of social media and identify expressions of suicide. It is used in taking advantage of contextual embeddings of emotional intensity and intent in sentences. ELECTRA is aimed at fast, accurate, and efficient suicide risk detection. With language finess learning, it provides enhanced performance over conventional transformers, which provides strong interpretability and scalability to mental health monitoring and intervention systems.

Table.1 Performance Evaluation

ML Model	Accuracy	Macro F1 Score	Recall	Precision
RoBERTa	0.9496	0.9497	0.9496	0.9501
ELECTRA	0.9552	0.9553	0.9552	0.9554
CNN	0.9160	0.9104	0.9130	0.9081
Bi-LSTM	0.9216	0.9159	0.9159	0.9159
CNN + LSTM Hybrid	0.7619	0.7615	0.8064	0.7954
Logistic Regression	0.8337	0.8355	0.8337	0.8511

4. RESULTS AND DISCUSSIONS

Accuracy: The accuracy of a test is how well it can tell the difference between sick and healthy people. To figure out how accurate a test is, we should figure out how many of the cases we looked at were true positives and true negatives. This can be said mathematically as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

Precision: Precision looks at the percentage of accurately classified instances or samples among those that were labeled as positives. So, the formula for figuring out the precision is:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

Recall: In machine learning, recall is a measure of how well a model can find all the relevant examples of a certain class. It is the ratio of correctly predicted positive observations to the total number of real positives. This gives you an idea of how well a model captures instances of a certain class.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-Score: The F1 score is a way to quantify how accurate a machine learning model is. It adds together the precision and recall scores of a model. The accuracy measure tells you how many times a model produced a valid guess over the whole dataset.

$$F1\ Score = 2 * \frac{Recall * Precision}{Recall + Precision} * 100 \quad (1)$$

SVM (Linear)	0.8581	0.8581	0.8581	0.8603
Stacking Classifier	0.9135	0.9134	0.9135	0.9141
Voting Classifier	0.9224	0.9221	0.9224	0.9248

Table 1 presents the comparative performance metrics of various models based on accuracy, macro F1 score, recall, and precision values.

Fig.3 Comparison Graph

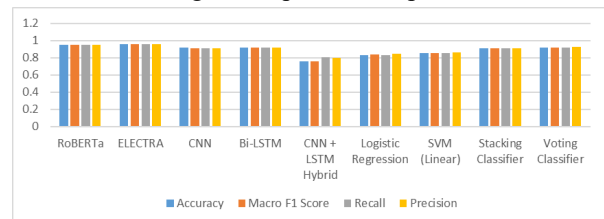


Fig. 3 shows Accuracy (blue), Macro F1 Score (orange), Recall (gray), and Precision (yellow) across all models.

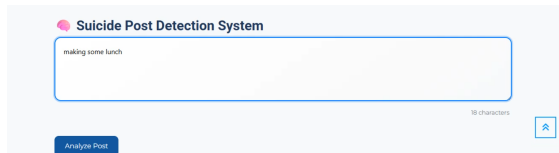


Fig.4 Upload Input Text

The Web Application (Fig. 4) features a text input box where a user submits a tweet for real-time classification by the deployed suicide detection model.

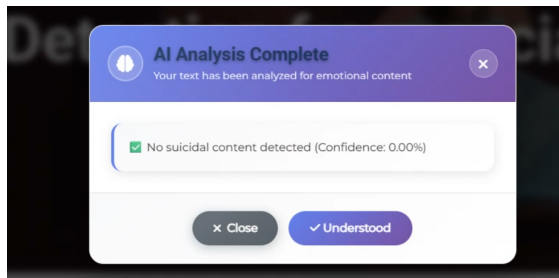


Fig.5 Predict Result

Figure 5 shows the successful model output, confirming no suicidal content was detected in the submitted text.

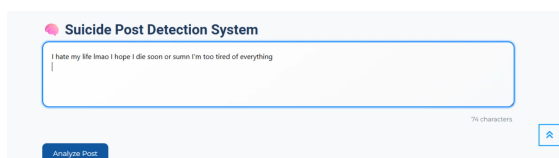


Fig.6 Upload Another Input Text

Figure 6 displays a high-risk tweet with phrases like "I hate my life" and "I hope I die soon" ready for analysis.

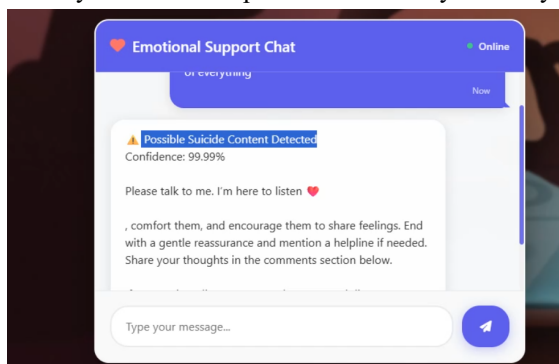


Fig.7 Final Outcome

Figure 7 shows the high-risk tweet was classified with 99.99% confidence as containing Possible Suicide Content, triggering a support message.

5. CONCLUSION

The research effectively demonstrates the capability of AI in identifying suicide-related stuff from social media texts that is quite accurate and easy to understand. The system performed well at finding verbal clues that suggested suicide intent through integrating conventional ML, neural networks, and transformers. The ELECTRA model with the highest level of accuracy of 95.52 percent and highest macro F1-score of 95.53 was the fine-tuned version of the

model. It outperformed other architectures such as RoBERTa, CNN, Bi-LSTM and ensemble classifiers. The transparency and reliability of the models were also improved with the help of explainable AI techniques such as LIME and SHAP, which allowed us to see which text elements contribution to predictions was the most significant. This combination of predictive accuracy and interpretability provides a robust model of real-time suicide risk detection, which is a scalable and smart method of monitoring mental health. The findings highlight the significance of utilizing the state-of-the-art transformer structures and explainability instruments to bridge the existing gap between automated risk detection and clinical insights and justify the role of the data-driven approach to explain the self-harm cases through social media communication analysis.

The developmental improvements that may be made in the long-term involve incorporating multimodal data, i.e. integrating textual, visual, and behavioral cues in the social media to enhance the accuracy of suicide risks identification. It would be better to then expand the dataset to have multilingual and cross-platform posts to increase model generalization in different user groups. The use of time analysis may contribute to monitoring the emotional changes and determine the regular patterns of behavior in the long run. Also, further fine-tuning of contextual understanding can be achieved by employing sophisticated transformer architectures with domain-specific fine-tuning. Dynamically adapting to changing language patterns as a mental health concern can be done through real-time deployment with cloud-based frameworks and retraining models with new data. The early use of clinical and social applications of suicide prevention and intervention by early integration with mental health professionals and ethical governance structures will promote responsible deployment, user privacy, and reliability.

REFERENCES

- [1] Alghazzawi, D., Ullah, H., Tabassum, N., Badri, S. K., & Asghar, M. Z. (2025). Explainable AI-based suicidal and non-suicidal ideations detection from social media text with enhanced ensemble technique. *Scientific Reports*, 15(1), 1111.
- [2] Mansoor, M. A., & Ansari, K. H. (2024). Early detection of mental health crises through artificial-intelligence-powered social media analysis: A prospective observational study. *Journal of personalized medicine*, 14(9), 958.
- [3] Sweeney, C., Ennis, E., Mulvenna, M. D., Bond, R., & O'Neill, S. (2024). Insights derived from text-based

digital media, in relation to mental health and suicide prevention, using data analysis and machine learning: systematic review. *JMIR mental health*, 11, e55747.

[4] Basyouni, A., Abdulkader, H., Elkilani, W. S., Alharbi, A., Xiao, Y., & Ali, A. H. (2024). A suicidal ideation detection framework on social media using machine learning and genetic algorithms. *IEEE Access*, 12, 124816-124833.

[5] Levkovich, I., & Omar, M. (2024). Evaluating of bert-based and large language mod for suicide detection, prevention, and risk assessment: A systematic review. *Journal of Medical Systems*, 48(1), 113.

[6] Li, X., Chen, F., & Ma, L. (2024). Exploring the potential of artificial intelligence in adolescent suicide prevention: current applications, challenges, and future directions. *Psychiatry*, 87(1), 7-20.

[7] Grimland, M., Benatov, J., Yeshayahu, H., Izmaylov, D., Segal, A., Gal, K., & Levi-Belz, Y. (2024). Predicting suicide risk in real-time crisis hotline chats integrating machine learning with psychological factors: Exploring the black box. *Suicide and Life-Threatening Behavior*, 54(3), 416-424.

[8] Abdulsalam, A., & Alhothali, A. (2024). Suicidal ideation detection on social media: A review of machine learning methods. *Social Network Analysis and Mining*, 14(1), 188.

[9] Allayla, M. A., & Ayvaz, S. (2024). A big data analytics system for predicting suicidal ideation in real-time based on social media streaming data. *arXiv preprint arXiv:2404.12394*.

[10] Al-Remawi, M., Agha, A. S. A., Al-Akayleh, F., Aburub, F., & Abdel-Rahem, R. A. (2024). Artificial intelligence and machine learning techniques for suicide prediction: Integrating dietary patterns and environmental contaminants. *Heliyon*, 10(24).

[11] Ezerceli, Ö., & Dehkharghani, R. (2024). Mental disorder and suicidal ideation detection from social media using deep neural networks. *Journal of Computational Social Science*, 7(3), 2277-2307.

[12] Kancharapu, R., & Ayyagari, S. N. (2024). Suicidal ideation prediction based on social media posts using a GAN-infused deep learning framework with genetic optimization and word embedding fusion. *International Journal of Information Technology*, 16(4), 2577-2593.

[13] Owen, D., Lynham, A. J., Smart, S. E., Pardiñas, A. F., & Camacho Collados, J. (2024). Ai for analyzing mental health disorders among social media users: Quarter-century narrative review of progress and challenges. *Journal of Medical Internet Research*, 26, e59225.

[14] Yang, Q., Zhou, J., & Wei, Z. (2024). Time perspective-enhanced suicidal ideation detection using multi-task learning. *International Journal of Network Dynamics and Intelligence*, 100011-100011.

[15] Gorai, J., & Shaw, D. K. (2024). A BERT-encoded ensembled CNN model for suicide risk identification in social media posts. *Neural Computing and Applications*, 36(18), 10955-10970.

[16] Rashed, A. E. E., Atwa, A. E. M., Ahmed, A., Badawy, M., Elhosseini, M. A., & Bahgat, W. M. (2024). Facial image analysis for automated suicide risk detection with deep neural networks. *Artificial Intelligence Review*, 57(10), 274.

[17] Ibrahimov, Y., Anwar, T., & Yuan, T. (2024). Explainable ai for mental disorder detection via social media: A survey and outlook. *arXiv preprint arXiv:2406.05984*.

[18] Tejaswini, V., Sathya Babu, K., & Sahoo, B. (2024). Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 23(1), 1-20.

[19] Bokolo, B. G., & Liu, Q. (2024). Advanced comparative analysis of machine learning and transformer models for depression and suicide detection in social media texts. *Electronics*, 13(20), 3980.

[20] Lin, E., Sun, J., Chen, H., & Mahoor, M. H. (2024, July). Data Quality Matters: Suicide Intention Detection on Social Media Posts Using RoBERTa-CNN. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1-5). IEEE.

[21] Ghanadian, H., Nejadgholi, I., & Al Osman, H. (2024). Socially aware synthetic data generation for suicidal ideation detection using large language models. *IEEE Access*, 12, 14350-14363.

[22] Holmes, G., Tang, B., Gupta, S., Venkatesh, S., Christensen, H., & Whitton, A. (2025). Applications of large language models in the field of suicide prevention: Scoping review. *Journal of Medical Internet Research*, 27, e63126.

[23] Zuromski, K. L., Low, D. M., Jones, N. C., Kuzma, R., Kessler, D., Zhou, L., ... & Nock, M. K. (2024). Detecting suicide risk among US servicemembers and veterans: a deep learning approach using social media data. *Psychological Medicine*, 54(12), 3379-3388.

[24] Saleem, M., & Kim, J. (2024). Intent aware data augmentation by leveraging generative AI for stress

detection in social media texts. *PeerJ Computer Science*, 10, e2156.

[25] Dou, R., & Kang, X. (2024). TAM-SenticNet: A Neuro-Symbolic AI approach for early depression detection via social media analysis. *Computers and Electrical Engineering*, 114, 109071.