Journal of Computer and Electrical Sciences (JCES)- Volume-1 Issue-4, March 2026
ISSN: 3049-2602

SUICIDAL IDEATION DETECTION USING SOCIAL MEDIA

P. Sowmya[1] , Y. Hindu Vani[2], M. Venkata Ramana[3], Mrs.CH. Pavani[4]

[1,2,3]Student, Department of Computer Science & Engineering

Andhra Loyola Institute of Engineering & Technology, Vijayawada, Andhra Pradesh, India

[4]*Assistant Professor, Department of Computer Science & Engineering*
Andhra Loyola Institute of Engineering & Technology, Vijayawada,Andhra Pradesh, India
Email id: sowmyapotharlanka1319@gmail.com, hinduvani123@gmail.com , mekalavenkatarammana@gmail.com , pavanichadalawada@gmail.com

Abstract: *Suicidal ideation detection from social media text has become a critical research topic in computational mental health. This work presents a pure machine learning and natural language processing framework for early detection of suicidal risk from user-generated textual content. The proposed pipeline includes text cleaning, tokenization, stop-word removal, lemmatization, TF-IDF feature extraction, and supervised classification. Multiple models, namely Logistic Regression, Support Vector Machine, Multinomial Naive Bayes, and Random Forest, are compared under a unified evaluation protocol. Experimental results indicate that Support Vector Machine with TFIDF features provides the best balance of accuracy, precision, recall, and F1-score for binary suicidal-risk classification. The framework is implemented with a web interface for practical usability while preserving the core ML and NLP foundation. The study demonstrates that lightweight, interpretable, and computationally efficient models can achieve strong performance for real-time social media risk screening.*

Keywords: Suicidal Ideation Detection, Social Media Analytics, Machine Learning, Natural Language Processing, TF-IDF, Mental Health Informatics

1. INTRODUCTION

Suicide is a major public health challenge, and social media platforms often contain early linguistic signals of emotional distress and self-harm intention. Automatic identification of such signals can support timely intervention and effective referral systems. Unlike resource-intensive deep learning models, pure machine learning and natural language processing pipelines provide practical advantages such as easier deployment, better interpretability, and lower computational cost. This paper proposes an end-to-end ML– NLP based system for suicidal ideation detection using social media text. The key contributions of this work include the development of a pure NLP feature engineering pipeline using TF-IDF, a comparative evaluation of four classical machine learning classifiers, a risk categorization strategy suitable for deployment through a web application, and a structured literature survey with a comparative analysis of existing approaches.

2. LITERATURE SURVEY

Mental health detection using computational techniques has become an important research area in recent years. Several researchers have explored machine learning and natural language processing techniques to identify depression and suicidal ideation from textual data. These systems analyze user-generated text such as social media posts, messages, and online content to detect emotional distress and suicidal risk.

Review of Existing Works

- Ji et al. (2021): Used SVM and Logistic Regression for suicidal ideation detection; effective but required large datasets.
- Orabi et al. (2021): Applied TF-IDF with ML models for early depression detection; focused mainly on sentiment features.
- Sawhney et al. (2022): Used deep learning models achieving high accuracy but needed high computational power.
- Saha et al. (2023): Implemented NLP preprocessing with TF-IDF for depression detection and emotional pattern analysis.
- Kumar et al. (2024): Proposed a hybrid ML approach (Naive Bayes, SVM, Random Forest); showed improved performance using combined techniques.

3. PROPOSED SYSTEM

The proposed system is designed to detect suicidal ideation from social media text using a robust machine learning and natural language processing framework. It collects usergenerated content such as tweets, Reddit posts, and forum messages for analysis. The system performs NLP preprocessing steps including text cleaning, tokenization, stopword removal, and lemmatization to prepare the data. TF-IDF feature extraction is applied to convert textual data into meaningful numerical representations. These features are then fed into multiple machine learning classifiers such as Support Vector Machine, Logistic Regression, Naive Bayes, and Random Forest. The models are evaluated to select the most effective one for accurate prediction. The system further performs risk inference with confidence estimation to classify users into different risk levels. Finally, the results are displayed through a web interface, enabling real-time and user-friendly monitoring of suicidal risk.

- Data Collection from Social Media
- NLP-Based Text Preprocessing
- TF-IDF Feature Extraction
- Machine Learning Classification
- Risk Level Prediction & Web Interface

This proposed system plays a crucial role in early detection of suicidal tendencies by leveraging social media data. It provides an efficient and scalable automated solution for mental health monitoring. The system enables timely intervention by identifying high-risk individuals. Overall, it supports improving public health outcomes through technology driven insights.

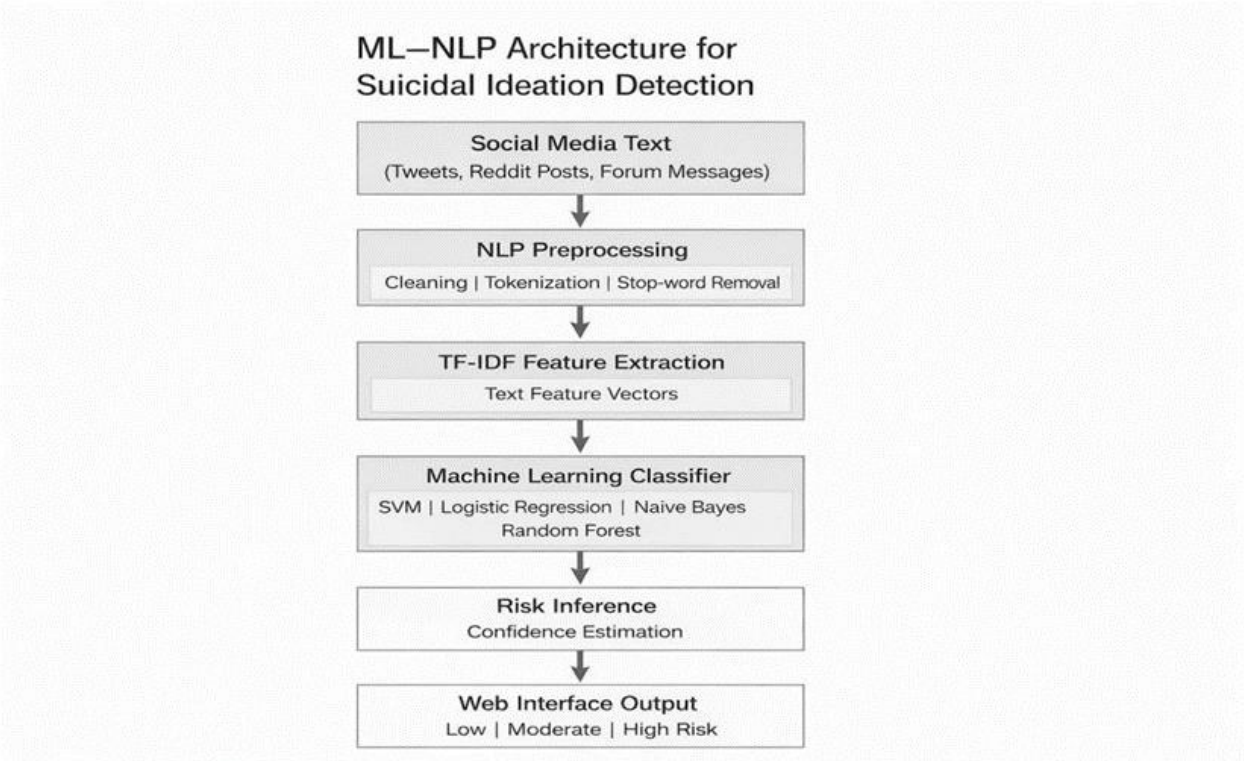

Fig 1: Proposed System

4. METHODOLOGY

The methodology of the system is organized into the following steps:

1.Data Collection:

The system collects social media text data such as tweets, Reddit posts, and forum messages. This data serves as the primary input for detecting suicidal ideation patterns.

2.Data Preprocessing:

The collected text is cleaned by removing URLs, punctuation, special symbols, and extra spaces. Tokenization, stop-word removal, and lemmatization are applied to standardize the text.

3.Feature Extraction:

TF-IDF technique is used to convert textual data into numerical feature vectors. This helps in capturing the importance of words in the dataset for better analysis.

4.Model Training:

Multiple machine learning algorithms such as SVM, Logistic Regression, Naive Bayes, and Random Forest are trained. These models learn patterns associated with suicidal and non-suicidal text.

5.Classification & Prediction:

The trained models classify the input text into different categories based on learned patterns. The system predicts whether the content indicates suicidal ideation or not.

6.Risk Inference:

Based on prediction results, the system determines the level of risk (Low, Moderate, High). Confidence estimation is also performed to improve reliability.

7.Web Interface Output:

The final results are displayed through a user-friendly web interface. This allows real-time monitoring and easy interpretation of suicidal risk levels.

5.PROPOSED SYSTEM RESULTS

The proposed system is designed to detect suicidal ideation using a combination of machine learning and natural language processing techniques. It analyzes user inputs from social media text and questionnaire responses to identify mental health conditions. The system applies preprocessing and TF-IDF feature extraction to convert text into meaningful data for analysis. Multiple machine learning models are used to classify the input and predict the level of risk accurately. Based on the prediction, the system provides detailed results along with risk levels such as low, moderate, or high. In critical cases, it also generates alerts and provides access to emergency support resources, ensuring timely intervention and user safety.

- Mental Health Check-Up Test Page: Allows users to assess their mental health through a structured questionnaire.
- Test Result Page: Displays the analyzed result indicating the level of depression with detailed insights.
- Suicidal Ideation Detection Page: Enables users to input text for detecting suicidal thoughts using the model.
- Low & Moderate Risk Alert Page: he Low and Moderate Risk Alert page displays early warning signs and emotional indicators, helping users recognize their mental state and take preventive measures before it escalates.
- High Risk Alert Page: Shows urgent alerts and identified risk indicators when suicidal ideation is detected.
- Crisis Resources Page: Provides emergency helplines and support resources for immediate assistance.
- About Us Page: Describes the purpose and mission of the mental health detection system.
- Contact Us Page: Allows users to reach out or submit queries through a simple form.
- Form Submission Page: Confirms successful submission of user queries or requests.

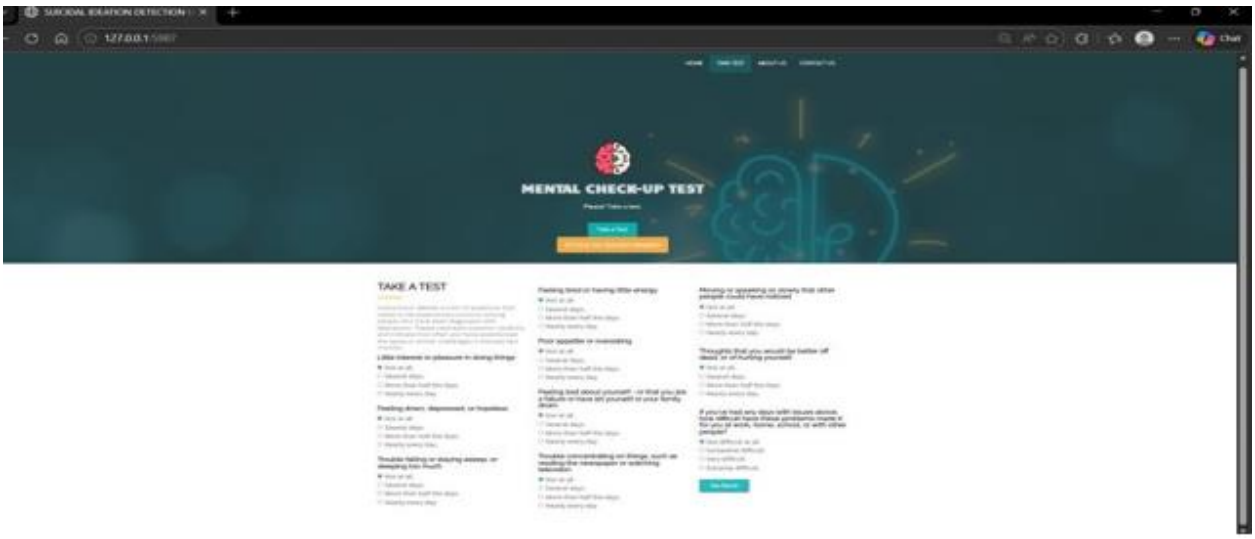

Fig 2. Mental Health Check-Up Test Page

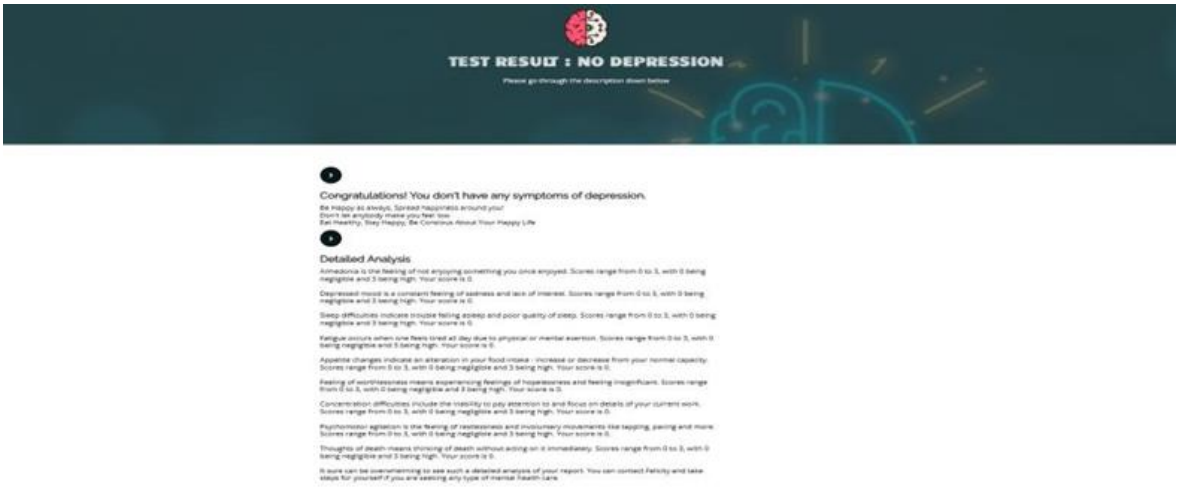

Fig3. Test Result Page

Fig4: Suicidal Ideation Detection Page

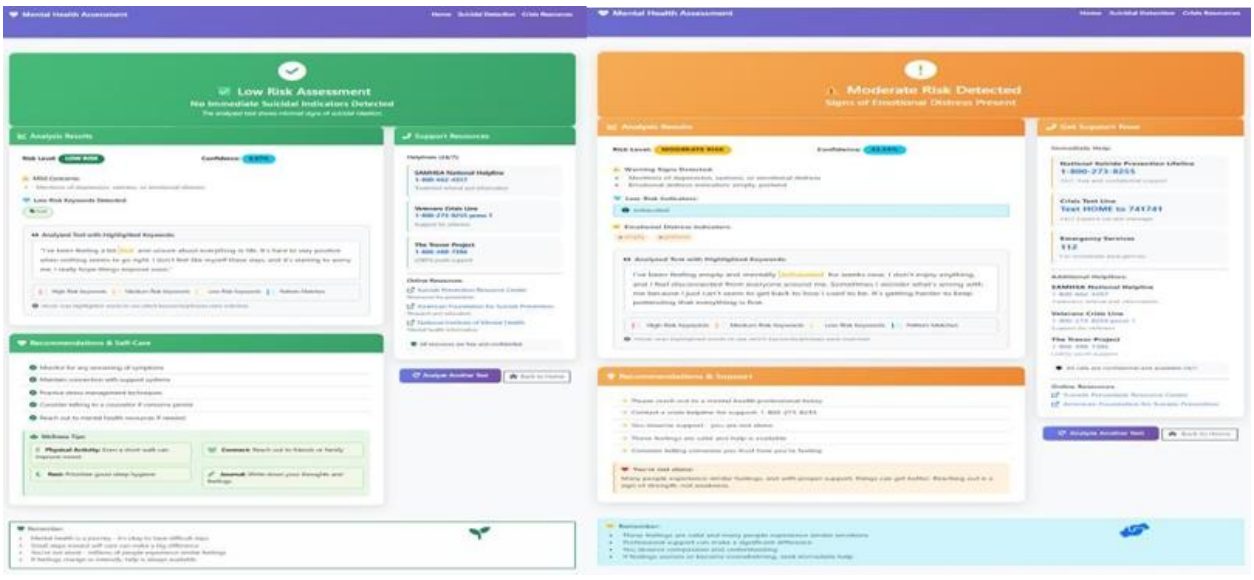

Fig5: Low & Moderate Risk Alert Page

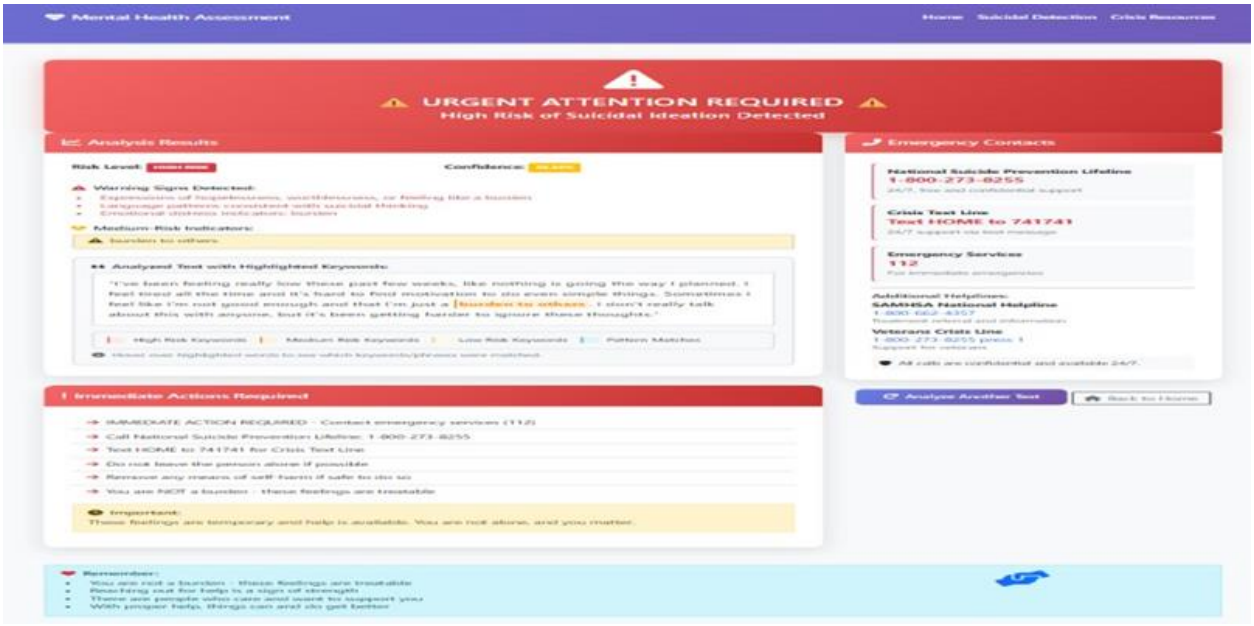

Fig 6: High Risk Alert Page

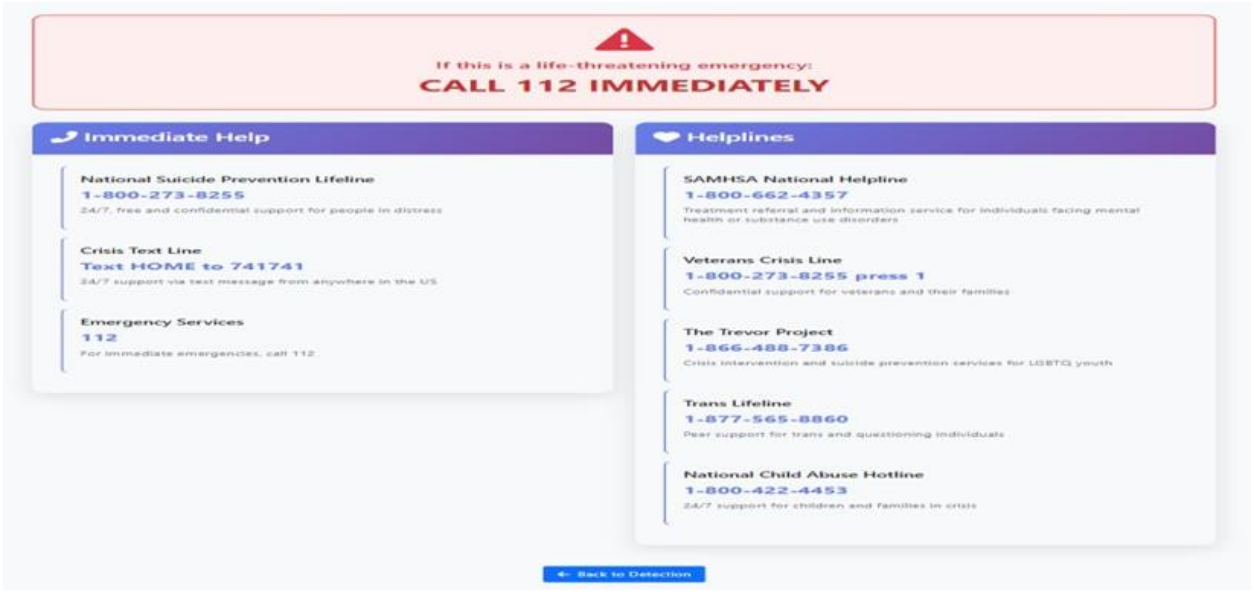

Fig7: Crisis Resources Page

Fig 8: About Us Page

Fig 9: Contact Us Page

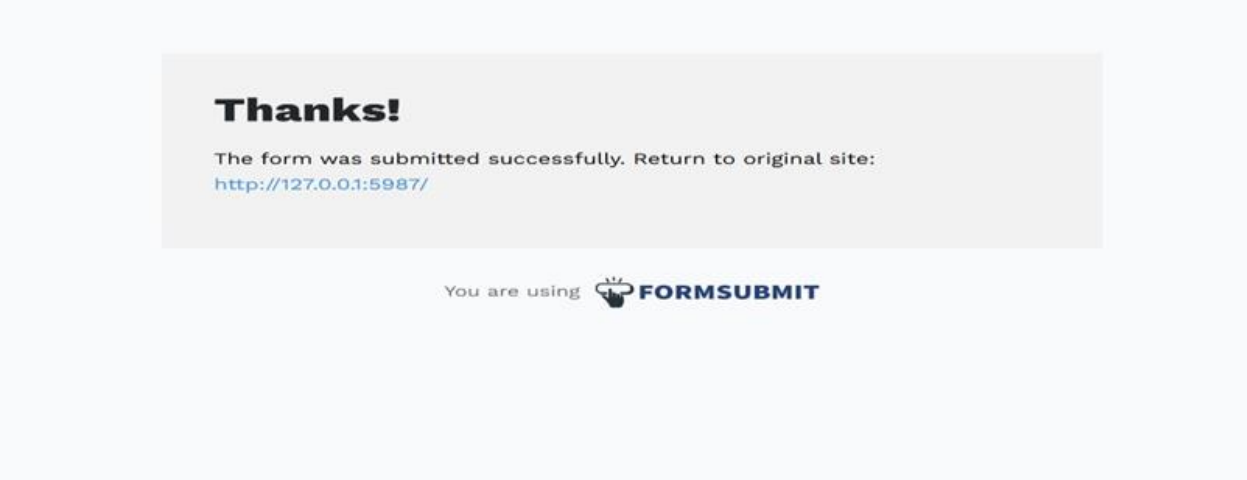

Fig 10: Form Submission Page

The proposed system effectively detects suicidal ideation and classifies user inputs into different risk levels such as low, moderate, and high. It provides accurate analysis with meaningful insights, enabling early identification of mental health concerns. Overall, the system supports timely intervention and enhances mental health monitoring through an efficient and user-friendly approach

6. CONCLUSION

This paper presented a pure machine learning and natural language processing based approach for suicidal ideation detection from social media. The proposed system combines standard NLP preprocessing, TF-IDF feature extraction, and supervised text classification. Experimental comparison shows that SVM outperforms other baseline models and can be integrated into a practical web-based screening interface. Future work can include multilingual expansion, transformer-based transfer learning for improved context understanding, active learning for low-resource labels, and fairness auditing across demographic and linguistic groups.

7. REFERENCES

[1] S. Ji *et al.*, "Suicide Ideation Detection Using Machine Learning Techniques," 2021.

[2] Orabi *et al.*, "Early Detection of Depression from Social Media Using Machine Learning," 2021.

[3] Y. Sawhney *et al.*, "Detection of Suicidal Ideation in Social Media Using Deep Learning," 2022.

[4] K. Saha *et al.*, "Depression Detection Using Natural Language Processing and Machine Learning," 2023.

[5] J. Kumar *et al.*, "Machine Learning Approaches for Suicide Risk Prediction from Text Data," 2024.

[6] M. De Choudhury et al., "Predicting Depression via Social Media," ICWSM, 2013.

[7] G. Coppersmith et al., "CLPsych 2015 Shared Task: Depression and PTSD on Twitter," ACL Workshop, 2015.

[8] Benton et al., "Multitask Learning for Mental Health Conditions," ACL, 2017.

[9] E. Shing et al., "Expert, Crowdsourced, and Machine Assessment of Suicide Risk via Online Posts," CLPsych, 2018.

[10] A. Turcan & K. Mc Keown., "Detecting Depression and Suicide Risk from Reddit," ACL, 2019.

[11] S. Ji et al., "A Survey on Suicide Detection via Social Media," IEEE Access, 2021.

[12] M. Matero et al., "Suicide Risk Assessment with Contextual Neural Networks," CLPsych, 2019.

[13] R. Sawhney et al., "Contextualized Suicide Ideation Detection on Social Media," EMNLP, 2020.

[14] World Health Organization., "Suicide Worldwide in 2019 – Global Health Estimates," 2021.

[15] American Psychiatric Association., "Diagnostic and Statistical Manual of Mental Disorders(DSM-5),"2013.