International Organization of Research & Development (IORD)
ISSN: 2348-0831
Vol 13 Special Issue 02 | 2026

Early Detection of Psychological Distress from Social Media Using AI-Driven Text Analytics

Mohammad Faizan[1], Khabbab Ullah Khan[2], Yogesh Katole[3]
[1,2]*Student, CSE, Siddhivinayak Technical Campus, Maharashtra, India*
[3]*Guide, CSE, Siddhivinayak Technical Campus, Maharashtra, India*

DOI: 10.5281/zenodo.19539095

ABSTRACT

Psychological distress, including depression, anxiety, and emotional instability, is a significant global health concern. Early detection plays a crucial role in preventing severe mental health conditions and enabling timely intervention. This research proposes an Artificial Intelligence-driven framework for detecting psychological distress using social media text analytics. Natural Language Processing techniques are used for preprocessing, feature extraction, and sentiment analysis, followed by machine learning classification. The system identifies linguistic and emotional indicators such as negative sentiment, social withdrawal, and repetitive negative expressions. Experimental results demonstrate high accuracy using Support Vector Machine classification. The proposed framework provides an automated and scalable solution for early mental health monitoring.

Keywords:- Psychological Distress, Machine Learning, NLP, Text Analytics, Sentiment Analysis, Mental Health

1. INTRODUCTION

Mental health disorders are increasing at an alarming rate across the world and are significantly impacting individuals' emotional stability, decision-making ability, productivity, and overall quality of life. Psychological distress, which includes symptoms such as persistent sadness, anxiety, hopelessness, and social withdrawal, often develops gradually and remains unnoticed in its early stages. One of the major challenges in mental healthcare is the lack of continuous and non-intrusive monitoring systems capable of identifying early warning signs before the condition becomes severe. As a result, many individuals do not receive timely support or intervention, leading to serious psychological and social consequences.

In recent years, social media platforms have become an integral part of daily life, where users frequently express their emotions, thoughts, and personal experiences through textual posts, comments, and interactions. These digital expressions serve as valuable behavioral indicators that can reflect a person's emotional and psychological condition over time. The large-scale availability of such user-generated textual data provides a unique opportunity for computational analysis of mental health patterns. Advanced Artificial Intelligence and Natural Language Processing techniques enable automated extraction of linguistic, emotional, and behavioral features from textual data, allowing early identification of psychological risk factors.

This research aims to develop an intelligent and data-driven system capable of detecting early signs of psychological distress using social media text analytics and machine learning techniques. The proposed approach focuses on analyzing emotional tone, sentiment patterns, and linguistic behavior to identify distress signals in a scalable, efficient, and non-intrusive manner. Such a system can assist researchers, healthcare professionals, and digital mental health platforms in early screening, continuous monitoring, and preventive mental healthcare, ultimately contributing to improved psychological well-being and timely intervention.

2. PROPOSED METHODOLOGY

The proposed framework aims to automatically detect early signs of psychological distress from user-generated social media text by leveraging Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques. The system is designed as a multi-stage pipeline consisting of data acquisition, preprocessing, feature extraction, model training, and classification. Initially, textual data is collected from publicly available social media platforms where users frequently express emotions, opinions, and psychological conditions through posts and comments. The collected dataset is examined for inconsistencies, duplicate entries, and missing values to ensure reliability and data integrity. Basic exploratory analysis is performed to understand textual distribution and class balance.

In the preprocessing stage, raw textual data is cleaned and normalized to improve quality and reduce noise. This

process includes tokenization, lower-case conversion, removal of stop-words, punctuation elimination, and filtering of URLs, emojis, and special characters. Stemming or lemmatization is applied to reduce words to their root form, improving semantic consistency and reducing dimensionality. These preprocessing steps enhance the efficiency and predictive capability of machine learning models.

Following preprocessing, the system performs feature extraction to transform textual data into meaningful numerical representations. TF-IDF (Term Frequency–Inverse Document Frequency) vectorization is employed to assign importance weights to words based on their occurrence across the corpus. In addition, sentiment analysis is applied to capture emotional polarity, while emotional keyword identification detects distress-related terms such as sadness, loneliness, anxiety, and hopelessness. Linguistic pattern analysis is also performed to identify behavioral indicators such as repetitive expressions, negative tone, and reduced social engagement signals. These combined features provide a comprehensive representation of psychological distress patterns present in textual data.

The extracted feature vectors are then used to train multiple supervised machine learning models, including Logistic Regression, Random Forest, and Support Vector Machine (SVM). The dataset is divided into training and testing subsets (typically 80:20) to evaluate model generalization on unseen data. Hyperparameter tuning and cross- validation techniques are applied to optimize model performance and prevent overfitting. Among the evaluated models, the Support Vector Machine demonstrated superior performance due to its effectiveness in handling high- dimensional sparse text data and its strong classification capability.

Finally, the trained system classifies incoming textual input into two categories: **Psychological Distress Detected** or **No Distress Detected**. The proposed framework provides an automated, scalable, and non-intrusive approach for early mental health monitoring and risk assessment. This system can support healthcare professionals and digital mental health platforms by enabling continuous and real-time psychological state analysis, with potential future enhancements through deep learning and multimodal data integration.

Stage	Component	Techniques / Methods	Objective
1	Data Acquisition	Social media text collection	Gather user-generated psychological and emotional text data
2	Data Preprocessing	Tokenization, Stop-word Removal, Lowercasing, Stemming, Noise Filtering	Clean and normalize raw textual data
3	Feature Extraction	TF-IDF, Sentiment Analysis, Emotional Keyword Detection, Linguistic Pattern Analysis	Convert text into meaningful numerical and psychological features
4	Model Training	Logistic Regression, Random Forest, Support Vector Machine	Train classifiers to identify distress patterns
5	Model Evaluation	Accuracy, Precision, Recall, F1-Score	Evaluate classification performance and reliability
6	Prediction Module	Distress / Non-Distress Classification	Detect early psychological distress
7	System Outcome	Automated Mental Health Monitoring	Provide scalable and non-intrusive distress detection

Table 3.1: Proposed Framework Components

3. ALGORITHM

3.1 Dataset Acquisition and Loading:

The textual dataset related to mental health discussions is collected from reliable social media platforms or publicly available repositories. The dataset is imported into the working environment and inspected for inconsistencies, duplicate records, and missing values. Basic exploratory data analysis is performed to understand class distribution and textual characteristics before further processing.

3.2 Text Preprocessing and Cleaning:

The raw textual data undergoes systematic preprocessing to enhance data quality and reduce noise. This step includes tokenization, lower-case normalization, removal of stop-words, punctuation, special symbols, hyperlinks, and irrelevant characters. Stemming or lemmatization techniques are applied to convert words into their root forms, ensuring dimensionality reduction and semantic consistency. This stage improves model efficiency and generalization capability.

3.3 Feature Extraction using TF-IDF:

The cleaned text corpus is transformed into numerical feature representations using Term Frequency–Inverse Document Frequency (TF-IDF) vectorization. TF-IDF assigns weights to terms based on their frequency within individual documents and across the entire corpus, thereby capturing discriminative keywords that contribute significantly to distress detection. This step converts unstructured textual data into structured feature vectors

suitable for machine learning algorithms.

3.4 Dataset Splitting:

The processed dataset is partitioned into training and testing subsets, typically using an 80:20 ratio. The training set is utilized to learn underlying patterns and classification boundaries, while the testing set is reserved for evaluating the model's predictive performance on unseen data. Stratified sampling may be applied to maintain balanced class distribution.

3.5 Model Training using Support Vector Machine (SVM):

A Support Vector Machine classifier is initialized and trained on the TF-IDF feature vectors of the training dataset. Kernel functions and hyperparameters such as regularization parameter (C) are optimized to enhance classification performance and prevent overfitting. SVM is selected due to its effectiveness in handling high-dimensional sparse text data and its robustness in binary classification tasks.

3.6 Distress Label Prediction:

After training, the optimized SVM model is used to predict distress labels for the testing dataset. The classifier assigns each text instance to predefined categories such as "Distressed" or "Non-Distressed" based on learned decision boundaries.

3.7 Performance Evaluation and Validation:

The effectiveness of the model is evaluated using standard performance metrics including Accuracy, Precision, Recall, and F1-score. These metrics provide comprehensive insight into classification reliability, error distribution, and overall system robustness. Additional analysis such as confusion matrix evaluation may be performed to examine false positives and false negatives.

4. RESULT AND DISCUSSION

To evaluate the effectiveness of the proposed mental health distress detection system, multiple machine learning classifiers were implemented and compared, including Logistic Regression, Random Forest, and Support Vector Machine (SVM). The models were trained using TF-IDF features extracted from preprocessed textual data and tested on unseen samples to measure generalization capability.

The experimental results indicate that the Support Vector Machine (SVM) classifier achieved the highest performance among all tested models. While Logistic Regression provided stable baseline results and Random Forest improved performance through ensemble learning, SVM demonstrated superior classification capability due to its effectiveness in handling high-dimensional sparse text data.

Model	Accuracy	Precision	Recall	F1-Score	Observations
Logistic Regression	85%	84%	83%	83.5%	Good baseline performance; may struggle with complex emotional patterns
Random Forest	88%	87%	86%	86.5%	Handles non-linearity better; reduced bias
Support Vector Machine	92%	91%	90%	90.5%	Best performance; effective in high-dimensional feature space

5. CONCLUSION

This research presented an AI-driven framework for the early detection of psychological distress using social media text analytics. The proposed system leverages natural language processing and machine learning techniques to analyze linguistic patterns, emotional tone, and behavioral indicators embedded within textual data. Experimental results demonstrate that the framework achieves high classification accuracy and effectively identifies signs of emotional distress such as negative sentiment, hopelessness, and social withdrawal. Furthermore, the system offers a scalable, automated, and cost-effective solution for continuous mental health monitoring, enabling timely identification and potential intervention. The proposed approach can assist healthcare professionals and support systems in proactive mental health assessment and may serve as a foundation for future intelligent mental health analytics platforms.

5. ACKNOWLEDGEMENT

The authors thank Prof. Yogesh Katole and the Department of Computer Science & Engineering, Siddhivinayak Technical Campus, for their valuable support and guidance.

6. REFERENCES

[1] World Health Organization, *Mental Health Atlas 2020*, Geneva: WHO, 2021.
[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space,"
[3] *Proc. ICLR*, 2013.
[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[5] S. Ji, C. Pan, X. Cambria, P. Marttinen, and S. Yu, "A Survey on Mental Health Detection Using Social Media Text," *IEEE Trans. Computational Social Systems*, vol. 8, no. 4, pp. 929–946, 2021.
[6] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2012.
[7] J. C. Eichstaedt et al., "Facebook Language Predicts Depression in Medical Records," *PNAS*, vol. 115, no. 44,
[8] pp. 11203–11208, 2018.
[9] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting Depression via Social Media," *Proc. ICWSM*, 2013.
[10] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying Mental Health Signals in Twitter," *Proc. CLPsych Workshop, ACL*, 2014.
[11] Benton, M. Mitchell, and D. Hovy, "Multi-Task Learning for Mental Health Using Social Media Text," *ACL Conference*, 2017.
[12] H. A. Schwartz et al., "Personality, Gender, and Age in the Language of Social Media," *PLOS ONE*, vol. 8, no. 9, 2013.
[13] World Health Organization, *Depression and Other Common Mental Disorders: Global Health Estimates*, WHO Press, 2021.
[14] T. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *JMLR*, vol. 12, pp. 2825–2830, 2011.
[15] K. Resnik et al., "Beyond LDA: Exploring Supervised Topic Modeling for Depression-Related Language in
[16] Twitter," *CLPsych Workshop, ACL*, 2015.