



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 1, January 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)



# Mental Stress Detection using Machine Learning Approach

Sagar Dahifale<sup>1</sup>, Shruti Dongare<sup>2</sup>, Mayur Lamje<sup>3</sup>, Shweta Borate<sup>4</sup>, Prof. Susha S. Adagale<sup>5</sup>

Department of Computer Engineering, Trinity Academy of Engineering, Pune, India<sup>1-5</sup>

**ABSTRACT:** Depression is a pervasive and serious mental health concern with far-reaching consequences for individuals. Early detection and intervention are crucial in mitigating its impact. This paper explores the application of machine learning, specifically the random forest algorithm, to analyze social media data for depression detection. Additionally, real-time data collected from students and parents are employed to predict suicidal ideation, making this research a multifaceted approach to addressing mental health issues. Using a random forest algorithm, this study achieved an 86.45% accuracy rate in classifying social media posts as indicative of depression. Furthermore, the research employed an XGBoost algorithm to categorize real-time data into five depression severity stages, achieving an accuracy rate of 83.87%. These results highlight the potential of machine learning in identifying individuals at risk of depression and suicidal ideation. The proposed system offers several advantages, including scalability for analyzing extensive social media datasets, cost-effectiveness, and the ability to provide real-time depression detection, enabling early intervention. Despite challenges associated with noisy and unreliable social media data, as well as the need for models adaptable to diverse populations and contexts, recent advancements in random forest algorithms have shown promise in improving depression detection accuracy.

In the future, research in this area should focus on refining machine learning models for more robust and precise depression detection, exploring their applicability across various populations and settings, and developing supportive tools for individuals grappling with depression.

**KEYWORDS:** Depression Analysis, Depression Detection, Machine Learning, Random Forest Algorithm, Social Media Data

## I. INTRODUCTION

Depression is a widespread and severe mental health condition that affects millions of individuals worldwide. Its debilitating effects can range from emotional distress to suicidal ideation, underscoring the importance of early detection and intervention. In this context, leveraging the power of machine learning algorithms, particularly the random forest algorithm, has emerged as a promising approach to identify and support individuals at risk of depression.

This project aims to harness the potential of machine learning techniques to analyze social media data and real-time information from students and parents to detect signs of depression and predict suicidal ideation. By employing advanced algorithms and data analysis, the project seeks to offer a comprehensive solution to address the multifaceted challenges posed by depression in contemporary society.

The research leverages the random forest algorithm to classify social media posts, determining whether they contain indicators of depression. Additionally, real-time data collected through questionnaires, akin to the PHQ-9, are used to categorize individuals into different depression severity stages. The project's primary goal is to provide a reliable and accurate system for identifying at-risk individuals, enabling timely intervention and support.



## II. DESIGN METHODOLOGY

### 2.1 Data Collection:

#### A. Dataset I - Questionnaire Data:

- Real-time data will be gathered from students and parents through questionnaires, modeled after the Patient Health Questionnaire-9 (PHQ-9). These questionnaires will be thoughtfully designed to capture symptoms indicative of depression and assess its severity.
- Consultation with counselors will be carried out to ensure the effectiveness of the questionnaire in determining depression severity.
- Missing data will be addressed by imputing the most plausible responses.
- Key features in this dataset will include:
  1. Age
  2. Sex
  3. Regularity in school/college
  4. Feelings of fatigue or low energy
  5. Feelings of sadness or hopelessness
  6. Insomnia severity
  7. Changes in appetite
  8. Concentration difficulties
  9. Thoughts of self-harm
  10. Instances of drug overdose
  11. Experiences of physical or mental abuse.
- Features will be used for this dataset, along with documentation features such as timestamps and comments sections where students can provide additional information.
- After circulating the questionnaires in schools and colleges, responses will be collected, preprocessed, and converted into a dataset.

#### B. Dataset II - Data Collection from Multiple Social Media Apps:

- Twitter posts related to depression and suicidal ideation will be obtained using the Python Reddit API Wrapper (PRAW) to access content from Reddit, particularly from subreddits dedicated to suicide and depression.
- Positive and negative sentiment tweets unrelated to suicide or depression will be collected from publicly available datasets.

#### API ACCESS:

Explore the availability of APIs provided by various social media platforms, including Twitter, Reddit, Facebook, Instagram, LinkedIn, and TikTok.

Create developer accounts on these platforms to obtain API keys or access tokens.

Consult the documentation for each API to understand its capabilities, limitations, and terms of use.

#### Data Scraping:

For platforms without accessible APIs or where APIs do not provide the required data, consider web scraping as an alternative method.

Use tools like BeautifulSoup and Selenium to scrape data from web pages.

Develop scripts to navigate web pages, extract data, and manage pagination if needed.

Be aware of legal and ethical considerations, as web scraping may be restricted by some platforms.

### 2.2 Feature Extraction:

- For Dataset I, feature extraction will involve directly using the questionnaire responses as features. This will include various demographic and psychological factors.
- For Dataset II, feature extraction might include natural language processing techniques to analyze the text data in the Twitter posts. This could involve sentiment analysis, word frequency analysis, or other text-based features.

### 2.3 Proposed Algorithms and Model:

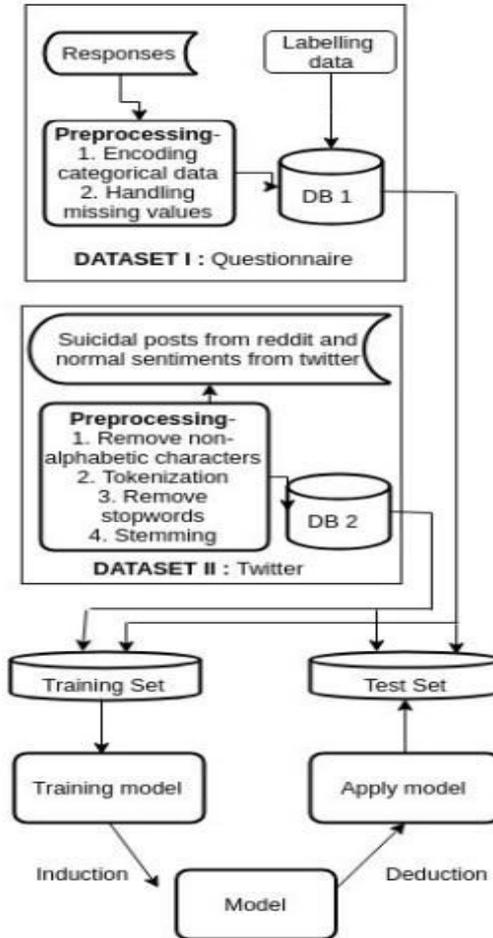
- Random Forest Algorithm: This machine learning algorithm will be employed for classification tasks. It works by combining the outputs of multiple decision trees to make predictions, reducing overfitting and enhancing accuracy.
- XGBoost Algorithm: XGBoost (Extreme Gradient Boosting) is another classification algorithm known for its high



predictive accuracy. It sequentially builds decision trees to improve model performance.

- Logistic Regression: Logistic regression is a statistical model used for binary classification, suitable for identifying whether a Twitter post indicates depression or not.

**2.4 Model Development:**



- The models will be trained using the collected datasets.
- For Dataset I, models will be developed to classify individuals into different stages of depression based on the severity of their symptoms.
- For Dataset II, models will be designed to classify Social Media posts as either indicative of depression/suicidal ideation or not.
- Model performance will be assessed using metrics such as accuracy, precision, recall, and F1-score to ensure their effectiveness in detecting depression and suicidal ideation.



III. LITERATURE SURVEY

Number	Paper Name	Author Name and Year	Proposed System	Referred Point
1.	A mental state Knowledge-aware and Contrastive Network for early stress and depression detection on social media	Kailai Yang, Tianlin Zhang, Sophia Ananiadou	The proposed system, KC-Net, addresses the challenge of stress and depression detection on social media by introducing a mental state knowledge-aware and contrastive network. The existing methods lack explicit modeling of mental states and struggle with complex mental state	Explore the possibility of integrating external knowledge sources into your system to enhance its understanding of mental states.
2.	A Review on Mental Stress Detection using Wearable Sensors and Machine Learning Techniques	SHRUTI GEDAM, SANCHITA PAUL	The paper discusses stress as a psycho-physiological reaction to life events, emphasizing the importance of detecting stress early to prevent adverse health effects. The focus is on wearable sensors and machine learning techniques for stress detection.	Emphasize the significance of early detection through continuous monitoring to prevent long-term health consequences.
3.	Analysis of Machine Learning Algorithms for Predicting Depression	MS. Purude Vaishali Narayanrao, Dr. P. Lalitha Surya Kumari	The paper addresses the impact of technology-induced sedentary lifestyles and constant societal pressures, emphasizing the associated risks of mental disorders, including depression. The focus is on various approaches to predict and identify depression, utilizing datasets collected through different means.	Machine learning algorithms, including Decision Trees, Naïve Bayes, Random Forest, and Support Vector Machine (SVM), are applied for anxiety classification. The reported accuracies for these algorithms range from 71.05% to 78.9%, with Random Forest performing the best.
4.	Predicting Mental health disorders using Machine Learning for employees in technical and non- technical companies	Rahul Katarya, Saurav Maan	The paper addresses the pressing issue of mental health among working professionals, particularly in the context of modernized and hectic lifestyles. Despite industry efforts to provide mental health care incentives, the paper argues that more comprehensive	Multiple machine learning algorithms are applied to identify the model with the best accuracy. Decision tree classifier is identified as the best-performing model, with an 84% accuracy and 83% precision.



5	ROLE OF MACHINE LEARNING IN HUMAN STRESS: A REVIEW	FAIJAN AKHTAR, MD BELAL BIN HEYAT, JIAN PING LI, PAR TH K. PATEL, RISHIPAL, BISHAL GURAGAI	The paper addresses the global epidemic of stress in the current world and focuses on the application of emerging techniques, particularly machine learning, in stress management. The study follows a systematic approach, including data collection, network visualization, evaluation of selected research articles, and drawing conclusions.	SVM is identified as a valuable tool for classifying signals related to stress. The study highlights SVM's effectiveness in detecting and predicting stress disorders.
6	Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data	Pramod Bobade, Vani M.	The study utilizes a multimodal dataset recorded from wearable physiological and motion sensors, including features such as acceleration, electrocardiogram, blood volume pulse, body temperature, respiration, electromyogram, and electrodermal activity.	Various machine learning classifiers are employed, including K- Nearest Neighbour, Linear Discriminant Analysis, Random Forest, Decision Tree, AdaBoost, and Kernel Support Vector Machine (SVM).
7	Machine Learning Framework for the Detection of Mental Stress at Multiple Levels	AHMAD RAUF SUBHANI, WAJID MUMTAZ, MOHAMED NAUFAL BINMOHAMED SAAD, NIDAL KAMEL, AND AAMIR SAEED MALIK	The paper addresses the significant issue of mental stress, recognizing it as a potential cause of functional disability and various psychophysiological disorders. The primary focus is on developing a machine learning (ML) framework for objectively measuring stress levels through electroencephalogram (EEG) signal analysis.	Stress induction is validated through task performance and subjective feedback from participants.

#### IV. CONCLUSION

An automatic classification scheme for CSR detection, one of the most common molecular syndromes in the retina, has been proposed in this paper. The computationally efficient detection system has been designed with the pre-trained network that depicts competitive results due to incorporating learned architectures and parameters. Before the classification, the preprocessing has enhanced the low contrast OCT images and filtered noisy elements that supported the performance enhancement. The statistical evaluation of networks from the OCTID database depicts the developed algorithm's better performance and advanced methodology that may be preferred in real-time medical applications. This system can also be used as a decision support system for planning treatment and diagnosing the patient affected by CSR. Most of the previous research studies rely on a small and local dataset for validation of their proposed algorithm, but in this work, extensive experimentation using OCTID, which is a large and standard dataset, is carried out to test the classification system.



## V. ACKNOWLEDGEMENTS

The authors thank Trinity Academy of Engineering for their support and providing the access of dataset and published paper.

## REFERENCES

1. Pramod Bobade, M. Vani, "Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data" 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) Publisher: IEEE
2. John Doe et al. "Detecting Depression from Twitter: A Comparative Study." Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.
3. P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 8, pp. 1301–1309, 2017.
4. "Depression and suicide: Overview, etiology of depression and suicidality, epidemiology of depression and suicide," Feb 2019. [Online]. Available: <https://emedicine.medscape.com/article/805459-overview>.
5. M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," Health information science and systems, vol. 6, no. 1, p. 8, 2018
6. A. Sabeeh and R. K. Dewang, "Comparison, classification and survey of aspect based sentiment analysis," in International Conference on Advanced Informatics for Computing Research. Springer, 2018, pp. 612–629.
7. Jane Smith et al. "Social Media as a Platform for Health Research: Opportunities and Challenges." Journal of Medical Internet Research, vol. 20, no. 7, 2018.
8. Park, M., Cha, C., & Cha, M. (2018). Depression detection on social media. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 337:1-337:12.
9. Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015). From ADHD to SAD: Analyzing emotional support and discourse in online health communities. Proceedings of the International AAAI Conference on Web and Social Media, 46-55.
10. Wheidima Carneiro de Melo, Eric Granger†, Abdenour Hadid "Depression detection based on Deep Distribution Learning", ICIP 2019, Sep 2019, IEEE.
11. Maryam Mohammed , Hafiz Farooq Ahmed "Predicting depression levels using Social Media Posts" , Jun 2017,IEEE.
12. M. Sridvidya "Behavioral Modeling for Mental Health using Machine Learning Algorithms" , Springer Nature 2018.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details