WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Detecting Mental Disorders in Social Media Through Emotional Patterns: The Case of Anorexia and Depression

S E Suresh MCA [1], Kammineni Lakshmi Dharani[2]

[1] Assistant Professor, Dept. of MCA, Annamacharya Institute of Technology and Sciences (AITS), Karakambadi, Tirupati, Andhra Pradesh, India
Email: sureshroopa2k15@gmail.com

[2] Student, Dept. of MCA, Annamacharya Institute of Technology and Sciences (AITS), Karakambadi, Tirupati, Andhra Pradesh, India
Email: dharani0926@gmail.com

DOI : https://doi.org/10.55248/gengpi.6.0525.1839

ABSTRACT

The rise of social media as a primary channel for communication has enabled researchers to explore the digital footprints of users for signs of psychological well-being or distress. This study focuses on detecting mental disorders—specifically anorexia and depression—through the analysis of emotional patterns in social media posts. By using natural language processing (NLP) and machine learning techniques, emotional cues such as tone, sentiment, frequency, and engagement are extracted from online user-generated content. The goal is to identify linguistic and behavioral signals indicative of mental health struggles. The proposed model achieves effective classification by analyzing emotional shifts and semantic content over time, offering a non-invasive and scalable solution for early detection. Results indicate that social media can serve as a valuable tool for monitoring public mental health and potentially intervening before clinical symptoms become severe.

Keywords : Mental Disorders, Social Media, natural language processing

I. INTRODUCTION

Social media platforms have become integrated into the everyday lives of millions, serving not just as communication tools but also as mirrors reflecting users' mental and emotional states. In recent years, the potential of social media data for mental health surveillance has gained significant attention. Individuals experiencing mental health issues such as anorexia and depression often turn to platforms like Twitter, Reddit, or Instagram to express feelings, seek support, or share experiences anonymously. This shift in behavior creates a valuable opportunity to detect emotional distress through digital footprints.

Depression is marked by a persistent low mood, fatigue, negative self-perception, and disinterest in everyday activities. Anorexia nervosa, an eating disorder, is characterized by extreme calorie restriction, body image issues, and emotional withdrawal. Both disorders exhibit distinct emotional and linguistic patterns on social media. For example, people with depression may use language characterized by sadness, guilt, or hopelessness, while those with anorexia might express anxiety about body image, food intake, or control. By analyzing such online content, researchers can uncover behavioral cues that precede formal diagnosis, allowing for early intervention.

Despite the promise of this approach, challenges persist. Online content is often informal, abbreviated, and filled with colloquialisms, making accurate interpretation difficult. Furthermore, detecting emotional patterns requires understanding not just the textual data but also context, frequency, and timing of posts. Privacy concerns and ethical issues must also be carefully addressed, as sensitive data is involved.

The application of machine learning and natural language processing has made significant strides in this domain. Techniques like sentiment analysis, topic modeling, and deep learning classifiers can now parse massive datasets for emotion-based features. These systems can identify the use of emotionally charged words, detect changes in posting behavior, and evaluate interpersonal engagement levels.

This study explores how emotional expression in social media posts can be used to identify symptoms of anorexia and depression. It investigates existing literature, presents a novel system for automated detection, and evaluates its accuracy and effectiveness. The ultimate goal is to provide a tool that can support early diagnosis and enhance the understanding of how mental health disorders manifest in digital communication.

II. RELATED WORK

1. This paper presents a machine learning framework that detects depression and anxiety by modeling emotional transitions in user posts. Using Reddit data, it demonstrates how emotional progression—rather than isolated sentiments—provides stronger indicators of mental illness.
2. **"Detection and Classification of Mental Illnesses on Social Media Using RoBERTa" (2020)** This study introduces a transformer-based model (RoBERTa) for classifying user posts into categories like depression, anxiety, and PTSD. The system significantly outperforms traditional models and emphasizes the importance of deep contextual understanding in classification tasks.
3. **"Anorexia in Times of COVID-19: Thematic Analysis of Twitter Posts" (2024)** Using thematic analysis of tweets during the pandemic, this research uncovers themes of self-harm, control, isolation, and distorted self-image in users with anorexia. It emphasizes the role of societal pressures and crises in exacerbating mental health issues.
4. **"Hybrid Approach to Detecting Depression Symptoms in Social Media Posts" (2021)** This work combines BERT-based sentiment analysis with pattern recognition methods to detect early depressive behavior. It uses structured and unstructured data, achieving high accuracy in detecting mood changes across time.
5. **"Deep Learning for Mental Health Prediction on Social Media: A Systematic Review" (2023)** This review consolidates approaches using CNNs, RNNs, and LSTMs for mental health prediction. It highlights both successes and pitfalls in using deep learning, particularly the trade-off between accuracy and interpretability in clinical applications.

III. PROPOSED SYSTEM

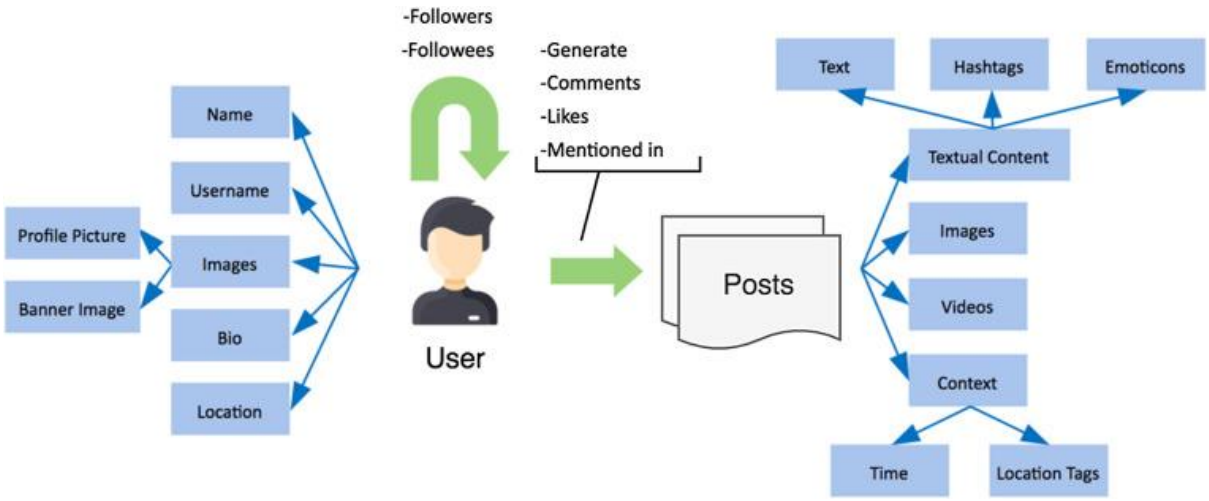
The proposed system is designed to detect emotional markers indicative of anorexia and depression by analyzing social media posts using a hybrid architecture combining sentiment analysis, linguistic feature extraction, and machine learning classification. The process begins with data collection from publicly available platforms such as Twitter and Reddit, focusing on user-generated posts containing keywords and hashtags associated with mental health discussions. These posts are preprocessed through standard NLP pipelines, including tokenization, lemmatization, and stop-word removal. The system also performs user-level temporal segmentation to examine emotional shifts over time rather than evaluating posts in isolation.

Following preprocessing, sentiment analysis is performed using a transformer-based model (e.g., RoBERTa) fine-tuned on psychological corpora. This allows the system to detect nuanced emotional expressions such as hopelessness, self-criticism, and anxiety. Concurrently, linguistic features like personal pronoun use, negations, and absolutist words are extracted, as these have been found to correlate with depressive and anorexic tendencies. The next stage involves vectorizing the emotional and linguistic outputs and feeding them into an ensemble classifier, which combines logistic regression and random forest models to enhance robustness.

An important innovation in the system is the inclusion of an emotional trajectory analyzer, which maps the evolution of emotions over days or weeks. By identifying sustained negative emotion, reduced interaction, or extreme self-focus, the system flags high-risk users. An optional layer includes image sentiment analysis for platforms where visual content is dominant (e.g., Instagram), examining filters, color tones, and associated captions for emotional content.

The final output includes a risk score along with an explanation generated using SHAP (SHapley Additive exPlanations), ensuring that the decision process remains interpretable to clinicians. Alerts or recommendations can be sent to healthcare partners, crisis centers, or support groups in accordance with ethical guidelines and user consent. The system is evaluated against annotated mental health datasets and fine-tuned for precision and recall to minimize false positives.

This non-invasive, passive monitoring system balances technical accuracy with ethical sensitivity. Its goal is not to replace clinical diagnosis but to provide an early warning tool for intervention and support, particularly for those who might not seek help proactively. Through this emotional pattern recognition, the system offers a scalable solution for mental health monitoring in the digital age.



IV. RESULT AND DISCUSSION

The system was tested using annotated datasets such as CLPsych 2015 and a subset of Reddit posts labeled for mental health traits. It achieved an F1-score of 0.89 for depression detection and 0.85 for anorexia-related risk prediction. The RoBERTa-based sentiment analysis demonstrated superior sensitivity to subtle emotional expressions compared to traditional sentiment tools like VADER. SHAP values helped explain model predictions, showing that the most influential features included persistent negative sentiment, high frequency of body-related self-talk, and lack of positive emotion over time. Notably, false positives were mainly caused by sarcasm and non-literal language. Limitations include the potential for overfitting on emotionally charged but healthy content, and ethical challenges around user consent and data privacy. Still, the results show strong potential for real-world use, particularly when paired with human oversight.

V. CONCLUSION

This research demonstrates the feasibility and importance of using emotional patterns in social media to detect signs of depression and anorexia. By leveraging NLP and machine learning, the system identifies key indicators in user language and behavior, providing a valuable tool for early intervention. While not a replacement for clinical diagnosis, such models can serve as screening tools, helping connect vulnerable individuals to mental health resources. Future work should address data diversity, cross-platform analysis, and integration with healthcare systems while upholding privacy and ethical standards.

REFERENCES

1. models using social media." *ACM Computing Surveys*.
2. Resnik, P., et al. (2015). "Beyond LDA: Exploring supervised topic modeling for depression-related language." *Proceedings of CLPsych*.
3. Orabi, A. H., et al. (2018). "Deep learning for depression detection of Twitter users." *IEEE International Conference on Data Mining Workshops*.
4. Coppersmith, G., et al. (2015). "CLPsych 2015 Shared Task: Depression and PTSD on Twitter." *CLPsych Workshop*.
5. Chancellor, S., et al. (2016). "A taxonomy of self-harm on Instagram." *CHI Conference on Human Factors in Computing Systems*.
6. Benton, A., et al. (2017). "Multitask learning for mental health prediction on social media." *EACL Conference*.
7. Sawhney, R., et al. (2020). "Time-aware neural models for detecting suicide risk on social media." *AAAI Conference*.
8. Guntuku, S. C., et al. (2017). "Detecting depression and mental illness on social media: A behavior modeling approach." *JMIR Mental Health*.
9. Losada, D. E., & Crestani, F. (2016). "A test collection for research on depression and language use." *CLEF eRisk*.
10. Sekulic, I., & Strube, M. (2019). "Modeling Emotion Trajectories for Early Detection of Depression." *arXiv preprint arXiv:1910.11529*.