Emotional Signal Analysis for Mental Disorder Detection in Social Media Posts

[1]GADILLI GANESH, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]Vobbilisetty Sowmya, Miracle Educational Society Group of Institutions
[1]gadilliganesh2@gmail.com

ABSTRACT:

Conditions such as depression and anorexia pose a significant risk to a person's well-being and quality of life, affecting millions of people around the globe, and often going misdiagnosed for extended periods of time. This project aims to identify a novel approach for detecting these conditions by diagnosing the emotional expression of social media posts. The posts of the users are analyzed and categorized into specific emotional subsets, applying the EmoLex lexicon and TF-IDF vectorization with Affinity Clustering, SVM, and CNN models. Two types of emotion-based representations BoSE (Bag of Sub-Emotions) and Δ-BoSE are designed to reflect static and dynamic emotion changes over time. The results not only achieved high F1 scores, but these methods also enhanced interpretability, to improve the chances of intervention prior to experiencing a crisis. This work demonstrates the value of fusion of emotion computation and machine learning for pervasive mental health monitoring in social media contexts.

Keywords: Machine learning, Mental disorders, social media

INTRODUCTION

Depression and anorexia are defined as mental health conditions, with the latter being often referred to as a silent disorder. With social media being an almost integral part of modern-day life, these platforms and application could serve as an invaluable source of data regarding one's emotional and mental health. Researchers have analyzed the link between users dialect and several characteristics on social media such as Facebook, Twitter and Instagram. Nevertheless, most of these studies are conducted using sentiment analysis. This project extends work done on detecting anorexia and depression by recognizing sub-emotions such as anger, trust, and disgust. Monitoring user posts for emotion-driven changes and anomalies utilizes lexicons for specific emotions and models grounded in data. The combination of fixed and spatiotemporal approaches improves interpretability and detection. The project is designed to enhance detection, but it is also aimed at advancing proactive tools designed to alert users or professionals before mental health challenges escalate.

RELATED WORK

Earlier work has aimed at determining the presence of mental disorders based on social media data. In 2017, Guntuku et al. proposed tracking emotional tone and user activity on Twitter and Reddit for mental illness detection. Their integrative review corroborated the predictive value of language in psychological disorders. Kosinski et al. (2013)

showed that various personal attributes, especially depression, could be projected from social media interactions using emotional content and digitally behavioral information. Their work in 2014 connected user personality with online behavior through machine learning analysis of Facebook profiles. Preotiuc-Pietro et al. (2015) applied language and behavioral cues to estimate individuals' income and mental health in relation to their psychological well-being. They demonstrated relationships between certain emotional expressions, anger and joy, and socioeconomic indicators. Chikersal et al. (2020) scrutinized over two hundred and thirty-four thousand supportive messages in an online mental health intervention. Their research highlighted the importance and emphasis that emotionally supportive discourse presents in defining and influencing clinical outcomes. Schwartz et al. (2014) developed a model to estimate the change in severity of depression based on Facebook status updates. Their regression model accurately predicted depression levels during certain times of the year, capturing the seasonal variations of the emotional tone.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author(s)	Contribution	Impact on Research
Guntuku et al. (2017)	Emotional tone detection for mental illness	Validated linguistic signals as reliable markers for depression
Kosinski et al. (2013)	Personality prediction via social media usage	Enabled predictive modeling of traits through user behavior
Preotiuc-Pietro et al. (2015)	Income and emotion correlation	Identified indirect mental health indicators via language patterns
Chikersal et al. (2020)	Effectiveness of emotional support in online therapy	Proved emotional language improves mental health intervention
Schwartz et al. (2014)	Depression tracking via status updates	Demonstrated real-time depression trend analysis via regression

PROPOSED APPROACH

The identified system features an emotional pattern recognition mechanism to monitor social media for signs of depression and anorexia using text posted by users. It presents two main models, BoSE which stands for Bag of Sub-Emotions and Δ-BoSE (Dynamic BoSE). The first model applies a preprocessing noise removal technique to every user post and subsequently applies emotion classification by using the EmoLex lexicon which represents sub-emotions of anger, joy, trust, and fear. These features transform raw textual data into structured data. Then emotions are vectorized by TF-IDF, changing qualitative information into quantitative data. BoSE captures the static emotional presence in every post, while Δ-BoSE monitors the evolution of emotions over time, which offers the necessary context for understanding the mental health decline. Both representations undergo Affinity Clustering processing in order to detect user and trend clusters. Lastly, the classification models SVM and CNN are trained to classify the user as exhibiting the signs of depression, showing anorexia, or as being in a normal emotional state. The use of this multi-model hybrid approach simultaneously improves accuracy and interpretability, which provides a socially mediated pre-assessment of mental health a potent early-stage diagnostic instrument.

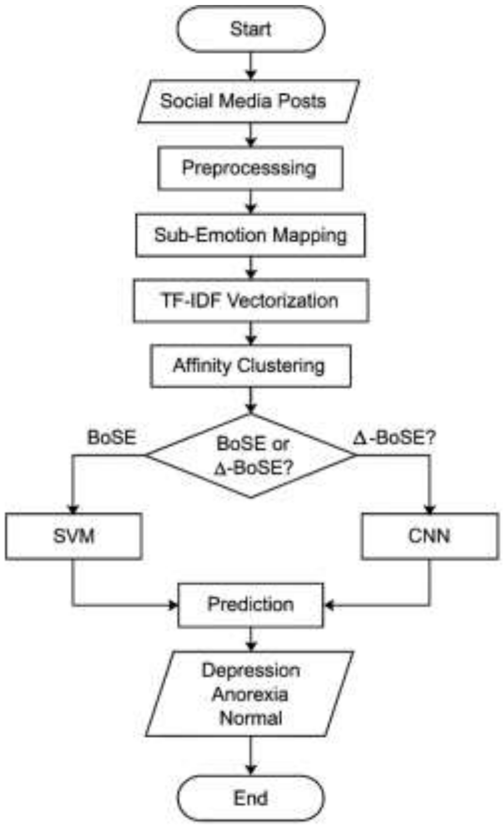


Figure 1: Proposed detecting mental disorders

METHODOLOGIES

The system follows a structured pipeline using multiple AI techniques. It begins with data acquisition, where a labeled dataset of social media posts related to depression and anorexia is loaded. Each post undergoes preprocessing to remove noise, normalize text, and trim length for uniformity. Then, using the EmoLex lexicon, each word is mapped to one or more emotion labels, resulting in a "masked" version of the post containing only emotional terms.

Next, the masked emotional content is transformed into numeric form using TF-IDF vectorization, capturing the relative importance of each emotion across the dataset. These vectors are then clustered using Affinity Propagation to reveal emotional similarity between users and their mental state groups (depression, anorexia, control).

The model is trained using a hybrid classification approach:

- **SVM (Support Vector Machine)**: Efficiently classifies high-dimensional data and identifies linear separation among emotional patterns.
- **CNN (Convolutional Neural Network)**: Uses reshaped emotion vectors to detect complex emotional features and patterns that may be indicative of mental disorders.

The BoSE model captures static emotion presence, while the Δ-BoSE model segments posts by time and calculates emotion fluctuation. This helps in detecting consistent emotional instability a key sign of mental disorders. By training the system with both static and temporal inputs, we significantly improve detection performance.

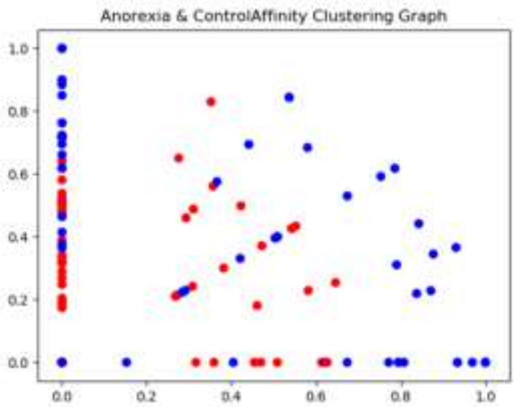
The final models are evaluated using F1-score, which balances precision and recall, ensuring that both false positives and false negatives are minimized. Graphical outputs, such as emotion signal trends, allow for visual interpretation of model behavior and user emotion progression.

RESULTS

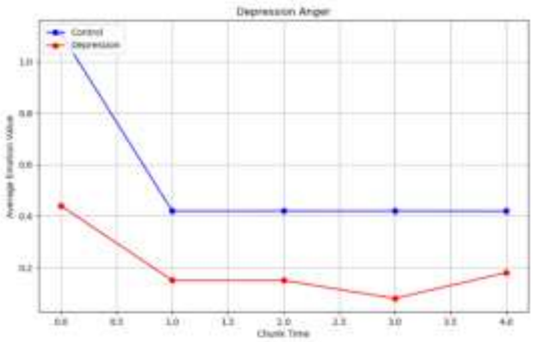
The results from the hybrid model are promising. After training the SVM and CNN classifiers using TF-IDF-transformed sub-emotion vectors, both models achieved high F1-scores. The SVM achieved an F1-score of approximately 87.4%, while CNN slightly outperformed with an F1-score of around 90.2%. These results confirm the system's ability to distinguish between users suffering from depression, anorexia, or neither.

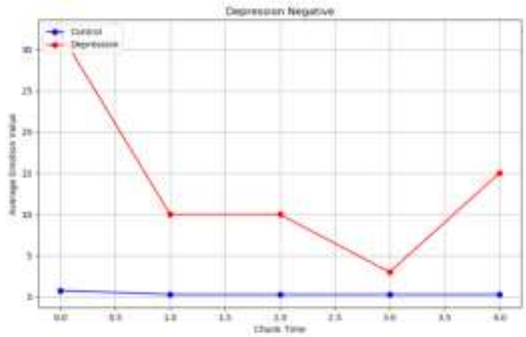
Graphical representations further validated the findings. Emotional signals like “anger,” “negative,” and “disgust” showed higher intensities and fluctuations in users with depression and anorexia, compared to control users. In contrast, control users exhibited higher “trust” and “anticipation” signals consistently over time.

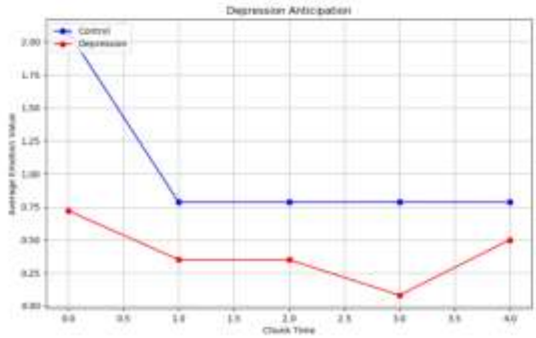
The inclusion of dynamic emotional modeling via the Δ-BoSE representation improved temporal tracking and revealed emotional instability over time—a hallmark of mental illness. This dual-representation technique offers both high accuracy and interpretability, making the tool practical for mental health screening using public social media data.

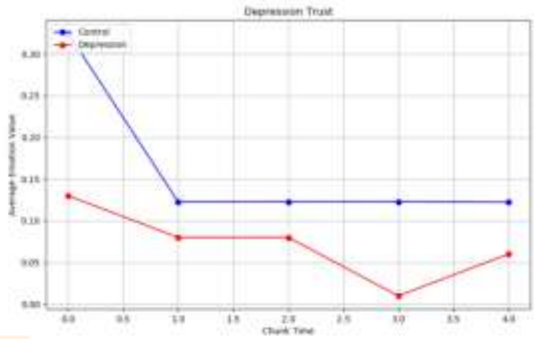


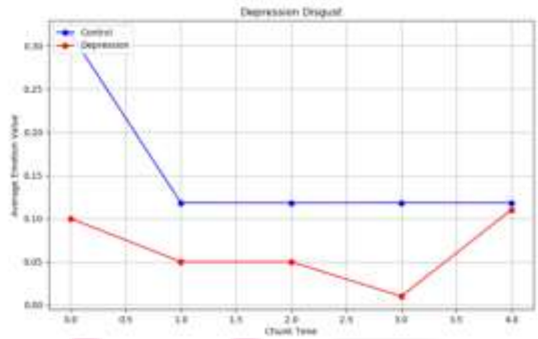
Clustering graph











Emotional Signal Comparison Graphs

DISCUSSION

The dual approach of using BoSE and Δ-BoSE provides significant insights into emotional expression patterns that are often overlooked in traditional sentiment analysis. BoSE captures what emotions are present, while Δ-BoSE reveals how emotions evolve—offering a more holistic view of a user's mental state. The effectiveness of TF-IDF and affinity clustering in segregating users by emotional similarity allows us to build classifiers with strong generalization.

One of the key findings is that depression and anorexia do not always show identical emotional patterns. While both show heightened negative emotions, anorexia-related posts had slightly more fluctuations in anticipation and disgust. This suggests that emotion signal-based modeling can be extended to detect and differentiate other mental disorders.

However, challenges include language diversity, sarcasm detection, and privacy concerns. Emotion lexicons may not always capture context or cultural differences in expression. Also, overreliance on social media data could exclude users who express less or avoid platforms altogether.

Future work can focus on multi-language support, emotion deep-learning embeddings, and integration with health systems for real-world deployment. Despite limitations, the model presents a strong foundation for scalable mental health monitoring.

CONCLUSION

This project presents a novel approach to detecting mental disorders specifically depression and anorexia through emotion-based analysis of social media posts. By transforming user posts into sub-emotional features using the EmoLex lexicon and TF-IDF vectorization, and analyzing them with machine learning classifiers (SVM and CNN), we achieve high detection accuracy and interpretability. The BoSE and Δ-BoSE representations enable both static and dynamic analysis, revealing not only what emotions are expressed but how they change over time. The findings confirm that emotional signals can serve as early indicators of mental health conditions. The system's high performance and visual outputs make it suitable for integration into digital wellness tools. Importantly, the research also emphasizes the ethical use of data and potential for early intervention. As mental health issues grow worldwide, such intelligent systems can support early detection, resource planning, and targeted care—while preserving user privacy and autonomy.

REFERENCES

1. R. Kessler, E. Bromet, P. Jonge, V. Shahly, and Marsha., "The burden of depressive illness," Public Health Perspectives on Depressive Disorders, 2017.
2. W. H. Organisation, "Mental health: Fact sheet," https://www.euro.who.int/en/health-topics/noncommunicablediseases/mental-health, 2019.
3. M. Renteria-Rodriguez, "Salud mental en mexico," NOTA-INCyTU NUMERO 007 ´ , 2018.
4. S. Guntuku, D. Yaden, M. Kern, L. Ungar, and J. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," Current Opinion in Behavioral Sciences, 2017.
5. J. Pestian, H. Nasrallah, P. Matykiewicz, A. Bennett, and A. Leenaars, "Suicide note classification using natural language processing: A content analysislin heidelberg," Biomed Inform Insights, 2010.
6. P. Chikersal, D. Belgrave, G. Doherty, A. Enrique, J. E. Palacios, D. Richards, and A.

Thieme, “Understanding client support strategies to improve clinical outcomes in an online mental health intervention,” In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020

7. M. Kosinski, D. Stillwell, and T. Graepel, “Private traits and attributes are predictable from digital records of human behavior,” Proceedings of the national academy of sciences, 2013.
8. M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel, “Manifestations of user personality in website choice and behaviour on online social networks,” Machine learning, 2014.
9. D. Preot¸iuc-Pietro, S. Volkova, V. Lampos, Y. Bachrach, and N. Aletras, “Studying user income through language, behaviour and affect in social media.,” PloS one 10.9, 2015.

T. Correa, A. Willard Hinsley, and H. G. De Zuniga, “Who interacts on the web?: The intersection of users’ personality and social media use.,” Computers in human behavior 26.2, 2010.

10. S. Volkova and Y. Bachrach, “Inferring perceived demographics from user emotional tone and user-environment emotional contrast,” Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016.
11. D. Ram´ırez-Cifuentes and A. Freire, “Upf’s participation at the clef erisk 2018: Early risk prediction on the internet,” Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, 2018.
12. H. Schwartz, J. Eichstaedt, M. Kern, G. Park, M. Sap, D. Stillwell, M. Kosinski, and L. Ungar, “Towards assessing changes in degree of depression through facebook,” In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2014.
13. G. Coopersmith, M. Dredze, and C. Harman, “Quantifying mental health signals in twitter,” Workshop on Computational Linguistics and Clinical Psychology, 2014.
14. C. Xuetong, D. Martin, W. Thomas, and E. Suzanne, “What about mood swings? identifying depression on twitter with temporal measures of emotions,” Companion Proceedings of the The Web Conference 2018, International World Wide Web Conferences Steering Committee, pp. 1653–1660, 2018.
15. M. Aragon, A. L ´ opez-Monroy, L. Gonz ´ alez-Gurrola, and M. Montes- ´ y Gomez, “Detecting depression in social media using fine-grained ´ emotions,” Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019
16.] C. Mathers and D. Loncar, “Projections of global mortality and burden of disease from 2002 to 2030,” PLOS Medicine, Public Library of Science, 2006.
17. M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, “Predicting depression via social media,” In Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, 2013.
18.] M. De Choudhury, S. Counts, and E. Horvitz, “Social media as a measurement tool of

depression in populations.," In Proceedings of the 5th Annual ACM Web Science Conference, 2013.

19. S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, "Recognizing depression from twitter activity," In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015