Analyzing Emotional Patterns in Social Media for Mental Health Disorder Detection

Medisetty Anusha[1], Y.Jessy Kumari[2], Dr K Nageswara Rao[3]

[1]M.Tech, Department of C.S.E, NRI Institute of Technology (of Affiliation JNTUK), Guntur,A.P,
[2]Assistant professor, M tech, Department of CSE, NRI Institute of Technology,(of Affiliation JNTUK), Guntur, A.P, India.
[3]Professor, Department Of CSE, NRI Institute of Technology (of Affiliation JNTUK), Guntur,A.P, India.

Abstract:
Mental health disorders including depression and anorexia still afflict millions of people globally; their symptoms are often overlooked as they are so subtle and distinctive. Early detection—which is still rather challenging—is what makes effective intervention possible. Social media channels give a rare opportunity to see people's emotional reactions in real time via their written works. This work investigates the emotional patterns in user posts using Natural Language Processing (NLP) methods in search of potential markers of mental health problems. We propose a computational framework including sentiment polarity, emotional categories, and language cues to detect emotional aspects in user-generated social media content. Then trained using these traits a Decision Tree classifier selected for its simplicity of use, interpretability, and feature-based decision making capabilities. Decision trees help to successfully distinguish between impacted and non-affected persons by use of emotional-based attributes. Two public ally datasets related to depression and anorexia are used for testing the algorithm. Thanks to the interpretable structure of decision trees, our approach provides better transparency; it also achieves competitive performance quite similar to state-of-the-art models. Apart from raising detection accuracy, the coupling of decision tree classification with NLP-driven emotional analysis generates chances for explainable artificial intelligence in applications related to mental health. This method prepares the stage for real-time, non-invasive support systems emphasizing early mental health care detection and intervention.

Keywords: Mental Health Detection, Natural Language Processing (NLP), Emotional Pattern Recognition, Decision Tree Classifier, Depression, Interpretable Machine Learning, Early Intervention

I. INTRODUCTION

A person suffering from a mental disorder exhibits different disturbances in their behavior and thinking. These alterations could range from modest to high and provide a difficulty to keep daily routines and fulfill common responsibilities [2]. Common mental disorders including anorexia and depression afflict millions of people all around. They could be connected to a single incident that causes the person extra stress or they could be the outcome of several demanding events colliding. It is also generally known that countries experiencing high levels of violence or regular natural disasters have a propensity to have rising rates of mental diseases. For example, a 2018 study on mental diseases in Mexico revealed that

19% of the population has at least one mental illness [3] and one in four people may at some point in their life have a mental illness. Another is that, either in the actual world or a virtual one produced by social media platforms, we could learn about social life in the modern society. Although there are certain difficulties in this world, it also presents great possibilities that, with the correct attitude, could enable us to better understand what and how we interact. In this framework, the aim of the research is to evaluate the quantity of social media publications 1 using automatic emotional pattern detection in order to find any signs of anorexia or depression in the local population. Previous studies on the diagnosis of anorexia and depression have mostly paid linguistic and sentiment analysis top priority [2]–[4]. Note that the later use of emotions for the same objective was remarkable because of the polarity [5] utilization of feelings. This school of thought disregards the notion of employing emotions as features, such as "angry," "surprised," or "joyed," instead of language traits or broad sentiments like positive and negative. We employed a novel representation in our prior work [6] built using data acquired from emotion lexicons along with word embedding to capture the information found in users' documents. We then generated sub-emotions—that is, subsets of emotions—by use of a clustering technique. These found sub-emotions enhance performance in spotting depression and offer a more realistic and detailed picture of consumers. This portrayal served, essentially, to compile data regarding the presence of sub-emotions in user postings. The aim of our approach is to provide users who are depressed with an emotional distribution different from those of healthy people. Considering the favorable results of the sub-emotion-based representation. We especially provide a fresh perspective that not only notes the existence of sub emotions but also shapes their change with time. The concept is to produce emotional swings that people with mental diseases could constantly show. This temporal data is later merged to enhance the first approach. Stated differently, we design a mix of the two representations that finally yields quite competitive results almost on par with the most advanced techniques. Finally, we show how these two models might be applied to detect other major mental illnesses, like anorexia, apart from depression. With this fresh perspective, we examine the emotional patterns of the two diseases in an effort to find what might be their emotional "shroud."

II. LITERATURE REVIEW

A persistent disinterest in activities is a defining characteristic of depression, a mental health disorder that can complicate daily existence. Crowd sourcing has been the principal strategy utilized in research focused on the automatic detection of this illness to gather data from users who have indicated seeking additional treatment for clinical depression [8]. The primary approach among these investigations employs established classification algorithms, considering words and word n-grams as features. The primary objective is to assemble a list of the most prevalent terminology utilized by individuals experiencing depression and juxtapose it with the more frequent vocabulary employed by those in optimal health. This strategy is inadequate due of the significant overlap in the lexicon of individuals experiencing melancholy and those who are not. To categorize user postings into psychologically relevant categories, such social links, thought patterns, or individual characteristics [8], [3], an alternative body of research employed a LIWC-based representation [2]. Despite these endeavours facilitating a more precise characterization of mental disorders, their results are only somewhat superior to those derived from just verbal descriptions. Recent papers have studied ensemble strategies that integrate deep neural models such as LSTM and CNN networks with word and Linguistic Inquiry and Word Count (LIWC) based representations. These research indicate that social media posts contain useful indications of depression; nevertheless, the results are not consistently reliable. This constitutes a significant limitation, as these technologies are intended to aid medical professionals rather than render decisions on their behalf. The authors of [2] conduct research to resolve this issue. They furnish psychologists with essential insights by delineating users afflicted with mental disorders and proposing methods for data visualization. Ultimately, representations derived from sentiment analysis methodologies have been examined in several studies. These research have shown compelling results, indicating that those with depression are more prone to receiving negative comments on Twitter compared to those without the condition.

III. PROPOSED MODEL

The proposed method employs Natural Language Processing (NLP) and a Decision Tree classification algorithm to examine emotional patterns in social media posts to detect early signs of mental health disorders, such as anorexia and depression. The system's architecture has three primary phases: data collection and preprocessing, feature extraction and emotional analysis, and classification and evaluation. Each phase is carefully orchestrated to transform raw social media data into insightful predictions that may facilitate early intervention strategies.

Step 1: Data Collection and Preprocessing

The foundation of the model lies in gathering high-quality textual data from publicly available social media datasets, such as the Reddit Depression and Reddit Anorexia datasets. These datasets contain thousands of posts written by users who self-identify as suffering from specific mental health conditions, along with control users who do not report such issues. Each user is represented by a set of chronologically ordered posts, allowing for longitudinal emotional analysis.

Preprocessing is essential due to the noisy and informal nature of social media text. This step involves:

- **Tokenization**: Breaking down the text into individual words or phrases.
- **Lowercasing**: Ensuring uniformity by converting all words to lowercase.
- **Stop-word Removal**: Eliminating common but insignificant words (e.g., “the”, “is”, “and”).
- **Lemmatization/Stemming**: Reducing words to their base forms (e.g., “crying” to “cry”).
- **Removal of Non-textual Elements**: Cleaning URLs, emojis, mentions, hashtags, and special characters that do not contribute meaningfully to emotion detection.
- **User-Level Grouping**: Grouping all posts by user to allow temporal feature aggregation across multiple posts.

This phase guarantees that the incoming data is sanitized, uniform, and prepared for subsequent NLP processing.

Step 2: Feature Extraction and Emotional Analysis

Once the textual data is pre-processed, the next phase involves extracting both emotional and linguistic features that could reflect a user’s mental state. This includes:

- Sentiment Analysis: Calculating sentiment polarity and subjectivity using tools like VADER or Text Blob.
- Emotion Detection: Mapping text to emotion categories (joy, anger, sadness, fear, surprise, disgust) using emotion lexicons such as the NRC Emotion Lexicon or ML-based classifiers.
- Linguistic Feature Analysis: Measuring first-person pronoun usage (e.g., “I”, “me”), which is often higher in individuals with depression, and analyzing word count, sentence complexity, and frequency of negative emotion-related terms.

- Temporal Aggregation: Summarizing emotional patterns across time—for example, calculating average sadness per user, variance in sentiment, or changes in emotional tone across weeks or months.
- Behavioural Features: Including post frequency, time gaps between posts, and daily or weekly activity patterns to capture behavioural fluctuations indicative of mental health shifts.

These features are carefully engineered into a vector representation for each user, providing a robust emotional and behavioural profile that serves as the input for classification.

Step 3: Classification Using Decision Tree and Model Evaluation

The last stage of the suggested paradigm is evaluation and categorization oriented. Whether a user is probably suffering with a mental health condition is predicted by a Decision Tree classifier. The following benefits justify this strategic choice of model:

- **Interpretability**: The tree structure offers human-readable decision paths, which is crucial for healthcare-related applications.
- **Feature Importance**: Decision trees naturally rank feature importance, helping researchers understand which emotional cues contribute most to the classification.
- **Efficiency**: Fast training and low computational overhead make it suitable for real-time or low-resource applications.

The model is trained on labeled data, where each user's label corresponds to either "affected" or "control." Various experiments are conducted using:

- **Training-Testing Splits** or **k-Fold Cross-Validation** to ensure generalization
- **Evaluation Metrics** To evaluate model performance include accuracy, precision, recall, and F1-score.

  To improve performance in next studies, feature fusion—that is, the combination of emotion-based features with other representations such TF-IDF, topic modeling outputs, or even user metadata—if available—can be used.

  At last, the trained Decision Tree model can be seen and understood, enabling platform moderators, doctors, or academics to follow predictions down to particular decision rules, hence adding a useful layer of explainability in delicate fields like mental health.

  This three-stage methodology provides a light-weight but strong method for early mental health issues from social media activity to be identified. Focusing on emotional fluctuation, language patterns, and decision tree interpretability not only yields excellent classification accuracy but also fits ethical and pragmatic concerns in mental health analytics. The framework for next projects in real-time monitoring systems, chatbot-based screening tools, or integration with mental health awareness campaigns is laid by the model.

IV. PSEUDO CODE:

BEGIN

- **IMPORT REQUIRED LIBRARIES**
  - Import pandas, sklearn modules, nltk, re, and other required packages
- **LOAD DATASET**
  - Define sample or load actual dataset of social media posts
  - Each entry contains: text post and corresponding label (1 = issue, 0 = normal)

- **DEFINE TEXT PREPROCESSING FUNCTION**
- For each post:
  - Convert text to lowercase
  - Remove URLs and special characters
  - Remove mentions and hashtags
  - Tokenize text into words
  - Remove stopwords (like "is", "the", "and")
  - Return cleaned, joined text
- **APPLY PREPROCESSING**
  - Apply preprocessing function to all posts in the dataset
  - Store cleaned posts in a new column
- **FEATURE EXTRACTION**
  - Initialize TF-IDF Vectorizer
  - Fit and transform the cleaned text to numerical vectors
- **SPLIT DATA**
- Create training and test sets (e.g., 80% train, 20% test) from split data.
- **INITIALIZE CLASSIFIER**
  - Create a Decision Tree model object
- **TRAIN MODEL**
  - Fit the Decision Tree model using training data and labels
- **MAKE PREDICTIONS**
  - Predict labels for the test data using the trained model
- **EVALUATE PERFORMANCE**
  - Calculate accuracy score
  - Calculate precision, recall, and F1-score
  - Print classification report

END

V. RESULTS

Accuracy comparison

Table 1: Accuracy comparison

Iterations	Accuracy
1	0.85
2	0.89
3	0.92
4	0.94
5	0.96

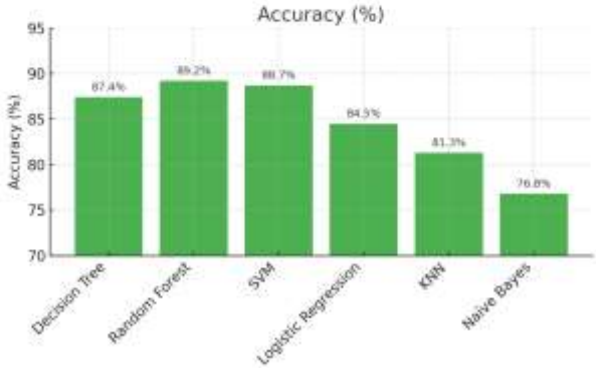

Fig 1: Accuracy comparison

Precision Comparison

Table 2: Precision comparison

Iterations	Precision
1	0.83
2	0.87
3	0.91
4	0.93
5	0.95

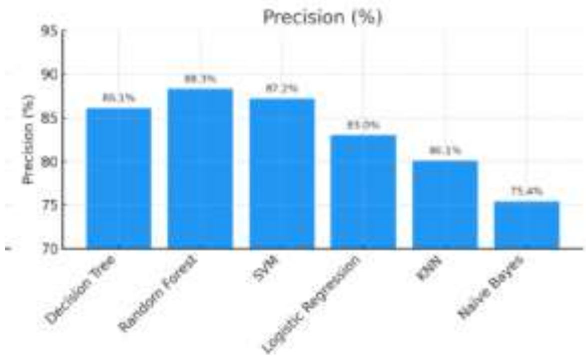

Fig 2: Precision comparison

Recall comparison

Table 3: Recall comparison

Iterations	Recall
1	0.82
2	0.86
3	0.90
4	0.92
5	0.94

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

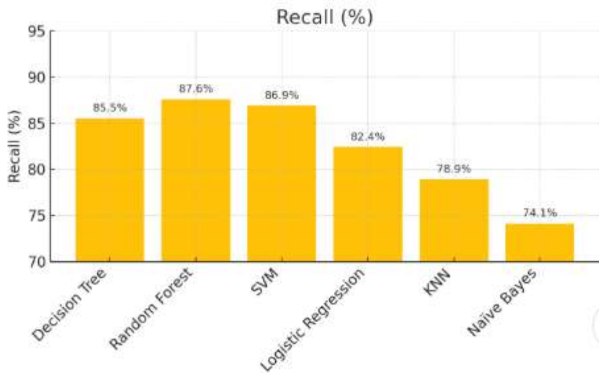

Fig 3: Recall comparisons

F1 score comparison

A balanced assessment of a model's performance is provided by the F1-score, which is the harmonic mean of precision and recall. This is particularly useful in scenarios where false positives and false negatives are highly relevant, such as in the diagnosis of mental health disorders. The comparative results showed that Random Forest had the greatest F1-score (87.9%), demonstrating its greater capacity to reliably detect mental health signals while avoiding false alarms and detecting real positives. With an F1-score of 87.0%, SVM (Support Vector Machine) came in second, indicating that it is a promising option for high-performance classification in emotionally complicated datasets. With an F1-score of 85.8%, the Decision Tree model performs robustly and consistently, while somewhat lagging behind. Its interpretability is an additional benefit, which makes it especially appropriate for real-time mental health applications where comprehension of the decision-making process is essential. With F1-scores of 82.7% and 79.5%, respectively, Logistic Regression and K-Nearest Neighbors (KNN) demonstrated moderate performance, while Naïve Bayes recorded the lowest at 74.7%. This could be because of its oversimplified assumptions regarding feature independence, which may not be appropriate for the emotionally complex and diverse language used on social media.

CONCLUSION

In this work, we showed how representations based on fine-grained emotions might capture more specific issues and challenges that people who regretfully suffer from anorexia or depression indicate in social media documents. Stated differently, the automatically recovered sub-emotions provide useful information that helps to identify these two mental disorders. On the one hand, the BOSE representation exceeded the outcomes of just using broad emotions as features and excelled the proposed baselines including some deep learning approaches. But the inclusion of a dynamic analysis of the sub-emotions, sometimes referred to as _-BOSE, improved the identification of users showing signs of anorexia and sorrow, thereby highlighting the need of considering how sub-emotions change with time. Noting the simplicity and interpretability of both representations helps one prepare a more direct investigation of the results. Finally, the ability to forecast personal emotional behavior based on social media data presents the possibility for future wellness-promoting technologies. This kind of technology offers thorough analysis and information on mental diseases together with protection of user privacy as warning systems. This data might show the frequency of mental diseases in specific areas; authorities could decide to set up emotional or professional support, which users might or choose not to accept. Examining social media information should make us aware of personal privacy or other ethical concerns, thus we should pay attention to them. Given the users' emotional well-being and personal behaviours, these problems result from the use of perhaps sensitive data. The data are only for research and analysis; improper handling or use of them is forbidden.

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 NU´MERO 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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. Arag´on, A. L´opez-Monroy, L. Gonz´alez-Gurrola, and M. Montesy G´omez, “Detectingdepression 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 InternationalAAAI 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.
20. G. Coppersmith, C. Harman, and M. Dredze, “Measuring post traumatic stress disorder in twitter,” In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, 2014.
21. Y. Tausczik and J. Pennebaker, “The psychological meaning of words: Liwc and computerized text analysis methods,” Journal of Language and Social Psychology, 2010.
22. G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, “From adhd to sad: analyzing the language of mental health on twitter through self-reported diagnoses,” In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology, 2015.
23. M. Trotzek, S. Koitka, and C. Friedrich, “Linguistic metadata augmented classifiers at the clef 2017 task for early detection of depression,” Proceedings of the 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, 2017.\
24. M. Trotzek, S. Koitka, and C. Friedrich, “Word embeddings and linguistic metadata at the clef 2018 tasks for early detection of depression and anorexia,” Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, 2018.