Detection of Depression in Social Media Posts using Emotional Intensity Analysis

M. Kiran Mayee
Department of Computer Science and Engineering, Ravindra College of Engineering for Women, Kurnool, Andhra Pradesh, India
kinu92@gmail.com (corresponding author)

R. Deepthi Crestose Rebekah
Department of CSE, Ravindra College of Engineering for Women, Kurnool, Andhra Pradesh, India
rdeepthicse@recw.ac.in

T. Deepa
Department of Computer Science & Engineering (AI & ML), P. V. P. Siddhartha Institute of Technology, India
deepayasoda@gmail.com

G. Divya Zion
Department of Artificial Intelligence & Data Science, Mother Theresa Institute of Engineering and Technology, Palamaner-517408, Chittoor, Andhra Pradesh, India
ziondivya@gmail.com@gmail.com

K. Lokesh
Department of Artificial Intelligence & Data Science, Mother Theresa Institute of Engineering and Technology, Palamaner-517408, Chittoor, Andhra Pradesh, India
klokesh280488@gmail.com

Received: 12 April 2024 | Revised: 6 May 2024 and 13 June 2024 | Accepted: 20 June 2024

ABSTRACT

Tapping into digital footprints on social media, this research focuses on providing new insights into detecting depression through textual analysis. Initially, emotional raw data found in social media posts, aimed particularly at the expressions of anger, fear, joy, and sadness, were collected and analyzed. These emotions, each scored by their intensity, offer a quantifiable view into the users' mental state, serving as possible depression markers. Central to the methodological framework adopted is the binary classification system, which classifies texts into depressive or non-depressive states, well founded by the patterns unearthed from the data. The proposed model rigorously trains Artificial Intelligence/Machine Learing (AI/ML) models to traverse through the complexities of natural language, concentrating on noticing delicate indications that signal depression. The introduced models are tested and measured with accuracy, precision, recall, and F1-score. RoBERTa, DistilBERT, and Electra are the transformer-based models emphasized in this research. Their performance is critically evaluated, with the results denoting particular capabilities in understanding and contextualizing language, which is the key advantage in the early identification of mental health issues. This research stands at the intersection of technology and mental health, revolutionizing mental health monitoring and intervention.

Keywords-sentimental analysis; emtional analysis; machine learning; natural language processing

I. INTRODUCTION

Depression, a widespread mental health issue, influences a vast number of people worldwide. This condition not only changes individuals' mood and behavior, but also affects main physical functions like sleep and appetite. Traditional methods for early depression identification, which are based on self-reports and clinical assessments, frequently face challenges in terms of reliability and accessibility. These issues emerge due to various social and personal factors that can impact the aforementioned methods' accuracy and have been increased owing to the broad social media use and the extensive amount of textual data. Thus, innovative ways for monitoring mental health and the progress of new intervention strategies need to be generated. Social media podiums, where individuals frequently express their emotions, thoughts, and experiences, have become a rich storage of data for detecting signs of depression. The textual content shared on these platforms acts as a mirror, reflecting the emotional state of its users. These data can provide early indicators of mental health issues and are investigated by the present study for potential depression markers to be uncovered. In specific, the current study systematically analyzed four distinct datasets, each representing different emotional expressions: anger, fear, joy, and sadness. These datasets have been carefully collected from social media posts. Their result lies in the textual content of the posts and additionally in the expansion of "Intensity Scores". These scores are particular as they quantify the reliability of the emotions expressed, giving a measurable dimension to the often subjective and delicate nature of the emotional content. The research work mainly focused on applying ML and Deep Learning (DL) techniques to review these datasets for signs of depression. ML, with its strengths in pattern diagnosis, classification, and prediction, is especially suited for evaluating complex, high-dimensional data such as text. This research work introduces a novel neural network model specifically designed for matching user queries with relevant research articles in the Scopus database, overcoming the limitations of traditional keyword-based methods. The innovative architecture enhances query-article matching accuracy by leveraging DL techniques. Additionally, the creation of a new dataset tailored for this task provides a valuable resource for future research. Comprehensive evaluations demonstrate the model's superior performance compared to existing methods, highlighting its potential to significantly improve academic search engines. This work advances the field by facilitating more efficient and accurate information retrieval for researchers.

II. LITERATURE REVIEW

In the realm of detecting depression through social media, several notable studies have been conducted. Authors in [1] implemented a delicate method to identify depression through textual analysis of social platform posts, using two different datasets from Reddit and Twitter. This methodology deployed a balanced acquisition of data, assuring an equal representation of depressive and non-depressive posts. An ensemble model of DL classifiers, merging Fasttext for word embedding with a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) model, termed "FCL," was proposed. In [2], the authors utilized a combination of social media data, including 400 forum posts by depressed individuals and control data from non-depression forums. They implemented Support Vector Machines (SVMs) with character and word n-grams, optimized through random search and 5-fold cross-validation. The data preprocessing included bag-of-n grams features, which were weighted by applying the BM25 algorithm. The highest F1 scores were achieved through the employment of this approach, with changing effectiveness based on the specificity of the implemented datasets. This work projected that a careful selection of source data is crucial for accurate depression detection in social media texts, and emphasized the use of Reddit as a valuable data source due to its anonymity and lack of length limitations in posts.

Authors in [3] adopted a multi-task multi-lingual technique for sentiment analysis, handling resource deficiency by using cross-lingual word embeddings. They implemented Recurrent Neural Network (RNN) models, Gated Recurrent Unit (GRU), LSTM, and Bidirectional LSTM (LSTM_Bi). In [4], 22,808 Reddit posts were collected over a period of three months, involving both anxiety-related and general posts, which were data pre-processed by removing URLs, HTML tags, and punctuation, and further pre-processed by deleting stop words and lemmatizing. Word2Vec, Doc2Vec, LDA, and LIWC features were utilized for feature generation, while logistic regression, SVM, and neural networks were implemented for classification. According to the results, up to 98% accuracy in classification was achieved. In [5], 22,355 clinical records in Mandarin Chinese were trained for depression to be detected. Data preprocessing steps included tokenization, digital encoding, sequence trimming, and padding. This study used BERT and CNN models for classification, fine-tuned with the continuation of 512 tokens. The BERTgeneral model achieved an AUC of 0.93, with high sensitivity and specificity, but varied in execution, which was noted between civilian (AUC = 0.91) and military samples (AUC = 0.79). Specialized models, namely BERT civilian and BERT military, were also implemented but showed no significant enhancement in execution. In [6], researchers evaluated Bengali text from blogs and open sources, implementing ML models like SVM, random forest, logistic regression, KNN, and NB, and DL models such as LSTM and GRUs for depression severity detection. The highest results were achieved with GRUs, gaining an accuracy of 81%. Authors in [7] presented a novel approach for depression detection by evaluating speech as a sequence of acoustic events, combining acoustic "words" with speech landmarks, deploying Natural Language Processing (NLP) methods. The flexible approach integrates diverse event types and allows for various fusion levels, exhibiting significant enhancement in F1 (depressed) scores of up to 15% and 13% compared to traditional acoustic approaches. Authors in [8] applied a systematic review, adhering to PRISMA guidelines and registered with PROSPERO, They evaluated 327 articles from medical databases using keywords related to ML and psychiatry, excluding non-English and some other types of publications. Out of these, 58 articles were included, focusing on themes like symptom extraction, illness severity classification, and therapy effectiveness derived from patient medical records and social media data.

In [9], various Python packages like Keras, Pandas, Gensim, and Matplotlib were utilized to implement a multilingual sentiment analysis model. The model, incorporating voice synthesis and cross-lingual word embeddings with 300-dimensional vectors, was tested using English-only synthesized speech. The evaluation included auditory assessments by native speakers and deployed CNN architecture for each voice sample. The results compared the implementation of static/dynamic embeddings in BERT and VectorMap across English, Japanese, and Mangarian, projecting significant improvements in precision, recall, and F-measure for the proposed cross-lingual sentiment estimation model over the baseline multilingual models. The model efficiency was demonstrated in both static and dynamic embedded settings across various languages.

In [10], Facebook was used as a platform for detecting depression through social media interactions. The K-nearest neighbor technique was adopted to analyze emotional states associated with depression. In [11], a sophisticated multi-kernel SVM-based model was utilized. This model was implemented to extract a variety of characteristics from social media profiles to accurately depict individuals' mental health conditions. Authors in [12] turned their attention to Twitter, applying algorithms, such as Random Forest, Naive Bayes, and Liblinear to detect signs of depression in tweets. Their study was successfully differentiated between 35 Twitter users experiencing depression and 62 who were not. Notably, the Liblinear algorithm emerged as the most effective one in that study.

In the current study, the application of ML models like Etiqa'a, which classifies WhatsApp messages with 81% accuracy to safeguard minors, showcases the versatility of the previously addressed technologies in preventive digital health monitoring. This model's approach to alerting on inappropriate content mirrors the proposed methodology's efficacy in using AI for emotional analysis, emphasizing the potential of machine learning in enhancing digital interactions and mental health safeguarding [13-15].

III. PROPOSED METHODOLOGY

This study deployed two primary datasets: an emotion-labeled text dataset and a comprehensive collection of tweets. The former included categories, such as anger, fear, joy, and sadness, providing a diverse range of sentiments for analysis. The tweet dataset, named tweets_combined_12.csv, was carefully arranged to include a wide range of public sentiments expressed on that social platform. Figure 1 displays the distribution of intensity scores across four different emotions: anger, fear, joy, and sadness. These histograms offer a visual representation of the frequency of various intensity levels for each emotion within the collected datasets.

A. Data Preprocessing

Figure 2 provides the details of the data preprocessing steps followed in the research. Text Normalisation: Mistakes were fixed, the text was converted to lowercase, and slinguistic variances were standarized in order for an overall standarization to be achieved. Noise Removal: To improve the quality of the data, unnecessary information, such as usernames, hashtags, URLs, and special characters were eliminated. Managing Missing Values: To ensure dataset integrity, sophisticated procedures were employed to impute or eliminate missing values. Lemmatization and tokenization: The text was divided into discrete parts, or tokens, and was reduced to either dictionary or base form.

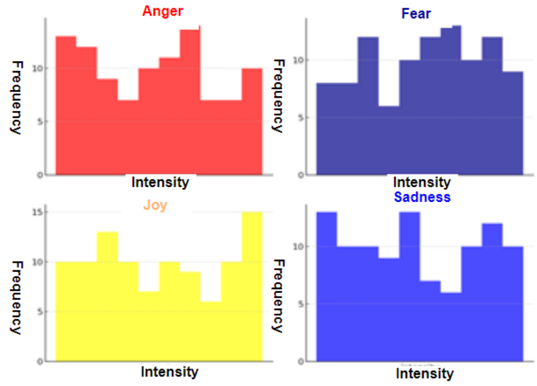

Fig. 1. Emotional intensity distribution across anger, fear, joy, sadness.

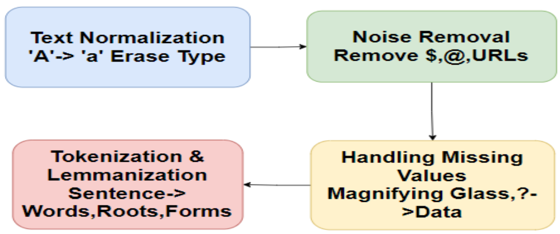

Fig. 2. Detail steps of data preprocessing.

B. Feature Extraction

This phase was pivotal in transforming raw text into analyzable features. The proposed approach combined classical NLP techniques with modern embedding methods.

- Vectorization: TF-IDF (Term Frequency-Inverse Document Frequency) was used to convert text data into a numerical format, highlighting the importance of words within the documents. Word Embeddings: Advanced embedding techniques like Word2Vec or GloVe were deployed to capture contextual meanings.
- Syntactic Parsing: Utilizing libraries like Spacy, sentence structures were analyzed, extracting grammatical and syntactic patterns valuable for understanding sentiment.

C. Model Development and Training

Figure 3 outlines the sequential workflow for detecting depression from textual data. It starts with datasets for specific emotions—anger, fear, joy, and sadness—likely sourced from social media. These datasets are then combined into one, which is processed by AI/ML models trained to identify depressive patterns. The final stage is the analysis output, where the model's predictions are evaluated to determine the presence of depression. This process highlights the transition from raw data to actionable insights in the context of mental health.

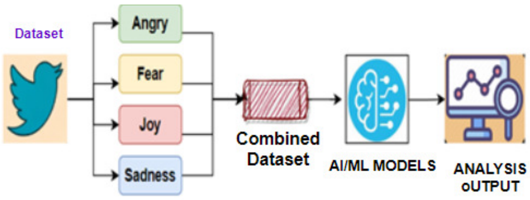

Fig. 3. Methodology: emotion-driven depression detection pipeline.

The diagram in Figure 4 offers a visual comparison of target class distributions across six cleaned Twitter datasets, labeled from df1 to df6. df0 is the generated tweeter csv file and the remaining files were obtained from open-source datasets. Each subplot represents one dataset, showcasing the frequency of tweets classified into two target groups. The '0' class, represented by the blue bar, likely corresponds to tweets that do not exhibit the condition being studied, such as a specific sentiment or behavioral indicator, whereas the '1' class, shown in orange, indicates tweets that do exhibit the condition under investigation.

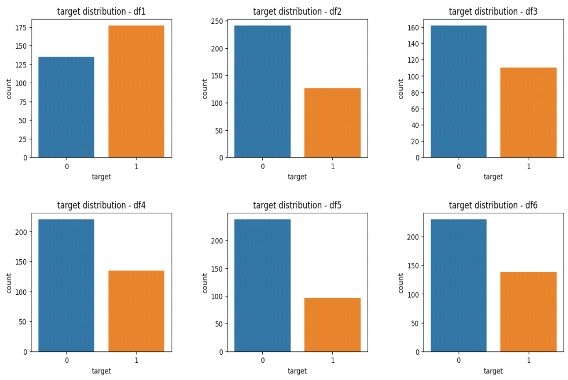

Fig. 4. Distribution of target variables across multiple twitter datasets.

D. Extracting and Analyzing Emotional Context

The extraction of emotional text from social platform posts includes a combination of NLP methods and sentiment analysis techniques. The process entails several key steps:

- **Text Preprocessing:** Social platform texts are cleared by removing noise such as links, symbols like hashtags, and special characters. The text is then tokenized, and stop words are removed for meaningful words to be retained.
- **Emotion Recognition**: A pre-trained emotion classification model is utilized to identify the primary emotion in each post. The model classifies emotions into common categories, such as anger, fear, joy, and sadness.
- **Intensity Scoring**: For each analyzed emotion, an intensity value is assigned deploying a sentiment analysis technique. The emotion score is measured on a scale from 0 to 1, with higher scores representing stronger expression.
- **Contextual Analysis:** To analyze the context, topic modeling is implemented on the text. This helps to identify the themes or topics associated with different emotional classes.
- **Depression Detection:** Based on the emotional text and intensity values, texts are classified as depressive or non-depressive states.

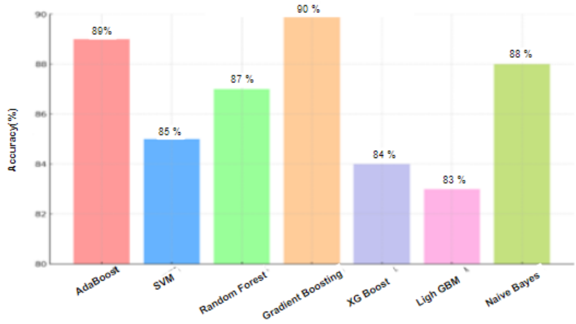

Fig. 5. Comparative analysis of machine learning model accuracies.

IV. EVALUATION

A. Basic Machine Learning Models

The evaluation framework was multi-dimensional, focusing on both accuracy and exact understanding of text sentiment. The chart in Figure 6 provides a detailed comparison of three NLP models—RoBERTa, DistilBERT, and Electra—across four crucial evaluation metrics: accuracy, precision, recall, and F1-Score. RoBERTa excels particularly in Recall (0.93) and F1-Score (0.91), indicating its strength in identifying true positive cases of depression and maintaining a balance between precision and recall. DistilBERT, while slightly trailing, presents robust scores with an accuracy of 0.86 and an F1-Score of 0.86, suggesting good overall performance but with some limitations in precision (0.77). Electra shows similar effectiveness with F1-Score of 0.84, but slightly lower precision and recall, indicating a modest trade-off between identifying true positives and avoiding false positives. This visual comparison helps in understanding the relative strengths and weaknesses of each model in handling text-based depression detection.

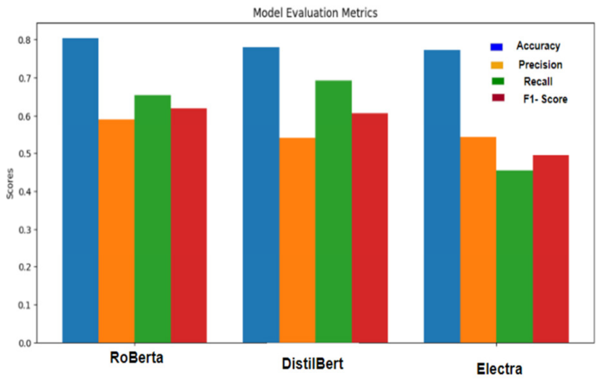

Fig. 6. Comparative evaluation of transformer-based NLP models.

V. CONCLUSION

The proposed model has achieved commendable results in the application of deep learning techniques for depression detection through social platform text analysis. By implementing advanced transformer-based models, such as DistilBERT, RoBERTa and Electra, it has reached high accuracy and placed the field in a new level. The initiation of emotional intensity values allocation and analysis has added depth to the introduced model, allowing for a greater understanding of individuals' mental strength. The presented evaluation framework entails the metrics of accuracy, precision, recall, and F1-score, ensuring a holistic estimation of the proposed model's performance. The latter's ability to discern subtle linguistic cues offers a promising direction for future research, particularly aimed at the development of more personalized mental health monitoring tools. Future studies could expand on this approach by exploring multimodal data and incorporating real-time analysis to further improve the timeliness and effectiveness of interventions.

VI. LIMITATIONS AND FUTURE DIRECTIONS

The findings of this study might not be practicable to all groups of people because the proposed model solely considered certain types of social platform users. In the future, a vast range of people and different social platforms should be explored to provide more real world applicable results. Also, future work should take into account other formats of data, namely video or audio to get a complete picture of depression. It will also help to conduct research over a longer time period to study how perfectly the models work in different stages and situations of depression.

REFERENCES

[1] V. Tejaswini, K. Sathya Babu, and B. Sahoo, "Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 1, Jan. 2024, Art. no. 4, https://doi.org/10.1145/3569580.

[2] I. Pirina and C. Coltekin, "Identifying Depression on Reddit: The Effect of Training Data," in *3rd Social Media Mining for Health Applications Workshop & Shared Task*, Brussels, Belgium, Oct. 2018, pp. 9–12, https://doi.org/10.18653/v1/W18-5903.

[3] B. Lalitha, V. Madhurima, C. Nandakrishna, S. B. Jampani, S. J. N. Chandra, and R. P. Venkat, "Data Augmentation based Cross-Lingual Multi-Speaker TTS using DL with Sentiment Analysis," *ACM Transactions on Asian and Low-Resource Language Information Processing*, pp. 1–17, Jul. 2023, https://doi.org/10.1145/3628428.

[4] J. H. Shen and F. Rudzicz, "Detecting Anxiety through Reddit," in *Fourth Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Vancouver, BC, Canada, Dec. 2017, pp. 58–65, https://doi.org/10.18653/v1/W17-3107.

[5] T.-Y. Chen, H.-T. Chu, Y.-M. Tai, and S.-N. Yang, "Performances of Depression Detection through Deep Learning-based Natural Language Processing to Mandarin Chinese Medical Records: Comparison between Civilian and Military Populations," *Taiwanese Journal of Psychiatry*, vol. 36, no. 1, pp. 32–38, Mar. 2022, https://doi.org/10.4103/TPSY.TPSY_9_22.

[6] M. K. Kabir, M. Islam, A. N. B. Kabir, A. Haque, and M. K. Rhaman, "Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques," *JMIR Formative Research*, vol. 6, no. 9, Sep. 2022, Art. no. e36118, https://doi.org/10.2196/36118.

[7] Z. Huang, J. Epps, D. Joachim, and V. Sethu, "Natural Language Processing Methods for Acoustic and Landmark Event-Based Features in Speech-Based Depression Detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 435–448, Feb. 2020, https://doi.org/10.1109/JSTSP.2019.2949419.

[8] M. Guan, S. Cho, R. Petro, W. Zhang, B. Pasche, and U. Topaloglu, "Natural language processing and recurrent network models for identifying genomic mutation-associated cancer treatment change from patient progress notes," *JAMIA Open*, vol. 2, no. 1, pp. 139–149, Apr. 2019, https://doi.org/10.1093/jamiaopen/ooy061.

[9] P. Nuthakki *et al.*, "Deep Learning based Multilingual Speech Synthesis using Multi Feature Fusion Methods," *ACM Transactions on Asian and Low-Resource Language Information Processing*, pp. 1–16, Jun. 2023, https://doi.org/10.1145/3618110.

[10] M. R. Islam, A. R. M. Kamal, N. Sultana, R. Islam, M. A. Moni, and A. ulhaq, "Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique," in *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering*, Rajshahi, Bangladesh, Feb. 2018, pp. 1–4, https://doi.org/10.1109/IC4ME2.2018.8465641.

[11] Z. Peng, Q. Hu, and J. Dang, "Multi-kernel SVM based depression recognition using social media data," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 43–57, Jan. 2019, https://doi.org/10.1007/s13042-017-0697-1.

[12] J. N. C. Sekhar, B. Domathoti, and E. D. R. Santibanez Gonzalez, "Prediction of Battery Remaining Useful Life Using Machine Learning Algorithms," *Sustainability*, vol. 15, no. 21, Jan. 2023, Art. no. 15283, https://doi.org/10.3390/su152115283.

[13] C. Magashi, J. Agbinya, A. Sam, and J. Mbelwa, "Prediction of SACCOS Failure in Tanzania using Machine Learning Models," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12887–12891, Feb. 2024, https://doi.org/10.48084/etasr.6696.

[14] D.-D. Le, V.-H. Pham, and T.-A. Bui, "Computational and Experimental Investigation of Thermal Generation in CNC Milling Machine Spindle Bearing with the Oil-Air Lubrication Method," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12900–12905, Feb. 2024, https://doi.org/10.48084/etasr.6603.

[15] S. M. Alanazi and G. S. M. Khamis, "Optimizing Machine Learning Classifiers for Enhanced Cardiovascular Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12911–12917, Feb. 2024, https://doi.org/10.48084/etasr.6684.