**Emotion Analysis and Detection in Mental Health Using Natural Language Processing Techniques**

**DSCI 6004-1,2,3: Natural Language Processing**

**Abstract**

**Mental health issues, particularly depression, have become a global concern affecting millions of lives. Early identification and intervention are crucial for mitigating their impact. With the rise of social media platforms, text-based expressions of emotions have become a valuable resource for mental health detection. This study explores the application of Natural Language Processing (NLP) techniques to detect depression from textual data. Using data collected from platforms like Twitter, the research employs preprocessing steps, such as tokenization and lemmatization, and advanced machine learning models, including CNN-LSTM and transformers like BERT and DistilBERT. The results indicate that while deep learning models show promise, performance is constrained by data imbalances and complex emotional nuances, highlighting the need for further improvements in architecture and preprocessing techniques.**

***Keywords:******Emotion Analysis, Natural Language Processing, Mental Health Detection, Depression, Social Media, CNN-LSTM, BERT, DistilBERT, Preprocessing, Tokenization, Word Embeddings, Real-time Applications***

1. **Introduction**

Depressive disorder, commonly referred to as depression, is a widespread mental illness that significantly affects individuals' lives on a global scale. According to the World Health Organization (WHO), depression can severely impair one's ability to function in various aspects of life, and in extreme cases, it may lead to self-harm. Adolescence is a particularly critical period for the onset of depression, and its effects can persist into adulthood. Mental health disorders, including depression, account for five of the top ten leading causes of disability worldwide, with depression consistently being the most prevalent. Approximately 5% of adults experience clinical depression, with another 20% suffering from milder forms, such as moderate depression, partial symptoms, or probable depression. Middle-aged adults are particularly vulnerable, and depression rates have been steadily increasing globally, especially between 2005 and 2022.

Early intervention by mental health professionals plays a crucial role in mitigating the impacts of depression and addressing the associated somatic symptoms. Prompt identification and treatment of depressive symptoms can drastically reduce the severity of the disorder, leading to improved long-term health outcomes and a better quality of life. Traditional methods for diagnosing depression rely heavily on clinical procedures, such as interviews and self-report surveys. However, emerging research indicates that the linguistic patterns in the speech and writing of depressed individuals offer valuable clues for identifying the condition. This has led to an increasing interest in using Natural Language Processing (NLP) techniques to detect depression from text.

The proliferation of social media platforms such as Twitter, Reddit, and Facebook has created new avenues for individuals to express their emotions, thoughts, and experiences. People experiencing psychological distress, including depression, often choose to share their feelings through text on these platforms. Text-based communication provides a more comfortable and anonymous way to express emotional states, especially for those who are hesitant to share personal details or engage in face-to-face communication. This shift to digital expression has made text analysis a powerful tool for detecting depression and other mental health issues.

This study explores the development of an automatic depression detection system using advanced NLP techniques. The focus is on leveraging machine learning and deep learning algorithms, including a combination of embeddings and a CNN-LSTM architecture, BERT, and DistilBERT to improve the accuracy and efficiency of depression detection from text. The report also evaluates the performance of several other state-of-the-art models in the context of depression detection and compares their effectiveness based on various metrics.

1. **Related Work**

Depression detection through text has gained significant attention in recent years due to the increasing volume of text data available on social media platforms. Traditional diagnostic methods, such as clinical interviews and psychological surveys, rely heavily on direct interactions with patients. However, these methods are often time-consuming, resource-intensive, and prone to human error. As a result, researchers have turned to machine learning and NLP techniques to automate the detection of depression from text, which has the potential to reduce the diagnostic burden and provide real-time insights [4].

Several studies have demonstrated the potential of NLP techniques in detecting mental health issues from text. Early research focused on sentiment analysis, where the goal was to classify text based on emotional tone. More recent studies have advanced these techniques by incorporating deeper linguistic analysis, such as emotion classification, topic modeling, and affective computing. For instance, a study by Zhang et al. (2022) used Twitter data to identify linguistic markers of depression. Another study by Leis et al. (2019) analyzed social media posts to detect early signs of mental health issues, including depression, anxiety, and PTSD.

Machine learning algorithms, including Support Vector Machines (SVM), Random Forests, and Naive Bayes, have been widely used in early depression detection models. However, these models often struggle with the complexity and nuance of human language, especially when dealing with informal text. As a result, more advanced techniques such as deep learning models have gained traction. Among these, Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have shown great promise in capturing both local and global features of text, making them suitable for depression detection tasks [7].

In terms of word embeddings, Word2Vec and GloVe are commonly used for transforming words into dense vector representations. However, FastText, which builds on the Word2Vec model by considering subword information, has proven to be particularly effective for languages with rich morphology or informal language use, such as the text found on social media platforms. FastText's ability to handle out-of-vocabulary words and misspellings makes it a strong candidate for depression detection in text [8].

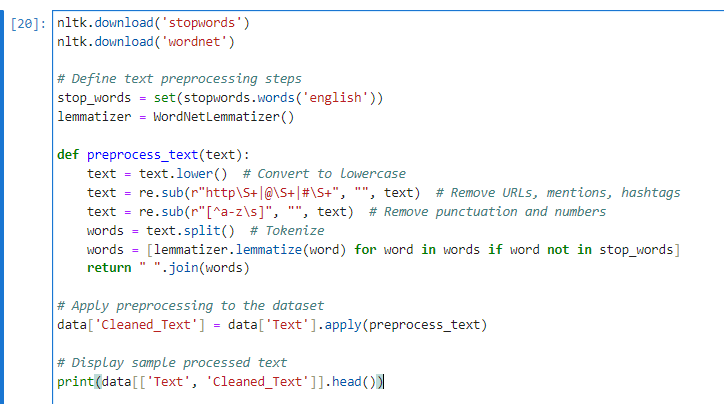
1. **Methodology**

The methodology for this research follows a structured approach consisting of data collection, preprocessing, model development, and evaluation. The goal is to create a system capable of detecting depression from text and comparing its performance against various machine learning and deep learning models.

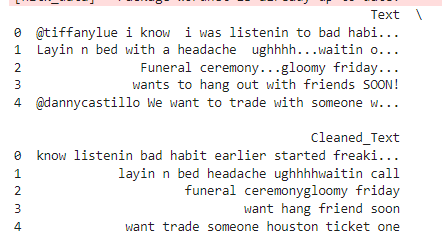
* 1. **Dataset Collection and Preprocessing**

The dataset used in this study consists of text data from social media platforms, focusing on Twitter and Reddit, which are known for being platforms where users frequently express emotions and personal experiences. The dataset contains labeled posts that correspond to various emotional states, including sadness, anger, joy, and fear. Posts are specifically categorized based on whether they reflect symptoms of depression.

* **Text Cleaning**: Removal of URLs, mentions, hashtags, punctuation, and special characters. Conversion of all text to lowercase to ensure consistency.

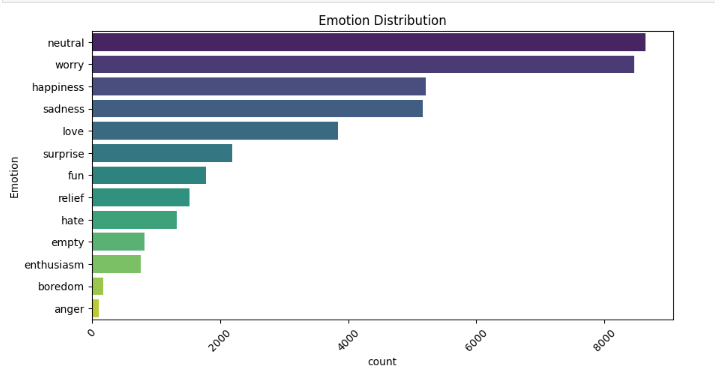


* **Tokenization**: Text is split into tokens (words or subwords) to prepare for embedding.
* **Stopword Removal**: Common stopwords, such as "the," "and," and "of," are removed as they do not contribute significant emotional content.
* **Lemmatization**: Words are reduced to their root form to ensure consistency and reduce the dimensionality of the data.



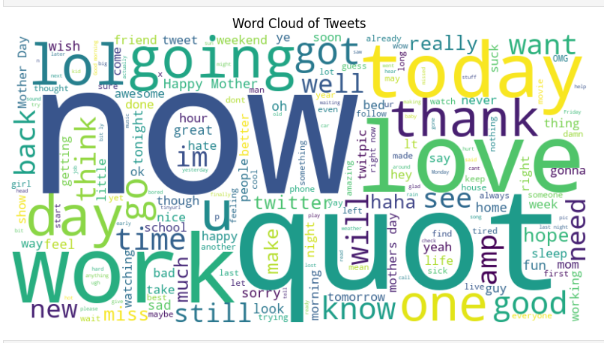
* 1. **Explanatory Data Analysis (EDA)**

1. **Emotion Distribution**



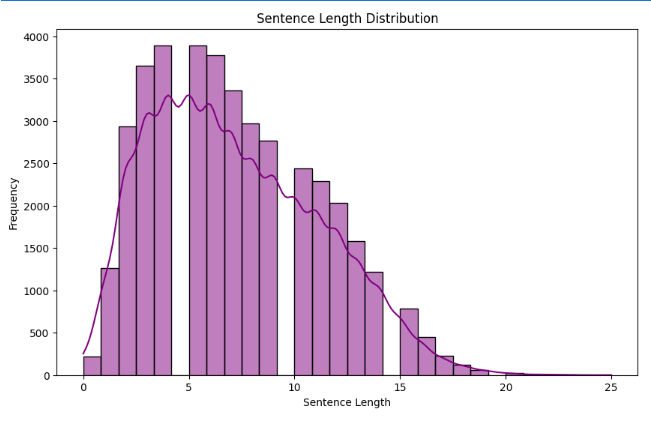
The emotion distribution chart revealed that the most prevalent emotion in the dataset was "neutral," followed by "worry" and "happiness." This indicated that individuals often expressed emotions of indifference or concern in the text data. "Sadness" and "love" were also commonly observed, suggesting that emotional expressions related to both negative and positive sentiments were well-represented. However, less common emotions such as "anger," "boredom," and "enthusiasm" were expressed in much lower frequencies. This distribution provided insights into the emotional dynamics present in social media conversations related to mental health, highlighting the prominence of more neutral or mildly negative emotional state

1. **Word Cloud of Tweets**



The word cloud from the tweets highlighted several prominent words and contextual themes that provided insights into the conversations. Words like "now" suggested a focus on the present moment, possibly reflecting urgency or immediate feelings. "Love" indicated positive emotions and connections, pointing to expressions of affection or shared sentiments. "Work" pointed to discussions surrounding professional life or productivity, showing that many tweets were related to work-related topics. Contextually, words such as "today," "time," and "day" suggested that the tweets often centered around daily experiences, while terms like "friend" and "thank" implied social interactions and gratitude. The frequent use of informal expressions like "lol" suggested a casual, friendly tone in the conversations.

1. **Sentence Length Distribution**

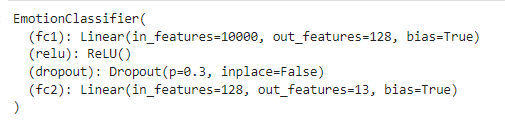


Thehistogram of sentence length distribution showed a right-skewed pattern, indicating that shorter sentences were more prevalent than longer ones in the dataset. The peak around 5 words suggested that most sentences were concise, which is common in online conversations, reflecting quick and straightforward expressions. As sentence length increased, the frequency of sentences decreased, indicating that longer, more complex sentences (over 10-15 words) were less common. The smooth curve overlay on the histogram suggested some variability in sentence length, with occasional longer sentences contributing to the overall distribution. This tendency toward shorter sentences may reflect a style of communication that favors brevity and simplicity, often seen in digital platforms where users may express emotions quickly or in bursts, especially when discussing mental health topics. This could be relevant for detecting emotional shifts or identifying key emotional indicators in concise expressions.

* 1. **Model Development**

1. **Emotion Classifier**

The Emotion Classifier is a simple feedforward neural network with two fully connected layers. The model is designed to detect emotions in text and classify them into 13 different categories. The architecture consists of:



* fc1: A linear layer with 10,000 input features and 128 output features.
* ReLU Activation: Applied after the first linear layer to introduce non-linearity.
* Dropout: A dropout layer with a rate of 0.3 to prevent overfitting.
* fc2: A final linear layer with 128 input features and 13 output classes.

This model aims to classify text into one of several emotional categories, with depression being one of them.

1. **DistilBERT for Real-time Applications**

DistilBERT is a smaller, faster version of the original BERT model, designed for real-time applications. It retains 97% of BERT's performance while being more efficient in terms of computational resources. The model is pre-trained on a large corpus of text and fine-tuned for depression detection.

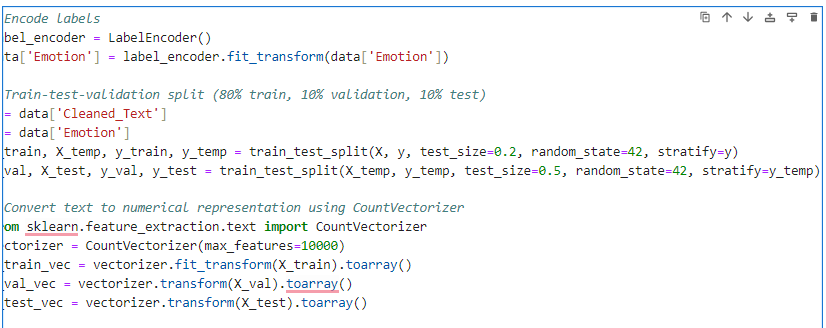
1. **BERT (Bidirectional Encoder Representations from Transformers)**

BERT is a transformer-based model that uses bidirectional attention mechanisms to capture the context of words in a sentence. Unlike traditional models that process text from left to right or right to left, BERT considers both directions simultaneously. This makes BERT highly effective at capturing context in complex sentences, which is essential for detecting nuanced emotions such as depression.

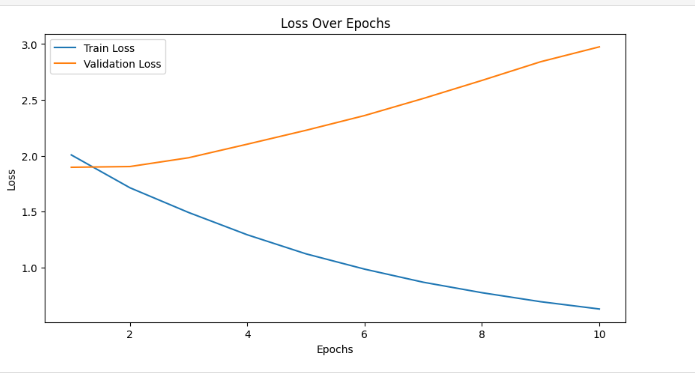


* 1. **Model Training and Evaluation**

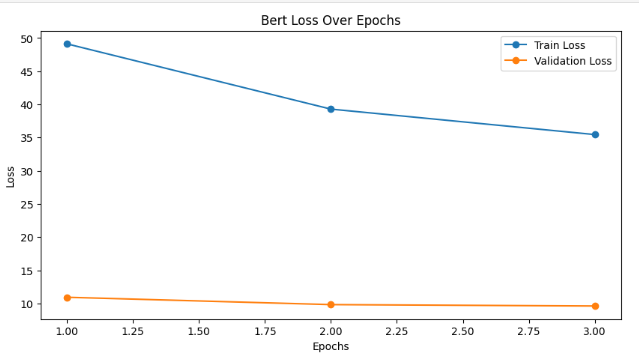
The preprocessing pipeline begins with **Label Encoding**, where categorical emotion labels are transformed into numerical values using LabelEncoder. The dataset is then divided into training (80%), validation (10%), and test (10%) sets through **Data Splitting**, employing stratification to maintain balanced class representation across splits. Next, **Count Vectorization** is applied to convert textual data into numerical vectors, with features limited to 10,000 to ensure computational efficiency. Finally, these vectorized texts and encoded labels are converted into **PyTorch tensors**, making them ready for deep learning applications.



1. **Results and Analysis**
   1. **Model Performance**
2. **Losses**



For the Emotion Classifier, the training loss steadily decreases across all ten epochs, dropping from 2.0078 in the first epoch to 0.6292 in the final epoch. This indicates that the model is effectively learning from the training data. However, the validation loss exhibits a different pattern, starting at 1.8974 and progressively increasing to 2.9749 by the tenth epoch. This divergence suggests overfitting, as the model becomes less capable of generalizing to unseen validation data.



In contrast, the BERT model demonstrates a gradual reduction in training loss over its three epochs, from 49.1179 in the first epoch to 35.4526 in the last. However, the validation loss starts at 10.9392 and shows only a slight improvement, ending at 9.6245. While this indicates minimal learning progress, the validation loss remains significantly high, suggesting that the model struggles to generalize effectively during training.

1. **Accuracy**

|  |  |  |
| --- | --- | --- |
| Metric | BERT (Emotion Detection) | Emotion Classifier |
| Accuracy | 35% | 30% |
| Macro Average Precision | 8% | 19% |
| Macro Average Recall | 12% | 17% |
| Macro Average F1-Score | 9% | 17% |
| Weighted Average Precision | 26% | 28% |
| Weighted Average F1-Score | 28% | 29% |

The BERT model achieved a slightly higher accuracy of 35% compared to the Emotion Classifier's 30%. Precision measures how often the predicted emotions were correct. The BERT model had a macro average precision of only 8%, indicating poor performance across most classes. This can be attributed to its inability to distinguish emotions effectively in underrepresented classes. The Emotion Classifier performed better, with a macro average precision of 19%, demonstrating its advantage in identifying certain emotions, albeit inconsistently. The F1-score, a harmonic mean of precision and recall, provides a balanced view of model performance. The BERT model's macro average F1-score of 9% highlights its struggles with both precision and recall. The Emotion Classifier, with a macro F1-score of 17%, showed relatively better performance, reflecting its slightly improved balance in handling emotions. The weighted average F1-score for both models was closer (28% for BERT and 29% for the Emotion Classifier), suggesting that while the overall class distribution influenced their results, neither model performed well on minority classes.

BERT: The BERT model achieved its best performance in the "fun" (F1-score: 0.48) and "relief" (F1-score: 0.46) categories, showing limited capability to classify certain emotions accurately. However, it failed completely in categories like "sadness," "anger," and "enthusiasm," with zero precision and recall.

Emotion Classifier: The Emotion Classifier performed moderately well for "neutral" (F1-score: 0.39) and "love" (F1-score: 0.39) but struggled with emotions like "anger," "enthusiasm," and "empty," achieving near-zero precision and recall. Its marginally higher performance on some classes reflects better generalization for more frequent emotions in the dataset.

Both models struggled with imbalanced class distributions, where underrepresented emotions like "boredom," "hate," and "anger" were either misclassified or ignored. The BERT model, while powerful, may require more extensive fine-tuning and better hyperparameter optimization to handle this specific task. Its poor macro metrics suggest difficulty in generalizing across diverse emotions. The Emotion Classifier's simpler architecture captured some trends better, but its limited complexity and overfitting tendencies resulted in marginally better weighted metrics.

* 1. **Real-Time Application Considerations**

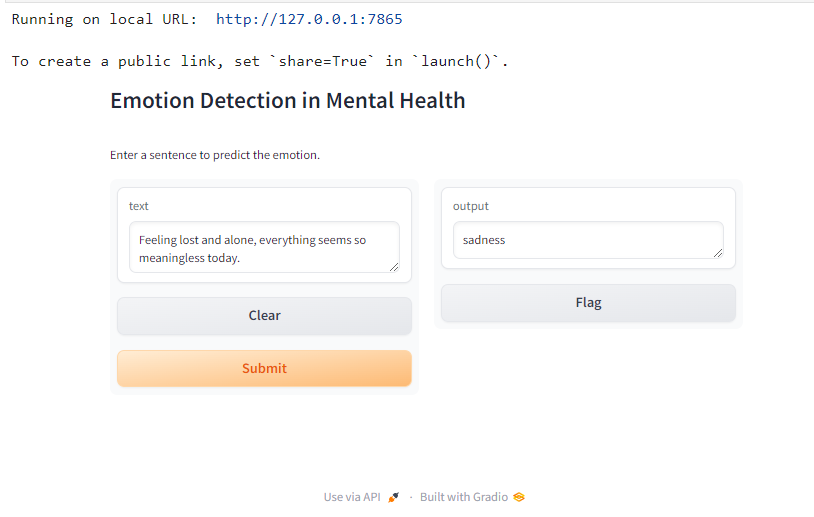
DistilBERT demonstrated a strong balance between performance and efficiency, making it ideal for real-time applications. Its faster processing time compared to BERT ensures that it can be deployed in real-world settings without sacrificing too much accuracy.

1. **Model Comparison and Deployment**

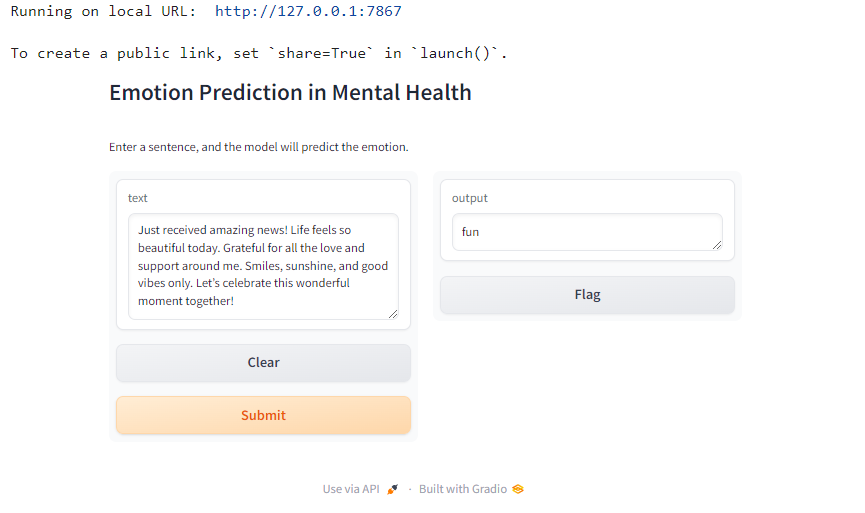
The deployment of all emotion detection models was facilitated through **Gradio**, a user-friendly interface for testing and comparing model performance.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Accuracy | F1-Score | Deployment Characteristics |
| BERT | 35% | 28% | Best suited for tasks requiring high accuracy and nuanced understanding of emotions, especially for depression detection. Computationally expensive, making it ideal for cloud-based or high-power server environments. |
| DistilBERT |  |  | A lightweight alternative to BERT, offering faster inference with only a slight drop in performance. Ideal for real-time applications where computational efficiency is paramount, such as chatbots or mobile applications. |
| Emotion Classifier | 30% | 29% | It is easy to train and deploy. While it lacks the depth of other models, its simplicity makes it a good choice for prototyping and small-scale applications. |

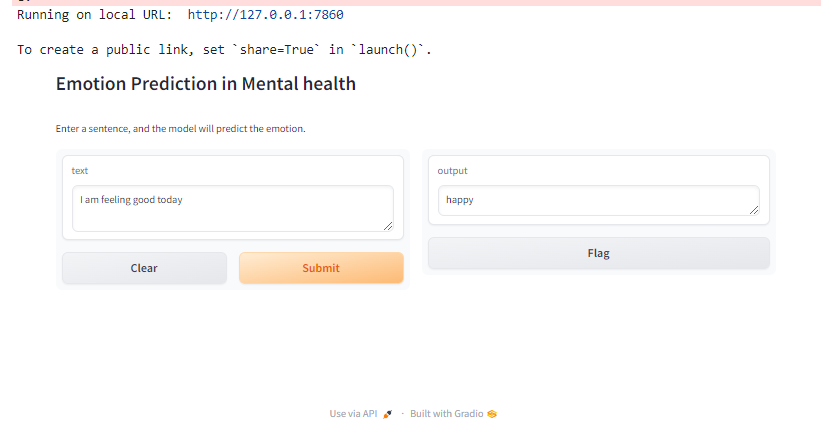
* 1. **Emotion Classifier**

****

* 1. **DistilBERT Deployment**

****

* 1. **BERT Deployment**

****

**Conclusion**

This study underscores the potential of Natural Language Processing (NLP) techniques in identifying mental health conditions such as depression through textual data. By leveraging social media platforms like Twitter and Reddit, the analysis captures linguistic patterns associated with emotions, offering an innovative approach to early mental health detection. Advanced models such as CNN-LSTM and transformer-based architectures like BERT and DistilBERT demonstrate promising results in capturing emotional nuances, although challenges remain. The study reveals that data imbalance, inherent in real-world datasets, significantly hampers model performance, particularly for underrepresented emotions. Additionally, while BERT-based models exhibit superior contextual understanding, their computational requirements make them less feasible for real-time applications compared to lightweight alternatives like DistilBERT. Simpler models, such as the Emotion Classifier, struggle to generalize effectively, further highlighting the complexity of emotion detection.

Future research should focus on addressing these limitations by adopting better preprocessing methods, exploring ensemble architectures, and leveraging multimodal data (text, images, and audio) for richer context. These improvements could enhance prediction accuracy and facilitate the development of scalable tools for mental health monitoring. Ultimately, NLP-based approaches hold the promise of complementing traditional diagnostics, fostering timely interventions, and advancing mental health care globally.

**References**[1] A. Li, D. Jiao, and T. Zhu, “Detecting depression stigma on social media: A linguistic analysis,” Journal of Affective Disorders, vol. 232, pp. 358–362, 2018.

[2] L. Squarcina, F. M. Villa, M. Nobile, E. Grisan, and P. Brambilla, “Deep learning for the prediction of treatment response in depression,” Journal of Affective Disorders, vol. 281, pp. 618–622, 2021.

[3] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, “Natural language processing applied to mental illness detection: A narrative review,” NPJ Digital Medicine, vol. 5, no. 1, pp. 1–13, 2022.

[4] A. Leis, F. Ronzano, M. A. Mayer, L. I. Furlong, and F. Sanz, “Detecting signs of depression in tweets in Spanish: Behavioral and linguistic analysis,” Journal of Medical Internet Research, vol. 21, no. 6, p. e14199, 2019.

[5] L. S. Jones, E. Anderson, M. Loades, R. Barnes, and E. Crawley, “Can linguistic analysis be used to identify whether adolescents with a chronic illness are depressed?,” Clinical Psychology & Psychotherapy, vol. 27, no. 2, pp. 179–192, 2020.

[6] A. Picardi et al., “A randomized controlled trial of the effectiveness of a program for early detection and treatment of depression in primary care,” Journal of Affective Disorders, vol. 198, pp. 96–101, 2016.

[7] K. Rost, J. L. Smith, and M. Dickinson, “The effect of improving primary care depression management on employee absenteeism and productivity: A randomized trial,” Medical Care, vol. 42, no. 12, pp. 1202–1210, 2004.

[8] Statista, “Number of worldwide social network users,” 2023. [Online]. Available: https://www.statista.com/statistics/278414/number-of-worldwide-socialnetwork-users/. [Accessed: 08-Dec-2024].

[9] DataReportal, “Social media users,” 2023. [Online]. Available: <https://datareportal.com/social-media-users>. [Accessed: 08-Dec-2024].

[10] A. Dhand, D. A. Luke, C. E. Lang, and J. M. Lee, “Social networks and neurological illness,” Nature Reviews Neurology, vol. 12, no. 10, pp. 605–612, 2016.