MULTIMODAL SENTIMENTAL ANALYSIS: DETECTING DEPRESSION FROM SOCIAL MEDIA

Team Member 1: Arda Kabadayi
NetID: akabaday
Team Member 2: Mike Chuvik
NetID: mochuvik

Introduction

Especially in today's world, depression is an important global health issue that affects millions of lives and often remains unnoticed and untreated. According to the World Health Organization (WHO), approximately 300 million people worldwide suffered from depression in 2015 which makes it one of the biggest health problems globally. Also, tragically around 800,000 lives are lost to suicide annually and many of them are linked to untreated depression (WHO, 2017). This terrifying situation highlights the urgent need of accurate diagnosis and early detection because misdiagnosis and late treatment can even worsen the disorder, diminishing an individual's quality of life and increasing economic and social problems.

The emergence of social media gives us a unique opportunity to address this issue because these platforms contain a lot of data, in both textual and visual way, that contains users' thoughts, emotions and social interaction. By analyzing the users' patterns in these social networks, researchers can identify early signs of depression, often before individuals pursue medical assistance. Additionally, early detection can reduce the severity of the condition.

As a result, depression is a critical health issue that requires effective solutions, and social media appears as a valuable resource to solve this problem. To contribute this, our plan is going to be, by using the recent advancements in both Natural Language Processing(NLP) and Computer Vision(CV), developing complex models which are efficient in detecting depression from social media.

Description of the Project (with Methodology)

As we said, this study focuses on detecting depression using multimodal social media data. Our data include users' tweets, profile pictures, and banners from Twitter. As today depression detection has mostly relied on traditional machine learning models and just textual data, but our approach plans to improve this by using advanced models and multimodal data.

Why are we planning to use advanced models?

Previous studies have often relied on traditional ML techniques, but this approach can be insufficient for such an important task. Because of this importance and commonality even a minor improvement in accuracy is significant. Especially decreasing the number of false-negative patients, the patients that have depression but it is not detected correctly, plays a huge role. By using advanced models, we prioritize accuracy over computational cost to address this issue effectively. So, we aim to show their capability to surpass traditional methods.

Why are we planning to use multimodal data?

Just relying on textual data can result in misleading interpretations because languages sometimes lack sufficient context. For example, a user with a profile picture of a devil and a tweet saying, "I like angels a lot 😂," might seem positive if we only consider the text. However, combining textual data with visual data from the profile picture reveals deeper allusions that are critical for understanding the user's mental state. As a result, this research uses both text and image data to capture these subtle meanings to ensure a comprehensive analysis.

Methodology

The methodology of this project starts with an evaluation of the individual effectiveness of computer vision models such as VGGNet and ResNet on visual data, including user profile pictures and banners. Simultaneously, we will evaluate the performance of advanced natural language models such as BERT and GPT on text data that contains user tweets. This phase aims to determine which models perform best with their respective modalities. Once we identify the optimal models, we will fine-tune them to get better accuracy. The next step will be integrating the features extracted from these visual and textual models into a unified representation. By combining the strengths of both modalities, our system will gain a broader understanding of user behavior, which is important for detecting depression. A binary classification model is then applied to this multimodal feature set to classify users as either showing or not showing symptoms of depression.

Objectives of the Project

We mostly talked about the objectives of the project in the description section but to make them clearer our 2 main objectives are outlined below:

1) Firstly, our goal is to show that the use of advanced models, especially transformer-based models, can lead to better performance compared to traditional models.

2) Secondly, by combining multiple data sources, such as text and images, we aim to demonstrate that multimodal models offer advantages over unimodal models.

Literature Review(with supporting references)

When we looked at the literature most of the work done about this field uses various basic machine learning techniques such as support vector machines(svm), decision trees and random forest. While few studies explore advanced models like Bert; however, these are just relying on textual data and do not consider visual data. Even fewer studies investigate the use of multimodal data combined with advanced models, leaving a significant gap in the research. So, we would like to contribute the field to fill that gap. Below, we have listed five supporting and influential papers that have guided our understanding and approach."

1) **A deep learning model for detecting mental illness from user content on social media:** This paper not just explore the detection of depression but also different mental health disorders as well such as anxiety, bipolar disorder, borderline personality disorder, schizophrenia, and autism. They collected their data using Reddit that includes numerous mental health related communities (or so-called subreddits), such as r/depression, r/bipolar, and r/schizophrenia. They utilized from 6 different subreddit and for each subreddit they developed six independent binary classification models. To classify, XGBoost and convolutional neural network (CNN) were employed. Their methodology involved first creating word embeddings using the Word2Vec model and then supplying these embeddings as inputs to convolutional layers in CNNs. After all they found out that this approach gives higher classification accuracy for CNN models compared to XGBoost across all the targeted subreddits which demonstrates the potential of CNNs in effectively identifying mental health symptoms based on social media text data. For us this paper is very influential because it demonstrates the potential of CNNs in effectively identifying mental health symptoms based on social media text data.[1]

2) **Harnessing the Power of Hugging Face Transformers for Predicting Mental Health Disorders in Social Networks:** This paper focuses on increasing the accuracy of detecting depression by using advanced transformer models(BERT), to address the limitations of traditional diagnostic processes. It compares four BERT models from Hugging Face with standard machine learning techniques such as support vector machines and decision trees. The researchers used Twitter users' tweets and bios for depression prediction and preprocess the data by removing hashtags and emojis. However, we think that this decision is debatable because these elements may

[1] Kim, Jina, et al. "A deep learning model for detecting mental illness from user content on social media." *Scientific reports* 10.1 (2020): 11846.

provide valuable insight into the data. Their findings demonstrated over a 6% improvement in metrics compared to prior research. This emphasizes the potential of pre-trained transformer models to detect mental health problems effectively. Finally, the study concludes that social media data is a valuable resource for identifying mental health issues and that transformer-based models offer a significant advancement in this field. This paper is influential because this is one of the first papers that shows advanced models can outperform traditional models in this field.[2]

3) **Exploring emotional patterns in social media through NLP models to unravel mental health insights:** This paper aims to create an ensemble framework for the classification of mental health disorders in social media posts. It utilizes the Reddit mental health dataset. In this case the problem is not a binary classification problem (indicating whether there is a symptom or not) but a multi-class classification problem, where each post is assumed to belong to one of 15 mental health categories, such as alcoholism, addiction, and autism. By using a voting ensemble technique, the study combines predictions from fine-tuned transformer models, including XLNet, RoBERTa, and ELECTRA. It shows that this ensemble technique achieves a better accuracy of 0.780 which surpasses the individual model performances (XLNet: 0.767, RoBERTa: 0.775, and ELECTRA: 0.755). So, the findings emphasize the effectiveness of ensemble approaches in advancing mental health disorder detection from social media data. This is influential because this is the only paper we saw that uses ensemble techniques.[3]

4) **Cooperative Multimodal Approach to Depression Detection in Twitter:** This study uses both advanced models and multimodal data which is a novel approach compared to traditional methods that rely solely on textual information. In addition to this novelity, the authors propose a multi-agent reinforcement learning framework called COMMA, where two policy gradient agents collaboratively select relevant text and image features. The extracted features are then used to evaluate the classification accuracy for depression detection. The methodology uses gated recurrent units (GRU) for text feature extraction and convolutional neural networks (CNN) for image feature extraction. COMMA's effectiveness is demonstrated through experimental results on a depression benchmark dataset that shows significant improvements in performance compared to traditional methods such as Naive Bayes. This work highlights the utility of combining textual and visual modalities for mental health analysis. This is influential because this is the only paper we saw that uses reinforcement learning.[4]

5) **Advancing Depression Detection in Social Media: A Multimodal Aspect-Level Sentiment Analysis Approach:** This paper explores advancing depression detection

---

[2] Pourkeyvan, Alireza, Ramin Safa, and Ali Sorourkhah. "Harnessing the power of hugging face transformers for predicting mental health disorders in social networks." *IEEE Access* 12 (2024): 28025-28035.

[3] Shetty, Nisha P., et al. "Exploring emotional patterns in social media through NLP models to unravel mental health insights." *Healthcare Technology Letters* (2025).

[4] Gui, Tao, et al. "Cooperative multimodal approach to depression detection in twitter." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. No. 01. 2019.

through multimodal aspect-level sentiment analysis. The study evaluates various pre-trained models, with TomBERT (aspect-level, multimodal) achieving the highest performance, with an accuracy of 89.67% and an F-score of 89.58% which is followed by mBERT (non-aspect-level, multimodal) and BERT (aspect-level, unimodal). The results emphasize the power of multimodal approaches in sentiment analysis and depression detection by showing the potential of models like TomBERT to provide more accurate insights than unimodal methods. This paper describes things close to what we are wanting to do. [5]

Evaluation Metrics

In our case we want to compare traditional model's vs unimodal model's vs multimodal models. When we looked at the literature, we say that always the same metrics are used and these ones are accuracy, precision, recall and F-1 score. We will also use these metrics to evaluate our idea. You can see the description of these techniques below.

1) **Accuracy:** It measures the proportion of correct predictions (both true positives and true negatives) out of the total predictions made. It gives an overall idea of how well the model performs but it may give wrong insight if the data is not balanced. Because of this reason we also need to use other metrics as well.

   Accuracy = (TP + TN) / (FP + FN + TP + TN)

   Where: TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives

2) **Precision:** It measures the proportion of correctly predicted positive instances out of all instances predicted as positive. It focuses on the quality of positive predictions.

   Precision = TP / (FP + TP)

3) **Recall(Sensitivity):** It measures the proportion of correctly predicted positive instances out of all actual positive instances. It indicates how well the model captures true positives.

   Recall = TP / (FN + TP)

[5] Liu, Yuxin. "Advancing Depression Detection in Social Media: A Multimodal Aspect-Level Sentiment Analysis Approach." *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*. Vol. 185. Springer Nature, 2024.

4) **F1-Score:** It is the harmonic means of precision and recall. It provides a balance between the two metrics, especially useful when there is an imbalance between classes.

F1-Score = 2 × Precision × Recall / (Precision + Recall)