Multimodal Analysis of Instagram Posts to Track Depression and Mood Progression

Sujith Thota
Master's in Computer Science
IDS 596: Independent Study in Information and Decision Sciences
University of Illinois Chicago
UIN: 664966783
Email: sthota@uic.edu

***Abstract*—This study investigates the use of multimodal analysis to detect and track behavioral signals of depression through Instagram posts. By leveraging the capabilities of the Gemini API and large language models (LLMs), we extract and analyze visual, textual from social media content created by individuals diagnosed with depression. These features are aligned with clinical indicators from the PHQ-9 questionnaire, enabling the identification of emotional and psychological cues embedded in posts. The proposed pipeline involves extracting multimodal features, filtering emotionally intense posts, validating them through clustering, and summarizing mood progression over time using an LLM. The resulting system offers a novel, AI-driven approach to monitoring mental health trends on social media platforms and generating personalized mood trajectories.**

I. Introduction

Depression is one of the most prevalent mental health disorders worldwide, affecting over 280 million people according to the World Health Organization. Early detection and continuous monitoring are crucial for effective intervention, yet traditional methods rely heavily on self-reporting and clinical assessments, which may not capture day-to-day emotional fluctuations. In recent years, the rise of social media platforms has created an opportunity to observe subtle behavioral patterns that could serve as digital biomarkers for mental health conditions.

Instagram, in particular, is a widely used, visual-centric platform where users regularly share aspects of their personal lives through photos, videos, and captions. These posts often reflect users' moods, routines, social engagement, and overall mental well-being. As such, Instagram offers a rich and underutilized source of data that can reveal changes in emotional state over time—especially in individuals experiencing psychological distress.

This study aims to develop an end-to-end, automated pipeline that can analyze Instagram content to detect depression-related signals and summarize mood progression. Leveraging recent advances in multimodal artificial intelligence, we use the Gemini API to extract emotional and semantic features from image, video, and text content. These features are then mapped to clinical categories based on the PHQ-9 depression questionnaire, a widely used tool for screening depressive symptoms.

Our goal is to generate interpretable, user-specific mood trajectories that reflect the evolution of emotional state over time. This not only supports retrospective mental health analysis but also opens up possibilities for early warning systems and personalized intervention tools, bridging the gap between passive digital observation and active mental health care.

II. Dataset and Tools

The dataset contains Instagram posts from approximately 75–78 individuals diagnosed with depression. Each user's data includes:

- Images and videos
- Captions and embedded text

We selected Google's Gemini API for multimodal feature extraction due to its ability to handle images, videos, and text together, and for being a cost-effective alternative to other APIs like ChatGPT.

III. Methodology

To detect depression-related cues and track mood progression, we developed a five-stage pipeline combining multimodal feature extraction, PHQ-9 scoring, post filtering, clustering validation, and temporal mood summarization. Each stage is described below.

A. Feature Extraction Using Gemini API

The first step involves extracting multimodal features from Instagram posts. Each post may contain images, videos, captions, and sometimes embedded or transcribed text. We used the **Gemini API**, a vision-language model developed by Google, which can process image, video, and textual inputs simultaneously. This API enables scalable and cost-effective extraction of emotional and contextual signals from user posts.

For each post, the following features were extracted:

- **Visual features**: Emotion expressed in faces, head tilt, presence of tears, eye redness, use of filters, brightness, contrast, saturation, presence of animals or people, and scene composition.
- **Textual features**: Sentiment polarity of captions, embedded text detected in images, and keyword analysis based on PHQ-9 symptom terms.

In total, over 80 features were computed per post, resulting in a rich multimodal representation suitable for psychological analysis.

B. PHQ-9 Scoring and Depression Indicator Mapping

To translate multimodal features into clinical insight, each post was evaluated against the **PHQ-9** symptom categories. These include:

- Loss of interest
- Feeling down or depressed
- Sleep disturbances
- Fatigue or low energy
- Eating issues
- Low self-esteem
- Concentration difficulty
- Psychomotor changes
- Thoughts of self-harm

For each category, the model assigned both **binary (yes/no)** and **multiclass (none, mild, moderate, severe)** labels to reflect the intensity of the symptom as inferred from the post content. This mapping was guided by heuristic rules and keyword sets derived from PHQ-9 definitions, adapted for social media language.

C. Filtering Emotionally Intense Posts

Not all posts in a user's timeline are emotionally significant or relevant to depression detection. Therefore, we filtered posts by selecting those with high PHQ-9 similarity scores across any of the nine categories. This step reduced noise and allowed the subsequent analysis to focus only on those posts most likely to carry meaningful emotional content.

This filtering stage is crucial to avoid misleading interpretations caused by neutral or off-topic posts (e.g., advertisements, memes, or celebratory content).

D. Clustering for Validation of Filtering

To validate the emotional filtering process, we applied **unsupervised clustering algorithms** (e.g., KMeans) on the filtered set of posts. The goal was to check whether emotionally intense posts naturally form distinct clusters in feature space.

These clusters were analyzed to confirm whether they corresponded to shared emotional themes or PHQ-9 symptom patterns. For example, posts with negative caption sentiment, darker color palettes, and isolated subjects often grouped together, validating the filtering logic.

E. Mood Progression Modeling via LLMs

The final stage involved modeling the temporal progression of each user's mood. For every user, the filtered and chronologically sorted posts were passed to a **large language model (LLM)** prompt that summarized the emotional evolution over time. The input included:

- Key image/video features
- Caption sentiment trends
- PHQ-9 scores per post

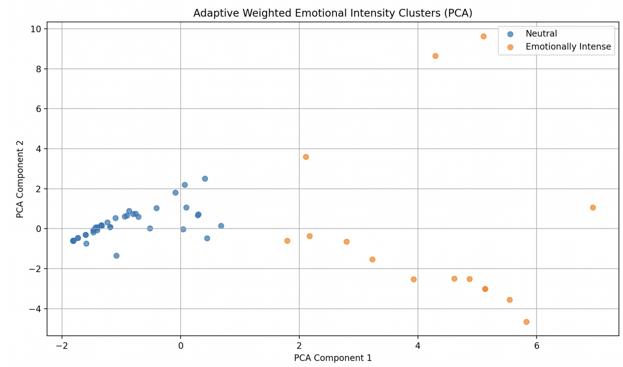

Fig. 1. KMeans clustering of emotionally intense Instagram posts. Clusters show grouping based on PHQ-9-related visual and textual features for a profile.

F. Prompt Design for Mood Summarization

To enable mood trajectory generation from emotionally intense posts, we designed a detailed prompt for the large language model (LLM). This prompt provides the model with image/video sentiment, embedded text, caption content, and timestamp information, guiding it to analyze temporal emotional progression and causality.

The prompt used is as follows:

> **Your Task:**
> Analyze the image/video, caption, and embedded text in conjunction with their timestamps, and perform the following:
>
> - Summarize the user's mood course and temporal emotional pattern over time.
> (Categorize each period using primary mood states: Happiness, Sadness, Fear, Disgust, Anger, Surprise.)
>
> - Describe the progression and evolution of emotional states over time, identifying potential trajectories of emerging or worsening mental illness if present.
>
> - Provide an overall analysis based on all features combined — including caption content, visual sentiment, embedded text, and notable emotional shifts across time.
> (Summarize the mood and emotional states into meaningful time periods, dynamically selecting time ranges based on observed significant life events. Emphasize clinically relevant topics such as health issues, diagnoses, loss-related events, or major life transitions.)
>
> - Additionally, model the emotional causality over time:
> Recognize that emotional states at a later time may be influenced by earlier impactful events.

For example: a health diagnosis or major life stressor at timestamp T1 may continue to affect the user's mood at timestamps T5, T6, and beyond, even if not explicitly mentioned.
Connect earlier events to later emotional shifts, building an implicit causal graph of the user's emotional journey.

- Identify and describe any emotional echoes or delayed emotional reactions, where earlier events resurface or influence future mood patterns.

```
          You are an AI psychiatrist
analyzing a user's Instagram posts
longitudinally over time.
Each post contains a Timestamp,
Image/Video content, Embedded Text,
and Caption.
```

Your Task:

```
Analyze the image/video, caption,
and embedded text in conjunction
with their timestamps, and perform
the following:

- Summarize the user's mood course
and temporal emotional pattern over
time.
(Categorize each period using
primary mood states: Happiness,
Sadness, Fear, Disgust, Anger,
Surprise.)

- Describe the progression and
evolution of emotional states
over time, identifying potential
trajectories of emerging or
worsening mental illness if
present.

- Provide an overall analysis
based on all features combined |
including caption content, visual
sentiment, embedded text, and
notable emotional shifts across
time.
(Summarize the mood and emotional
states into meaningful time
periods, dynamically selecting
time ranges based on observed
significant life events. Emphasize
clinically relevant topics such
as health issues, diagnoses,
loss-related events, or major life
transitions.)

  - Additionally, model the emotional
  causality over time:
  Recognize that emotional states at
  a later time may be influenced by
  earlier impactful events.
  For example: a health diagnosis or
  major life stressor at timestamp T1
  may continue to affect the user's
  mood at timestamps T5, T6, and
  beyond, even if not explicitly
  mentioned.
  Connect earlier events to later
  emotional shifts, building an
  implicit causal graph of the user's
  emotional journey.

  - Identify and describe any
  emotional echoes or delayed
  emotional reactions, where earlier
  events resurface or influence
  future mood patterns.
```

This prompt was fed with chronologically sorted post summaries (caption, embedded text, and visual sentiment) per user.

The LLM then generated a short narrative describing the user's mood trajectory across time. For example, it might highlight a shift from social, vibrant posts to solitary, emotionally negative content, reflecting a potential onset of depression.

This summarization allowed for intuitive, personalized insight into mental health trends, providing a bridge between raw data and clinically relevant interpretations.

IV. Results and Observations

The proposed pipeline was successfully implemented to extract, analyze, and summarize emotional patterns from Instagram posts using multimodal AI techniques. The following observations highlight the effectiveness of the system.

A. Multimodal Feature Extraction

Using the Gemini API, we were able to extract meaningful features from all available modalities — including images, videos, captions, and embedded text. These features included visual sentiment indicators (e.g., facial expressions, lighting, color palette), audio sentiment (for videos), and textual cues from both captions and image-embedded text. The average number of features per post exceeded 80, offering a rich representation for psychological analysis.

B. PHQ-9-Based Filtering and Clustering

Posts were successfully scored against PHQ-9 depression indicators in both binary and multiclass formats. This enabled the filtering of emotionally intense posts, eliminating irrelevant or emotionally neutral content. The filtered posts were further validated using KMeans clustering, where clusters revealed coherent emotional themes — such as sadness with low brightness and isolation, or happiness with group photos and vibrant colors.

C. Mood Progression Generation via LLM

Filtered posts were sorted chronologically for each user and passed into a large language model (LLM) with a structured prompt designed to simulate psychiatric analysis. The model generated mood summaries that captured emotional trajectories over time, including the influence of prior events on future emotional states.

D. Example Summary: Crystal Parlove89

One user, identified as *crystal_parlove89*, exhibited a rich and evolving emotional journey, as captured by the LLM:

- **Period 1 – Pre-New Year:** Signs of sadness with emerging hope. Posts reflect depression and anxiety, with captions indicating efforts at self-care.
- **Period 2 – New Year's Eve:** Mixed emotions, combining sadness, anger, and optimism. Grief over her mother's death is evident.
- **Period 3 – January Slump:** Return of sadness, self-care actions like makeup are framed as efforts to improve mood.
- **Period 4 – Medication Conflict:** Anger and fear arise from conflict with psychiatric treatment and physical health decline.
- **Period 5 – Post-Hospitalization:** Fluctuating mood with signs of resilience. Mentions of fatigue and attempts to regain control via public engagement.
- **Period 6 – Anniversary Grief:** Intense sadness resurfaces due to emotional echoes around her mother's death anniversary.

E. Emotional Causality and Longitudinal Insights

The model was also able to detect emotional causality — linking major life events (e.g., the death of the user's mother or a hospital stay) with delayed emotional reactions. These insights offer potential for building emotional timelines or causal graphs that aid psychological interpretation.

F. Overall Findings

The pipeline was effective in:

- Mapping multimodal post content to PHQ-9 indicators.
- Identifying emotionally intense periods in user timelines.
- Generating interpretable summaries of mood evolution.
- Revealing connections between early traumatic events and long-term emotional consequences.

This system demonstrates the feasibility of using social media and LLMs for passive, non-invasive mental health monitoring and highlights the promise of AI in supporting digital psychiatry.

V. Limitations and Future Work

- Social media posts are curated and may not reflect true mental states.
- Some posts lacked audio or descriptive captions.
- PHQ-9 detection relies on heuristic mapping, not clinical assessment.

Future work includes:

- Use of time-series models (e.g., LSTM) for dynamic mood prediction
- Clinical validation of AI-generated summaries
- Extension to other platforms like TikTok and Twitter

VI. Code and Repository

All code, processed data, and result summaries associated with this study are available in a public GitHub repository for transparency and reproducibility.

- **Repository:** https://github.com/Sujith-Nihar/Depression-detection-based-on-instagram-posts

The repository includes:

- Python scripts for multimodal feature extraction using the Gemini API
- PHQ-9 scoring and emotional filtering code
- Clustering notebooks for post-validation
- Prompt design and LLM-based mood summarization scripts
- Sample mood summaries and analysis reports per user

The repository ensures that the entire methodology—from feature extraction to mood progression generation—can be reviewed, reused, or extended by the research community.

References

[1] R. Chandrasekaran, S. Kotaki, and A. H. Nagaraja, "Detecting and tracking depression through temporal topic modeling of tweets: insights from a 180-day study," *npj Mental Health Research*, vol. 3, no. 62, 2024. Available: https://doi.org/10.1038/s44184-024-00107-5

[2] S. Yadav, C. Caragea, C. Zhao, N. Kumari, M. Solberg, and T. Sharma, "Towards Identifying Fine-Grained Depression Symptoms from Memes," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Toronto, Canada, Jul. 2023, pp. 8890–8905. Available: https://aclanthology.org/2023.acl-long.495

[3] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The PHQ-9: Validity of a brief depression severity measure," *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, Sep. 2001.