

Mental Health Meme Classification

NLP Project Proposal

Group-12

Manan Aggarwal 2022273 manan22273@iiitd.ac.in	Shobhit Raj 2022482 shobhit22482@iiitd.ac.in	Souparno Ghose 2022506 souparno22506@iiitd.ac.in
--	---	---

1 Problem Statement

Memes have become a popular medium for expressing emotions on social media, often reflecting mental health struggles like Anxiety and Depression.

This project focuses on the following classification tasks:

- 1. Depression Symptom Classification (Multi-Label):** Identifies one or more depression symptoms per meme using the RESTORE dataset (Yadav et al., 2023).
- 2. Anxiety Classification (Single-Label):** Assigns each meme a single anxiety-related category using the AxiOM dataset (Mazhar et al., 2025).

To ensure robust evaluation, we use macro-F1 and weighted-F1 scores, assessing overall performance and handling class imbalance effectively.

2 High-Level Plan

Our approach follows a structured pipeline to process and classify meme content:

- 1. OCR and Pre-processing:** Extract text from meme images and apply necessary cleaning.
- 2. Figurative Reasoning:** Infer contextual meaning from images in textual form.
- 3. Textual Representation:** Encode extracted text using a BERT-based Transformer for rich contextual embeddings.
- 4. Visual Embedding:** Extract visual features from memes using a ViT-based Encoder.
- 5. Multimodal Attention Mechanism:** Assign importance to textual, visual, and figurative reasoning features.
- 6. Final Classification Model:** Pass embeddings for category prediction.

3 Approach

- 1. OCR and Pre-processing:** Extract text via Google Docs OCR using Google App Script, followed by removal of misaligned classes.
- 2. Figurative Reasoning:** Use LLaVA (Liu et al., 2023) to generate textual interpretations of meme context.
- 3. Textual Representation:** Process OCR and figurative reasoning text via a BERT-based Transformer for contextual embeddings.
- 4. Visual Embedding:** Extract spatial and semantic features using a ViT-based Encoder.
- 5. Multimodal Attention Mechanism:** Compute attention weights to emphasize key features across text, image, and figurative reasoning inputs.
- 6. Final Classification Model:** Concatenate multimodal embeddings and classify using a Transformer-based Classifier, trained separately on AxiOM and RESTORE datasets.

References

- Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2023. [Visual instruction tuning](#). *Preprint*, arXiv:2304.08485.
- Abdullah Mazhar, Zuhair hasan shaik, Aseem Srivastava, Polly Ruhnke, Lavanya Vaddavalli, Sri Keshav Katragadda, Shweta Yadav, and Md Shad Akhtar. 2025. [Figurative-cum-commonsense knowledge infusion for multimodal mental health meme classification](#). *Preprint*, arXiv:2501.15321.
- Shweta Yadav, Cornelia Caragea, Chenye Zhao, Naincy Kumari, Marvin Solberg, and Tanmay Sharma. 2023. [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)*, pages 8890–8905, Toronto, Canada. Association for Computational Linguistics.