Mental Health Meme Classification

Abhilash Jacob Mathew
IIIT Delhi
New Delhi, India
abhilash24006@iiitd.ac.in

Dhruvkumar Patel
IIIT Delhi
New Delhi, India
dhruvkumar24032@iiitd.ac.in

Harsh Pala
IIIT Delhi
New Delhi, India
harsh24119@iiitd.ac.in

Abstract

Mental health conditions such as depression and anxiety are often expressed through internet memes, making automated classification an important step in understanding and addressing online mental health discourse. This project focuses on the classification of mental health symptoms in memes by leveraging extracted textual content. The dataset consists of two tasks: **Depression Multi-Label Classification**, where each meme can belong to one or more of seven predefined depression categories, and **Anxiety Single-Label Classification**, where each meme is assigned one of eight predefined anxiety categories.

1 Introduction

The rise of social media has revolutionized the way people communicate and express emotions, particularly among younger populations. A unique phenomenon emerging from this cultural shift is the use of memes as a medium for discussing personal mental health struggles. These often humorous yet deeply emotional images offer a lens into the collective psyche of online communities, making them valuable resources for analyzing psychological states such as anxiety and depression.

The subtlety and ambiguity inherent in meme language pose challenges to traditional Natural Language Processing (NLP) and computer vision models. This project addresses these challenges by exploring the classification of mental health symptoms expressed in memes through multimodal analysis—focusing on depression and anxiety classification using both textual (OCR) and visual (image) features.

Our system performs two tasks: **multi-label classification for depression symptoms** and **single-label classification for anxiety symptoms**, using a novel hybrid approach combining contextual reasoning, retrieval-augmented generation, and vision-language fusion.

2 Related Work

Our work builds heavily on the research titled *Figurative-cum-Commonsense Knowledge Infusion for Multimodal Mental Health Meme Classification* (Mazhar et al., 2025). The paper introduces the AxiOM dataset and the M3H model, emphasizing the need for commonsense reasoning and figurative understanding in classifying mental health symptoms. Prior efforts in meme analysis have largely focused on harmful content or humor detection, often ignoring the emotional and psychological layers that memes express.

M3H leverages GPT-4o for figurative reasoning and incorporates Retrieval-Augmented Generation (RAG) to improve model comprehension. The study presents the RESTORE dataset for depressive symptom classification and AxiOM for anxiety symptoms, setting a precedent for hybrid approaches combining vision and language models. Our research extends this line of work by experimenting with different models and proposing an advanced fusion model—HyCore-M3Net—that balances grounded visual understanding with high-level contextual reasoning.

3 Methodology

We experimented with three distinct models:

Model 1: OCR + BERT (Baseline)

This model uses only the OCR-extracted text from memes as input. The text is processed using the EasyOCR toolkit, and then passed into a pretrained BERT model for classification. This provides a baseline for understanding how well text alone can represent the psychological state depicted in the meme.

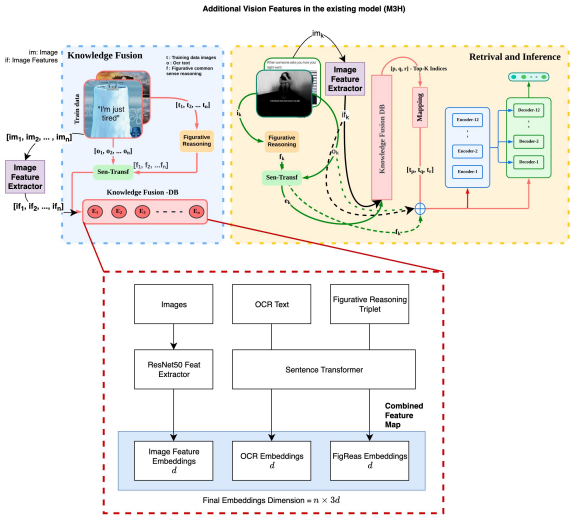


Figure 1: Architecture diagram of Mental-BART + Vision Enhanced Fusion + M3H RAG model.

Model 2: Mental-BART + Vision Enhanced Fusion + M3H RAG

This model integrates three sources of information: OCR text, figurative reasoning, and visual content. The textual content is embedded using the BAAI/bge-m3 sentence transformer. Figurative reasoning is extracted using Qwen2.5-VL-7B-Instruct to derive cause-effect, metaphorical, and emotional representations, which are also embedded. Visual features are extracted via a modified ResNet-50 network. These features are concatenated to form a 3072-dimensional multimodal vector for each meme. This architecture builds on the M3H pipeline and enhances it through vision-enriched reasoning.

Model 3: HyCore-M3Net

The HyCore-M3Net employs a two-stream architecture. The first stream processes OCR text and detected visual regions using LXMERT, leveraging cross-modal attention for grounded interpretation. The second stream uses a MentalBART encoder to process figurative reasoning text along with retrieved examples (via RAG) for enriched context. The outputs of both streams are fused using a Multi-Layer Perceptron. This fusion allows the model to jointly understand explicit visual-text interactions and implicit semantic content.

4 Dataset Description

We utilized two curated datasets for fine-grained mental health symptom classification:

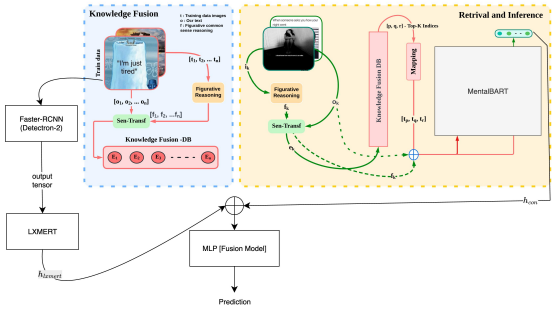


Figure 2: Architecture diagram of the proposed HyCore-M3Net model.

- **RESTORE (Depression):** Based on PHQ-9 categories, includes labels like Self-Harm, Lack of Interest, and Feeling Down.
- **AxiOM (Anxiety):** Based on GAD symptoms, with labels such as Nervousness, Excessive Worry, and Restlessness.

OCR was applied uniformly, with the depressive dataset containing pre-extracted text.

4.1 Dataset Statistics

Train Dataset Statistics (Anxiety):

- **Restlessness:** 463
- **Nervousness:** 424
- **Impending Doom:** 417
- **Difficulty Relaxing:** 407
- **Lack of Worry Control:** 378
- **Excessive Worry:** 368

Test Dataset Statistics (Anxiety):

- **Restlessness:** 115
- **Nervousness:** 106
- **Impending Doom:** 105
- **Difficulty Relaxing:** 102
- **Lack of Worry Control:** 94
- **Excessive Worry:** 92

Depressive Dataset:

4.1.1 Train Dataset (Depression):

- **Feeling Down:** 2085
- **Eating Disorder:** 1939
- **Sleeping Disorder:** 1562
- **Self-Harm:** 1516
- **Low Self-Esteem:** 855
- **Concentration Problem:** 595
- **Lack of Interest:** 471
- **Lack of Energy:** 122

4.1.2 Test Dataset (Depression):

- **Feeling Down:** 218
- **Low Self-Esteem:** 114
- **Eating Disorder:** 92
- **Self-Harm:** 81
- **Sleeping Disorder:** 79
- **Lack of Interest:** 71
- **Concentration Problem:** 66

4.1.3 Validation Dataset (Depression):

- **Feeling Down:** 195
- **Low Self-Esteem:** 85
- **Self-Harm:** 61
- **Lack of Energy:** 54
- **Eating Disorder:** 49
- **Sleeping Disorder:** 45
- **Lack of Interest:** 45
- **Concentration Problem:** 42

4.2 Visualizations

The following visualizations provide an overview of the dataset splits and model performance:

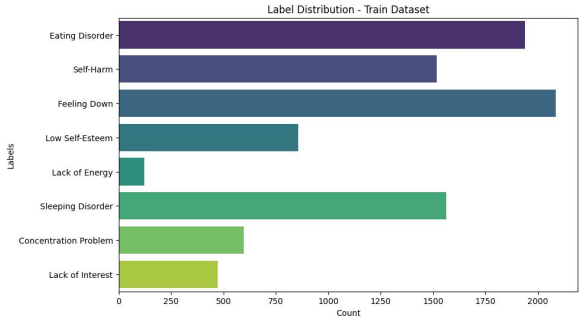


Figure 3: Distribution of the Train, Validation, and Test datasets for the Depression dataset.

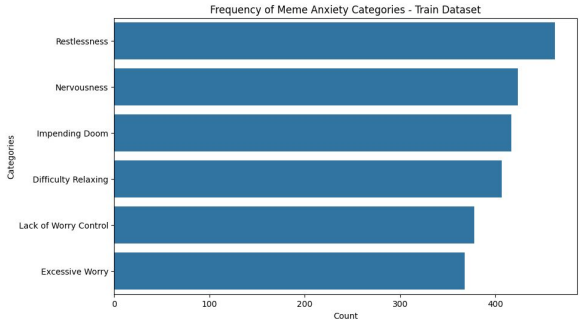


Figure 4: Distribution of the Train and Test datasets for the Anxiety dataset.

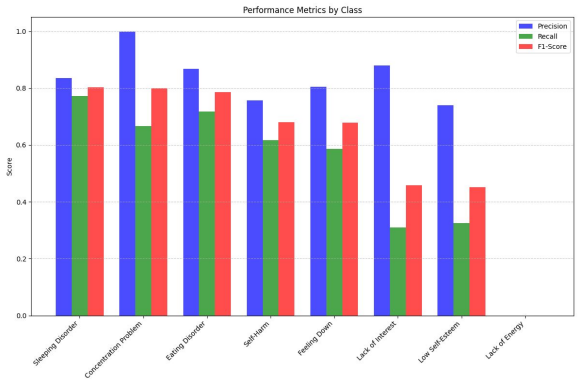


Figure 5: Performance of OCR + BERT on Depression dataset.

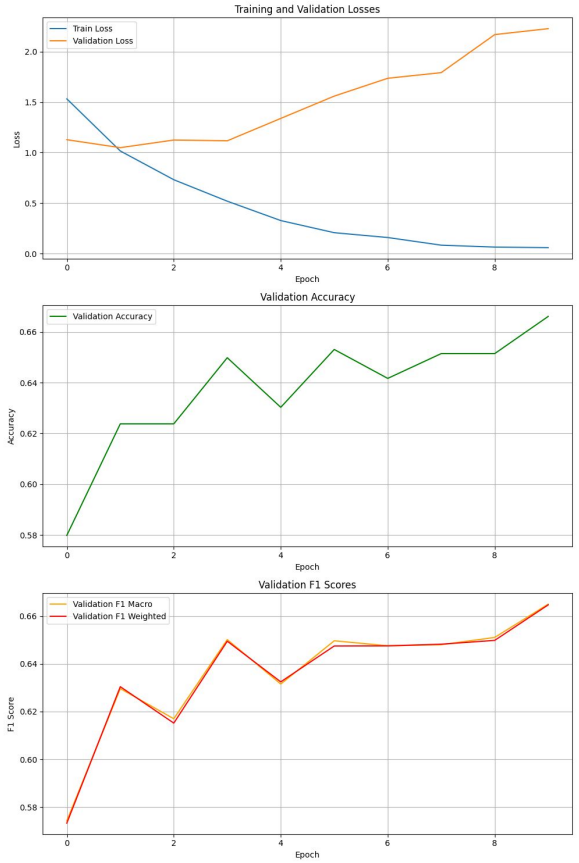


Figure 6: Performance of OCR + BERT on Anxiety dataset.

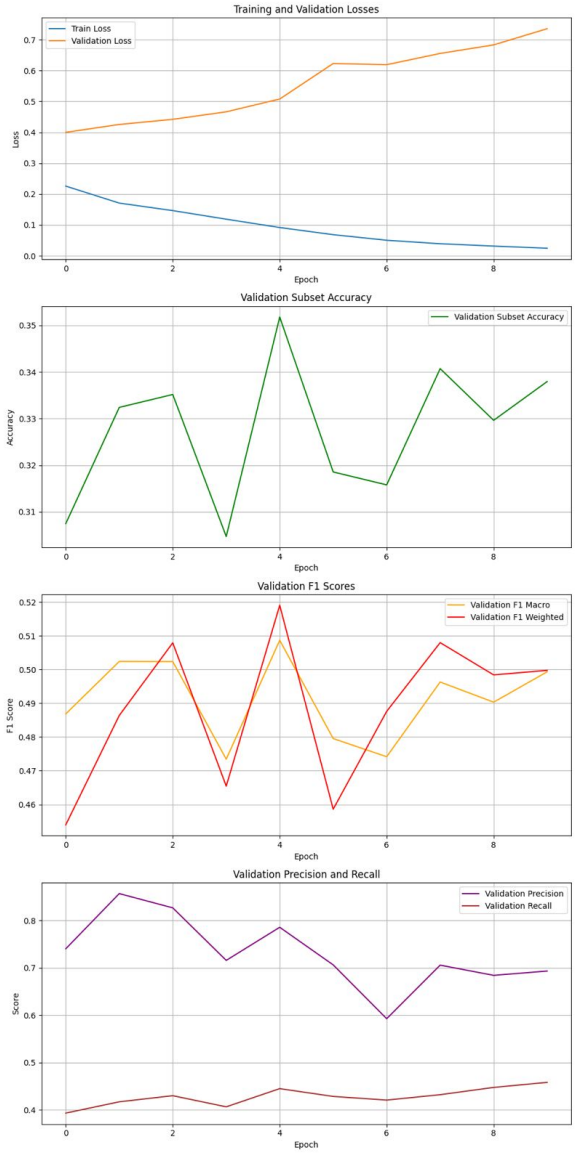
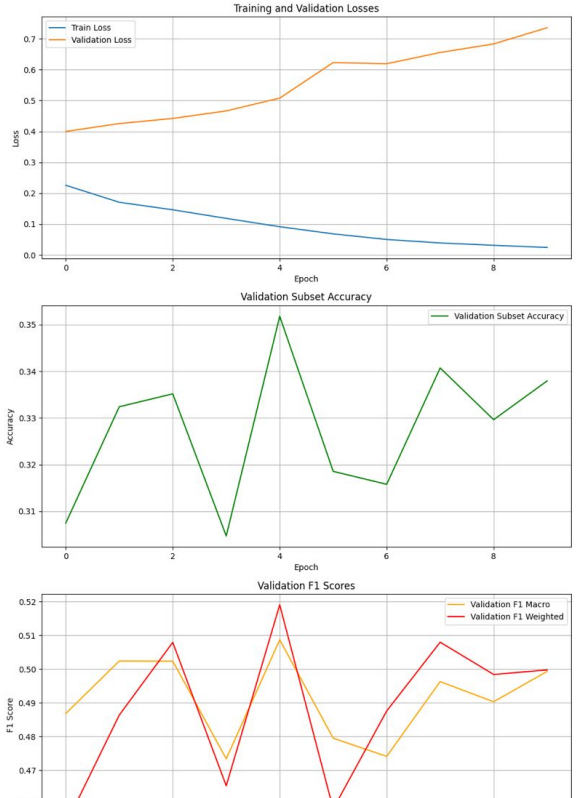


Figure 7: Performance of HyCore-M3Net on Depression dataset.



5 Results

We evaluate the models across two datasets: Depression and Anxiety, using Macro F1 and Weighted F1 scores. The results are as follows:

- **OCR + BERT:**
  - **Anxiety:**
    - Macro F1 = 0.4746
    - Weighted F1 = 0.5548
  - **Depression:**
    - Macro F1 = 0.5527
    - Weighted F1 = 0.6236
- **Mental-BART + Vision Enhanced Fusion + M3H RAG:**
  - **Anxiety:**
    - Macro F1 = 0.6512
    - Weighted F1 = 0.6510
  - **Depression:**
    - Macro F1 = 0.5822
    - Weighted F1 = 0.6597

6 Analysis

The results clearly indicate the benefit of multimodal integration. The baseline text-only model falls short in accurately identifying symptoms due to the lack of visual and figurative cues. Models incorporating figurative reasoning and vision (especially HyCore-M3Net) outperform others significantly. The late-stage fusion in HyCore-M3Net allows complementary information to reinforce semantic understanding. Common label confusion in the anxiety dataset (e.g., Impending Doom vs. Lack of Worry Control) suggests overlap in meme expression patterns.

7 Conclusion

This research highlights the effectiveness of combining textual, visual, and commonsense reasoning for mental health meme classification. The HyCore-M3Net outperforms both traditional and enhanced multimodal baselines by leveraging a hybrid architecture designed for grounded and contextual understanding.

8 Future Work

Future research directions include:

- Expanding datasets across cultural boundaries
- Incorporating generative models for classification explanation
- Embedding emotion and sarcasm detection modules
- Building real-time systems for online mental health surveillance

References

Anees Mazhar and 1 others. 2025. Figurative-cum-commonsense knowledge infusion for multimodal mental health meme classification. *arXiv preprint arXiv:2501.15321*.