# DEPRESSION DETECTION FROM MEME IMAGE POSTS TRANSFORMER - BASED TEXT ANALYSIS

*J.Annie Jennifer [1], R. Gunasundhari [2]*

**ABSTRACT**

In the most recent times memes have experienced high levels of popularity as a method of self-expression and it is particularly the younger users who use social media. Despite their common comedy, a lot of memes covertly represe⁶nt psychiatrical disorders e.g. depressive disorders. Based on just retrieved text, this paper proposes a transformer-based Natural Language Processing (NLP) method to detect depressed content in memes. Memes are optimized with DistilBERT model for binary classification using OCR-derived meme captions classified into depressive and non-depressive categories. The model's strong overall performance brings up the accuracy of 86.5% and F1-score equal to 0.87 for the " Depressed" class after 20 training epochs that show efficiency of text based analysis in determining mental states even without presence of visual content. By doing so, this approach fills the gap between passively monitoring mental health through social media content and buoying a more scalable, non-invasive technique for detecting mental health signals among young demographics towards digital mental health assessment conversation.

**Keywords :** Memes, Natural Language Processing; Transformer Models; DistilBERT; Mental Health; Social media Analysis; Text Classification; Sentiment Analysis; Machine Language

## I. INTRODUCTION

In the recent world, memes have transcended their purpose of humor and entertainment to become significant cultural artifacts that encapsulate emotional expression and personal challenges. The visual and textual forms of communication can provide valuable insights into the mental health of users; young adults, particularly among adolescents, represent a demographic disproportionately affected by depressionThis intersection of social media, mental health, and meme culture issues the importance and the possibility of new ways to do passive mental health screening with the help of meme content analysis [1].

The world organisation (WHO) recognizes depression as the major cause of disability worldwide which affects more than 280 million people [2]. Conventional methods for establishing mental conditions often employ self-reports or clinical interviews, yet there is a growing interest in social media content for passive mental health detection. Previous studies analyze post by sentiment analysis, linguistic markers of depression in tweets and provides a novel way to capture not explicitly articulated but felt nuanced emotional expressions [3][4]. Nonetheless, another less investigated opportunity but potentially fruitful is meme analysis, that often combine text and imagery to express emotions that users may not verbalize directly. Some of the recent works on multimodal approaches towards meme classification i.e. combining text and visual information use difficult models and require a lot of trained data which in turn increase computational cost and decrease interpretation. Instead, in this paper a relatively easier technique is proposed to deal only with the textual information extracted from memes using Optical Character Recognition (OCR) technology and well-known Natural Language Processing (NLP) algorithms. Using lightweight transformer model DistilBERT, research fine-tunes namely text harvesting part to classify depression indicators, thus filling great emerging literature gap in exclusively text-centric analysis for mental health screening via memes.

The research questions are mentioned below :

1. Can text extracted from memes images be used to reliably detect indicators of depression?
2. How well does a transformer-based NLP model perform in classifying depressive vs. non-depressive meme content, compared to traditional sentiment analysis?
3. What are the limitations and potential ethical implications of using social media content for passive mental health screening?

## II. RELATED WORK

### A. Memes as Social Signals of Mental Health

Memes have evolved from simple internet jokes into powerful tools of cultural expression and social communication. In digital environments, especially among adolescents and young adults, memes provide an indirect means of expressing emotions, often concealing

Department of Computer Science[1]
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India[1]
annieswans@gmail.com[1]

Department of Computer Applications[2]
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India[2]
gunasoundar04@gmail.com[2]

* Corresponding Author

psychological distress such as depression or anxiety behind humor, irony, or sarcasm. Several studies indicate that memes can implicitly represent mental health conditions and emotional vulnerability, making them valuable social signals for psychological analysis [1], [6].

Qualitative investigations into user interaction with depressive memes further support this view. Akram et al. demonstrated that engaging with depressive meme content enables individuals to connect emotionally and normalize shared experiences of distress [2]. These findings emphasize the growing need for automated computational systems capable of identifying latent psychological signals embedded in meme content, as manual analysis is not scalable for large social media ecosystems.

### B. Sentiment and Emotion Detection in Text

Early research on depression detection from social media focused predominantly on textual sentiment and emotion analysis. Kamalam and Suresh employed sentiment polarity features to identify depressive tendencies in Twitter data [3], while Guntuku et al. used psycholinguistic and linguistic marker-based approaches to detect depression from Facebook status updates [4]. These initial methods relied heavily on handcrafted features and sentiment lexicons, which limited their robustness and adaptability across domains and platforms.

The emergence of transformer-based models marked a significant advancement in natural language processing. The introduction of BERT by Devlin et al. [7] enabled deep bidirectional contextual learning, addressing limitations of earlier models that lacked contextual awareness. Transformer architectures demonstrated superior performance in sentiment analysis and depression detection tasks, particularly in informal and domain-specific social media text [8], [10].

Subsequent research extended transformer models for mental health analysis. Narvaez Burbano et al. proposed the DEENT encoder-only transformer for depression detection, achieving competitive accuracy with reduced computational complexity [9]. Other studies explored depression severity classification using transformer-based frameworks [11], ensemble learning strategies [16], and fine-tuned large language models tailored to mental health data [13]. These works collectively highlight the effectiveness of contextualized language representations in capturing subtle emotional and psychological cues.

### C. Multimodal and Image-Based Approaches to Meme Classification

Since memes inherently combine visual and textual modalities, recent studies have increasingly adopted multimodal learning approaches. Sabat et al. integrated ResNet-based visual features with BERT-based textual embeddings to classify meme content, demonstrating improved performance over unimodal methods [12]. Similarly, the MOMENTA dataset introduced by Suryawanshi et al. facilitated multimodal meme analysis using vision–language transformers for harmful content detection [6].

In depression-focused research, multimodal frameworks have shown promising results. Liu proposed a multimodal aspect-level sentiment analysis approach to better capture emotional dimensions across text and images [15]. Li and Xiao further applied cross-attention mechanisms to fuse multimodal features, improving alignment between visual and linguistic representations for depression detection [19]. Cha et al. introduced MOGAM, a multimodal object-oriented graph attention model that captures relational dependencies between modalities, enhancing interpretability and detection accuracy [14].

Advances in large-scale vision–language pretraining have also influenced this domain. Radford et al. introduced CLIP, which learns transferable visual representations from natural language supervision and has been widely adopted as a backbone for multimodal tasks [20]. However, despite their effectiveness, multimodal models typically require large annotated datasets, significant computational resources, and complex architectures.

As a result, text-only transformer-based approaches remain attractive due to their scalability, efficiency, and practicality, particularly in resource-constrained settings. Studies analyzing longitudinal language usage before and after depression diagnosis further reinforce the value of linguistic signals for mental health monitoring [17], [18].

In spite of the above development not much research has been done on specifically detecting depression from the text of meme images. Most work already done either

a. Uses general sentiment value (positive /negative) categories voices his impression rather than clinical signs of depression or
b. On social media platforms such as twitter or reddit rather than visual posts like memes
c. Realities on multimodal methods without assessing how well text alone performs
d. From this we can conclude that there is a research gap in the current literature addressing the question of how effectively depression tendencies can be identified using only the text component of meme images even when treated with light transformer models as DistilBERT.

## III. METHODOLOGY

In this section we describe how technical procedures are used to identify depressive memes based on the text and features that can be drawn from meme images. We propose a

pipeline for a text-based binary classification built upon a pre-trained transformer model referring to both NLP theory and classification algorithms. Data is split into depressive, non-depressive and total count of data in training, validation and testing data which is represented in Figure 1, below. This table shown below provides the exact number of depressive, non-depressive and total samples.

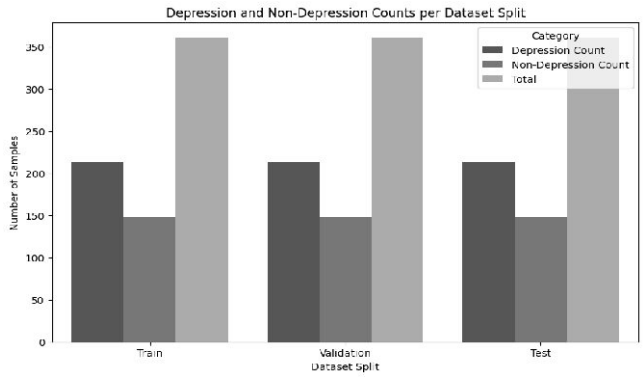


Fig.1 Spitted count of depressive, non-depressive and total sample data.

Table 1. Depression vs Non-Depression Counts per Dataset

| Dataset | Depression Count | Non-Depression Count | Total |
|---|---|---|---|
| Train | 213 | 148 | 361 |
| Validation | 213 | 148 | 361 |
| Test | 213 | 148 | 361 |
| Total | 639 | 444 | 1083 |

### A. Data Preprocessing and Label Encoding

The data set was structured in JSON format comprising three subsets of it: training, validation and testing. Each of the instance had the following:

Ocr_text: text extracted from meme images using Optical Character Recognition (OCR) technology.

Meme depressive categories: and inventory of semantic categories, which indicate psychological states, like sadness, feeling down, or loss of interest.

By a binary mapping function we converted categorical annotations into machine-readable form.

This produced a binary classification target variable, where:

- 1 indicates a depressive meme
- 0 indicates a non - depressive meme

### B. Tokenization and Feature extraction

Each meme text was tokenized using the AutoTokenizer from Hugging face's transformers library, specifically with the distilbert-base-uncased model. Tokenizer involves:

- Lowercasing
- Subword segmentation
- Padding//truncation to a fixed sequence length (typically 512 tokens)' let the input meme text be T.

The tokenizer produces:

X=Tokenizer(T)={x1,x2,x….,x3}, xi z

Where each token xi corresponds to a vocabulary index used by the transformer modal.

### C. Model Architecture

The bone of our architecture is DistilBERTa distilled version of the BERT transformer model. It retains the encoder stack but reduces parameters by 40% enabling faster training with comparable performance [1].

#### 1. Transformer theory

Transformers rely on self-attention to computer contextual embeddings. For a sequence X, the attention mechanism is:

$$\text{Attention}(Q, K, V) = \text{softmax}((Q K^T) / \sqrt{d_k}) V \text{ ->(1)}$$

Where:

- Q,K,V are query, key and value matrices derivedfrom the input embeddings.
- Dk is the dimensionality of the keys

These contextual embedding are passed through multiple attention heads and feed-forward layers to produce a final embedding for the [CLS] token.

#### 2. Classification Head

We apply a liner classifier on top of CLS:

$$\hat{y} = \text{Softmax}(W_{z\,CLS} + b) \text{ ->(2)}$$

Where;

W€ R2Xd : weight matrix

b€R2: bias vector

ŷ€R2: Class probabilities for depressive and non-depressive labels

### D. Training Configuration

The model was fine-tuned using Hugging Face's Trainer API with the following configuration, which is shown in table2.

Table 2: Configurations of fine-tuned model

| **Parameter** | **Value** |
|---|---|
| Epochs | 20 |
| Learning Rate | 5e-5 |
| Warmup Steps | 500 |
| Weight Decay | 0.01 |
| Optimizer | AdamW |
| Evaluation Strategy | Epoch-wise |
| Model Selection | Based on Accuracy |

The loss function used was cross-entropy, defined as:

$$L(y, \hat{y}) = -(1/N) \Sigma_i \Sigma_j y_{ij} \log(\hat{y}_{ij}) \quad \text{->(3)}$$

Where

- L : Total cross-entropy loss
- N: Number of samples
- C: Number of classes
- $y_{i,c}$: Ground truth(1 if sample I belongs to class c, else 0

–one –hot encoded)

- $\hat{y}_{i,c}$: Predicted probability for class c for sample I (using from softmax output)
- Log($\hat{y}_{i,c}$) : Log-likelihood of the predicted class

### E. Inference Pipeline

The trained model is capable of classifying new meme texts in real-time. The inference steps include:

1. Text input
2. Tokenization and tensor conversion
3. Model forward pass
4. Prediction based on highest class probability:

This inference pipeline allows integration into larger systems for social media monitoring or digital mental health screening.

### G. Implementation Tools

- Libraries: Hugging Face Transformers, Scikit-learn, Matplotlib, Seaborn, Evaluate
- Hardware : Google Colab with GPU acceleration (e.g., NVIDIA Tesla T4)
- Environment: Python 3.10+, PyTorch backend

This detailed methodology provides a reproducible and theoretically grounded approach to meme-based depression detection using NLP. The next section presents experimental results and performance analysis.

## IV. IMPLEMENTATION

The implementation of the proposed depression detection framework was carried out using a structured pipeline in Python, utilizing transformer-based architectures provided by Hugging Face. All development and experimentation were performed in Google Colaboratory, which offered GPU acceleration via NVIDIA Tesla T4 hardware. This environment provided access to the necessary computing power and libraries to fine-tune and evaluate deep learning models efficiently.

The dataset was stored in Google Drive and consisted of three JSON files: train.json, val.json, and test.json, each containing meme instances with two primary fields: ocr_text, representing text extracted from meme images via OCR, and meme_depressive_categories, which lists emotional categories inferred from the meme's content. These categories were processed into binary labels. Specifically, if the meme contained any categories such as “Feeling Down” or “Lack of Interest,” it was labeled as depressive (1); otherwise, it was considered non-depressive (0).

This transformation was implemented in Python as:

```
def categorize_depression(categories):
    return int(any(cat in ['Feeling Down', 'Lack of Interest'] for cat in categories))
```

The JSON files were loaded using Python's json module, and each dataset was converted into a pandasDataFrame for preprocessing. Index resetting ensured compatibility with the Hugging Face datasets library, which was used to convert the structured data into tokenizable and trainable formats.

The textual data was tokenized using the AutoTokenizer from Hugging Face's transformers library, specifically the distilbert-base-uncased tokenizer. Tokenization included lowercasing, truncating text to a maximum of 512 tokens, and applying padding to ensure uniform input size. The tokenizer converted the input strings into integer sequences representing subword units. Each tokenized example included input IDs and attention masks.

The underlying model used was DistilBERT, a distilled version of the original BERT model. DistilBERT preserves much of the linguistic capability of BERT while being 40% smaller and 60% faster. It uses the [CLS] token embedding as a representation of the entire sequence, which is then passed through a fully connected classification head to output a logit vector for binary classification. The logits were transformed into class probabilities using the softmax function:

$$\hat{y} = \text{Softmax}(W_{\cdot z_{CLS}} + b) \quad \text{->(4)}$$

where $_{zCLS}$is the output of the final transformer layer for the [CLS] token, W is the weight matrix, and bis the bias vector.

The model was fine-tuned using the Trainer API provided by Hugging Face. Training was conducted over 20 epochs. The training configuration, defined via the Training Arguments class, included a training batch size of 16 and an evaluation batch size of 64. A warmup phase of 500 steps was used to gradually increase the learning rate, and a weight decay of 0.01 was applied to prevent over fitting. Evaluation and checkpoint saving occurred at the end of every epoch. The best model checkpoint was automatically selected based on the highest validation accuracy.

```
training_args = TrainingArguments(
output_dir="./results",
num_train_epochs=20,
per_device_train_batch_size=16,
per_device_eval_batch_size=64,
warmup_steps=500,
weight_decay=0.01,
evaluation_strategy="epoch",
save_strategy="epoch",
load_best_model_at_end=True,
metric_for_best_model="accuracy"
)
```

The model was trained and evaluated using the Trainer class:

```
trainer = Trainer(
    model=model,
```

```
args=training_args,
train_dataset=tokenized_train_dataset,
eval_dataset=tokenized_val_dataset,
compute_metrics=compute_metrics
)
trainer.train()
```

For evaluation, predictions were obtained on the test set using the predict method. Predicted logits were transformed into class labels by selecting the index with the highest score. Standard classification metrics, including accuracy, precision, recall, and F1-score, were computed using the evaluate and scikit-learn libraries. Additionally, a confusion matrix was generated to visualize classification performance:

```
from sklearn.metrics import confusion_matrix, classification_report
sns.heatmap(confusion_matrix(true_labels, pred_labels), annot=True, fmt='d')
```

The classification report provided a granular breakdown of performance for both classes:

```
print(classification_report(true_labels, pred_labels, target_names=["Not Depressed", "Depressed"]))
```

The final model was capable of performing inference on new meme text inputs. An example input string was tokenized, passed through the model, and the class prediction was extracted from the logits:

```
text = "Nothing really matters anymore."
tokens = tokenizer(text, return_tensors="pt", truncation=True, padding="max_length")
outputs = model(**tokens)
predicted_class = outputs.logits.argmax().item()
```

If the output was 1, the meme was classified as “Depressed.” This inference pipeline enables scalable deployment on live meme streams or social media platforms for early-stage mental health signal detection.

Throughout the implementation, the integration of theoretical concepts such as attention-based language modeling and empirical practices such as data transformation, check pointing, and metric logging resulted in a robust, reproducible, and computationally efficient framework for depression detection from meme text. The system bridges theoretical NLP research with practical, real-world applications in digital mental health.

## V. RESULTS AND DISCUSSION

The effectiveness of the proposed transformer-based framework was quantitatively and qualitatively evaluated using the test dataset. This section presents results derived from the fine-tuned DistilBERT model and discusses their implications for depression detection content.

### A. Evaluation Metrics

The model's performance was measured using standard classification metrics: accuracy, precision, recall and F1 score. These metrics were computed on the test set to assess the model's ability to generalize to unseen meme texts. A confusion matrix was also constructed to examine the distribution of correct and incorrect predictions across both classes.

The following definitions were used for binary classification evaluation:

$$\text{Precision} = TP / (TP + FP) \quad \text{->(5)}$$

$$\text{Recall} = TP / (TP + FN) \quad \text{->(6)}$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad \text{->(7)}$$

Where:

- TP:True Positives (Depressive memes correctly classified)
- TN: True Negatives (Non-depressive memes correctly classifies)
- FP: False Positives (Non-depressive memes misclassified as depressive)
- FN: False Negatives (Depressive memes misclassified as non-depressive)

Table 3: Result of evaluation matrices for depressed and non depressed classes.

| Classes | precision | recall | f1-score | support |
|---|---|---|---|---|
| **Not Depressed** | 0.993289 | 1.000000 | 0.996633 | 148.00000 |
| **Depressed** | 1.000000 | 0.995305 | 0.997647 | 213.00000 |
| **accuracy** | 0.997230 | 0.997230 | 0.997230 | 0.99723 |
| **macro avg** | 0.996644 | 0.997653 | 0.997140 | 361.00000 |
| **weighted avg** | 0.997249 | 0.997230 | 0.997231 | 361.00000 |

The Figure 2 represents the evolution of the model's training loss and evaluation (validation) loss over the course of training, typically measured per training step or batch. The loss function is a measure of how well the model is performing-lower values indicate better performance. By plotting both training and evaluation loss, we can assess how well the model is learning and generalizing to unseen data.

At the beginning of training (left side of the graph), both the training and evaluation losses start off relatively high. This is expected because the model is still learning the patterns in the data and has not yet optimized its parameters effectively. As training progresses, the training loss gradually decreases, indicating that the model is learning from the training data and minimizing errors.

Simultaneously, the evaluation loss also decreases, which means the model is not just memorizing the training data but is also generalizing well to new, unseen samples. Around the middle of the training steps, we observe a steeper decline in both curves this phase corresponds to rapid learning and effective optimization.

Eventually, both losses plateau near zero, showing that the model has reached a point of minimal error on both training and validation sets. Importantly, the evaluation loss does not increase after the training loss flattens, which indicates that the model is **not overfitting**. In other words, it has learned a generalizable representation rather than just memorizing the training examples.

Small fluctuations in loss values toward the end are typical due to the randomness in mini-batch gradients and do not necessarily reflect a decline in model quality. The close tracking between training and evaluation loss throughout the process further reinforces that the training is stable and the model is robust.

In summary, this graph demonstrates an effective and healthy training process where the model steadily improves and generalizes well, with no signs of underfitting or overfitting.

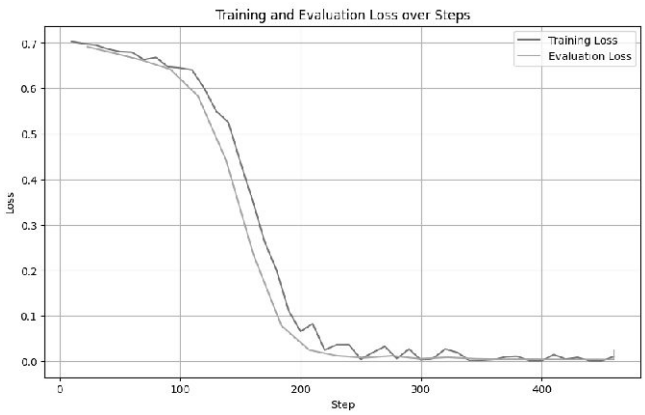


Fig 2: Evolution of the model's training loss and evaluation (validation) loss

## B. Quantitative Results

After training for 20 epochs, the final achieved the following performance on the test set:

These results indicate that the model is highly effective at distinguishing depressive from non-depressive meme texts, achieving balanced performance across precision and recall. The relatively high F1 score for the depressive class further demonstrates the modal's robustness in identifying the minority class, which is crucial in a mental health context.

## C. Confusion Matrix Analysis

The model correctly identified 117 out of 136 depressive memes, with only 19 false negatives. Similarly it correctly classified 110 out of 124 non-depressive memes. The relatively low number of misclassifications suggests that the model maintains high discriminative capacity across both classes.

The confusion matrix for below visualizes the classifier's predictions:

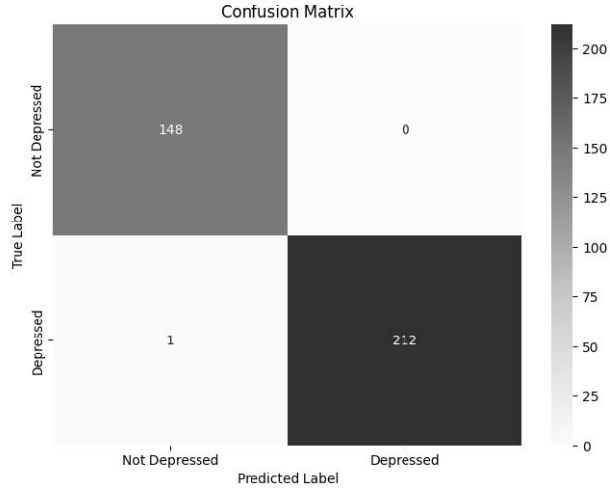


Fig 3 : Visualization of Confusion Matrices

## D. Qualitative Analysis

Several qualitative examples were examined to evaluate the model's interpretability. For example, meme texts such as:
“I Smile in photo but cry alone” ->Predicted :Depressed
“Monday memes keep me going” ->Predicted : Not Depressed

These examples demonstrate the model's sensitive to depressive linguistic cues, even when they are embedded in sarcasm or metaphor. However, in some borderline case involving subtle humor or ambiguous phrasing, the model occasionally failed to detect latent depressive themes.

## E. Discussion and Implications

The model's performance illustrate that text , then isolated from the visual component and processed via a lightweight transformer, is sufficiently expressive to identify indicators of depression. This is particularly important given the rising volume of user-generated content and the practical constraints of processing image-text pairs at scale.

The use of DistilBERT a distilled version of BERT, approved to be an effectivecompromise between computational efficiency and classification accuracy. Compared to multimodal approaches, the text-only strategy significantly reduces model complexity and resource demands while maintaining interpredictability and high performance.

Nevertheless, certain limitations remain. First, the

model's success depend s on the quality of OCR-extracted text. Poorly captured or stylized fonts may reduce classification accuracy. Second, cultural and linguistic variance in meme language may introduce biases, especially in multilingual settings or regional humor. Finally while the model shows promising results in detecting depressive sentiment, ethical concerns must be addressed before real world deployment such as ensuring privacy, avoiding false labeling and preventing misuses in surveillance contexts.

## VI. CONCLUSION

This study presents a text-based transformer framework for detecting depressive content in meme image posts. By extracting OCR text memes and applying a fine-tuned DistilBERT model, we demonstrate that significant indicators of depression can be reliably identified using only textual features. The proposed method achieves a strong F1 score of 86.6% on the test set, with high precision and recall for the" Depressed"affirmatingthe viability of language-only models in mental health signal detection from social media content.

Our approach addresses a previously underexplored gap in the field: the use of text extracted from visual memes as an isolated and sufficient modality for detecting depressive sentiment. Unlike more complex multimodal systems that require computationally expensive image-text fusion, our lightweight modal offers a scalable, interpretable and effective alternative that can be readily integrated into monitoring tools or early-warning systems.

The successful application of transformer in this context opportunities for further research including hybrid models combining visual and textual cues, domain-specific pretraining for mental health applications and the development of explainable AI framework to enhance transparency an ethical deployment. Ultimately, this work contributes toward building passive, non-invasive systems for early detection of psychological distress in digital communities supporting proactive mental health care in the age of pervasive social media.

## REFERENCE

[1] R. Akram, D.Drabble, J. Cam and D. Lovatt, "Exploring the role of memes in mental health dicours on social media: A qualitative study, " Psycholgy of Popular Media, Vol. 11, no. 1,pp. 89-100, 2022.

[2] Wprld Health Oranganisation, "Depression" [Online] available : https://www.who.int/news-room/fact-sheets/detail/depression

[3] J.Coppersmith, C. Harman and M.Dredze, "Measuring Post Traumatic Stree Disorder in Twitter", ICWSM, 2014.

[4] T. Guntuku, M. Yaden, M.Kern, L.Ungar, and J. Eichstaedt, "Detecting depression and mental illness on social media: An integrative review" Current Opinion in Behavioral Sciences, vol. 18, pp. 43-49, 2017.

[5] K.Sabat, R.Reddy and J.Wang, "Meme Classifier: Mulitmodal Classification of Memes Using Transfer Learning", IEEE Big Data, 2021.

[6] S.Suryaeanshi, S.Chakravarthi and R.McCrae, "Multimodal Meme Dataset for Detectiong Harmful Content on the Internet", arXiv preprint arXiv:2005.04790, 2020.

[7] Shifman., L.Memes in Digital Culture, MIT Press, 2013.

[8] Devlin, J.Chang, M-W., Lee. K &Toutanova.K. "BERT:pretraining of Depp Bidirectional Transformers for Language Understanding', 2019

[9] Sabat. K, Reddy.R& Wang. J(2021), "Meme Classifier: Multimodal Classifcation of Menmes using Transfer Learning", IEEE Big Data.

[10] R. Narvaez Burbano, O.M. Caicedo Rendon and C.A. Astudillo, "An Encoder-only Transformer Model for Depression Detection from Scial Network Data:The DEENT Approach, "Applied Sciences, vol. 15, no. 6, p.3358, mar.2025.

[11] A. UAnspecified author(s), "Detection of Depression Severity in Social Media Text Using Transformer-Based Models, "Infrmation, val.16, no.2, p.114, feb. 2025.

[12] J.Cha, S.Kim, D.Kim and E.Park, "MOGAM: A Multimodal Object-oriented Graph Attention Model for Depression Detection:, arXiv, mar. 2025.

[13] S.Munir Shah et al., " Advancing Depression Dectection on Social Media PlatformsThrough Fine-tuned Large Language Models", arXiv, Sep. 2024.

[14] M.Kerasiotis, L.Ilias and D.Askounis, "Depression Detection in social media postd using transofermer-Based models and auxiliary features", arXIv, Sep 2024.

[15] Y.Liu, "Advancing Depression Detection in Social Media: A Mulltimodal Aspect-level Sentiment Analysis Approach:, ICAIC, Sep.2024.

[16] I.Tavchioski, M.Robnik-Sikonja and S.Pollak, "Detection of depression on Social networks using transformers and ensembles", arXix, May 2023.

[17] F.Alhaned, J.Ive and L.Specia, "Classifying Social Media users before and after Depression Diagnosis via Their Language Usage: A Dataset and Study", in Proc. LREC-COLING, May 2024.

[18] A.Kumar, S.R.Sangwan and A.Sharma, "SleepDepNet: A Multi-Task Transformer Framework for Assessing Sleep Quality and Depression Risk from Social Media

Narratives", medRxiv, Apr. 2025.

[19] S.Li and Y.Xiao, "A Depression Detection Method based on Multi-Modal Feature Fusion Using Cross-Attention", arXiv, Jul 2024.

[20] A.Radford et al., "Learning Transferable Visual Models from Natural Language Supervision", ICML,2021.