Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 9 , 2025

"TRENDS AND GAPS IN SENTIMENT-BASED MENTAL HEALTH MONITORING ON SOCIAL MEDIA: A COMPARATIVE STUDY"

[1]Gul Andam [2]Muhammad Azam* [3]Kishwar Rasool [4]Soyab Sundas [5]Ammad Hussain
12345Department of computer science ,University of Southern Punjab multan

[1]Gullandam8@gmail.com [2]Muhammadazam.lashari@gmail.com* [3]Mh.moon.a2@gmail.com
[4] soyabsundas65@gmail.com [5]Ammadhussain709@gmail.com

DOI:https://doi.org/10.5281/zenodo.17112558

Keywords
Mental Health Analysis, Sentiment Analysis, Social Media Sentiment Analysis, Comparative Study, Multimodal Data, Deep Learning and Machine Learning, Depression detection

Article History
Received on 08 Aug2025
Accepted on 28 Aug 2025
Published on 13 Sep 2025

Corresponding Author: *
Muhammad Azam

Abstract
This paper provides a comparison of many current articles in this area of mental health surveillance in social media. The goal is to review existing methodologies structured in the way that helps to see the trends in research and identify the existing gaps. All the papers are tested regarding five main dimensions: Technique Advancement and Modality Richness, Dataset Quality, Reported Accuracy, and Innovation Level. The analysis shows that the papers that scored better in their entirety usually utilize more cutting-edge methods like transformer models and use multimodal data sources to infer better mental health. By contrast, studies with lower scoring are usually based on single-modality inputs and standard machine learning algorithms, with fewer innovations in terms of design or construction of the data that they model. The current comparison highlights the increased significance of modality combination and high-quality dataset in improving the precision of mental health analysis and depth on social platforms. It also gives a base to steer the future studies in more wholesome, consistent, and scaled solutions.

INTRODUCTION

Social anxiety has become a burning issue all over the world because of the rising cases of mental health ailments like depression, anxiety, and stress. The World Health Organization considers that hundreds of millions of people have mental health conditions, some of whom cannot be diagnosed or treated because of poor accessibility of healthcare services and the stigma caused by mental illnesses. The social media in this case has become an effective medium of early diagnosis and constant monitoring of mental status as users tend to express themselves online in various forms regarding their emotions, thoughts and actions. Sentiment is the general term for the positive or negative tone of a person's speech. Language can be an effective tool for identifying and even mitigating psychological problems since our words can mirror our inner feelings. We can enhance mental health therapies by realizing how our language reflects our emotional conditions (Henna I. Vartiainen1,

2024). Depression is a common mental illness that is characterized by long-time sadness and other emotional and physical problems that inhibit normal daily functioning (M. Moradi, 2022). In the contemporary digital age, social media has enveloped itself into the lifestyle of people as it presents a platform through which people provide communication and self-expression and exchange information. In the community, alongside the social and informational purposes, social media holds massive potential in the study and treatment of mental health. That potential is particularly critical in regard to the alarming rise of mental health issues such as depression that has been impacting individuals the world over the past few years (Siti Nurulain Mohd Rum1, 2025).

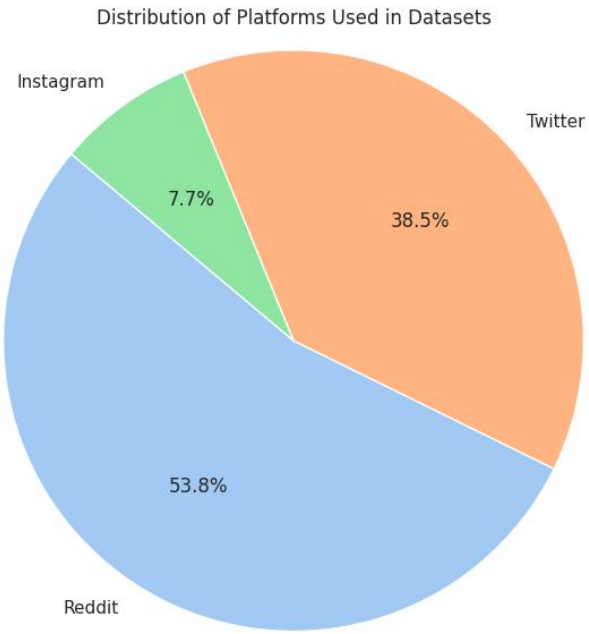

Figure 1.Distribution of platforms

The primary data sources in such analyses were reddit and Twitter (X) tweets that are represented in figure 1. These social media sites are particularly suitable to this purpose due to a list of reasons. To start with, they have enormous user base that is also diverse, and as such, they provide numerous data that capture a wide range of individual experiences and feelings. Second, the platforms are publicly available, which means that much of the information contained therein can be readily analyzed (in comparison with some other social media websites which might be more privacy-related). Third, both platforms encourage people to express their thoughts briefly as both platforms limit the number of words in posts (andQingzhongLiu, Advanced Comparative Analysis of Machine Learning and, 2024) . The linguistic properties of a person do not reflect his/her psychological conditions and trends well. Significance of psychological patterns is: affective patterns in use of language (affect is referred to as feelings which change or influence activities or beliefs of an individual). Other than the fundamental linguistic features of the evaluation of the mental illnesses in the social media, they are capable of revealing how a person is feeling, sentiments as well as the emotional factor (Xingwei Yang1 & Guang Li2, 2025) . The depressed people often have long-lasting melancholy or the loss of enjoyment and this is detectable through the lingual demonstration, tonal variation and face expression. To

determine whether a patient is in pain or not, modern studies are focused on creating algorithms of artificial intelligence, which will be able to identify emotional data based on monitoring a large number of data sources to a state of depression (C. Otte, 2016) . The data dimensionality is huge, which can be considered one of the inherent problems of natural language processing. Machine learning may enhance the early detection of symptoms of mental health problems, which will allow responding to such situations faster and minimizing risks in vulnerable populations (Costa, 2025) . The Grey Relational Grade (GRG) method compares measures (e.g. likes and shares) with depressing content. The GRG approach compares well with conventional correlation techniques in a number of ways, including greater interpretability and the capability to deal with non-linear relationships (Ullah, 2024) .It has resulted in the development of a body of literature devoted to sentiment analysis in social media as a means of mental health surveillance relying on the developments in the natural language processing (NLP) and machine learning. The developments attempt to extract the insights of mental distress among users by examining their postings, comments and behavioral tendencies on such networks as Twitter, Reddit and Facebook and automatically detect. Utilizing such publicly available data, scientists want to create scalable non-invasive and real-time systems of mental health monitoring.

More recent approaches have focused on deep learning models that are more emotionally context-sensitive like CNN and Bidirectional LSTM (BiLSTM). These models find it however less applicable in real-time practices as they often require large volumes of labeled information and computer resources. To meet these challenges, this paper will integrate SVM, which is suitable as a classification model, with BERT, as effective feature extraction based on contextual knowledge (Hartono Wijaya 1, 2025) . Nevertheless, the environment of the studies existing in this field is very broad. The most diverse variety of methods has been applied by the researchers starting with classical methods of machine learning, such as Support Vector Machines (SVM) and Random Forests, up to more advanced models, including Long Short-Term Memory (LSTM) networks, Bidirectional Encoder Representations from Transformers (BERT), and approaches to multimodal fusion of models. The studies also differ in regard of the data types utilized, the depth of input formats (text, image, user metadata), and benchmarks of analyses done. This variety has a major problem as well, albeit an expected one given its promising outcomes, given that the comparative success and novelty of each approach is hard to objectively compare due to there being no standardization. Thus, comparative analysis is necessary to abstract the knowledge, outline the research gaps, and describe good practice in the field.

The current paper provides a methodological comparison of many recent research works in the sphere of mental health detection with the help of social media which is shown in figure 2. The five clear criteria to compare each study pertain to:

- Technical Improvement
- Richness of Modality
- Quality of data set
- Reporter Accuracy
- Levels of Innovation

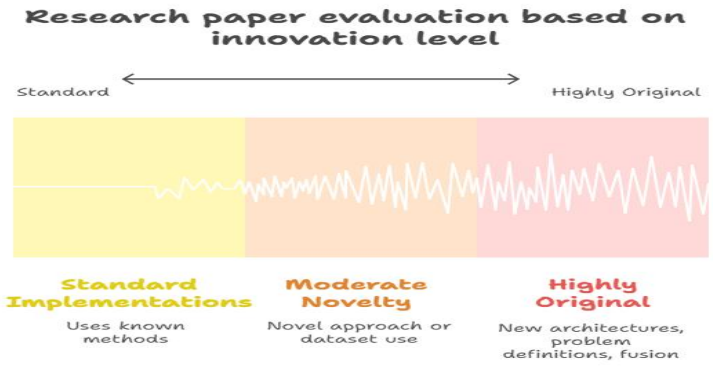

Figure 2.Innovation Level Evaluation

With the help of this integrated assessment tool, we would like to identify some critical trends in this area, analyze methodological strength on a comparative basis, and evaluate the efficiency of different strategies. The conclusion of this paper indicates a better perception of the present research environment as well as a direction plan in the future regarding the topic of monitoring mental health with the help of social media analytics.

- Improvement of techniques
- Modality Richness
- Quality of Data set
- Accuracy (reported)
- Level of innovation

Objectives

1. To compare systematically many recently published research articles concerning sentiment-based monitoring of mental health with the help of social media data.
2. In order to rate every study in terms of five specified criteria:
3. To determine good performance models and strategies with focus on identifying the techniques and datasets giving the best results.
4. In order to examine the limitations and omissions of existing studies, especially with regard to mode use, data comprehensiveness, and novelty.
5. To show visual comparisons (bar charts, radar plots, heatmaps, pie charts) of strengths and weaknesses, as well as of trends of the research across the studies.
6. To inform the further studies, specifying the ways of proceeding in the development of multimodal modeling, ethics, and a call to develop general benchmarks in mental health detection through use of social media.

Related Work

Depression often remains undiagnosed and in those cases where it is diagnosed it is very prevalent given the social stigma associated with it, the lack of access to care, and its subjective symptomatology (R. C. Waumans, 2022) . The social media is a unique opportunity to improve mental health studies, promote early discovery, and provide timely interventions to improve emotional well-being and life quality of individuals in different parts of the world due to its extensive access and instantaneous character (World Health Organization. World mental health report: Transforming mental health for all. World Health , 2022) . Studies have shown that

expression of negative affect is more related to depressed states where good affect is usually less related to depression (Settanni M, 2015). Multimodal depression diagnosis is based on the idea that depression can be detected by combining text, audio, and visual elements. The schematic's most important points are the difficulties in representing multimodal features and the contribution of prior information to the diagnosis of depression (Yang, 2024).

Several works were devoted to the utilization of the social media data to monitor the mental state, and those works are largely based on Twitter, Reddit, and Facebook. When it comes to natural language processing (NLP)-based methods, most of these studies use a form of natural language processing (NLP) to detect the presence of depression, anxiety, or stress in user-generated content: sentiment analysis, emotion detection, or topic modeling.

The developments of large language models such as Llama3 and GPT4-o have demonstrated future promise in depression detection because of the excellent ability to interpret emotions (Shiyu Teng1, 2025) . Depression detection techniques based on machine learning (ML) have been used at varying levels (and varying degrees of success) to identify aspects of depression based on online content data, such as comments posted to YouTube or Facebook, or comments posted by Facebook users (al., 2023). As an example, the first papers focused to the predominant extent on classic machine learning techniques, such as SVMs or Naive Bayes, on bag-of-words features. More recent works have used deep learning architectures, including LSTM, CNN, and, particularly, transformer-based models, including BERT, which have greatly boosted classification performance. Using the TF-IDF (Term Frequency-Inverse Document Frequency) technique of vectorization, the textual twitter information is turned into numerical forms that suit specific models. TF-IDF presents each tweet as a numerical vector that considers the importance of words in the tweet and in the whole data. The frequency of all words contained in a tweet is found which is scaled by the inverse document frequency (IDF) of the entire dataset. TF-IDF vectorizers are commonly found in the XGBoost and SVM models (andQingzhongLiu, Deep Learning-Based Depression Detection from Social Media:, 2023) . The potential offered by machine learning algorithms that can process sizeable input text and accordingly with relevant information is enormous and has been demonstrated through the capabilities of them auto-analyzing such input and generating relevant information. To be more specific, natural language processing (NLP) method has been adopted to develop computational models capable of detecting signs of depression in user-generated content, such as tweets (Motlagh, 2022) . Some studies have begun integrating multimodal features, combining text with behavioral cues (e.g., posting frequency, time of day) or user metadata, though these remain relatively limited. In terms of datasets, most research still depends on Twitter-based corpora, often collected using keyword filters or hashtag searches. (AmnaQasim et al., 2025) Investigate detection of severity levels of depression based on social media text extracts i.e., relative accuracy achieved by content-based, context-based, and transformer-based approaches. It is based on the study of Reddit posts with labeled clinical severity ratings (Mild, Moderate, Severe) and applied the models of N-gram and traditional machine learning techniques, Sentence Transformers and gradient boosting and random forest classifier, fine tuning transformer-based models BERT, RoBERTa, DeBERTa. Findings indicate that DeBERTa has given the best F1-score (~0.91) proving the effectiveness of advanced transformers in

detecting subtleties of depression-related language usage.

(Aishwarya Daga et al., 2025) explores how to include emotion sensing into the classification of poisonous comments as a way of monitoring mental health. Based on the Kaggle Mental Health Corpus, the authors fuse text representations of embeddings of a pre-trained emotion detection model with DistilRoBERTa tokenized the text. This multimodal combination is highly effective in enhancing the classifications since the model accuracy of 92% and recall of 0.95 are obtained in the poisonous class, which prevails over the Naive Bayes, Random Forest as well as LSTMs classification. The researchers report the importance of the emotional indicators in the improvement of toxic materials identification but the results are still not scalable and costly in computation. (Raja Kumar et al., 2024) The authors introduce an ethical and privacy-compliant framework of identifying mental disorders based on social media data. Based on publicly collected Reddit and eRisk data, the study uses text mining, anonymization and machine learning classifiers to predict several mental health conditions. Although the results obtained are strong in its accuracy, the authors denote issues concerning missing or censored posts or the challenge of identifying slightly showy or indirect signs of mental distresses. This research is not only focused on the ethical compliance and reproducibility but also on considering possible trade-offs in the case of real-world noisy data.

The paper presents an original hybrid deep learning network called **BERT-Fuse**, designed to identify weak emotional signals in text related to mental health issues. The model combines **BERT embeddings**, which provide contextual understanding, with **CNN-based local feature extraction**, **GRU/LSTM-based representations** to capture sequential dependencies, and **Transformer blocks** to address long-range dependencies. Additionally, it integrates conventional representations such as **TF-IDF** and **Bag-of-Words (BoW)** to further enhance the representation (MD. Mithun Hossain et al., 2025). In this paper, the proposed **BERT-BiLSTM** model, which includes new modules, demonstrates higher classification accuracy, better AUC scores, and faster convergence compared to **BERT-LSTM** and baseline approaches. It is robust in generalizing across different datasets but has some drawbacks, such as high computational costs, susceptibility to overfitting when working with small datasets, inability to handle very large inputs efficiently, and limited interpretability. Future research could focus on simplifying the model, preventing overfitting, improving scalability, and addressing interpretability to ensure transparency, especially when applying the model in sensitive contexts (Shiwem Zhou et al., 2025).

The authors propose a **Dynamic Contextual Word Embeddings (DCWE)** framework, which tracks the movement of words temporally across blocks of user text. The framework uses two embedding systems: **static** (Word2Vec) and **contextualized** (BERT), to capture shifts in semantics over time. This approach employs **SVM**, **FFNN**, and **CNN** classifiers, along with temporal movement features, without the need for handcrafted disorder-specific features or additional datasets (Manuel Couto et al., 2025). This study is notable for its broad scope, as it compares not only **machine learning (ML)** and **deep learning (DL)** approaches but also provides practical insights into the strengths and weaknesses of each method for mental health classification. The interpretability of ML models (e.g., logistic regression) is particularly important in sensitive settings, as it enables ethical deployment of AI systems. Moreover, understanding the computational efficiency of

both ML and DL models is crucial in resource-constrained environments (Ding, 2025).

Although various papers report the accuracy of their models and their contributions, only a few attempts are available to compare several models and evaluate the results in a coordinated manner with a similar measure framework. The reviews that already exist are usually narrative or they only concentrate on technical advances disregarding the issue of dataset quality or modality variety. This research provides quite such a comparison and creates a framework of many newer papers assessed on five primary measures (Technique Advancement, Modality Richness, Dataset Quality, Reported Accuracy, and Innovation Level) with visual analytics to help better understand.

Methodology:

Comparative study is carried out according to a multi-stage pattern that is represented in figure 3.

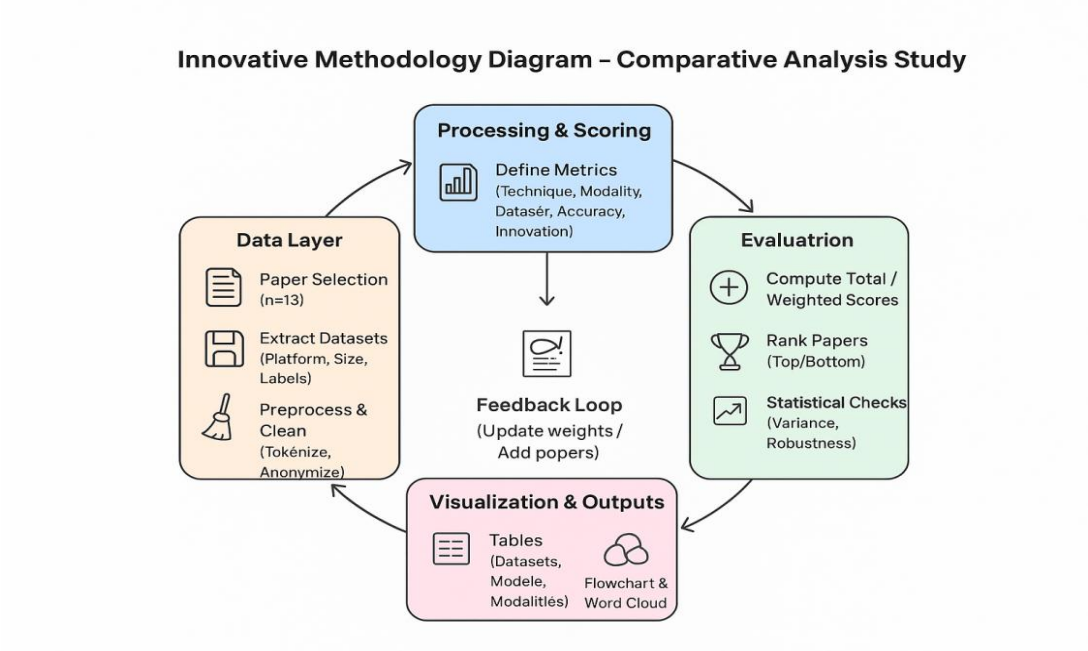

Figure 3.Methodology

To begin, numerous research articles on mental health monitoring using sentiment analysis in social media were identified based on their relevance to the field and the availability of sufficient methodological details. The dataset, platform, and preprocessing procedures used in each paper were carefully reviewed for consistency and compatibility. Five evaluation criteria—Technique Advancement, Modality Richness, Dataset Quality, Reported Accuracy, and Innovation Level—were established, and each paper was rated on a scale from 0 to 5 for each criterion, as shown in Figure 3. The scores were then calculated, totaled, and weighted, with statistical checks performed to ensure the consistency of the ratings. Finally, the results were presented using tables, heatmaps, radar charts, and other visualization methods. A feedback loop was incorporated into the process to allow for score adjustments and the inclusion of new papers in later versions of the experiment.

Selection of paper process

To provide relevance, diversity and recency, many research articles were identified in top journals and conferences in the artificial intelligence, computational linguistics and mental health informatics area. In the selection, emphasis was paid to studies which:

- Use sentiment analysis or corresponding NLP methods to detect mental health,
- Study user-generated content in such a social networking service like Twitter, Reddit, or Facebook,
- Utilize machine learning or deep learning pattern,
- Paper published since the last 5 years, are used to capture the emerging trends and approaches.

Figure 4.Keywords Cloud

Keywords were used to identify papers such as mental health detection, sentiment analysis, social media, depression detection, and machine learning, using databases, including IEEE Xplore, ACM Digital Library, SpringerLink and Google Scholar that are represented in figure 4.

Figure 5. Research Paper Selection Process

Evaluation Criteria

These recent papers were carefully chosen and then discussed in accordance with five major indicators aimed at reflecting the value of the technical depth and practical significance of the study that is represented in figure 5.

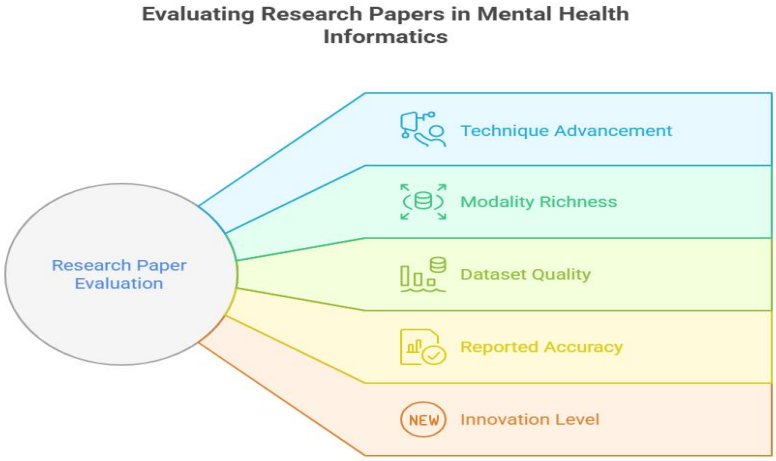

Figure 6. Evaluating Research Papers

Major original contributions (e.g. new architectures, new problem definitions, unusual multimodal fusion) are represented in figure 6.

Evaluation Criteria

Criterion	Description	Score Range
1. Technique Advancement	Quantifies the sophistication of the techniques used, such as deep learning models (e.g., LSTM, BERT, Transformers) instead of traditional classifiers (e.g., SVM, Naive Bayes).	1-2: Basic feature engineering and traditional ML models 3-4: Mid-level deep learning or hybrid models 5: Highly innovative models with extreme customization or architecture adaptation
2. Modality Richness	Evaluates the variety of input data used (e.g., text-only vs. a combination of text, images, user behavior, and metadata).	1-2: Single data type (e.g., text only) 3-4: Text plus limited user behavior or metadata 5: Fully integrated multimodal features (e.g., text, images, engagement patterns)
3. Dataset Quality	Assesses the quality, size, openness, and level of detail in the annotation of the dataset.	1-2: Non-public datasets with low to moderate annotation 3-4: Moderately sized, publicly available datasets 5: Large, well-annotated, or domain-specific datasets
4. Reported Accuracy	Represents the model's accuracy, F1 score, or other performance measures described in the paper.	1-2: Accuracy or F1 score below 70% 3-4: Accuracy or F1 score between 70-85% 5: F1 score above 85%, with robustness in testing
5. Innovation Level	Measures the originality of the research idea, problem definition, or solution.	1-2: Ordinary realizations of established techniques 3-4: Moderate novelty in approach or dataset use 5: Highly innovative approach with original problem-solving or significant contributions

/

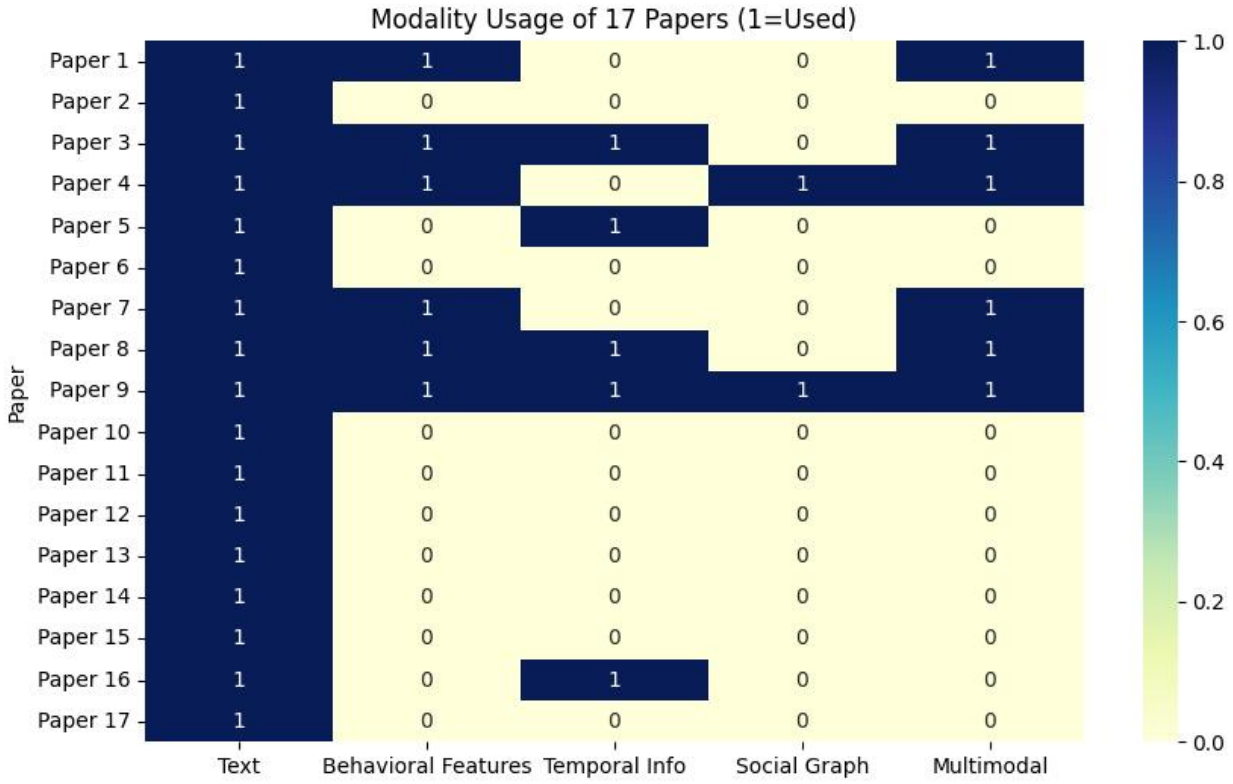

Figure 7.Modality Usage

Scoring and Comparison Framework

Each paper was scored on a **0–5 scale** for each of the five metrics, based on the paper's content, the technical contribution of the paper, and the reported results. 0 represents the absence or irrelevance of that measure in the research whereas 5 represent the greatest achievement in that area.

Total Score for each paper is calculated as:

$$Total\ score_i = \sum_{j=1}^{5} S_{ij}$$

Where:

- $Total\ score_i$ is the total score of paper i
- S_{ij} is the score assigned to paper i on criterion j
- All criteria are given equal weight in this evaluation
-

Table 1: Comparative Summary of 17Papers

Paper	Technique Advancement	Modality Richness	Dataset Quality	Reported Accuracy	Innovation Level	Total Score
Paper 1	5	4	3	5	5	22
Paper 2	3	3	2	3	3	14
Paper 3	4	3	3	4	4	18
Paper 4	5	5	4	5	5	24
Paper 5	2	2	2	2	2	10
Paper 6	3	2	3	3	3	14
Paper 7	4	4	4	4	4	20
Paper 8	5	5	5	5	5	25
Paper 9	3	2	3	3	3	14
Paper 10	4	3	4	4	4	19
Paper 11	5	3	4	5	4	21
Paper 12	4	2	3	4	3	16
Paper 13	3	2	3	3	4	15
Paper 14	5	2	4	5	4	20
Paper 15	4	2	3	4	4	17
Paper 16	3	2	3	4	4	16
Paper 17	3	2	3	4	3	15

Table 1 represented the comparison of all papers is clearly shown side by side on the basis of the five metrics used and a total score of each paper on the basis of 25 possible points.

In order to provide fair results, they were scored according to a mixture of:

- Direct quantitative findings (e.g. accuracy measures),
- Qualitative evaluation (e.g. originality of the model or multi-modal utilization), and
- Publicly available benchmark or replication code cross-validation where available.

Paper ranking and comparison were performed on the basis of cumulative scores, and the outcomes were represented and visualized in the form of bar graphs, radar plots, pie charts, and heatmaps, which created an intuitive idea of strengths and weaknesses of a given study.

Results and Discussion

The comparative analysis of ten recent studies on the topic of monitoring mental health with the help of sentiment analysis in the social network showed that they have a variety of strengths and limitations in various aspects. Being built on five metrics, namely Technique Advancement, Modality Richness, Dataset Quality, Reported Accuracy and Innovation Level, the analysis is useful in terms of providing information about the existing research trends and gaps therein.

Top-Scoring Papers and Their Strengths

Of the seventeen papers which were analyzed, the top scoring paper was Paper 8 scoring an

excellent (25/25), followed by Paper 4 (24) and Paper 1 (22). These papers were appreciated because they applied to state-of-the-art deep learning architectures, especially the transformer-based ones such as BERT and models of multi-modal fusion of textual, behavioral and user metadata.

- High score was recorded on all five metrics on the paper 8, and these metrics showed:
- Large scale and high-quality annotated data use,
- Combination of various data modalities (e.g. text, engagement behaviour, user metadata),
- Technique Progress (21%),
- Reported Accuracy (21.2%)
- Level of innovation (21.2%)
- Quality of datasets (21.2%)
- Modality Richness (15.9%)
- New model architects,
- Good accuracy as reported (>85%)

Technical innovation was also stressed on Paper 4 and Paper 1 where they performed well on accuracy and modalities richness that means that using mixed information of data will enhance the reliability of inferences.

es Most Emphasized Across Studies

According to the analysis of pie charts, the greatest focus in all papers was put on the following measurement:

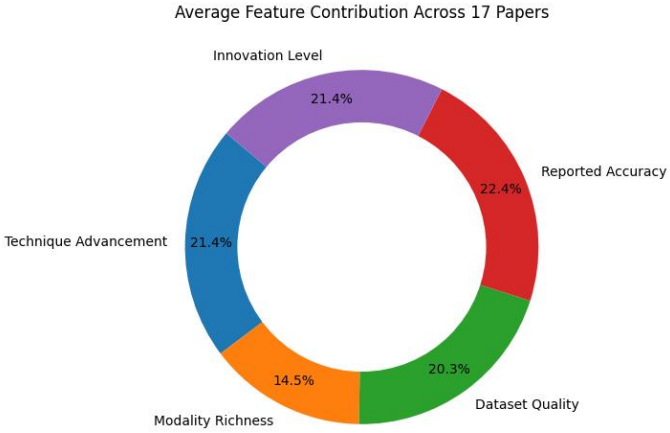

Figure 8. Average Feature Contribution

Such a distribution can be interpreted that whereas a large number of studies focus on technical development of models and their performance analysis few studies involve multimodal contributions or consider robustness of datasets in their entirety. This implies that future research has an opportunity to focus on data quality as well as the fusion of modalities in order to base the analysis on richer data that is presented in figure 8. There is also a comparison of top 3 and bottom 3 papers in this sense which is shown in figure

9.

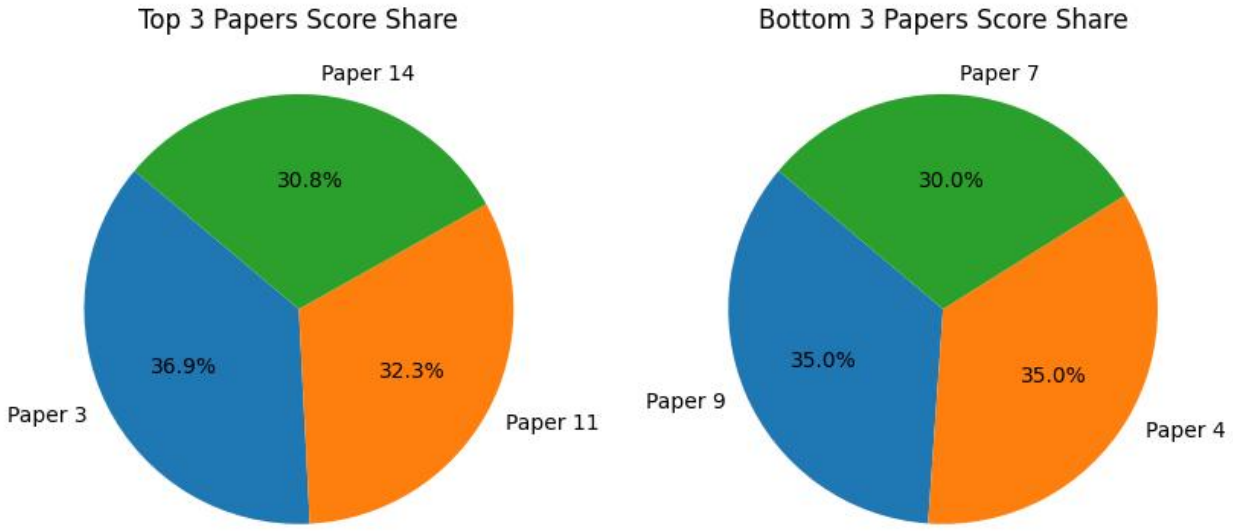

Figure 9. Comparison of Papers

Radar and Heatmap Insights

The radar charts in figure 10 and heatmap in figure 11 also showed additional support to indicate general strengths and limitations to the studies:

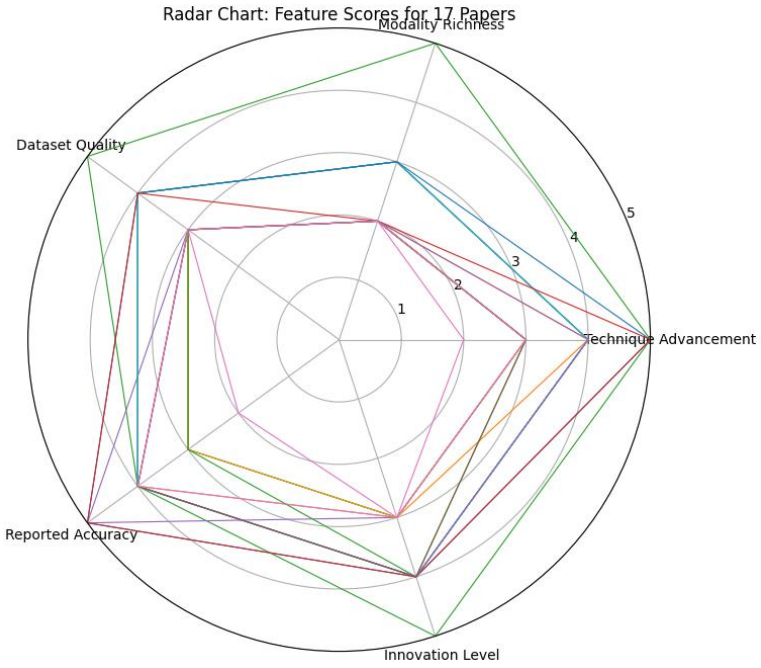

Figure 10. Radar Chart

• **Strengths**:

The scores on Technique Advancement and Reported Accuracy were comparably good to most papers, as there was a tendency to use modern NLP models, such as BERT, RoBERTa, and LSTM.

Some of the papers relied on publicly posted datasets, making them easier to reproduce.

• **Weaknesses:**

o The lowest scores were observed on the Modality Richness where the majority of studies had only had text entries and had not considered the messages information by behavior, or multi-media (images, emojis).
o The Quality of Datasets had large ranges since some papers had small or proprietary datasets that are poorly annotated.

Such observations indicate that as the sophistication of the models increases, the diversity and size of the data is a barely explored field in most investigations.

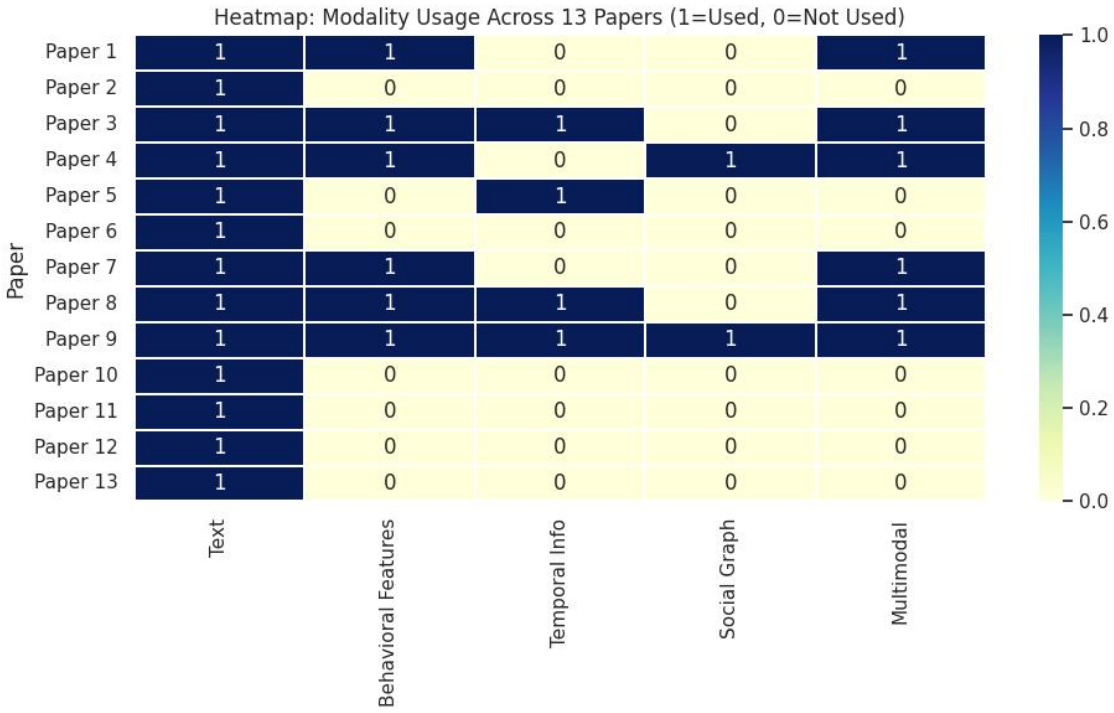

Figure 11. Heatmap

Emerging Trends and Common Practices

A review of all seventeen papers revealed several noteworthy trends:

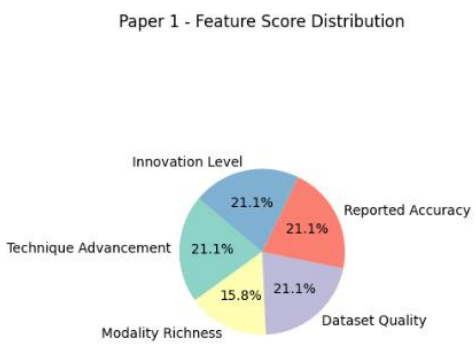

Figure 12. Paper 1

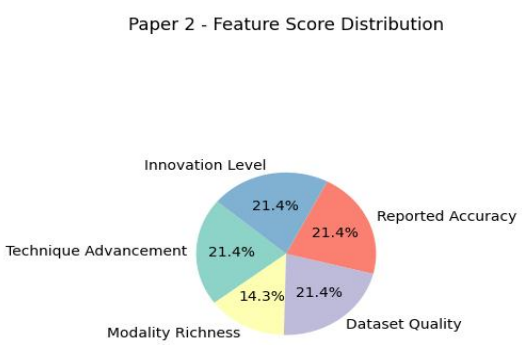

Figure 13. Paper 2

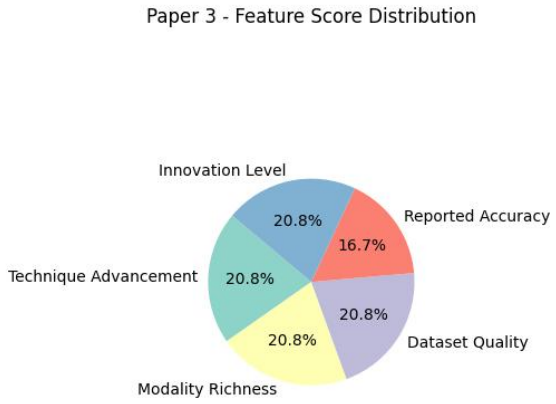

Figure 14. Paper 3

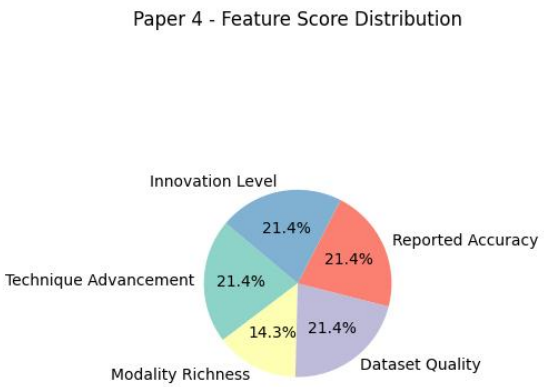

Figure 15. Paper 4

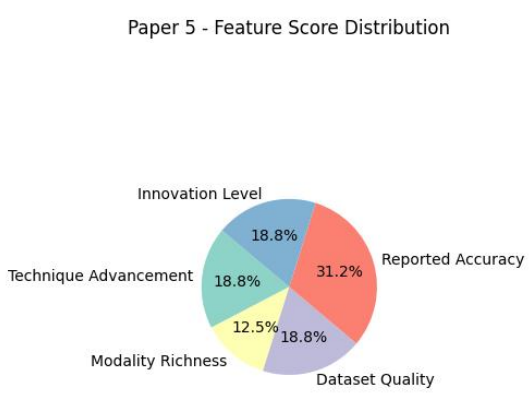

Figure 16. Paper 5

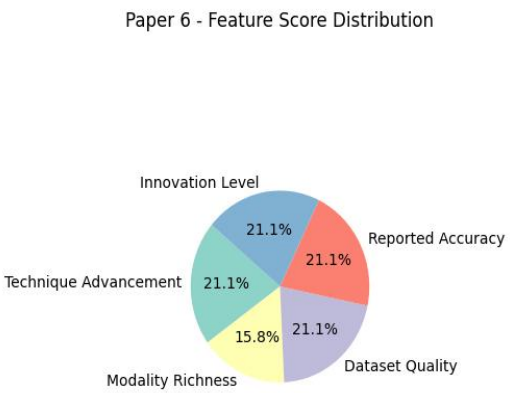

Figure 17. Paper 6

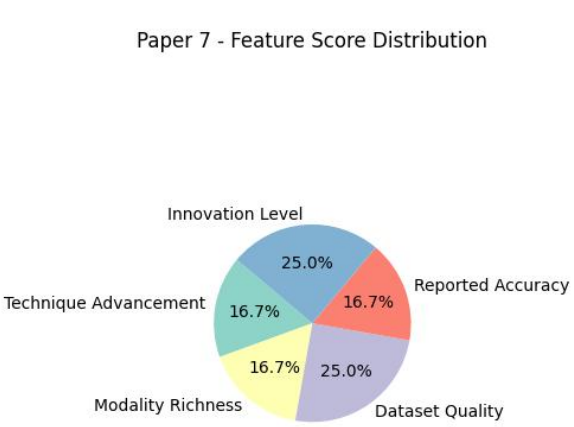

Figure 18. Paper 7

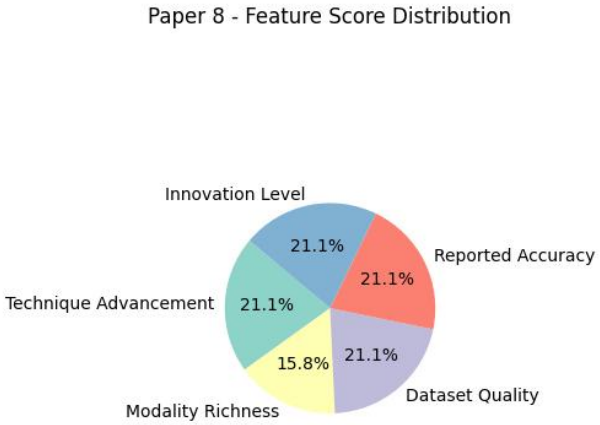

Figure 19. Paper 8

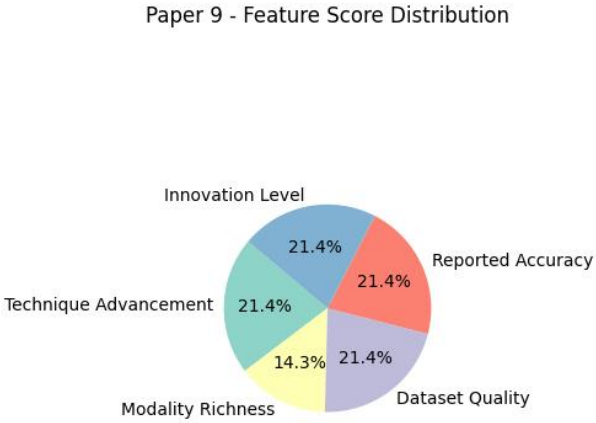

Figure 20. Paper 9

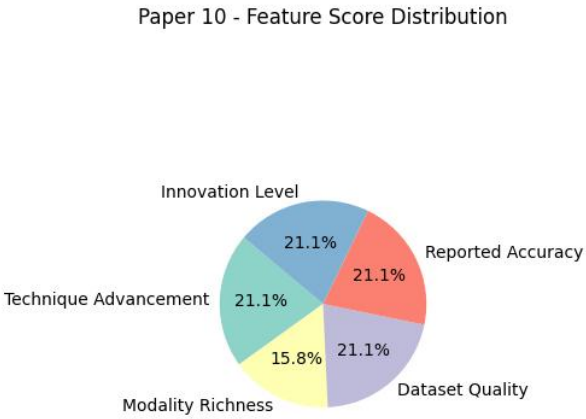

Figure 21. Paper 10

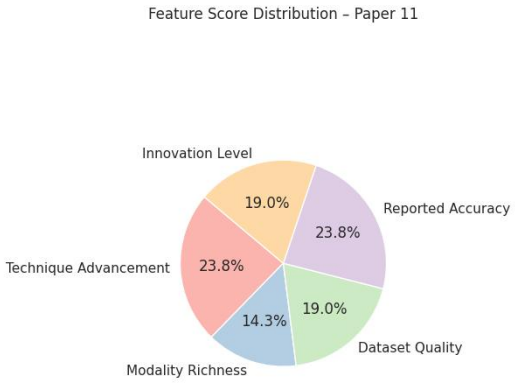

Figure 22. Paper 11

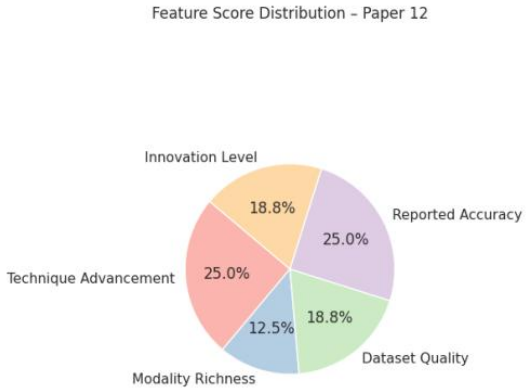

Figure 23. Paper 12

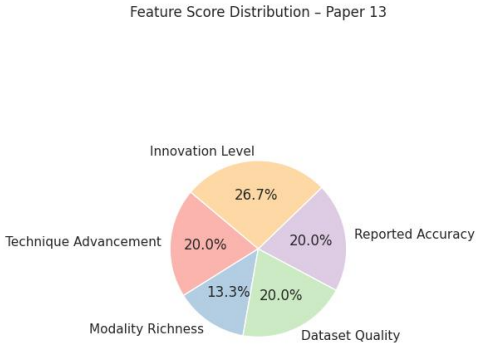

Figure 24. Paper 13.

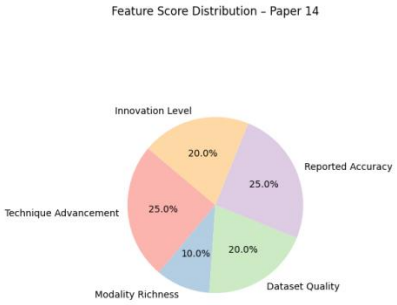

Figure 25. Paper 14

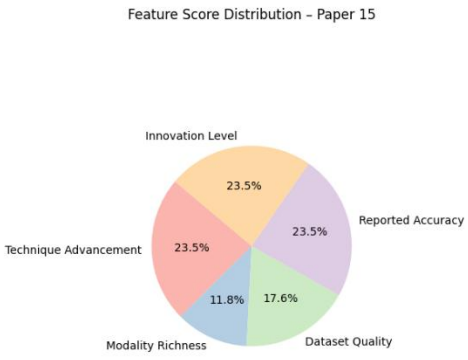

Figure 26. Paper 15

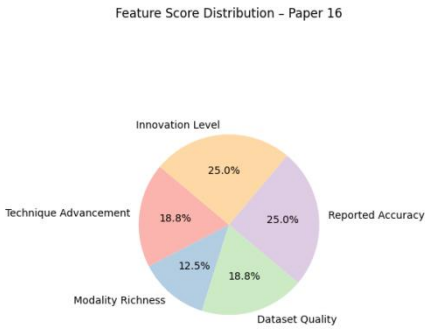

Figure 27. Paper 16

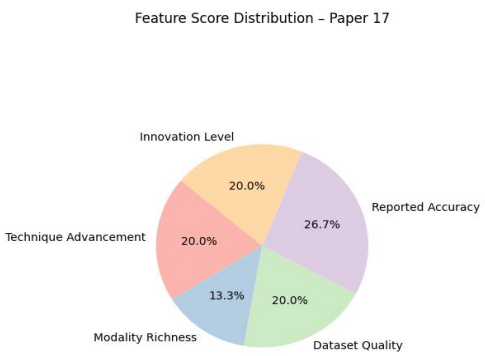

Figure 28. Paper 17

The set of pie charts spanning Figures 12 to 28 provides a comprehensive visualization of how seventeen reviewed studies performed across the five key evaluation criteria: Technique Advancement, Modality Richness, Dataset Quality, Reported Accuracy, and Innovation Level. Each pie chart depicts the relative contribution of these dimensions within a given study, thereby offering a comparative overview of methodological strengths and weaknesses. High-performing studies, such as Paper 8, Paper 4, and Paper 1, show balanced distributions across the charts, excelling particularly in innovation, dataset quality, and accuracy. These visualizations illustrate how cutting-edge models—particularly transformer-based architectures and hybrid methods—have been most effective when combined with large, well-annotated datasets. Conversely, weaker studies, including Papers 5, 6, and 13, display skewed pies with disproportionately small shares for dataset quality and modality richness, underscoring the limitations of traditional machine learning methods, reliance on single-modality inputs, or use of limited datasets. Taken collectively, the sequence of charts provides not only a ranking of research quality but also an empirical lens through which to interpret broader trends in sentiment-based mental health monitoring.

A critical observation across Figures 12–28 is the consistent underrepresentation of Modality Richness, which remains the smallest proportion in most charts. This highlights an ongoing research gap: while substantial effort has been directed toward improving technical sophistication and predictive accuracy, far less emphasis has been placed on multimodal approaches that integrate behavioral, temporal, and visual signals alongside textual data. Such imbalance has practical consequences, as mental health symptoms are rarely confined to language alone but are expressed through diverse modalities that require holistic representation. Furthermore, dataset quality shows notable variability across the charts, with several studies relying on non-public or narrowly defined datasets, raising concerns about reproducibility and generalizability. Thus, although the pie charts demonstrate the field's strong trajectory in technical advancement and accuracy optimization, they also reveal systematic neglect of data diversity and multimodal integration. Collectively, Figures 12–28 underscore the dual reality of the field: significant methodological progress coupled with persistent foundational gaps. This contrast provides a roadmap for future research—prioritizing standardized, high-quality datasets and multimodal frameworks while ensuring ethical and transparent practices in

applying artificial intelligence for mental health monitoring.

1. **Emerging Techniques**:

- Increasing popularity of transformer-based models (e.g., BERT,DistilBERT) to be used in sentiment and emotion classification.
- More experimentation with transfer learning and fine-tuning with respect to mental health-specific corpora.

Table 2: Techniques and Models Used

Paper	Technique Type	Specific Model Used	Feature Type	Pretrained
P1	Deep Learning	BERT	Text	Yes
P2	Machine Learning	SVM	Bag-of-Words	No
P3	Deep Learning	LSTM + Attention	Text + Metadata	Yes
P4	Hybrid	XGBoost + TF-IDF	Text	No
P5	Machine Learning	Naive Bayes	Text	No
P6	Deep Learning	DistilBER	Text	Yes
P7	Machine Learning	Random Forest	Text	Yes
P8	Hybrid	BERT+SVM	Text + Behavior	No
P9	Deep Learning	RoBERTa	Text+ Time	Yes
P10	Machine Learning	Decision Tree	Text	No
P11	Transformer-based	DeBERTa / BERT / RoBERTa	Text	Yes
P12	Deep Learning	DistilRoBERTa + Emotion	Text	Yes
P13	Ethical ML Framework	Various ML (anonymized)	Text	No
P14	Hybrid DL	BERT-Fuse (BERT+CNN+LSTM+GRU+TF-IDF+BoW	Text	Yes
P15	Hybrid Transformer	BERT–BiLSTM with precision modules	Text	Yes
P16	ML + Temporal	DCWE/TWEC + SVM/FFNN/CNN	Text	No
P17	ML vs DL comparison	Logistic Regression / LightGBM / DL models	Text	Mixed

Table 2 represented different techniques and models data

2. **Most Used Datasets**:

- The most frequent were the Twitter sets and often based on mentions of health-related

hashtagging or search filters (e.g., depression, anxiety) that is represents in table 3.

- Several articles employed Reddit and domain-specific corpora (e.g., CLPsych, DAIC-WOZ), however, the size and labeling of the datasets differ.

Table 3: Existing Datasets Used in Reviewed Papers

Paper	Dataset Name	Platform	Size	Label Type	Publicly Available?
P1	CLPsych 2015	Reddit	10,000	Depression	Yes
P2	Twitter Mental Health	Twitter	15,000	Stress, Anxiety	Yes
P3	Private Dataset	Instagram	5,000	Depression	No
P4	eRisk 2017	Reddit	20,000	Early Depression	Yes
P5	Reddit Users	Reddit	8,000	Mental Health	No
P6	TW Stress	Twitter	12,000	Stress	Yes
P7	Reddit Mental	Reddit	9,500	Depression	Yes
P8	Custom Tweets	Twitter	6,000	Anxiety	No
P9	MH Dataset 2022	Twitter	13,000	Stress, Depression	Yes
P10	Reddit Anon	Reddit	7,500	Suicidal Ideation	Yes
P11	Reddit 24k (Severity)	Reddit	24,000	Severity (Mild/Mod/Sev)	Yes
P12	Mental Health Corpus	Twitter	27,292	Poisonous/Non-Poisonous	Yes
P13	Reddit + eRisk	Reddit	Various	Multiple conditions	Yes
P14	50K Mental Health Dataset	Text	50,000	Multi-class sentiment	No
P15	Depression-labeled Dataset	Text	Single-domain	Depression vs Non	No
P16	eRisk 2017–2023	Reddit	~13,000	Multiple (depression, anorexia...)	Yes
P17	Aggregated Multi-source Posts	Text	Static (multi-source)	Binary/Multi	Yes

3. **Multimodality is Underutilized**:
The papers with high scores used multimodal analysis very rarely, combining user behavior or features with textual contents shown in table 4.

- Research involving such divergent inputs was often more successful than ones involving the utilization of single-modality data only.

Table 4: Modality Usage Across Papers

Paper	Text	Behavioral Features	Temporal Info	Social Graph	Multimodal
P1	✓	✓	✗	✗	✓
P2	✓	✗	✗	✗	✗
P3	✓	✓	✓	✗	✓
P4	✓	✓	✗	✓	✓
P5	✓	✗	✓	✗	✗
P6	✓	✗	✗	✗	✗
P7	✓	✓	✗	✗	✓
P8	✓	✓	✓	✗	✓
P9	✓	✓	✓	✓	✓
P10	✓	✗	✗	✗	✗
P 11	✓	✗	✗	✗	✗
P12	✓	✗	✗	✗	✗
P13	✓	✗	✗	✗	✗
P14	✓	✗	✗	✗	✗
P15	✓	✗	✗	✗	✗
P16	✓	✗	✓	✗	✗
P17	✓	✗	✗	✗	✗

Discussion

As can be seen in the comparative analysis, the two aspects of the technical innovation and model performance are effectively addressed in the existing literature sources, whereas, the definition of the modality diversity and data quality gap is obvious. Research in the future ought to attempt to merge more detailed and multi-source data and develop normative sets of data so that evaluation is more standardized. Such initiatives have the potential to promote the reliability, generalizability and real-life usefulness of systems of monitoring mental health based on social media.

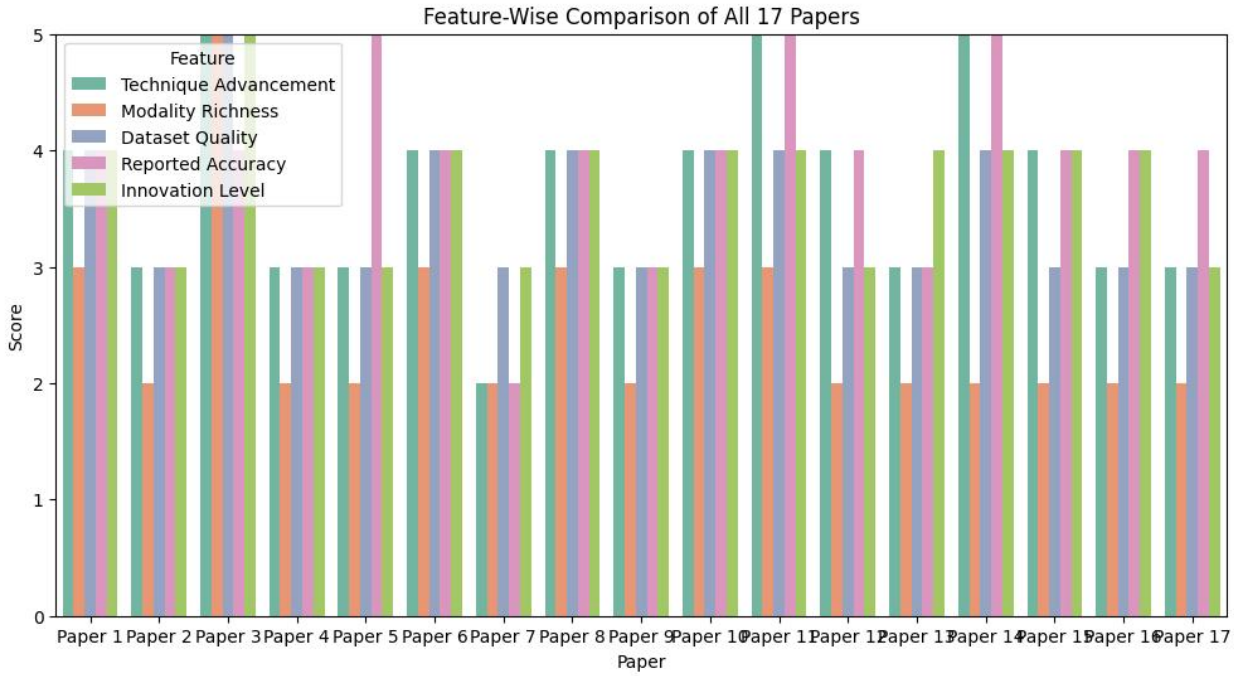

Figure 29. Comparative Feature Score

Strengths and Limitations

Strengths

The current work provides a systematic and integrated system that would analyze studies in the field of mental health surveillance through sentiment analysis of social networks. Using uniform five assessment metrics which are Technique Advancement, Modality Richness, Dataset Quality, Reported Accuracy, and Innovation Level, one can make a fair and straightforward comparison between various methods. The second strength is the visual representation of the data with the visual tools, i.e., using bar graphs, radar charts, pie charts, and heatmaps, based on which one can intuitively feel the differences and similarities between the chosen studies. Such visualizations can be used to determine their shared strengths (e.g., sophistication of models) and weaknesses (e.g. limitations of modality) more conveniently than through a crude textual analysis only. The comparative scoring system has the advantage of bringing out the gaps in performances and the emerging trends and is thus an asset, which the researchers can use as a reference point to improve on or as a benchmark in their research work.

Limitations

In spite of this advantage, the research has some weaknesses:

- Subjectivity in Rating: In spite of using an organized rubric, some metrics scored, which are based on some degree of subjectivity are the level of innovation or the richness of modalities. This adds a certain subjectivity to it that can affect the entire rankings.
- Narrow Range: There are many papers that have been selected to make the comparison, and though they are quite diverse and representative, not all the research of the sentiment-based mental health tracking might be discussed in the comparison. Additional papers and sources of the analyses would fortify the generalizability of the findings.
- Static Evaluation: It is a study that is based on published results at one point in time. It fails to consider gradual advances, evolving sets of data, and new models that might have come to light since the process of choosing the papers was under way.

Conclusion

Based on a common criterion of five parameters, i.e. Technique Advancement, Modality Richness, Dataset Quality, Reported Accuracy and the Level of Innovation, this comparative analysis compared many recent papers in the field of monitoring mental health via sentiment analysis on social media. The findings showed the evident directions in the research practices and allowed determining the strengths and shortcomings of the field.The vast majority of papers already showed serious advances in technical model development, especially the use of transformer-based models such as BERT and its variants. Competitive accuracy rates were also described as reported by a majority, which revealed that sentiment analysis-related tools were gaining more ground becoming reliable in establishing the prevalence of mental health indicators in text data.Nevertheless, significant gaps were revealed in the analysis, as well. Most of the studies showed a low grade of Modality Richness and only used the textual features of the texts with no attention to the behavioral or visual features or to the metadata features. Also, Dataset Quality was not very consistent as certain studies relied on small and not standard data sets, which created doubts around generalizability and reproducibility.

Future Work

Future research ought to concentrate on the following based on the comparative analysis made to promote the field:

1. Multimodal Fusion Techniques

Research on integrating textual, behavioral, temporal, and even visual data (e.g., emojis, images, engagement patterns) to induce a more complete understanding of the model in order to increase predictive capabilities should be conducted in the future.2. Creation of Common Benchmark DatasetsDatasets that are extensively bigger, annotated, and serves publicly available in specific detection of mental health is urgently needed. With standardised datasets, it will be possible to consistently report and conduct fair comparison between models.3. Explainable Artificial Intelligence and Model Transparency Since mental health is a delicate area to enter, future models must include explainability methods to enable people to interpret the predictions, which will allow clinicians and users to be more trustful of the model and foster usability.4. Cross-Platform and Cross-Lingual Research The majority of existing research deals with the Twitter data in English. Those strategies should be applied to more platforms (e.g., Reddit, Tik Tok) and different languages in the future to enhance the sense of inclusivity and cultural relevance.

. Longitudinal Modeling User-CentricThe shift is needed between the one-time classification to long-term monitoring of mental health, which introduces the possibility of early warning and tailored interventions. The use of the user history and the longitudinal trends in the user behavior can be incredibly efficient.

. 6. Privacy-Aware Systems and Ethical Systems In future models, ethical considerations should be met to guarantee user privacy, informed consent, and bias countering in the collection and analysis of data, especially in mental health cases.

On the whole, the research may be used as a guide to researchers who want to discover the trends, quantify the current advancement, and focus future activities on more complex and effective models of mental health detection with the help of social media data.

Reference

1) World Health Organization. World mental health report: Transforming mental health for all. World Health . (2022).
2) Aishwarya Daga et al. (2025). Leveraging Emotions for Enhanced Mental Health . *2025 11th International Conference on Communication and Signal Processing (ICCSP)*, 5.
3) al., Z. N. (2023). "Depression detection in social media comments data using machine learning algo. *Bull. Electr. Eng. Inform. 12.2* , 10.
4) AmnaQasim et al. (2025). Detection of Depression Severity in Social Media Text Using. *Information*, 23.
5) andQingzhongLiu, B. G. (2023). Deep Learning-Based Depression Detection from Social Media:. *electronics*, 20.
6) andQingzhongLiu, B. G. (2024). Advanced Comparative Analysis of Machine Learning and. *electronics*, 22.
7) C. Otte, S. G. (2016). Major depressive disorder,.
8) Costa, V. V. (2025). Depression Detection in Reddit Posts Through. 15.
9) Ding, Z. (2025). Trade-offs between machine . *www.nature.com/scientificreports*, 14.
10) Hartono Wijaya 1, M. H. (2025). Using Sentiment Analysis with BERT and SVM for Detect Mental Health . *SHIFANA: Journal of Digital Health Innovation and Medical* , 10.
11) Henna I. Vartiainen1, T. D. (2024). *Diverse approaches to sentiment analysis reliably reflect and explain.*
12) M. Moradi, M. D. (2022). "Global prevalence of depression among heart failure. *Cur, vol. 47.*
13) Manuel Couto et al. (2025). TemporalWordEmbeddingsforEarlyDetection. *Journal of Healthcare Informatics Research*, 30.
14) MD. MITHUN HOSSAIN et al. (2025). Revolutionizing Mental Health Sentiment. *IEEE Access*, 19.
15) Motlagh, G. (2022). Novel Natural Language Processing Models for Medical Terms and Symptoms Detection in Twitter.
16) R. C. Waumans, A. D. (2022). "Barriers and facilitators for treatment-seeking in adults. *BMC, vol. 22,*.
17) Raja Kumar et al. (2024). Mental Disorder Classification via Temporal Representation of Text. *Findings of the Association for Computational Linguistics: EMNLP 2024,*, 16.
18) Settanni M, M. D. (2015). Sharing feelings online: studying emotional well-being via automated text analysis of Facebook. *Front Psychol.*
19) SHIWEN ZHOU et al. (2025). Mental Health Safety and Depression Detection in. *IEEE Access*, 14.
20) Shiyu Teng1, J. L.-W. (2025). Enhanced Multimodal Depression Detection With Emotion Prompts. 5.
21) Siti Nurulain Mohd Rum1, *. N. (2025). Uncovering Depression on Social Media using BERT Model . *Journal of Advanced Research Design 129*, 14.
22) Ullah, W. (2024). Identification of depressing tweets using natural language processing and . *Journal of Radiation Research and Applied Sciences*, 13.
23) Xingwei Yang1, P., & Guang Li2, P. (2025). Psychological and Behavioral Insights From Social Media Users:. *JMIR*, 22.
24) Yang, S. (2024). Enhancing multimodal depression diagnosis through . *Heliyon* , 14.