Mathematical Modelling of Engineering Problems
Vol. 12, No. 9, September, 2025, pp. 3289-3295
Journal homepage: http://iieta.org/journals/mmep

A Comparative Study of BERT and RoBERTa for Sentiment Analysis on Twitter Data Related to Mental Health

Mohammad I. Fahmi[1*], Adli A. Nababan[2]

[1] Department of Information System, Faculty of Science and Technology, Universitas Prima Indonesia, Medan 20118, Indonesia
[2] Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta 11480, Indonesia

Corresponding Author Email: mohammadirfanfahmi@unprimdn.ac.id

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/mmep.120931

Received: 9 July 2025
Revised: 6 September 2025
Accepted: 12 September 2025
Available online: 30 September 2025

Keywords:
BERT, Robustly Optimized BERT Pretraining Approach (RoBERTa), sentiment classification, mental health, Twitter, natural language processing, social media analysis

ABSTRACT

The increasing prevalence of mental health issues exacerbated by the COVID-19 pandemic has sparked a surge in related discussions on social media, particularly Twitter (now known as X). These discussions serve as a valuable data source for analyzing public sentiment regarding mental health. This study aims to compare the performance of two widely used transformer models, BERT and Robustly Optimized BERT Pretraining Approach (RoBERTa), in classifying sentiment in tweets related to mental health. A dataset of 10,000 English-language tweets was collected through the Twitter API using the keywords "depression," "anxiety," "stress," and "mental health." After preprocessing, the data were categorized into three sentiment classes: positive, negative, and neutral. Both models were fine-tuned with the same parameters and evaluated using accuracy, precision, recall, and F1-score. The results showed that RoBERTa consistently outperformed BERT on all metrics, achieving 90.45% accuracy, 89.78% precision, 91.02% recall, and 90.39% F1-score. In comparison, BERT only achieved 87.60% accuracy and 85.74% F1-score. RoBERTa's superiority is attributed to its larger pretraining corpus and the removal of the Next Sentence Prediction (NSP) task, resulting in better understanding of informal language and emotional expressions on social media. This study confirms RoBERTa's potential as a more effective Natural Language Processing tool for real-time monitoring and early detection of mental health conditions through social media analysis.

1. INTRODUCTION

One global issue that requires serious attention is the increasing social stress and isolation triggered by the COVID-19 pandemic. According to the World Health Organization (WHO), more than 970 million people worldwide suffer from mental disorders, and this number continues to rise annually [1]. Mental health disorders are now a global issue that is gaining increasing attention. In this context, social media plays a crucial role as a platform for people to express their feelings, share their experiences, and discuss their psychological conditions, including symptoms of stress, depression, and anxiety [2]. One of the most frequently used platforms is X (formerly Twitter), which generates large amounts of data and can be used to capture public perceptions and sentiments regarding mental health issues directly [3].

However, data from social media such as Twitter/X is noisy and informal, often using slang, acronyms, and specific social contexts that are difficult for traditional text analysis methods to understand [4]. Social media has even been viewed as a passive sensor of mental health, as online interaction patterns can reflect an individual's psychological state. However, this approach still faces ethical and privacy challenges [5]. A recent systematic review emphasized that Natural Language Processing (NLP) approaches, particularly transformer-based ones, are urgently needed to handle the informal and noisy characteristics of social media text [6].

Previous research has shown that deep learning approaches are able to effectively detect depression-related posts on social media, thus supporting faster and more appropriate mental health interventions [7]. Deep learning-based NLP continues to experience rapid development, with recent reviews confirming that transformers are the backbone of current language models [8]. Among various models, BERT is still widely researched, and comparative evaluations show its strengths as well as limitations when compared to lightweight variants such as DistilBERT and more advanced models such as RoBERTa [9]. To address these limitations, a new RoBERTa-based approach has been developed. The hybrid RoBERTa model has been shown to consistently outperform BERT in sentiment classification tasks [10].

To address this, the Robustly Optimized BERT Pretraining Approach (RoBERTa) was born in 2019. This model removes the task of predicting the next sentence, trains with a much larger dataset, and uses a longer training time. As a result, RoBERTa consistently outperforms BERT on various NLP

benchmarks [11, 12].

Although both models have been widely used in various domains, there is still limited research that specifically compares the performance of BERT and RoBERTa in sentiment analysis related to mental health on social media [13]. While both are widely used in various fields, research comparing the performance of BERT and RoBERTa specifically in sentiment analysis about mental health on social media is still limited. This type of analysis requires a model that is highly sensitive to users' emotional expressions and distinctive language [14].

Based on this research gap, this study aims to conduct a comparative study between BERT and RoBERTa in sentiment classification on Twitter data related to mental health issues. The results are expected to aid the development of more accurate NLP models for online mental health monitoring and serve as the foundation for early detection systems that benefit health institutions, social organizations, and policymakers.

2. LITERATURE REVIEW

Research on social media-based mental health detection is growing as awareness grows that platforms like Twitter can be reflection of people's psychological well-being. According to Di Cara et al. [15], linguistic expressions recorded on social media can reflect psychological states in real time, enabling these platforms to function as "passive sensors" for mental health. However, as Rafieivand et al. [16] emphasize, these platforms can also be used to assess psychological well-being, the use of large-scale language models (LLMs) also faces ethical challenges, transparency, and risks of misuse, which require a careful research approach.

Transformer-based models, particularly BERT, have been shown to provide strong performance in sentiment analysis. Bello et al. [17] demonstrated BERT's effectiveness in classifying sentiment from tweets, demonstrating that the model is capable of capturing the nuances of informal language well. However, in terms of social impact, Shannon et al. [18] have successfully improved the accuracy of emotion classification by considering the identity of the speaker in a conversation. However, these studies have not specifically focused on detecting mental health from tweets and have emphasized validation in general medical domains or structured conversations [19].

In addition, the review of Salas-Zárate et al. [20] found that the majority of research still focuses on English-language Twitter and tends to evaluate models using traditional approaches. This limits the generalizability of the results, both linguistically and across platforms. They recommend adopting transformer-based approaches such as BERT or RoBERTa, which have been shown to be more adaptive in understanding emotional context. Recent studies on the classification of conversational emotions [21] demonstrates the potential of a hybrid approach. Additionally, temporal topic models on Twitter demonstrate that tracking depressive symptoms over time can provide important insights into their developmental patterns [22]. Lexicon-based approaches also remain useful in detecting depression in certain online communities, such as university students [23].

Based on this gap, this study offers a novel contribution by comparing the performance of BERT and RoBERTa specifically on detecting mental health symptoms from tweets. This focus is important because Twitter is a medium that best represents spontaneous emotional expression but is also full of the complexity of everyday language. By comparing two transformer architectures, this study not only tests the effectiveness of the models in detecting signs of depression, anxiety, and stress but also provides insight into which model is more optimal for the task of social media-based mental health detection. Recent research also shows that multimodal transformer models are beginning to be developed for mental health applications. These models combine text analysis with other data such as audio and images, providing a more comprehensive understanding of psychological conditions [24].

In this context, a comparative study of BERT and RoBERTa is important because they represent two dominant transformer architectures with distinct characteristics. BERT, while effective, still has limitations in Next Sentence Prediction (NSP) tasks, which can slow down context understanding. In contrast, RoBERTa optimizes by removing NSP, training with a larger dataset, and adjusting the tokenization process to make it more robust against informal language on social media. Therefore, comparing these two models in the Twitter-based mental health domain contributes not only to evaluating technical performance but also to understanding which model is more appropriate for detecting spontaneous psychological expressions in digital spaces.

3. RESEARCH METHOD

This study uses a transformer-based representation capable of capturing word context more deeply than traditional methods. The contextual embedding approach has been shown to improve sentiment analysis performance, particularly on Twitter data, which is dynamic and full of linguistic variation [25].

As cross-border social media usage increases, recent research underscores the importance of developing transformer models capable of conducting multilingual mental health analyses to achieve more representative results [26]. Through this approach, we evaluate both algorithms based on accuracy, precision, and computational efficiency in the context of a large and complex dataset.

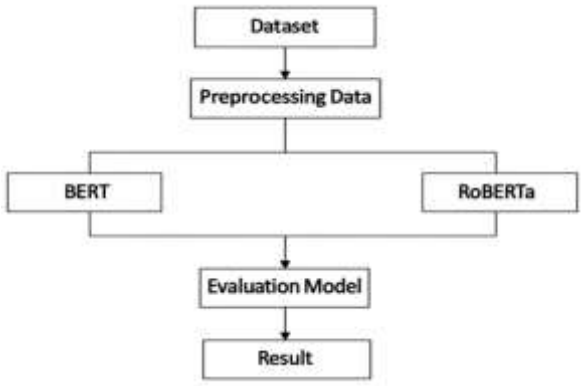

Figure 1. Research flow

To provide a clearer understanding of the methodology used in this study, the research flow in Figure 1 is depicted in the following diagram. This flowchart summarizes the main stages involved, starting with data collection and preprocessing, followed by fine-tuning the BERT and

RoBERTa models, and finally evaluating model performance. By presenting the flowchart systematically, this diagram aims to demonstrate a structured research approach that supports the validity and reliability of the results.

The data was collected from platform X (formerly known as Twitter) using the Twitter API. The keywords used for data retrieval were mental health, anxiety, depression, and stress. The data collection was conducted over a period from January to May 2025. Additionally, only English-language tweets were considered, and a data cleaning process was applied to remove spam, retweets, and tweets containing irrelevant links or media. The final dataset used in this study consisted of 10,000 tweets as shown in Table 1.

Table 1. Dataset

	conversation_id_str	created_at	favorite_count	full_text
0	1881850000000000000	Tue Jan 21 23:38:23 +0000 2025	0	1 in 5 U.S. adults face mental health challenges yearly affecting their well-being. If you or a loved one are impacted tune in for helpful remedies. Watch now: https://t.co/5LIA3bdWab #MentalHealth #Wellness #YouAreNotAlone
1	1881850000000000000	Tue Jan 21 23:27:56 +0000 2025	0	Aye @Ravens do yall have a mental health and wellness coach?? Along the lines of @Ericthomasbtc or @garyvee to get over and thru those big moments?? Our mistakes simply seem mental. The talent and coaching have been there the last 2 seasons.
2	1881850000000000000	Tue Jan 21 23:25:39 +0000 2025	1	One of Wagga Wagga's leading mental health advocates has called on all levels of government to support the establishment of a men's wellness centre in town https://t.co/wxDUOBAxJv
3	1881840000000000000	Tue Jan 21 23:00:23 +0000 2025	1	Ka Niyah Baker 13 was murdered and two teenage girls 15 and 16 have been arrested. Experts say major changes are needed to curb teen homicides. https://t.co/qK8BOaOw1F
4	1881830000000000000	Tue Jan 21 22:17:18 +0000 2025	1	A 13-year-old's brutal murder the teen girls who were arrested and what parents can learn https://t.co/7L2beRoleK @usatoday
5	1881820000000000000	Tue Jan 21 22:00:04 +0000 2025	0	Show your commitment to #employees by investing in proactive #mentalhealth initiatives. #TeksMed is excited to partner with Dr. Steve Conway to offer valuable mental wellness #workshops. Discover our available workshops through https://t.co/VLCK2PmgfE https://t.co/IR2lNOSyLO
6	1881820000000000000	Tue Jan 21 22:00:02 +0000 2025	1	Take the right steps for your mental wellness. Protect your peace of mind when feeling stressed anxious depressed or lonely. Professional support is readily available. You can #TalkItOut Visit https://t.co/i2A7XSgTod or https://t.co/9lpt88uefV for more resources. https://t.co/dBm1Cao8bv
⋮	⋮	⋮	⋮	⋮
10699	1928170000000000000	Fri May 30 21:59:45 +0000 2025	0	@thehonestlypod @jaketapper They pushed harder against Trump's perceived dishonesty about his health and mental acuity than they did this very real issue with Biden's. This is the problem and I say this as a non-Trump supporter but I most certainly have gained sympathy and it's mostly due to the media.

The mathematical formulation of the Transformer-based models used in this study, specifically BERT and RoBERTa, is presented below to provide a clearer explanation of the internal representations and attention mechanisms involved in processing input sequences.

$$\text{Attention}\,(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$

where:

Q = Query Matrix
K = Key Matrix
V = Value Matrix
d_k = Dimension of the key vector

Before being used in the sentiment classification model training process, Twitter data needs to go through a preprocessing stage to remove elements that are distracting and do not have meaningful semantic contributions. The initial stage in this process is cleaning the text from various non-text symbols, such as emojis, hashtags, mentions, and external links. Emojis or emoticons such as 😊 or 😂 often convey emotional expressions, but they cannot be processed directly by text-based models, and need to be removed. Likewise, hashtags containing certain topics (e.g. #mentalhealth), user mentions (@username), and links (e.g. http://...) are also removed to maintain data consistency and reduce noise. In addition, text normalization is carried out to simplify words that are written excessively or non-standard, such as "heelllppp" which is adjusted to "help," in order to improve readability by the model.

After the data is cleaned, the process continues with tokenization, which is breaking the text into small units or tokens. This tokenization is carried out using the default tokenizer from the pre-trained model used; for example, BERT uses the bert-base-uncased tokenizer, while RoBERTa uses the roberta-base tokenizer. The tokens are then converted into numeric representations, such as token IDs and attention masks, so that they can be processed by the model. Given that the length of each tweet varies, padding techniques are used to equalize the length of shorter inputs, and truncating to cut off inputs that exceed the maximum length (usually 128 tokens). The tokenized and normalized data is then used in the fine-tuning process of Transformer models such as BERT and RoBERTa. After training is complete, the model performance is evaluated using a few standard metrics, such as accuracy,

precision, recall, and F1-score, to assess the effectiveness of the model in classifying sentiment from social media texts that have been systematically processed.

In addition to the tokenization process, the next important step is creating an attention mask, which distinguishes between original tokens and padded tokens. The attention mask is 1 for tokens that are truly from the original text, while it is 0 for padded tokens. This allows the model to only pay attention to the relevant parts of the text without being distracted by padding. This process is crucial because in Twitter data, text length varies, and without a masking mechanism, the model could misinterpret the context of the input.

Next, the tokenized numerical representation is fed into the transformer architecture. At this stage, the initial embedding is processed through several layers of multi-head self-attention, allowing the model to capture relationships between words even when they are far apart in a sentence. For example, in a tweet containing an emotional phrase like "I feel hopeless today," the model is able to understand the relationship between the words "hopeless" and "feel" despite the distance between them. This is the advantage of BERT and RoBERTa over traditional NLP approaches, which tend to ignore long-term dependencies.

Finally, after passing through the transformer encoder layer, the output, a contextual vector, of each token is processed to produce a sentence representation. In BERT, a specific representation is derived from the [CLS] token, while in RoBERTa it uses the <s> token. This representation is then sent to the classification head to determine the sentiment category: positive, negative, or neutral. In this way, the model not only reads the text literally but also understands the emotional nuances contained in the tweet. Evaluation using metrics such as accuracy, precision, recall, and F1-score ensures that the model's performance can be objectively measured in the context of mental health detection.

Furthermore, the fine-tuning process applied to both models involves adjusting the pre-trained weights to better suit the mental health domain. At this stage, the model re-learns from the Twitter dataset by adjusting internal parameters through an optimization algorithm such as AdamW. This process is typically performed over multiple epochs with a carefully set learning rate, as too large a value can cause the model to fail to converge, while too small a value can slow down the training process.

Another equally important aspect is regularization strategies, such as the use of dropout and early stopping. Dropout is applied to neural network layers to prevent overfitting by "ignoring" some neurons during the training process. Meanwhile, early stopping is implemented by stopping training when performance on validation data begins to decline, even though accuracy on training data continues to increase. This way, the resulting model is more general and can perform better when faced with new, previously unseen data.

In practical implementation, the designed pipeline also takes computational efficiency into account. Training is performed using GPUs to accelerate the process, given the large number of parameters in BERT and RoBERTa. Furthermore, the batch size and sequence length are adjusted to ensure they do not exceed the memory capacity of the hardware used. This consideration is crucial because in real-world applications, NLP models for mental health detection often need to be implemented in real-time monitoring systems, requiring accuracy and efficiency in terms of time and resources. The training workflow comparison between BERT and RoBERTa is illustrated in Figure 2.

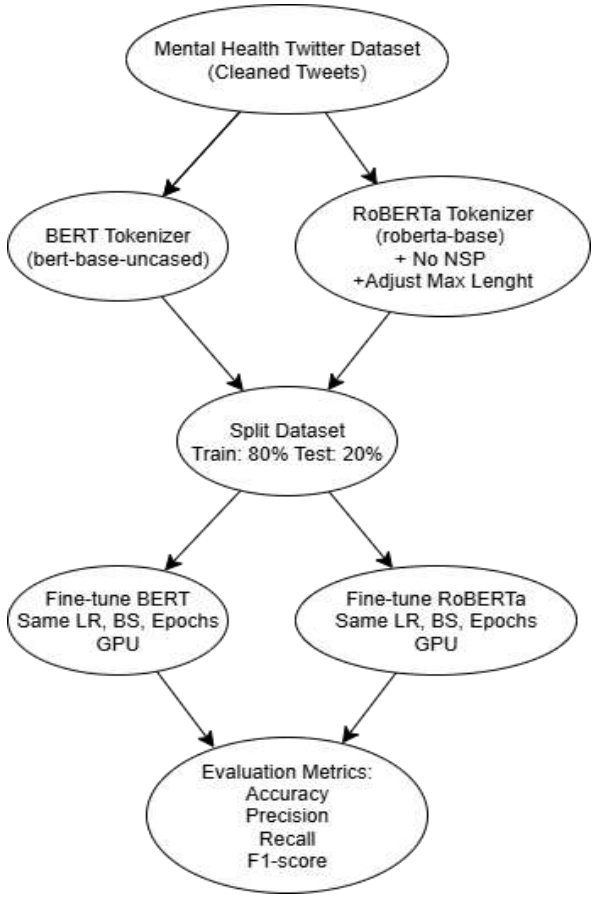

Figure 2. BERT vs. RoBERTa training workflow

The algorithm used in this study follows a fine-tuning approach based on Transformer architecture models: BERT and RoBERTa. Both models were imported from the Hugging Face Transformers library, using the English pre-trained versions: BERT-base-uncased and RoBERTa-base, respectively. The process began by tokenizing the tweets using each model's tokenizer. The cleaned dataset was then split into training and testing sets with an 80:20 ratio and adapted for three-class sentiment classification (positive, negative, and neutral).

The final dataset in this study consisted of 10,000 clean tweets. The dataset was divided into 80% training data (8,000 tweets) and 20% testing data (2,000 tweets). Both models, BERT and RoBERTa, were trained on the same training data to ensure fair comparisons. However, during the testing phase, there was a slight difference in the number of samples. BERT successfully processed 950 test samples, while RoBERTa only processed 948 test samples. This difference was caused by two samples from the Neutral class whose token length exceeded the maximum limit set by the RoBERTa tokenizer and was therefore automatically discarded during preprocessing. We transparently report this difference, but the difference in the two samples does not affect the validity of the performance comparison because the class proportions remain consistent across both models [27].

Although both algorithms share a fundamentally similar architecture, there are several key differences in the implementation of BERT and RoBERTa. Notably, RoBERTa

does not use NSP during pretraining, resulting in slight differences in the tokenizer and input pipeline, including how special tokens and padding are handled. Additionally, RoBERTa requires consistent input processing for longer token lengths, so the maximum sequence length was adjusted accordingly for this model. Here is a comparison table of BERT and RoBERTa models in Table 2.

Table 2. Comparison BERT and RoBERTa

Aspect	BERT	RoBERTa
Pretraining Task	Masked LM + NSP	Masked LM only (no NSP)
Pretraining Corpus	BooksCorpus + English Wikipedia	CC-News, OpenWebText, Stories, Books
Tokenizer Type	WordPiece	Byte-Pair Encoding (BPE)
Input Sequence Length	Max 512 tokens	Max 512 tokens (longer processed more consistently)
Special Tokens	[CLS], [SEP]	<s>, </s>
Padding Strategy	Tokenized using WordPiece-aware padding	BPE-aware padding

To evaluate model performance, four main metrics are used: Accuracy, Precision, Recall, and F-score (F1-score). The mathematical formula for each metric is shown below:

(1) Accuracy:

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

Shows the proportion of correct predictions (positive and negative) to the total test data.

(2) Precision:

$$Precision = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

Measures how accurate the model is in classifying positive data, that is, of all positive predictions, how many are positive

(3) Recall:

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

Measures the model's ability to find true positive data, that is, of all positive data, how many were successfully detected by the model

(4) F1-Score:

$$F1 - Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

It is the harmonic meaning between Precision and Recall, providing a balance between the two especially in unbalanced data

- True Positive (TP): the number of positive data points predicted as positive.
- True Negative (TN): the number of negative data points predicted as negative.
- False Positive (FP): the number of negative data points incorrectly predicted as positive.
- False Negative (FN): the number of positive data points incorrectly predicted as negative.

In this experiment, both models were trained using the same training parameters including learning rate, batch size, and number of epochs to ensure a fair comparison of evaluation results. All training processes were conducted using Graphic Processing Unit (GPU) processing units to improve time efficiency. The performance of each model was then evaluated using accuracy, precision, recall, and F1-score metrics to assess their effectiveness in classifying sentiment in tweets related to mental health.

The performance of each algorithm will be evaluated based on several metrics: accuracy, which indicates the percentage of correctly classified emails; precision, which measures the proportion of true positive predictions among all positive predictions; and recall, which reflects the proportion of true positive predictions among all actual positive data. Additionally, the F1-score is used as the harmonic mean of precision and recall providing a comprehensive overview of the model's performance. The evaluation also includes computational time, referring to the duration required for the training and testing processes.

Based on the research conducted, it is evident that the RoBERTa model outperforms BERT across all key performance metrics. Although RoBERTa requires slightly longer computational time (approximately 14 seconds more), its performance in sentiment classification is significantly superior. RoBERTa demonstrates a greater ability to capture sentence context more accurately due to its optimized pretraining architecture. The removal of the NSP task, along with training on a larger dataset and for a longer duration, enables RoBERTa to be more adaptive to linguistic diversity on social media, including slang and informal language styles.

Table 3. Performance evaluation

Evaluation Matrix	BERT	RoBERTa
Accuracy	87.60%	90.45%
Precision	86.12%	89.78%
Recall	85.37%	91.02%
F1-Score	85.74%	90.39%
Computational Time	142 Sec	156 Sec

Based on the evaluation metrics as shown in Table 3, RoBERTa demonstrates superior performance compared to BERT across all key measures, including accuracy (90.45% vs. 87.60%), precision (89.78% vs. 86.12%), recall (91.02% vs. 85.37%), and F1-score (90.39% vs. 85.74%). These results indicate that RoBERTa is more effective and consistent in identifying and classifying sentiment. Although RoBERTa requires slightly longer computational time (156 seconds compared to BERT's 142 seconds), its improved accuracy and enhanced ability to capture sentence context make it a more effective choice for sentiment classification tasks.

4. CONCLUSIONS

The results of this study show that RoBERTa consistently outperforms BERT across all evaluation metrics, including accuracy, precision, recall, and F1-score. This superiority aligns with the theory that pre-training optimization, the use of a larger dataset, and the removal of NSP make RoBERTa more adaptable to informal language often found on social media. These findings corroborate previous research reports that also indicated RoBERTa's superiority in sentiment analysis tasks

[10, 11].

Furthermore, the small difference in computation time (14 seconds longer for RoBERTa) is relatively insignificant compared to the performance improvements achieved. This confirms that the trade-off between efficiency and effectiveness is still within reasonable limits for practical applications, especially in the context of social media-based mental health monitoring, which requires high accuracy.

However, this study also identified limitations that warrant attention. First, the dataset used only included 10,000 English-language tweets, so generalizability to other languages or cultural contexts is limited. Second, this study only evaluated text as the sole modality, whereas mental health is often also reflected in non-verbal signals such as images, videos, or online interaction patterns. This opens opportunities to integrate multimodal approaches, as suggested by Cha et al. [23].

Then, this study not only provides empirical evidence regarding the performance of BERT and RoBERTa but also highlights the need to develop models that are more inclusive and adaptive to linguistic, temporal, and multimodal variations in social media-based mental health detection.

5. SUGGESTION

This study confirms that RoBERTa has significant advantages over BERT in sentiment classification on Twitter data related to mental health. This superiority is reflected in improved accuracy, precision, recall, and F1-score, making RoBERTa more effective in capturing the emotional expressions of social media users. Although requiring slightly longer computational time, these results demonstrate that efficiency is not a major barrier when the performance gains are significant.

The primary contribution of this study is providing a clear comparative evaluation between two popular transformer models in the social media-based mental health domain, while also filling a previously limited literature gap. Practically, these findings support the use of RoBERTa as part of an online mental health monitoring system, potentially used by healthcare professionals and policymakers for early detection.

However, this study also suffers from limitations in the size and scope of the dataset, as well as its limited use of monolingual data. Therefore, future research directions could focus on:

(1) Expanding the dataset to a larger, multilingual dataset.

(2) Incorporating multimodal approaches that integrate text, audio, and images.

(3) Developing a more computationally efficient model for large-scale implementation in real-time.

ACKNOWLEDGMENT

The authors would like to acknowledge Universitas Prima Indonesia and Bina Nusantara University for providing research facilities and institutional support.

Mohammad I. Fahmi conceptualized the research design, performed data preprocessing, and implemented the sentiment analysis models. Adli A. Nababan contributed to the methodological framework, supervised the model evaluation process, and revised the manuscript for technical and academic rigor.

DATA AVAILABILITY

The dataset used in this study, titled Mental Health Twitter Dataset for Sentiment Analysis (Jan–May 2025), is publicly available on Zenodo at https://doi.org/10.5281/zenodo.17144715.

REFERENCES

[1] World Health. Organization. Mental health. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response.

[2] Chancellor, S., De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. NPJ Digital Medicine, 3(1): 43. https://doi.org/10.1038/s41746-020-0233-7

[3] González Moreno, A., Molero Jurado, M.D.M. (2024). Presence of emotions in network discourse on mental health: Thematic analysis. Psychiatry International, 5(3): 348-359. https://doi.org/10.3390/psychiatryint5030024

[4] Huang, X., Wang, S., Zhang, M., Hu, T., Hohl, A., She, B., Li, Z. (2022). Social media mining under the COVID-19 context: Progress, challenges, and opportunities. International Journal of Applied Earth Observation and Geoinformation, 113: 102967. https://doi.org/10.1016/j.jag.2022.102967

[5] Grundke, A., Stein, J.P., Appel, M. (2023). Improving evaluations of advanced robots by depicting them in harmful situations. Computers in Human Behavior, 140: 107565. https://doi.org/10.1016/j.chb.2022.107565

[6] Xu, Q.A., Chang, V., Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. Decision Analytics Journal, 3: 100073. https://doi.org/10.1016/j.dajour.2022.100073

[7] Liu, D., Feng, X.L., Ahmed, F., Shahid, M., Guo, J. (2022). Detecting and measuring depression on social media using a machine learning approach: Systematic review. JMIR Mental Health, 9(3): e27244. https://doi.org/10.2196/27244

[8] Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., Fu, H. (2023). Transformers in medical imaging: A survey. Medical Image Analysis, 88: 102802. https://doi.org/10.1016/j.media.2023.102802

[9] Joshy, A., Sundar, S. (2022). Analyzing the performance of sentiment analysis using BERT, DistilBERT, and RoBERTa. In 2022 IEEE International Power and Renewable Energy Conference (IPRECON), Kollam, India, pp. 1-6. https://doi.org/10.1109/IPRECON55716.2022.10059542

[10] Semary, N.A., Ahmed, W., Amin, K., Pławiak, P., Hammad, M. (2023). Improving sentiment classification using a RoBERTa-based hybrid model. Frontiers in Human Neuroscience, 17: 1292010. https://doi.org/10.3389/fnhum.2023.1292010

[11] Chauhan, A., Sharma, A., Mohana, R. (2025). An enhanced aspect-based sentiment analysis model based on RoBERTa for text sentiment analysis. Informatica, 49(14): 193-202. https://doi.org/10.31449/inf.v49i14.5423

[12] Albladi, A., Uddin, M.K., Islam, M., Seals, C. (2025). TWSSenti: A novel hybrid framework for topic-wise

sentiment analysis on social media using transformer models. arXiv preprint arXiv:2504.09896. https://doi.org/10.48550/arXiv.2504.09896

[13] Zanwar, S., Wiechmann, D., Qiao, Y., Kerz, E. (2022). Exploring hybrid and ensemble models for multiclass prediction of mental health status on social media. In Proceedings of the 13th International Workshop on Health Text Mining and Information Analysis (LOUHI), Abu Dhabi, United Arab Emirates, pp. 184-196, https://doi.org/10.18653/v1/2022.louhi-1.21

[14] Bokolo, B.G., Liu, Q. (2023). Deep learning-based depression detection from social media: Comparative evaluation of ml and transformer techniques. Electronics, 12(21): 4396. https://doi.org/10.3390/electronics12214396

[15] Di Cara, N.H., Maggio, V., Davis, O.S., Haworth, C.M. (2023). Methodologies for monitoring mental health on twitter: Systematic review. Journal of Medical Internet Research, 25: e42734. https://doi.org/10.2196/42734

[16] Rafieivand, S., Moradi, M.H., Sanat, Z.M., Soleimani, H.A. (2023). A fuzzy-based framework for diagnosing esophageal mobility disorder using high-resolution manometry. Journal of Biomedical Informatics, 141: 104355. https://doi.org/10.1016/j.jbi.2023.104355

[17] Bello, A., Ng, S.C., Leung, M.F. (2023). A BERT framework to sentiment analysis of tweets. Sensors, 23(1): 506. https://doi.org/10.3390/s23010506

[18] Shannon, H., Bush, K., Villeneuve, P.J., Hellemans, K.G., Guimond, S. (2022). Problematic social media use in adolescents and young adults: Systematic review and meta-analysis. JMIR Mental Health, 9(4): e33450. https://doi.org/10.2196/33450

[19] Nababan, A.A., Sutarman, Zarlis, M., Nababan, E.B. (2024). Multiclass logistic regression classification with PCA for imbalanced medical datasets. Mathematical Modelling of Engineering Problems, 11(9): 2377-2387. https://doi.org/10.18280/mmep.110911

[20] Salas-Zárate, R., Alor-Hernández, G., Salas-Zárate, M.D.P., Paredes-Valverde, M.A., Bustos-López, M., Sánchez-Cervantes, J.L. (2022). Detecting depression signs on social media: A systematic literature review. Healthcare, 10(2): 291. https://doi.org/10.3390/healthcare10020291

[21] Ajlouni, N., Özyavaş, A., Ajlouni, F., Takaoğlu, F., Takaoğlu, M. (2025). Enhanced hybrid facial emotion detection & classification. Franklin Open, 10: 100200. https://doi.org/10.1016/j.fraope.2024.100200

[22] Chandrasekaran, R., Kotaki, S., Nagaraja, A.H. (2024). Detecting and tracking depression through temporal topic modeling of tweets: Insights from a 180-day study. NPJ Mental Health Research, 3(1): 62. https://doi.org/10.1038/s44184-024-00107-5

[23] Cha, J., Kim, S., Park, E. (2022). A lexicon-based approach to examine depression detection in social media: The case of Twitter and university community. Humanities and Social Sciences Communications, 9(1): 1-10. https://doi.org/10.1057/s41599-022-01313-2

[24] Kumar, S., Srivastava, A., Maity, R. (2024). Modeling climate change impacts on vector-borne disease using machine learning models: Case study of Visceral leishmaniasis (Kala-azar) from Indian state of Bihar. Expert Systems with Applications, 237: 121490. https://doi.org/10.1016/j.eswa.2023.121490

[25] Liu, Y., Zhi, T., Shen, M., Wang, L., Li, Y., Wan, M. (2022). Software-defined DDoS detection with information entropy analysis and optimized deep learning. Future Generation Computer Systems, 129: 99-114. https://doi.org/10.1016/j.future.2021.11.009

[26] Xue, Z., He, G., Liu, J., Jiang, Z., Zhao, S., Lu, W. (2023). Re-examining lexical and semantic attention: Dual-view graph convolutions enhanced BERT for academic paper rating. Information Processing & Management, 60(2): 103216. https://doi.org/10.1016/j.ipm.2022.103216

[27] Salau, A.O., Markus, E.D., Assegie, T.A., Omeje, C.O., Eneh, J.N. (2023). Influence of class imbalance and resampling on classification accuracy of chronic kidney disease detection. Mathematical Modelling of Engineering Problems, 10(1): 48-54. https://doi.org/10.18280/mmep.100106