DEPRESSION DETECTION USING SENTIMENT ANALYSIS OF SOCIAL MEDIA TEXT

Supported by tetfund
TERTIARY EDUCATION TRUST FUND

Zubairu Onimisi Isa, Sunday Eric Adewumi, Victoria Ifeoluwa Yemi-Peters
Department of Computer Science
Federal University Lokoja, Kogi State
Corresponding Author: isazubairu52@gmail.com

Received: February 7, 2025, Accepted: April 1, 2025

Abstract: Depression is a mental health disorder that negatively affects the thinking, feeling, and actions of an affected person. Depression can reduce an individual's ability to perform daily activities and leads to health colleges like hypertension and even suicide. Detecting depression early enough can help health workers to take necessary actions to reduce complications. The traditional ways of detecting depression are no longer effective as they are time-consuming and prone to error. Recent research on depression detection has advanced significantly by utilizing more sophisticated techniques like machine learning and deep learning. However, some gaps still exist in the previous research, which include the application of transformer-based models and deep learning-based models on larger datasets for depression detection through sentiment analysis. This research addresses this gap by building depression detection models through sentiment analysis of social media text focusing on 10,000 twitter datasets. Three algorithms were utilized in this research, which includes two pre-trained algorithms (BERT and RoBERTa) and one deep learning algorithm (LSTM). The study result shows that BERT outperforms RoBERTa and LSTM in terms of accuracy, stability generalization, and efficiency after three epochs. While RoBERTa also performs similarly to the BERT after three epoches, the LSTM shows signs of overfitting after constant improvement in accuracy as the number of epochs increases. With accuracies of 99.9% for BERT and 99.7% for RoBERTa, future research could explore more diverse datasets and a hybrid of pre-trained and deep learning models to improve the contextual understanding and performance of models.

Keywords: Depression, algorithms, deep learning, social media, sentiment analysis.

Introduction

Depression is a major depressive disorder that negatively affects the thinking, feeling, and actions of an affected person. People living with depression experience constant feelings of sadness and have no more interest in activities they always eyed most (Mustafa *et al*., 2020). Depression can reduce an individual's ability to perform daily activities and can develop into different forms of emotional and physical problems. Depression symptoms can include constant feelings of sadness, loss of interest, changes in appetite, fatigue, inability to concentrate, thoughts of death, and sleeping disorder (Lin *et al.,* 2020). Millions of people worldwide are suffering from depression today, and the numbers continue to increase. This continuous increase in the number of people suffering from depression is a major concern. The perfect place to identify people at risk of depression is the social media space.

Social media is an online platform where people communicate and interact daily. It offers opportunities for researchers to extract early signs of mental health issues such as depression, from users' comments and posts on social media platforms. Researchers can analyze large amounts of information from social media using machine learning algorithms to detect early signs of depression (Shitole *et al*., 2024). Data from social media platforms can also give researchers insights into people's mental health through their posts and comments, providing useful information into the problem of mental health.

Sentiment analysis is the process of extracting sentiments from information. Over the years, sentiment analysis has advanced significantly in the field of natural language processing. However, despite the major advancement of sentiment analysis, its application on the detection of mental health is still under-researched. Researchers find it difficult to detect mental health issues with the use of sentiment analysis. Social media provides a chance to track users' mental health through text data analysis (Wankhade *et al*., 2022). Some researchers have utilized machine learning algorithms to detect sentiments from social media data, but these sentiment analysis algorithms often fall short in identifying the subtle emotional cues linked to mental health problems. The existing literature lacks a thorough comparative analysis of different models in the sentiment analysis of mental health problems. Leveraging on resent reviewed research gaps which include the application of transformer-based models and deep learning-based models on larger datasets for depression detection using sentiment analysis resulting to overfitting and non-generalization of the models, This Research tend to improve on that by using more of transformer-based models and traditional deep learning models on larger dataset from Twitter for Depression detection through sentiment analysis.

Literature review

Over the years, different methods have been applied by different researchers on mental health detection. Some of the research carried out on early detection of mental health includes:

Colledani et al. (2025) developed a machine learning-based model for identifying anxiety, depression, and social anxiety. The study utilized the CAT algorithm on a dataset containing 564 records collected with the use of GAD-D, PHQ-9, and SAD-D scales. The result shows that the CAT achieved high diagnostic agreement, ranging between 75.4% and 87.7%.

Abilkaiyrkyzy et al. (2024) researched the analysis of mental health status using the concept of a Digital Twin. The study utilized the pre-trained BERT algorithm, which was fine-tuned on the E-DAIC dataset. The experimental result shows

that the proposed model was able to achieve 69% accuracy in detecting mental health severity and also have an acceptable and usability score of 84.75%.

Yang & Liu (2024) conducted research that uses big data analytics to develop a dynamic mental health monitoring system for vocational. The research was conducted on different demographic factors. The research collected mental health data of vocational students, using SCL-90 symptom self-rating scale. The experimental result shows that the total score distribution followed a normal distribution.

Yang (2024) carried out research to analyze factors that affect college students' mental health, with a focus on emotion detection based on the facial expressions of students. The long short-term memory (LST was used to process the 2-dimensional images. While the computer vision was used for image recognition. The dataset that was used in the research is a video of students' facial expressions captured. The experimental result shows an 80.3% accuracy on the general feature recognition, 89.3% accuracy on the emotion recognition, and 84.4% accuracy on the feature recognition.

Alshanketi (2024) carried out research to investigate depression from social media data. The research focused on the development of a depression detection model using social media datasets. The research utilized the MDHAN (Multi-Aspect Depression Detection with Hierarchical Attention Network) model to classify Twitter data. The proposed model was compared with state-of-the-art models like the CNN, SVM, MDL, and MDHAN models. The experimental result shows that the proposed MGADHF model was able to achieve 99.19% accuracy, outperforming the compared state-of-the-art models.

Fernandes et al. (2024) conducted research that focused on the detection of abnormalities in elderly people using machine learning techniques. The research utilized Local Outlier Factor (LOF), One-Class SVM, Robust Covariance, and Isolation Forest algorithms on ADL (Activities of Daily Living) data, which was collected from elderly people. The experimental result shows that the LOF-based model outperforms all other compared modes, achieving the best precision of 96% and F1-score of 96%.

Oryngozha et al. (2024) developed a text-based stress detection model using a machine learning approach. The research utilized NLP for text analysis, Bag of Words (BoW) feature extraction, and LR for classification. The dataset is the Reddit dataset and student posts collected from academic subreddits. The result shows that the proposed model achieved 77.78% accuracy and 79% F1-score on the Dreaddit dataset and 72% accuracy on the academic subreddit posts.

Saha et al. (2023) developed a suicidal and mental health issues detection model using social media data and machine learning algorithms. The research proposed the SVM algorithms compared to other state-of-the-art algorithms. The result of the experiment shows that the proposed model outperformed all other compared state-of-the-art models, achieving the highest accuracy of 88.6%.

Liputo et al. (2023) utilized the machine learning approach to develop a mental health prediction model using the Social Emotional Health Survey-Secondary (SEHS-S). The research used the DT algorithm for mental health prediction, and the K-Fold Cross-Validation (8-fold) was used for model evaluation. The result shows that the DT algorithm achieved 78.61% accuracy when the cross-validation was set to 8-fold

Based on the reviewed papers, we observed that the existing systems present a significant advancement in mental health care, as different advanced machine learning techniques have been applied, however, there are still some gaps identified in these studies.

Several studies focus primarily on mental health status but do not consider specific mental health issues like depression. Although Colledani et al. (2025) focused on identifying anxiety, depression, and social anxiety, the research utilized the machine learning-based CAT algorithm without considering the potential of deep learning architecture like the LSTM and pre-trained algorithms like BERT and RoBERTa. Additionally, while existing research focused on mental health classification, none of this research considered depression detection through text sentiment analysis. This study addresses these gaps by developing depression detection models through sentiment analysis of social media text, with the use of two pre-trained algorithms (BERT and RoBERTa) and a deep learning-based algorithm (LSTM).

Methodology

The method used to develop the proposed model follows a pattern similar to the Chrisp-DM methodology. This method has different unique phases, which are

Data Collection and Exploration

In the data collection and exploration stage, we collected the research dataset, which contains text discussions related to mental health. From different social media platforms. The dataset was loaded into the coding environment (Google Colab), and data exploration was carried out by understanding the data distribution and visualizing frequent words from the dataset through the word cloud visualization to identify the common words in the dataset, as shown in Figure 1. We observed that words like *depression, love, thank,* and *life* are common words in the dataset.

Figure 1: Word cloud visualization of frequent words in the dataset

Data Preprocessing

In the data preprocessing phase, we carry out the data cleaning by removing stop words and punctuations and also perform tokenization. At this phase, we also analyzed sentiment labels to check for class imbalance. The dataset sentiment distribution is shown in Figure 2, and the class imbalance in the sentiment the dataset is imbalance, with a labels is shown in Figure 3. This figure shows that there is an imbalance in the dataset with a large part of neutral/positive tweets (label 0) as compared to the negative tweets (label 1). We decided to use the oversampling of the minority class techniques to address the class imbalance.

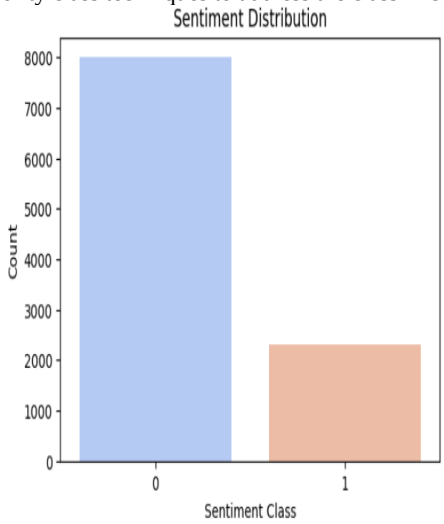

Figure 2: Sentiment Distribution

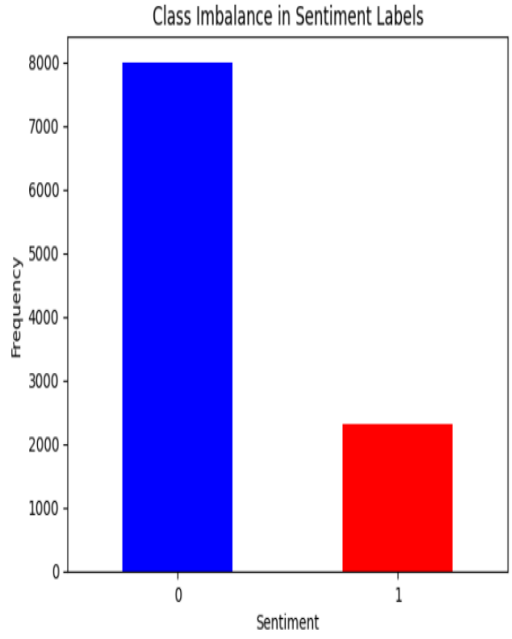

Figure 3: Class imbalance in sentiment labels

Data Splitting and Feature Engineering

After the data preprocessing phase, we further split the dataset into training (80%). And testing (20%) sets, which makes the tsining size to be 8251 samples and the testing size to be 2063 samples as shown in figure 3. We further used word embeddings to represent the text data during the feature engineering. The embedding layer converts each word into a dense vector representation, which is then fed into the model.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Display the sizes of the split datasets
print("Training set size:", X_train.shape[0])
print("Testing set size:", X_test.shape[0])
```

```
Training set size: 8251
Testing set size: 2063
```

Figure 3: Training and testing sets

Model Training and Architecture

In order to achieve a robust sentiment classification in the mental health detection model. We built and trained three models: two pre-trained models (BERT and RoBERTa) and one deep learning model (LSTM), with each of the models designed to leverage different strengths of NLP techniques.

BERT Model

We fine-tuned the BERT model for sentiment classification, following the architecture of a pre-trained BERT-based model with a classification head on top. Specifically, for the extraction of the contextual word embedding, we utilized a pre-trained BERT encoder. For classification, we utilized a fully connected den.se layer. To produce the probability for each of the sentiment classes, we utilized the SoftMax activation. To ensure optimal coverage, we trained the BERT with the use of cross-entropy loss**,** Adam optimizer**,** and a learning rate scheduler. Also, during the training, the batch size was set to 16 over multiple epochs.

RoBERTa Model

The RoBERTa, which is a variant of the BERT model, was also fine-tuned for sentiment classification. The ROBERTa advantage over the BERT includes dynamic masking to improve training performance, longer sequences for better contextual understanding, and byte-level tokenization for handling out-of-vocabulary words. Similar to the BERT model, we fine-tuned the RoBERTa model but with larger batch size and learning rate for stability.

LSTM Model

The **LSTM,** which is a deep learning-based model, was implemented as one of the models. The LSTM's architecture included an embedding layer whose purpose is to convert text into dense vector representations, an LSTM layer with the purpose to capture sequential dependencies in the text, a dropout layer to prevent overfitting, and a dense output layer with three classes for sentiment classification. The LSTM models' summary is presented in Table 1. We trained the LSTM model with the use of categorical cross-entropy loss and the Adam optimizer**.**

Table 1: LSTM model summary

Layer (type)	Output Shape	Param
Embedding	(None, 100, 100)	500,000
LSTM	(None, 128)	117,248
Dropout	(None, 128)	0
Dense	(None, 3)	387

Total Parameters: 617,635 (2.36 MB)
Trainable Parameters: 617,635 (2.36 MB)
Non-Trainable Parameters: 0 (0.00 B)

Model Evaluation

After the training of the three model, the models were further evaluated to know how the effective they are in sentiment classification. The models were evaluated based on the accuracy, validation accuracy, training loss, validation loss, loss, and time per epoch. These evaluation metrics provides insights into the models' ability to correctly classify sentiments in the dataset. Detailed results of this evaluation are presented in the results section, where the performance of each model is discussed in depth.

Results

BERT Model Results

The BERT model shows an incredible performance by achieving an almost perfect accuracy score with little training loss and validation loss within three epochs. The BERT model's performance is shown in Table 2, showing a constantly high accuracy and low loss values, which indicate BERT's effectiveness in sentiment classification.

Table 2: BERT Model Performance Across Three Epochs

Epoch	Training Loss	Validation Loss	Accuracy
1	1.3500%	0.5427%	99.9031%
2	0.0100%	0.4990%	99.9515%
3	0.4200%	0.4656%	99.9515%

RoBERTa Model Results

RoBERTa also shows an incredible performance, close to that of the BERT model in terms of accuracy. But it has a slightly higher validation loss in the third epoch, which shows a little overfitting. Table 3 shows the summary of the RoBERTa model.

Table 3: RoBERTa Model Performance Across Three Epochs

Epoch	Training Loss	Validation Loss	Accuracy
1	1.2000%	0.8503%	99.9031%
2	0.9100%	0.7821%	99.9031%
3	0.4500%	1.5193%	99.7092%

LSTM Model Results

The LSTM model shows a different learning pattern, with a gradual improvement in the accuracy over multiple epochs. But its validation loss increased after the initial epochs, showing a potential overfitting. Table 4.3 presents the LSTM performance metrics.

Table 4.3: LSTM Model Performance Across Ten Epochs

Epoch	Accuracy	Loss	Val Accuracy	Val Loss	Time per Epoch
1	34.62%	109.92%	31.00%	1.1009	12s
2	38.99%	107.86%	30.50%	1.1058	5s
3	86.53%	95.00%	32.50%	1.1438	6s
4	91.083%	57.56%	23.00%	1.3777	9s
5	96.54%	19.43%	29.50%	1.5883	6s
6	99.01%	06.39%	25.00%	1.8049	5s
7	99.75%	03.30%	27.00%	1.834	6s
8	99.66%	02.84%	30.00%	2.1162	9s
9	1.00%	00.82%	28.00%	2.2297	7s
10	99.87%	00.67%	25.00%	2.1965	5s

Discussion of the Result

The evaluation of the pre-trained models (BERT and RoBERTa) and deep learning (LSTM) provided us with significant insights into how well they performed the sentiment analysis task. The BERT model shows optimal performance throughout the training process, outperforming the other compared model with consistently stable validation loss and high accuracy across the three epochs. The stability in the validation loss shows that the BERT model effectively generalized to unseen data. The RoBERTa also performed excellently, with results very close to that of BERT, but has slightly higher fluctuations in validation loss. The validation loss variation in the RoBERTa model shows that the model is sensitive to changes in data distribution compared to BERT. Despite the fluctuations in validation loss in the RoBERTa model, it still maintains an excellent overall performance, achieving the second-best model in terms of

accuracy and generalization. On the other hand, the LSTM model shows constant improvement in accuracy as the number of epochs increases but shows signs of overfitting. The sign of overfitting shown is the increasing validation loss as the number of epochs increases, showing that while the model performs well on the training data, it struggles to maintain the same level of performance on the testing data. The result makes it clear that transformer-based models (BERT and RoBERTa) perform better in terms of stability, generalization, and training efficiency, with more resilience against overfitting than the deep learning model (LSTM).

Conclusion

In this study, we developed three depression detection models through the sentiment analysis of social media text data. This study was motivated by the constant increase in mental health-related issues and the potential of AI to serve as a warning system against it. This research leverages NLP and machine learning techniques to classify text-based sentiments. Three different algorithms were trained in this study, these algorithms include two pre-trained algorithms (BERT and RoBERTa) and one deep learning algorithm (LSTM). The dataset used in the study consists of text about depression discussions collected from an online platform. This study aligned with the research objective of developing a depression detection model using sentiment analysis on social media data. The results demonstrated that transformer-based models, BERT and RoBERTa, outperformed the LSTM model in terms of accuracy, generalization, and stability. One of the most significant findings is the high classification accuracy achieved by BERT (99.9%) and RoBERTa (99.7%), reinforcing their effectiveness in sentiment classification tasks like depression detection. The exhibition of overfitting by the LSTM suggests that traditional recurrent neural networks may struggle with sentiment classification when handling complex linguistic patterns and large-scale social media datasets.

This study contributes significantly to the field of AI-driven mental health assessment, with an emphasis on the potential of NLP in sentiment monitoring.

Future research could explore more diverse and larger datasets to enhance generalization and reduce biases in models' sentiment classification. Also, research could focus on hybrid models that combine transformer-based models like BERT and RoBERTa with recurrent architectures like LSTMs and GRUs, and more advanced AI models to improve contextual understanding and performance of models.

Conflict of Interest

The authors have no conflict of interest

References

Abilkaiyrkyzy, A., Laamarti, F., Hamdi, M., & Saddik, A. el. (2024). Dialogue System for Early Mental Illness Detection: Toward a Digital Twin Solution. *IEEE Access*, *12*. https://doi.org/10.1109/ACCESS.2023.3348783

Alshanketi, F. (2024). Revolutionizing Generalized Anxiety Disorder Detection using a Deep Learning Approach with MGADHF Architecture on Social Media. *International Journal of Advanced Computer Science and Applications*, *15*(1). https://doi.org/10.14569/IJACSA.2024.0150192

Colledani, D., Barbaranelli, C. & Anselmi, P. (2025) Fast, smart, and adaptive: using machine learning to optimize mental health assessment and monitor change over time. *Sci Rep* **15(1)**, 6492. https://doi.org/10.1038/s41598-025-91086-w

Fernandes, A., Leithardt, V., & Santana, J. F. (2024). Novelty detection algorithms to help identify abnormal activities in the daily lives of elderly people. *IEEE Latin America Transactions*, *22*(3). https://doi.org/10.1109/TLA.2024.10431423

Lin, C., Hu, P., Su, H., Li, S., Mei, J., Zhou, J., & Leung, H. (2020). SenseMood: Depression detection on social media. *ICMR 2020 - Proceedings of the 2020 International Conference on Multimedia Retrieval*. https://doi.org/10.1145/3372278.3391932

Liputo, S., Tupamahu, F., Hasyim, W., Sabiku, S. A., Parman, R., & Hanapi, A. (2023). Prediction of Elementary School Students' Mental Health using Decision Tree Algorithm with K-Fold Cross-Validation in Bone Bolango Regency, Gorontalo Province. *Journal La Multiapp*, *4*(6). https://doi.org/10.37899/journallamultiapp.v4i6.1005

Mustafa, R. U., Ashraf, N., Ahmed, F. S., Ferzund, J., Shahzad, B., & Gelbukh, A. (2020). A Multiclass Depression Detection in Social Media Based on Sentiment Analysis. *Advances in Intelligent Systems and Computing*, *1134*. https://doi.org/10.1007/978-3-030-43020-7_89

Oryngozha, N., Shamoi, P., & Igali, A. (2024). Detection and Analysis of Stress-Related Posts in Reddit's Acamedic Communities. *IEEE Access*, *12*. https://doi.org/10.1109/ACCESS.2024.3357662

Saha, S., Dasgupta, S., Anam, A., Saha, R., Nath, S., & Dutta, S. (2023). An Investigation of Suicidal Ideation from Social Media Using Machine Learning Approach. *Baghdad Science Journal*, *20*. https://doi.org/10.21123/bsj.2023.8515

Shitole, S., Lattoo, S., & Chillal, A. (2024). Predicting Depression Levels Using Social Media Posts: A Comprehensive Survey. *International JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, *08*(01). https://doi.org/10.55041/ijsrem28147

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7). https://doi.org/10.1007/s10462-022-10144-1

Yang, H., & Liu, Q. (2024). RETRACTED ARTICLE: Innovative research of dynamic monitoring system of mental health vocational students based on big data. *Personal and Ubiquitous Computing*, *28*(S1). https://doi.org/10.1007/s00779-021-01644-y

Yang, W. (2024). Extraction and analysis of factors influencing college students' mental health based on deep learning model. *Applied Mathematics and Nonlinear Sciences*, *9*(1). https://doi.org/10.2478/amns.2023.2.00773