ISSN 2007-9737

Adaptation of Transformer-Based Models for Depression Detection

Olaronke O. Adebanji, Olumide E. Ojo,
Hiram Calvo*, Irina Gelbukh, Grigori Sidorov

Instituto Politécnico Nacional, Centro de Investigación en Computación,
Mexico

{olaronke.oluwayemisi, olumideoea, ir.gelbukh}@gmail.com,
{hcalvo, sidorov}@cic.ipn.mx

Abstract. Pre-trained language models are able to capture a broad range of knowledge and language patterns in text and can be fine-tuned for specific tasks. In this paper, we focus on evaluating the effectiveness of various traditional machine learning and pre-trained language models in identifying depression through the analysis of text from social media. We examined different feature representations with the traditional machine learning models and explored the impact of pre-training on the transformer models and compared their performance. Using BoW, Word2Vec, and GloVe representations, the machine learning models with which we experimented achieved impressive accuracies in the task of detecting depression. However, pre-trained language models exhibited outstanding performance, consistently achieving high accuracy, precision, recall, and F1 scores of approximately 0.98 or higher.

Keywords. Depression, bag-of-words, word2vec, GloVe, machine learning, deep learning, transformers, sentiment analysis.

1 Introduction

Depression is a profound psychological disorder that affects people in different regions of the world, regardless of their gender, age, or social status [1].

It is a psychiatric condition characterized by persistent feelings of boredom, negativity, and sadness during daily activities. People experiencing depression often face challenges in interpersonal relationships, occupational performance, and maintaining healthy bonds, which ultimately affect their overall well-being. It is estimated that more than 280 million people worldwide struggle with depression, making it the main cause of disability on a global scale [6].

Despite its prevalence and impact, depression frequently evades detection and remains untreated. Early detection of depression presents challenges due to people who do not seek professional help or are unaware of their symptoms.

The presence of social media (SM) and online forums provides researchers with a unique opportunity to analyze online expressions of thoughts and emotions, potentially uncovering indications of depression.

Therefore, it is essential to develop accessible and efficient techniques that can help identify people at risk of depression, allowing them to receive the necessary support. SM platforms have become popular for communication and self-expression [8, 7, 47].

Sentiment analysis (SA) identifies and extracts subjective information from text, such as SM posts, reviews, and news articles. By analyzing the language used in SM texts [41, 42, 4, 31, 17, 2, 28], SA algorithms can determine the general sentiment of the text.

SM serves as a valuable tool to connect with people who might be susceptible to depression or who face challenges related to mental well-being. Considering the extensive use of SM platforms such as Facebook, Twitter, and Instagram by billions of people around the world to communicate and share information, it has evolved into an integral aspect of contemporary communication

doi: 10.13053/CyS-28-1-4691

methods. Studies have shown that people with depression use SM as a coping mechanism, seeking support and validation from others through online interactions [3, 9, 36].

Environmental and social factors, among others, have a great influence on the development of depression. The process of analyzing large volumes of text derived from SM using natural language processing (NLP) techniques allows the identification of language patterns that can indicate depression [24, 48, 52, 49].

Machine Learning (ML) techniques are revolutionizing the field of NLP [19, 26, 43, 30, 35, 10, 31, 32, 27]. These techniques have enabled researchers to build sophisticated models that can analyze and understand complex human language, including sentiment, syntax, and semantics. These algorithms use statistical rules to discover patterns in the data and use them to inform decisions as a result of this learning.

Deep learning (DL), a subset of ML, involves training artificial neural networks with many layers to recognize patterns and make decisions. Transformers are a type of DL architecture that has become popular for its ability to process sequential data, such as text.

Fine-tuned pre-trained language models are a specific application of these DL techniques. These models are pre-trained on massive volumes of text data before being fine-tuned for specific tasks such as named entity recognition or sentiment analysis.

By fine-tuning these models on a specific task, researchers can leverage the pre-existing knowledge encoded in the model and achieve state-of-the-art performance on the task at hand.

In this paper, we performed various experiments to test the efficiency of traditional ML and DL techniques, including fine-tuned pre-trained transformer models, for the detection of depression in social media texts.

We conducted an in-depth investigation of these models and examined their performance. Our goal is to provide information on how well these models can detect depression and highlight areas for future research and development.

The findings of our research improved our understanding of the potential use of these sentiment analysis techniques to detect depression and inform the development of targeted interventions that can reduce the burden of depression on society as a whole. Our research contributes to the existing literature as follows:

1. A thorough analysis of depression was carried out, as well as the exploration of the possible use of social media as a tool to express depression traits, and how machine learning can help detect depression on social media data.
2. We applied different feature representations with machine learning and deep learning algorithms for depression detection and evaluated the performance of the models using accuracy, recall, precision, and F1 scores.
3. We evaluated pre-trained language models and show that they exhibit outstanding performance by consistently achieving high accuracy, precision, recall, and F1 scores.
4. Context, feature extraction, and pre-training all had a significant impact on the models' performance as far as depression detection is concerned.

2 Literature Review

SA approaches have gained interest as a promising method of identifying patterns in text that can serve as indicators of depression. These approaches involve classifying the sentiment expressed in a given text to identify potential signs of depression symptoms.

In a study by Haque et al. [18], machine learning algorithms were employed to develop models capable of effectively identifying depression in children. The findings revealed that the Random Forest Classifier exhibited the highest efficiency in detecting depression.

Furthermore, the study identified 11 specific questions that can be used to detect depression in children and adolescents, helping to early diagnosis and treatment of this condition while understanding the contributing factors.

Another study by Reece et al. [38] used machine learning techniques to analyze Instagram data to identify possible indicators

Table 1. Related Studies on Detecting Depression

Model	Reference	F1 Score	Accuracy	Year
MNB	S.G. Burdisso et al.	0.96	0.96	2019
MLP	I. Fatima et al.	0.92	0.92	2019
CNN	J. Kim	0.79	0.75	2020
RFC	A Priya et al.	0.77	0.80	2020
Char CNN	K. Cornn	0.94	0.93	2020
SVM	H.S. AlSagri et al.	0.79	0.83	2020
Sense Mood	C. Lin et al.	0.94	0.88	2020
3D-CNN	H. Wang et al.	0.64	0.77	2021
RFC	EM de Souza Filho et al.	0.89	0.89	2021
LSTM	M. Muzammel et al.	0.95	0.95	2021
SBERT CNN	Z. Chen	0.86	0.86	2023

of depression. The study involved evaluating more than 43,000 Instagram photos and extracting statistical features such as color analysis, metadata components, and face identification.

Interestingly, their algorithm outperformed general practitioners in diagnosing depression, highlighting the potential of computational analysis of visual social media data as a scalable approach to detecting mental illnesses. In the study conducted by Cornn K. [14], a combination of various machine learning algorithms and neural networks was used to classify depression within social media text.

The most successful model was a CNN model, achieving an impressive accuracy of 92.5%. The one-dimensional convolutional layer played a vital role in noise reduction and was regarded as the most crucial component of the model.

Interestingly, the use of Word embedding proved to be ineffective in representing the text used in this particular study. In another work by Ziwei et al. [54], an application was developed to differentiate between depressive and non-depressive tweets using a classification function.

The application also provided a visualization of the user's depression status through a web interface. The research emphasized the importance of early detection of depression and highlighted the potential of social media platforms in predicting mental and physical illnesses.

However, the application faced limitations imposed by Twitter's API, such as the constraint of analyzing only a limited number of tweets. In a study conducted by De Choudhury et al. [15], sentiment analysis techniques were used to analyze Facebook data to detect symptoms of depression.

The findings revealed that individuals with depression symptoms tended to use a higher frequency of first-person pronouns, express negative emotions through their choice of words, and display a reduced use of terms associated with happiness in their Facebook posts, compared to individuals without symptoms.

Chen et al. [11] conducted a data analysis on Reddit data to identify people with depression. They proposed a hybrid deep learning model that combined a pre-trained sentence BERT (sBERT) with a convolutional neural network (CNN) to effectively identify individuals with depression based on their Reddit posts.

Interestingly, the model exceeded previously reported state-of-the-art results in the literature, achieving an accuracy of 0.86 and an F1 score of 0.86. The improved hybrid model was also applied to other text analysis tasks, showcasing its versatility and efficacy.

The research carried out by Wen et al. [51] used social media data to detect depression among users. Through the development of a classification model specifically designed to identify depression in tweets, the authors achieved remarkable results, with a high test accuracy of 98.94% and an F1 score of 99.04%.

The study highlights the effectiveness of analyzing the language used on social media platforms as a valuable approach for the early detection of depression among individuals. In a related study, Hosseini et al. [21] explored the integration of psychological and psychoanalytical insights to improve the identification of individuals with depression.

By combining traits observed in both depressed and non-depressed groups, the researchers created a bipolar feature vector. They successfully improved their models and achieved an impressive F1 score of 82.75% using a modified Bayesian classifier to classify social media users into depressed and non-depressed groups. In the research conducted by Wang et al. [50], a method to improve the features was introduced as input to a 3D CNN speech emotion recognition

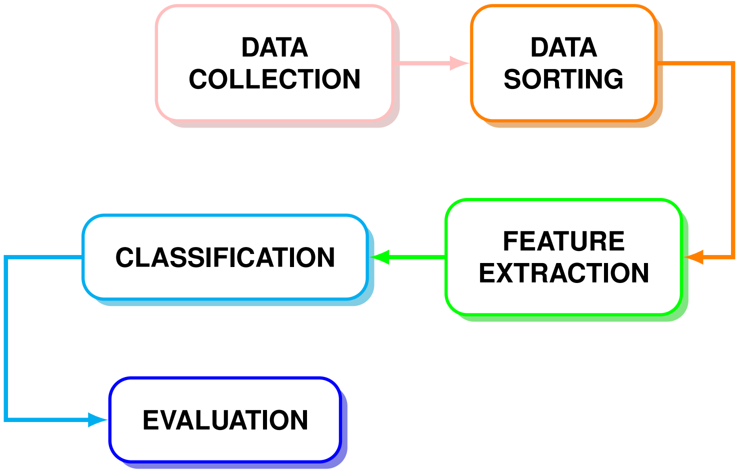

Fig. 1. Depression detection process flow chart

model, with the aim of identifying depression in its earliest stages.

The experiments carried out demonstrated that the combination of the enhanced feature and the model significantly improved the ability to detect and recognize depression.

Additionally, their study emphasized the necessity for future investigations to incorporate more detailed levels of analysis and extract additional features from speech signals to enhance detection accuracy.

Muzammel et al. [25] conducted experiments on depression detection by integrating multimodal features and selecting the optimal fusion strategy. The authors proposed two unimodal representations based on RNN and CNN networks.

These networks were utilized to acquire dynamic temporal representations of multimodal data, allowing for a comprehensive understanding of depression.

These investigations indicate that supervised learning techniques can be effective in identifying depression through the analysis of social media data. The summarized research findings related to depression detection are presented in Table 1.

However, there are some limitations to some of these methods that highlight the need to continue developing and fine-tuning these techniques to improve their accuracy and effectiveness. Figure 1 illustrates the steps involved in our classification method.

3 Methodology

3.1 Data

The dataset used in this experiment was sourced from Kaggle [5], a widely used platform known for hosting diverse datasets and machine learning competitions for individuals and organizations. It consists of depression-related text, acquired from Reddit, a highly popular social media platform worldwide, using web scraping techniques.

The datasets includes a total of 7,731 posts, which we divided into train and test sets to ensure accuracy and consistency in the analysis. The sentiment classes are ('1') or non-depression ('0'), which indicate whether the text contained expressions of depression or not.

Table 2 presents an example of text labeled with sentiment classes denoting depression and non depression. The dataset was divided into two parts: the training set and the testing set, consisting of 6539 and 1192 text inputs, respectively.

Table 3 presents the statistics of the text indicating depression and non-depression in both the training and the testing sets.

Table 2. Sample Text with Sentiment Classes

Text	Label
i ve lost everything i lost my best friend a community of people who were my only social outlet i m a failure i m i ve never been in a relationship i couldn t graduate college i m stuck working at a job which doesn t pay enough for me to afford rent so i have to live with my retirement age parent i can t find a job anywhere else i started cutting myself today never did it a a teenager but i did it now and it feel great i don t want to die but i don t see any other solution i can not afford help to me being in debt is worse than death i ve lost so much i can t go on	1
I ve been feeling really depressed lately and find myself with no one to talk I have these cry spell whenever i m alone and convinced that i m worthless and not worth anyone s time it s getting harder to pick myself up from the floor bed and be productive or practice self care my friend live far away and emotionally at arm length my family understands that i m depressed but not how much it debilitates me with no one to talk to i feel trapped i m hoping finding online support can help me understand how to go on so i m kinda new to this how does this thread help you	1
am i really just that awful no one want to be my friend my old friend abuse me i hate everything but especially myself when will it get better	1
Our membership had expired and to renew them, we have to do a new induction which can’t happen until next Tuesday	0
bored of sims for today and still thinking of a name for me and like youtube account to post our awesome new video on idea people	0
hetty christ heh yeah i shakily conquered the ladder pointless job though we are too far away to receive digital signal with antenna	0

To comprehensively evaluate the effectiveness and reliability of our depression detection models, we conducted extensive experiments by combining intelligent pre-trained transformer models with traditional machine learning techniques.

By integrating diverse feature representations and transformer architectures, we obtained valuable insights into the performance and suitability of various approaches for depression classification.

The availability of this dataset on Kaggle makes it easier for other researchers to replicate this experiment and build on the work done in this research.

3.2 Models

The traditional machine learning algorithms included Multinomial Naive Bayes (MNB) [37], Stochastic Gradient Descent (SGD) [53], Logistic Regression Classifier (LRC) [40], Decision Tree Classifier (DTC) [45], Random Forest Classifier (RFC) [33], K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP) [46].

These algorithms are commonly used in text classification tasks and are well established in the field of machine learning [34, 43, 29, 44]. Furthermore, we also used fine-tuned pre-trained language models for depression detection. The models used in the study included BERT [16], RoBERTa [23], XLM-RoBERTa [13],

Table 3. Statistics of depression and non-depression in the train and test datasets

Data	Instances	Label
Train	3,239	1
	3,300	0
Test	592	1
	600	0
Total	3,831	1
	3,900	0

DistilBERT [39], ALBERT [22], DistilRoBERTa [23] and ELECTRA [12]. These models are capable of capturing semantic and syntactic relationships between words, and the efficiency and effectiveness of these techniques make them often used for a wide range of applications, including language generation, machine translation and text classification.

4 Results

For this study, we evaluated different machine learning and pre-trained language models to detect and evaluate signs of depression. We extracted meaningful features from the text of social media to represent language patterns associated with depression.

The accuracy, precision, recall, and F1 evaluation metrics were used to assess the performance of the depression detection models. The features used in our experiments include bag-of-words (BoW), Word2Vec, and GloVe embeddings. By analyzing these results, we shed light on the profound influence of these distinct features on the overall performance of the models.

4.1 Experiment with Traditional Machine Learning Models and BoW

The BoW model represents text data as a collection of individual words and converts them into numerical representations that can be used by various machine learning algorithms. Machine learning models are trained on labeled datasets, where each text sample is associated with labels indicating the presence or absence of depression. The models learn to identify patterns and associations between the extracted BoW features and the corresponding labels.

The trained models are then evaluated using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. Our findings are presented in Tables 4 - 7, which provide a comprehensive overview of the results of our experiments. In the experiment, several models were evaluated using the Bag-of-Words (BoW) feature representation, and their performance scores were recorded.

According to the results, the LRC model achieved the highest performance in all metrics, with an average accuracy, precision, recall, and F1 score of 0.96. This indicates that the model excelled at accurately classifying depression. The SGD and SVM also demonstrated strong performance with average scores of 0.94 and 0.95 respectively.

These models showed excellent overall performance in terms of accuracy, precision, recall, and F1 score. However, MNB, DTC, RFC, and MLP achieved good performance, with average scores ranging from 0.83 to 0.91. Although these models did not achieve as high scores as the top performers, they still exhibited reasonably good results.

The KNN model had the lowest performance among the evaluated models, with an average score of 0.74. This suggests that the model faced challenges in accurately classifying instances related to depression compared to the other models.

Table 4. Result of machine learning models using the BoW feature representation

Model	Average	Accuracy	Precision	Recall	F1
MNB	macro avg	0.83	0.87	0.83	0.83
	weighted avg		0.87	0.83	0.83
SGD	macro avg	0.94	0.94	0.94	0.94
	weighted avg		0.94	0.94	0.94
LRC	macro avg	0.96	0.96	0.96	0.96
	weighted avg		0.96	0.96	0.96
DTC	macro avg	0.86	0.87	0.86	0.86
	weighted avg		0.87	0.86	0.86
RFC	macro avg	0.93	0.93	0.93	0.93
	weighted avg		0.93	0.93	0.93
KNN	macro avg	0.74	0.78	0.74	0.73
	weighted avg		0.78	0.74	0.73
SVM	macro avg	0.95	0.96	0.95	0.95
	weighted avg		0.96	0.95	0.95
MLP	macro avg	0.91	0.91	0.91	0.91
	weighted avg		0.91	0.91	0.91

4.2 Experiment with Traditional Machine Learning Models and Word2Vec

Unlike the BoW model, Word2Vec captures not only the frequency of words, but also their semantic meaning and contextual relationships. The Word2Vec model learns dense vector representations by analyzing large corpora of text data.

It represents each word in a high-dimensional vector space, where words with similar meanings or contextual usage are located closer to each other. The text data were preprocessed by tokenizing the text into words and removing any stop words or irrelevant characters.

Each word is then replaced by its corresponding Word2Vec vector representation obtained from the pre-trained model. This transforms the text data into numerical vectors, where each word is represented by a dense vector of fixed length.

The Word2Vec vectors are subsequently used as input features for machine learning models to detect depression. The models learn to identify patterns and associations between Word2Vec embeddings and the corresponding labels and are evaluated using accuracy, precision, recall, and F1 score.

Using Word2Vec word embeddings, the models effectively capture semantic and contextual information within the text data, resulting in improved accuracy and more meaningful predictions. The findings of the analysis, using Word2Vec as feature representations, are presented in Table 5.

Table 5 presents notable insights into the performance of different machine learning models using Word2Vec features. Among these models, the MLP model stands out with an impressive accuracy of 0.94. Both the RFC and SVM models consistently demonstrated moderate performance

Table 5. Result of machine learning models using the Word2Vec feature representation

Model	Average	Accuracy	Precision	Recall	F1
MNB	macro avg	0.52	0.53	0.52	0.48
	weighted avg		0.53	0.52	0.48
SGD	macro avg	0.81	0.83	0.81	0.80
	weighted avg		0.84	0.81	0.80
LRC	macro avg	0.87	0.87	0.87	0.87
	weighted avg		0.87	0.87	0.87
DTC	macro avg	0.82	0.82	0.82	0.82
	weighted avg		0.82	0.82	0.82
RFC	macro avg	0.91	0.92	0.91	0.91
	weighted avg		0.92	0.91	0.91
KNN	macro avg	0.80	0.83	0.80	0.80
	weighted avg		0.83	0.80	0.80
SVM	macro avg	0.91	0.91	0.91	0.91
	weighted avg		0.91	0.91	0.91
MLP	macro avg	0.94	0.94	0.94	0.94
	weighted avg		0.94	0.94	0.94

with accuracy, precision, recall, and F1 scores hovering around 0.91. The SGD, KNN, LRC, and DTC models performed adequately, albeit at a slightly lower level. The MNB model exhibited poor performance, as indicated by lower accuracy, precision, recall, and F1 scores.

4.3 Experiment with Traditional Machine Learning Models and GloVe

In order to conduct further analysis, we employed the use of GloVe embedding representation to capture the semantic relationships between words. These vector representations are derived from the co-occurrence statistics of words in a corpus.

By encoding information about word meaning and context, these embeddings enable machine learning models to benefit from this knowledge. Using pre-trained GloVe embeddings, each word in the text is mapped to its corresponding vector representation.

These word vectors are then concatenated to create document-level representations, which are subsequently used to train the machine learning models. The results of our experiments using GloVe embeddings are presented in Table 6. The results of the experiment using the GloVe feature representation for machine learning models are summarized in Table 6.

The SGD, LRC, and SVM models consistently outperformed all other models, achieving high accuracy, precision, recall, and F1 scores of approximately 0.94, 0.96, and 0.95, respectively. The RFC model also exhibited strong performance, with accuracy, precision, recall, and an F1 score of around 0.93.

The KNN, DTC and MLP models yielded good performance, yielding an accuracy, precision, recall, and F1 score of approximately 0.74, 0.86, and 0.91, respectively. On the other hand, the MNB model showed relatively lower performance compared to the other models, with

Table 6. Results of machine learning models using the GloVe feature representation

Model	Average	Accuracy	Precision	Recall	F1
MNB	macro avg	0.60	0.64	0.60	0.58
	weighted avg	-	0.64	0.60	0.58
SGD	macro avg	0.94	0.94	0.94	0.94
	weighted avg	-	0.94	0.94	0.94
LRC	macro avg	0.96	0.96	0.96	0.96
	weighted avg	-	0.96	0.96	0.96
DTC	macro avg	0.86	0.87	0.86	0.86
	weighted avg	-	0.87	0.86	0.86
RFC	macro avg	0.93	0.93	0.93	0.93
	weighted avg	-	0.93	0.93	0.93
KNN	macro avg	0.74	0.78	0.74	0.73
	weighted avg	-	0.78	0.74	0.73
SVM	macro avg	0.95	0.96	0.95	0.95
	weighted avg	-	0.96	0.95	0.95
MLP	macro avg	0.91	0.91	0.91	0.91
	weighted avg	-	0.91	0.91	0.91

an accuracy, precision, recall, and F1 score of approximately 0.60. These findings indicate that when combined with these specific models, the GloVe feature representation can be highly valuable for the analysis and classification of depression in textual data.

4.4 Experiment with Transformer Architectures

Pre-trained language models have demonstrated remarkable success in various NLP tasks [31, 20]. Initially trained on vast amounts of text data from the Internet, these models acquire a contextual understanding of words and sentences.

To apply pre-trained language models for depression detection, we fine-tuned them by training them on labeled data. Labeled data consist of text samples annotated with depression-related labels.

Through the fine-tuning process, the pre-trained language models learn to capture significant linguistic patterns and contextual cues associated with depression.

Using the fine-tuned models, we classify new text samples as either indicating depression or not. We evaluated the models' performance using various metrics and found that the ELECTRA and Roberta-large models outperformed others, achieving the highest cumulative scores across all metrics.

Notably, these models achieved an F1 score of 0.99 each after only 10 epochs. Table 7 presents the performance of the transformer models. Table 7 displays the results of an extensive evaluation of various pre-trained language models in the depression detection task.

Transformer models showcased strong performance across all metrics evaluated. BERT achieved an accuracy, precision, recall, and F1 score of 0.97, demonstrating its effectiveness in detecting depression.

Furthermore, models such as RoBERTa, XLM-RoBERTa, DistilBERT, ALBERT and ELECTRA consistently achieved high scores, with accuracy, precision, recall, and F1 scores around or above 0.98.

Table 7. Results of the Transformer models in the experiment

Model	Feature	Accuracy	Precision	Recall	F1
BERT	Transformer Embedding	0.97	0.97	0.97	0.97
RoBERTa	Transformer Embedding	0.99	0.99	0.99	0.99
XLM-RoBERTa	Transformer Embedding	0.98	0.98	0.98	0.98
DistilBERT	Transformer Embedding	0.98	0.98	0.98	0.98
ALBERT	Transformer Embedding	0.98	0.98	0.98	0.98
DistilRoBERTa	Transformer Embedding	0.96	0.97	0.96	0.96
ELECTRA	Transformer Embedding	0.99	0.99	0.99	0.99

These models exhibited robustness and reliability in the capture and comprehension of complex language patterns. While DistilRoBERTa slightly underperformed compared to other models, it still achieved an accuracy of 0.96. Both the ELECTRA and Roberta large models achieved the highest F1 scores of 0.99 each.

This underscores its exceptional potential for accurately detecting depression in this experiment. Incorporating these models could potentially revolutionize the identification and treatment of depression, leading to early detection and treatment.

5 Discussion

Pre-trained language models can continuously improve and adapt as they encounter new data, enhancing their diagnostic accuracy and generalization capabilities.

In this research, an investigation was conducted into the effectiveness of a variety of machine learning models in detecting depression in social media data, including pre-trained language models such as BERT, RoBERTa, XLM-RoBERTa, DistilBERT, ALBERT, DistilRoBERTa, and ELECTRA.

These models were assessed based on their accuracy, precision, recall, and F1 scores. Throughout the test, all transformer models achieved high accuracy and F1 scores, with RoBERTa and ELECTRA as the best performers.

This high performance of pre-trained transformer models suggests that they can effectively identify depression in text data.

Furthermore, these models can provide institutions responsible for the prevention of depression with a cost-effective alternative to their traditional methods of recognizing depression.

With the use of pre-trained language models and social media data for depression detection, significant advancement has been made in this study, emphasizing the potential of pre-trained language models and social media analysis for depression treatment and prevention.

6 Conclusions

As pre-trained language models continue to evolve, they hold the potential to revolutionize the field of depression prevention and treatment. The key strength of pre-trained language models lies in their ability to learn from vast amounts of diverse textual data, enabling them to discern subtle linguistic cues indicative of depression across different languages. Our study demonstrates the high effectiveness of pre-trained language models in detecting depression in English text from social media sources.

In our experiments, the pre-trained language models with which we experimented obtained very good accuracy, precision, recall, and F1 values. However, more research is needed to determine whether they are generalizable to larger, more diverse datasets and a different language. Real-world application challenges such as model biases, interpretability, and scalability still need to be addressed.

Our findings still underscore the need to leverage these pretrained language models to detect and address depression at scale. Through continued development, these models can contribute significantly to early detection and improved well-being for individuals suffering from depression.

Acknowledgments

The work was carried out with partial support from the Mexican government through grant A1-S-47854 of CONACYT, Mexico, and by the Secretaria de Investigacion y Posgrado of the Instituto Politécnico Nacional, Mexico, under grants 20211884, 20220859 and 20230140, EDI; and COFAA-IPN.

The authors thank CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercomputo of the INAOE, Mexico, and acknowledge the support of Microsoft through the Microsoft Latin America PhD Award.

References

1. **Abdul-Razzak, H., Harbi, A., Ahli, S. (2019).** Depression: Prevalence and associated risk factors in the United Arab Emirates. Oman Medical Journal, Vol. 34, No. 4, pp. 274–282. DOI: 10.5001/omj.2019.56.

2. **Adebanji, O. O., Gelbukh, I., Calvo, H., Ojo, O. E. (2022).** Sequential models for sentiment analysis: A comparative study. Proceedings of the 21st Mexican International Conference on Artificial Intelligence. Advances in Computational Intelligence, pp. 227–235. DOI: 10.1007/978-3-031-19496-2_17.

3. **Ali, F., Tauni, M. Z., Ashfaq, M., Zhang, Q., Ahsan, T. (2023).** Depressive mood and compulsive social media usage: The mediating roles of contingent self-esteem and social interaction fears. Information Technology and People. DOI: 10.1108/itp-01-2021-0057.

4. **Armenta-Segura, J., Núñez-Prado, C. J., Sidorov, G. O., Gelbukh, A., Román-Godínez, R. F. (2023).** Ometeotl@Multimodal hate speech event detection: Hate speech and text-image correlation detection in real life memes using pre-trained BERT models over text. Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text, pp. 53–59.

5. **Banachewicz, K., Massaron, L., Goldbloom, A. (2023).** The Kaggle book: Data analysis and machine learning for competitive data science. Pack Publishing, pp. 530.

6. **Bhatt, S., Devadoss, T., Jha, N. K., Baidya, M., Gupta, G., Chellappan, D. K., Singh, S. K., Dua, K. (2023).** Targeting inflammation: A potential approach for the treatment of depression. Metabolic Brain Disease, Vol. 38, No. 1, pp. 45–59. DOI: 10.1007/s11011-022-0 1095-1.

7. **Braddock, J., Heide, S., Spaniardi, A. (2023).** Introduction to the virtual world: Pros and cons of social media. Teens, Screens, and Social Connection: An Evidence-Based Guide to Key Problems and Solutions, pp. 31–48. DOI: 10.1007/978-3-031-24804-7_3.

8. **Bui, H. Q., Tran, T. T. (2023).** CMC users' positive and negative emotions: Features of social media platforms and

users' strategies. In Multidisciplinary Applications of Computer-Mediated Communication. pp. 188–210. DOI: 10.4018/978-1-6684-7034-3.ch010.

9. **Buodo, G., Moretta, T., Santucci, V. G., Chen, S., Potenza, M. N. (2023).** Using social media for social motives moderates the relationship between post-traumatic symptoms during a COVID-19-related lockdown and improvement of distress after lockdown. Behavioral Sciences, Vol. 13, No. 1, pp. 53. DOI: 10.3390/bs13010053.

10. **Calvo, H., Carrillo-Mendoza, P., Gelbukh, A. (2018).** On redundancy in multi-document summarization. Journal of Intelligent and Fuzzy Systems, Vol. 34, No. 5, pp. 3245–3255. DOI: 10.3233/jifs-169507.

11. **Chen, Z., Yang, R., Fu, S., Zong, N., Liu, H., Huang, M. (2023).** Detecting reddit users with depression using a hybrid neural network. Proceedings of the 11th IEEE International Conference on Healthcare Informatics, pp. 610–617. DOI: 10.1109/ICTACS56270.2022.9988489.

12. **Clark, K., Luong, M. T., Le, Q. V., Manning, C. D. (2020).** ELECTRA: pre-training text encoders as discriminators rather than generators. International Conference on Learning Representations, pp. 1–18. DOI: 10.48550/ARXIV.2003.10555.

13. **Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V. (2019).** Unsupervised cross-lingual representation learning at scale. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 8440–8451.

14. **Cornn, K. (2020).** Identifying depression on social media. Stanford University Stanford.

15. **De Choudhury, M., Gamon, M., Counts, S., Horvitz, E. (2021).** Predicting depression via social media. Proceedings of the International AAAI Conference on Web and Social Media, Vol. 7, No. 1, pp. 128–137. DOI: 10.1609/icwsm.v7i1.14432.

16. **Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2019).** BERT: Pre-training of deep bidirectional transformers for language understanding. North American Chapter of the Association for Computational Linguistics, Vol. 1, pp. 4171–4186.

17. **García-Mendoza, C. V., Gambino, O. J., Villarreal-Cervantes, M. G., Calvo, H. (2020).** Evolutionary optimization of ensemble learning to determine sentiment polarity in an unbalanced multiclass corpus. Entropy, Vol. 22, No. 9, pp. 1020. DOI: 10.3390/e22091020.

18. **Haque, U. M., Kabir, E., Khanam, R. (2021).** Detection of child depression using machine learning methods. Public Library of Science One, Vol. 16, No. 12, pp. e0261131. DOI: 10.1371/journal.pone.0261131.

19. **Hernández-Castañeda, A., Calvo, H., Gelbukh, A., García-Flores, J. J. (2016).** Cross-domain deception detection using support vector networks. Soft Computing, Vol. 21, No. 3, pp. 585–595. DOI: 10.1007/s00500-016-2409-2.

20. **Hoang, T. T., Ojo, O. E., Adebanji, O. O., Calvo, H., Gelbukh, A. (2022).** The combination of BERT and data oversampling for answer type prediction. Proceedings of the Central Europe Workshop, Vol. 3119.

21. **Hosseini-Saravani, S. H., Besharati, S., Calvo, H., Gelbukh, A. (2020).** Depression detection in social media using a psychoanalytical technique for feature extraction and a cognitive based classifier. Proceedings of the 19th Mexican International Conference on Artificial Intelligence. Advances in Computational Intelligence, Vol. 12469, pp. 282–292. DOI: 10.1007/978-3-030-60887-3_25.

22. **Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R. (2019).** ALBERT: A lite BERT for self-supervised learning of language representations.

Proceedings of the International Conference on Learning Representations. Conference Blind Submission. DOI: 10.48550/arXiv.1909.11942.

23. **Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019).** RoBERTa: A robustly optimized BERT pretraining approach. Proceedings of the International Conference on Learning Representations. Conference Blind Submission. DOI: 10.48550/ARXIV.190 7.11692.

24. **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. Proceedings of the 17th International Conference on Information Technology New Generations, pp. 659–662. DOI: 10.1007/978-3-030-43020-7_89.

25. **Muzammel, M., Salam, H., Othmani, A. (2021).** End-to-end multimodal clinical depression recognition using deep neural networks: A comparative analysis. Computer Methods and Programs in Biomedicine, Vol. 211, pp. 106433. DOI: 10.1016/j.cmpb.2021.106433.

26. **Ojo, O., Adebanji, O., Calvo, H., Dieke, D., Ojo, O., Akinsanya, S., Abiola, T., Feldman, A. (2023).** Legend at ArAIEval shared task: Persuasion technique detection using a language-agnostic text representation model. Proceedings of ArabicNLP, pp. 594–599.

27. **Ojo, O. E., Adebanji, O. O., Gelbukh, A., Calvo, H., Feldman, A. (2023).** MedAI dialog corpus (MEDIC): Zero-shot classification of doctor and AI responses in health consultations.

28. **Ojo, O. E., Gelbukh, A., Calvo, H., Adebanji, O. O. (2021).** Performance study of n-grams in the analysis of sentiments. Journal of the Nigerian Society of Physical Sciences, Vol. 3, No. 4, pp. 477–483.

29. **Ojo, O. E., Gelbukh, A., Calvo, H., Adebanji, O. O., Sidorov, G. (2020).** Sentiment detection in economics texts. Proceedings of the 20th Mexican International Conference on Artificial Intelligence. Advances in Computational Intelligence, pp. 271–281. DOI: 10.1007/978-3-030-60887-3_24.

30. **Ojo, O. E., Gelbukh, A., Calvo, H., Feldman, A., Adebanji, O. O., Armenta-Segura, J. (2022).** Language identification at the word level in code-mixed texts using character sequence and word embedding. Proceedings of the 19th International Conference on Natural Language Processing: Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts, pp. 1–6.

31. **Ojo, O. E., Ta, H. T., Gelbukh, A., Calvo, H., Adebanji, O. O., Sidorov, G. (2023).** Transformer-based approaches to sentiment detection. Vol. 2, pp. 101–110. DOI: 10.1007/978-3-031-23476-7_10.

32. **Ojo, O. E., Ta, T. H., Gelbukh, A., Calvo, H., Sidorov, G., Adebanji, O. O. (2022).** Automatic hate speech detection using deep neural networks and word embedding. Computación y Sistemas, Vol. 26, No. 2, pp. 1007–1013. DOI: 10.13053/cys-26-2-410 7.

33. **Parmar, A., Katariya, R., Patel, V. (2019).** A review on random forest: An ensemble classifier. International Conference on Intelligent Data Communication Technologies and Internet of Things, Vol. 26, pp. 758–763. DOI: 10.1007/978-3-030-03146-6_86.

34. **Peng, J., Feldman, A., Jazmati, H. (2015).** Classifying idiomatic and literal expressions using vector space representations. Proceedings of the International Conference Recent Advances in Natural Language Processing, pp. 507–511.

35. **Peng, J., Feldman, A., Vylomova, E. (2014).** Classifying idiomatic and literal expressions using topic models and intensity of emotions. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 2019–2027. DOI: 10.3115/v1/D14-1216.

36. **Popat, A., Tarrant, C. (2022).** Exploring adolescents' perspectives on social media and mental health and well-being – A qualitative literature review. Clinical Child Psychology and Psychiatry, Vol. 28, No. 1, pp. 323–337. DOI: 10.1177/13591045221092884.

37. **Raschka, S. (2014).** Naive Bayes and text classification I - Introduction and theory. arXiv. DOI: 10.48550/arXiv.1410.5329.

38. **Reece, A. G., Danforth, C. M. (2017).** Instagram photos reveal predictive markers of depression. European Physical Journal of Data Science, Vol. 6, pp. 15. DOI: 10.1140/epjds/s13688-017-0110-z.

39. **Sanh, V., Debut, L., Chaumond, J., Wolf, T. (2019).** DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. Proceedings of the 5th Edition Co-located with Neural Information Processing Systems, pp. 1–5. DOI: 10.48550/arXiv.1910.01108.

40. **Schein, A. I., Ungar, L. H. (2007).** Active learning for logistic regression: an evaluation. Machine Learning, Vol. 68, No. 3, pp. 235–265. DOI: 10.1007/s10994-007-5019-5.

41. **Shahiki-Tash, M., Armenta-Segura, J., Ahani, Z., Kolesnikova, O., Sidorov, G., Gelbukh, A. (2023).** LIDOMA at HOMO-MEX2023@IberLEF: Hate speech detection towards the mexican spanish-speaking LGBT+ population. The importance of preprocessing before using BERT-based models. Proceedings of the Central Europe Workshop and Iberian Languages Evaluation Forum, Vol. 3496.

42. **Shahiki-Tash, M., Armenta-Segura, J., Ahani, Z., Kolesnikova, O., Sidorov, G., Gelbukh, A. (2023).** LIDOMA@ DravidianLangTech: Convolutional neural networks for studying correlation between lexical features and sentiment polarity in Tamil and Tulu languages. Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages, pp. 180–185.

43. **Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., Gordon, J. (2012).** Empirical study of machine learning based approach for opinion mining in tweets. Proceedings of the 11th Mexican International Conference on Advances in Artificial Intelligence, Vol. 7629, pp. 1–14. DOI: 10.1007/978-3-642-37807-2_1.

44. **Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., Chanona-Hernández, L. (2014).** Syntactic n-grams as machine learning features for natural language processing. Expert Systems with Applications, Vol. 41, No. 3, pp. 853–860. DOI: 10.1016/j.eswa.2013.08.015.

45. **Swain, P. H., Hauska, H. (1977).** The decision tree classifier: Design and potential. IEEE Transactions on Geoscience Electronics, Vol. 15, No. 3, pp. 142–147. DOI: 10.1109/TGE.1977.6498972.

46. **Taud, H., Mas, J. F. (2018).** Multilayer perceptron (MLP). Geomatic approaches for modeling land change scenarios, pp. 451–455. DOI: 10.1007/978-3-319-60801-3_27.

47. **Tovar, M., Rosillo, M., Spaniardi, A. (2023).** Social media's influence on identity formation and self expression. Teens, Screens, and Social Connection: An Evidence-Based Guide to Key Problems and Solutions, pp. 49–61. DOI: 10.1007/978-3-031-24804-7_4.

48. **Trotzek, M., Koitka, S., Friedrich, C. M. (2018).** Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. IEEE Transactions on Knowledge and Data Engineering, Vol. 32, No. 3, pp. 588–601. DOI: 10.1109/tkde.2018.2885515.

49. **Uddin, M. Z., Dysthe, K. K., Følstad, A., Brandtzaeg, P. B. (2021).** Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Computing and Applications, Vol. 34, No. 1, pp. 721–744. DOI: 10.1007/s00521-021-06426-4.

50. **Wang, H., Liu, Y., Zhen, X., Tu, X. (2021).** Depression speech recognition with a three-dimensional convolutional network. Frontiers in Human Neuroscience, Vol. 15. DOI: 10.3389/fnhum.2021.713823.

51. **Wen, S. (2021).** Detecting depression from tweets with neural language processing. Journal of Physics: Conference Series, Vol. 1792, No. 1, pp. 12058. DOI: 10.1088/1742-6596/1792/1/012058.

52. **Wolohan, J. T., Hiraga, M., Mukherjee, A., Sayyed, Z. A., Millard, M. (2018).** Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with NLP. Proceedings of the 1st International Workshop on Language Cognition and Computational Models, pp. 11–21.

53. **Zhang, T. (2004).** Solving large scale linear prediction problems using stochastic gradient descent algorithms. Proceedings of the 21st International Conference on Machine Learning, pp. 116. DOI: 10.1145/1015330.1015332.

54. **Ziwei, B. Y., Chua, H. N. (2019).** An application for classifying depression in tweets. Proceedings of the 2nd International Conference on Computing and Big Data, pp. 37–41. DOI: 10.1145/3366650.3366653.

Article received on 30/08/2023; accepted on 16/11/2023.
** Corresponding author is Hiram Calvo.*