International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 06 | Jun 2026 www.irjet.net p-ISSN: 2395-0072

An Intelligent BERT Based Transfer Learning Model for Depression Identification in Online Social Media Platform

Rakhi Soni[1], Bhanu Pratap Singh[2]

[1]*Research scholar, Aditya College of Technology and Science, Satna, Madhya Pradesh, India*
[2] *HOD, CSE, Aditya College of Technology and Science, Satna, Madhya Pradesh, India*

Abstract - *Depression is a critical mental health disorder affecting millions of individuals worldwide, often remaining undiagnosed due to limitations in traditional clinical assessment methods. With the rapid growth of social media platforms, large volumes of user-generated textual data provide an opportunity for early detection of depressive behavior. This study proposes an intelligent transfer learning-based approach using BERT Base Uncased for depression identification from social media text. The model leverages deep contextualized embeddings to capture semantic and linguistic patterns associated with depressive expressions. The proposed system includes data preprocessing, tokenization, model fine-tuning, and evaluation phases. Experiments were conducted on a benchmark dataset containing 7,731 social media comments, achieving competitive performance in terms of accuracy, precision, recall, and F1-score. Comparative analysis demonstrates that the BERT-based model outperforms traditional and conventional deep learning approaches in handling complex textual patterns. The results highlight the effectiveness of transformer-based models for early and reliable depression detection. Future work aims to incorporate multimodal data such as images and voice to further enhance prediction accuracy and robustness.*

Key Words: Depression Detection, BERT Base Uncased, Transfer Learning, Natural Language Processing (NLP), Social Media Analysis, Deep Learning, Sentiment Analysis, Mental Health Prediction

1. INTRODUCTION

Depression is a sizeable mental contamination that impacts hundreds of thousands of people around the sector. Depression is characterised through loss of hobby, longer unhappiness, and numerous emotional and bodily problems that have an effect on each day functioning and exceptional of existence consistent with the WHO [1], one of eight people in the world suffer from all types of intellectual disease, and the most of the populace is unable to utilize powerful cures. Techniques traditionally used to diagnose despair encompass clinical tests and questionnaires stuffed by means of patients whose patients are met, as questions are usually now not looking for help, rather than the entire individual affected by depressive signs and symptoms. With the arrival of numerous social media websites, along with X (formerly Twitter), Instagram, Reddit, fb, and greater, we've opened the streets to locate human conduct and moods in actual time. People often share their ideas, feelings and matters they do in this social media platform each day and generate big statistics which could show statistics about their intellectual contamination. As an end result, the concept of using social media facts to recognize melancholy has come to be famous. The machine learning and Deep Learning techniques have proven to be extremely promising in regions along with natural language processing (NLP), predictive analytics, and photo recognition. The advanced technologies can predict consequences very appropriately via reading big amounts of unstructured statistics and shooting hidden patterns. When figuring out depression the usage of social media posts, ML and DL models can be educated to perceive verbal data, behavioural styles, and other signs of melancholy symptoms. Wald (RF) [7], [8], [9]. These early methods established the capability of MLS on this area, however have been often restrained by using the variety and complexity of human language. There are awesome advances in the creation of deep mastering models. To extract information from text, researchers used complicated fashions along with the repeating neural network (RNNS), [11], [12] and folding community (CNNS) [13] [15] [16]. This discipline included multimodal information [17], [18], [19] to file further enhancements, combining text, images and user conduct to provide a more whole information of the user's intellectual country. Advances in learning transmission and growing overstated language fashions have enabled a greater green and powerful depression detection gadget. Our paper represent the work would include having performance evaluation with a balanced dataset as well as include more deep learning models and incorporating advanced pre-trained word embedding's using GloVe and BERT to extract features

2. RELATED WORK

Recent studies have explored a variety of machine learning and deep learning techniques for depression detection using social media data. In study [20], an innovative BERT-RF

feature engineering approach was introduced, combining contextualized embeddings with probabilistic features to enhance textual representation. Similarly, study [21] proposed a hybrid deep learning model integrating CNN and LSTM architectures, further improved through a Two-State LSTM (TS-LSTM) with an attention mechanism. This model demonstrated strong performance, achieving high accuracy, precision, recall, and F1-score. However, its limitations include reduced effectiveness on long text sequences and lack of integration with advanced pre-trained models such as ELMo and BERT.

In study [22], depression intensity classification was performed using a labeled Twitter dataset, where multiple transformer-based lightweight models were evaluated. The ESG model outperformed others, including DistilBERT, achieving an F1-score of 89% with reduced training time. Nonetheless, the model's applicability is limited to short-text data such as tweets, indicating the need for training on more diverse datasets like Reddit to improve generalization.

A hybrid machine learning framework named D2X was proposed in study [23], utilizing sentiment analysis on social media platform X (Twitter) data. The model integrates Support Vector Machine (SVM) and Random Forest techniques to process multimodal inputs such as text, emoticons, and images. Despite its effectiveness, challenges remain in handling heterogeneous data types and addressing cultural and linguistic biases, as the dataset is limited to Thai language content.

In study [24], a deep learning architecture combining CNN and BiLSTM with an attention mechanism (CBA) was introduced and evaluated on the CLEF2017 dataset. The model achieved superior performance compared to existing approaches. However, the study highlights the need for balanced datasets, inclusion of additional deep learning models, and the integration of advanced embeddings such as GloVe and BERT to further enhance performance.

Study [25] proposed a novel framework leveraging fastText embeddings, TF-IDF features, and XGBoost classifier to distinguish between depression and suicidal content on Reddit. While the approach improved classification accuracy and contextual understanding, it lacks the incorporation of diverse emotional features and multiclass classification for varying depression intensities.

Finally, study [26] presented an attention-based BiLSTM-CNN model for detecting depressive content in Bangla social media text. The model achieved high accuracy and demonstrated cross-lingual effectiveness when applied to English data. However, limitations include higher false-positive rates and the need for more diverse datasets, particularly for low-resource languages.

Overall, existing research highlights the effectiveness of hybrid and transformer-based models in depression detection. However, challenges such as handling long-text data, dataset diversity, cross-lingual adaptability, and incorporation of advanced contextual embeddings remain open for further investigation.

3. METHODOLOGY

This section outlines the generic architecture of a depression detection system, detailing each component involved in the process, from data acquisition to evaluation. The architecture of the proposed model is shown below in figure below:

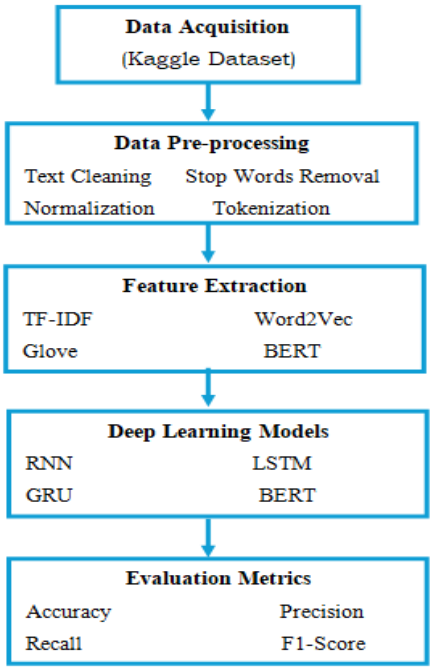


Fig-1: System architecture

BERT Base Uncased is a transformer-based NLP model that understands word context by analyzing both left and right surroundings (bidirectional learning). It is a smaller, efficient version of BERT with 12 layers, 768 hidden units, 12 attention heads, and about 110 million parameters. The "uncased" variant ignores capitalization, treating words like "apple" and "Apple" as the same, making it suitable for tasks where case sensitivity isn't important.

It is pre-trained on large text datasets (Books Corpus and Wikipedia) using two key objectives:

- **Masked Language Modeling (MLM):** Predicts missing words in a sentence.
- **Next Sentence Prediction (NSP):** Determines if two sentences logically follow each other.

This pre-training allows BERT Base Uncased to perform effectively across various NLP tasks such as text classification, question answering, and sentiment analysis.

Basic steps for proposed BERT model includes following:

- **Initialize Model**
  Load BERT Base Uncased model and tokenizer.
- **Preprocess Text**
  Tokenize input, add [CLS] and [SEP], convert to token IDs, apply padding, and create attention masks.
- **Convert to Tensors**
  Prepare input IDs and attention masks as model inputs.
- **Fine-Tune Model (Optional)**
  Train on labeled data using a suitable loss function and optimizer.
- **Run Inference**
  Pass processed input through the model to obtain logits.
- **Generate Predictions**
  Apply SoftMax/Sigmoid and select output labels (or spans for QA, token labels for NER).
- **Post-Process Output**
  Convert predictions into readable format (labels, entities, or answers).
- **Evaluate**
  Measure performance using accuracy, precision, recall, and F1-score.

4. RESULT

The datasets were considered from Kaggle. This dataset has 7731 comments from social media platforms in which 3900 comments are not depressed while 3831 comments are depressed represented by 0 and 1 respectively. Distribution of classes in dataset is shown below:

Fig-2: Class distribution

Confusion matrix for proposed model is shown below:

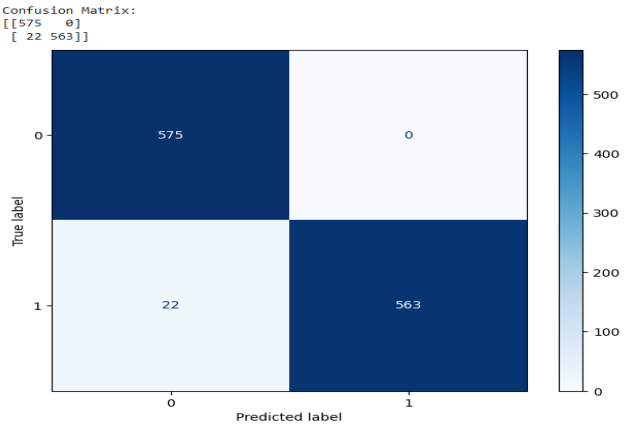


Fig-3: Confusion matrix

Testing accuracy is shown below:

Test Accuracy: 0.9810

	precision	recall	f1-score	support
0	0.9631	1.0000	0.9812	575
1	1.0000	0.9624	0.9808	585
accuracy			0.9810	1160
macro avg	0.9816	0.9812	0.9810	1160
weighted avg	0.9817	0.9810	0.9810	1160

Fig-4: Classification report

Chart below represents evaluation of accuracy for proposed model with many existing models:

Chart-1: Accuracy comparison
Comparison char6t for precision is shown below:

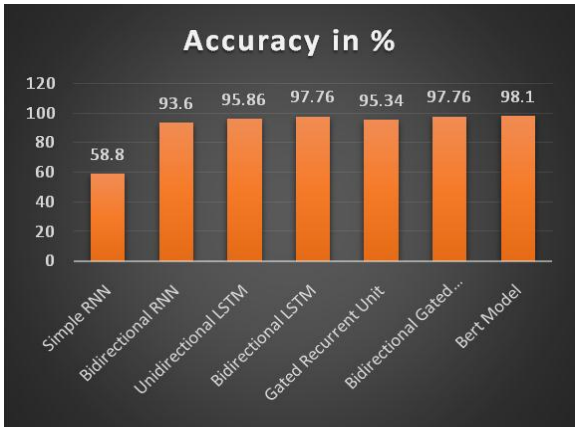
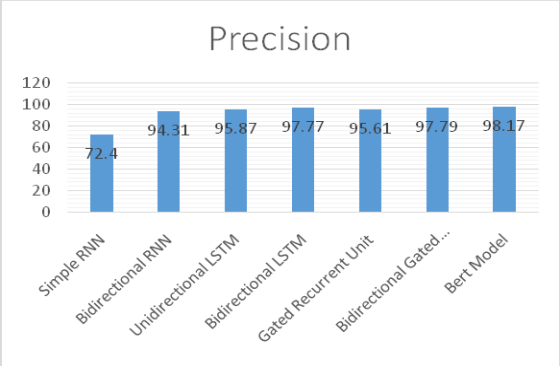


Chart-2: Precision comparison



5. CONCLUSIONS

This work focuses on developing accurate models for early depression detection using NLP techniques. We conducted a comparative study of various RNNs and Transfer models variants with social media text data, finding that they all performed similarly in accuracy. However, the BERT model variant showed slight advantages in handling large datasets and complex phrases. Notably, Bidirectional LSTM and GRU models slightly outperformed their unidirectional counterparts, indicating that all NLP tasks require bidirectional analysis. From this work, we can safely conclude that this work contributes to more reliable and accurate depression diagnosis models through NLP techniques, with practical applications for mental health support. In future, we plan to integrate multimodal data, such as images and voice, to create more comprehensive and accurate models for early depression detection

REFERENCES

[1] World Health Org. (2022). Mental Disorders. Accessed: Jun. 6, 2022. [Online]. Available: https://www.who.int/newsroom/factsheets/detail/mental-disorders.

[2] N. A. Asad, Md. A. Mahmud Pranto, S. Afreen, and Md. M. Islam, "Depression detection by analysing social media posts of user," in Proc. IEEE Int. Conf. Signal Process., Inf., Commun. Syst. (SPICSCON), Nov. 2019, pp. 13–17.

[3] Fatima, B. U. D. Abbasi, S. Khan, M. Al-Saeed, H. F. Ahmad, and R. Mumtaz, "Prediction of postpartum depression using machine learning techniques from social media text," Expert Syst., vol. 36, no. 4, Aug. 2019, Art. no. e12409, 10.1111/exsy.12409.

[4] Y. Ding, X. Chen, Q. Fu, and S. Zhong, "A depression recognition method for college students using deep integrated support vector algorithm," IEEE Access, vol. 8, pp. 75616–75629, 2020.

[5] R. Skaik and D. Inkpen, "Using Twitter social media for depression detection in the Canadian population," in Proc. 3rd Artif. Intell. Cloud Compute. Conf., New York, NY, USA, Dec. 2020, pp. 109–114, doi: 10.1145/3442536.3442553.

[6] H. S. Alsagri and M. Ykhlef, "Machine learning-based approach for depression detection in Twitter using content and activity features," IEICE Trans. Inf. Syst., vol. E103.D, no. 8, pp. 1825–1832, Aug. 2020, doi: 10.1587/TRANSINF.2020EDP7023.

[7] Z. N. Vasha, B. Sharma, I. J. Esha, J. Al Nahian, and J. A. Polin, "Depression detection in social media comments data using machine learning algorithms," Bull. Electr. Eng. Informat. vol. 12, no. 2, pp. 987–996, Apr.2023.[Online]Available:https://beei.org/index.php/EEI/article/view/4182/3162.

[8] K. Sabaneh, M. A. Salameh, F. Khaleel, M. M. Herzallah, J. Y. Natsheh, and M. Maree, "Early risk prediction of depression based on social media posts in Arabic," in Proc. IEEE 35th Int. Conf. Tools Artif. Intell. (ICTAI), Nov. 2023, pp. 595–602.

[9] Helmy, R. Nassar, and N. Ramdan, "Depression detection for Twitter users using sentiment analysis in English and Arabic tweets," Artif. Intell. Med., vol. 147, Jan. 2024, Art. no. 102716.

[10] H. Uddin, D. Bapery, and A. S. M. Arif, "Depression analysis from social media data in Bangla language using long short term memory (LSTM) recurrent neural network technique," in Proc. Int. Conf. Compute., Commun., Chem., Mater. Electron. Eng. (IC4ME2), Jul. 2019, pp. 1–4.

[11] H. Kour and M. K. Gupta, "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM," Multimedia Tools Appl., vol. 81, no. 17, pp. 23649–23685, Jul. 2022, doi: 10.1007/s11042-022-12648-y.

[12] S. R. Narayanan, S. Babu, and A. Thandayantavida, "Detection of depression from social media using deep learning," J. Positive School Psychol., vol. 6, no. 4, pp. 4909–4915, 2022.

[13] Husseini Orabi, P. Buddhitha, M. Husseini Orabi, and D. Inkpen, "Deep learning for depression detection of Twitter users," in Proc. 5th Workshop Compute. Linguistics Clin. Psychol., Keyboard Clinic, K. Loveys, K. Niederhoffer, E. Prud'hommeaux, R. Resnik, and P. Resnik, Eds., Jun. 2018, pp. 88–97. [Online]. Available: https://aclanthology.org/W18-0609.

[14] P. K. Gamaarachchige and D. Inkpen, ‘‘Multi-task, multi-channel, multi-input learning for mental illness detection using social media text,’’ in Proc. 10th Int. Workshop Health Text Mining Inf. Anal. (LOUHI), E. Holderness, A. Jimeno Yepes, A. Lavelli, A.-L. Minard, J. Pustejovsky, and F. Rinaldi, Eds., Nov. 2019, pp. 54–64. [Online]. Available: https://aclanthology.org/D19-6208.

[15] Lin, P. Hu, H. Su, S. Li, J. Mei, J. Zhou, and H. Leung, ‘‘Sense Mood: Depression detection on social media,’’ in Proc. Int. Conf. Multimedia Retr., New York, NY, USA, Jun. 2020, p. 407, doi: 10.1145/3372278.3391932.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ‘‘BERT: Pre-training of deep bidirectional transformers for language understanding,’’ in Proc. Conf. North Amer. Chapter Assoc. Compute. Linguistics, Hum. Lang. Technol., vol. 1, J. Burstein, C. Doran, and T. Solorio, Eds., Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423.

[17] M. An, J. Wang, S. Li, and G. Zhou, ‘‘Multimodal topic-enriched auxiliary learning for depression detection,’’ in Proc. 28th Int. Conf. Compute. Linguistics, Barcelona, Spain, D. Scott, N. Bel, and C. Zong, Eds., 2020, pp. 1078–1089. [Online]. Available: https://aclanthology.org/2020.coling-main.94.

[18] Y. Chiu, H. Y. Lane, J. L. Koh, and A. L. P. Chen, ‘‘Multimodal depression detection on Instagram considering time interval of posts,’’ J. Intell. Inf. Syst., vol. 56, no. 1, pp. 25–47, Feb. 2021, doi: 10.1007/s10844-020- 00599-5.

[19] Zafar, D. Aftab, R. Qureshi, Y. Wang, and H. Yan, ‘‘Multi-explainable temporalNet: An interpretable multimodal approach using temporal convolutional network for user-level depression detection,’’ in Proc. IEEE/CVF Conf. Compute. Vis. Pattern Recognition. Workshops (CVPRW), Jun. 2024, pp. 2258–2265.

[20] M. A. Abbas, K. Munir, A. Raza, N. A. Samee, M. M. Jamjoom and Z. Ullah, "Novel Transformer Based Contextualized Embedding and Probabilistic Features for Depression Detection From Social Media," in *IEEE Access*, vol. 12, pp. 54087-54100, 2024, doi: 10.1109/ACCESS.2024.3387695.

[21] S. Almutairi, M. Abohashrh, H. H. Razzaq, M. Zulqarnain, A. Namoun and F. Khan, "A Hybrid Deep Learning Model for Predicting Depression Symptoms From Large-Scale Textual Dataset," in *IEEE Access*, vol. 12, pp. 168477-168499, 2024, doi: 10.1109/ACCESS.2024.3496741.

[22] M. Rizwan, M. F. Mushtaq, U. Akram, A. Mehmood, I. Ashraf and B. Sahelices, "Depression Classification From Tweets Using Small Deep Transfer Learning Language Models," in *IEEE Access*, vol. 10, pp. 129176-129189, 2022, doi: 10.1109/ACCESS.2022.3223049.

[23] T. Angskun, S. Tipprasert, A. Thippongtorn and J. Angskun, "D_2X: Depression Detection System through X Using Hybrid Machine Learning," in *IEEE Access*, vol. 12, pp. 172820-172831, 2024, doi: 10.1109/ACCESS.2024.3502241.

[24] Joel Philip Thekkekara, Sira Yongchareon, Veronica Liesaputra, “An attention-based CNN-BiLSTM model for depression detection on social media text”, Expert Systems with Applications, Volume 249, Part C, 2024, 123834, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2024.123834.

[25] Sayani Ghosal, Amita Jain, “Depression and Suicide Risk Detection on Social Media using fastText Embedding and XGBoost Classifier”, Procedia Computer Science, Volume 218, 2023, Pages 1631-1639, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.141.

[26] Tapotosh Ghosh, Md. Hasan Al Banna, Md. Jaber Al Nahian, Mohammed Nasir Uddin, M. Shamim Kaiser, Mufti Mahmud, “An attention-based hybrid architecture with Explainability for depressive social media text detection in Bangla”, Expert Systems with Applications, Volume 213, Part C, 2023, 119007, ISSN 0957-4174,https://doi.org/10.1016/j.eswa. 2022. 119007.