

## ABSTRACT

Depression impacts around 280 million people worldwide. It is defined by enduring sadness and a persistent loss of interest. Limited access to treatment due to high costs and availability issues, highlights the need for affordable early detection methods. Machine learning has shown promise in detecting depression, especially using text data from social media, where users share emotions openly. This study investigates the use of BERT, a transformer model, combined with the Grey Wolf Optimizer (GWO) to detect depression in tweets, by applying a professionally re-labelled Kaggle dataset to enhance early detection. The optimized parameters include pre-trained models, batch sizes, and learning rates. This study reveals that the GWO significantly enhances the performance of BERT in text-based depression detection. The best performance is achieved using BERT optimized by GWO; it is outperforming when using BERT alone. The best parameter combination, which achieves the best validation f1-score, is a model name called bert-base-cased-finetuned-mrpc, batch size of 64, and learning rate of 0.0001. The testing set results an accuracy of 0.8400 and precision, recall, and f1-score of 0.8356.

**Keywords:** depression detection, text classification, BERT, GWO