CHAPTER 1

INTRODUCTION

1.1 Background

Depression is a mental illness. It is defined by enduring sadness and a persistent loss of interest, setting it apart from ordinary mood changes [1]. In 2019, it affected about 3.59% of the global population or roughly 280 million people [2]. Depression can greatly affect individuals on various areas of life, such as personal relationships, work, and school. People who have gone through trauma, major loss, or other stressful situations are more likely to develop depression [1].

Generally, depression is a treatable condition, yet only a limited number of individuals have access to treatment [3]. The process of receiving a diagnosis for depression involves seeking assistance from a psychiatric specialist, which frequently involves considerable expenses per session. Based on a report by CNBC, conventional therapy sessions lasting one hour cost about $65 to 250$ in the USA [4]. Consequently, therapy may not be affordable for everyone. Therapy may also not be easily accessible due to several factors such as location, availability of mental health professionals, and stigma surrounding seeking treatment [5] [6]. However, it has been shown through several studies that early recognition and treatment of depression can lead to improvements in the negative impacts of the disorder [7–9]. Hence, there is a need to investigate alternative and affordable methods for early detection of depression that are more widely accessible.

One of the alternative approaches to detecting depression is by using machine learning methods. This method can be implemented in diverse forms of data, including text, images, and audio, and has demonstrated promising results [10–15]. Uddin, et al.. studied depression detection using text data in 2022. They utilized a Norwegian information website as the source of their dataset. TThe model's performance is outstanding, achieving an accuracy of 0.9988. This high accuracy is obtained when using 21,470 data along with 5-fold cross-validation, robust features, and RNN [14]. Amanat et al. study another text-based depression detection in 2022. By combining LSTM and RNN, they succeeded in attaining high accuracy score of 99.66% with the use of 10-fold validation [11].

People use social media to share different types of information, such as text, pictures, audio, and more. On social media, individuals are known to share their emotions and experiences openly [16]. Consequently, social media platforms provide a bunch of data that can be utilized for the purpose of identifying depression, as evidenced by several studies on using social media data for depression detection [17–21]. Furthermore, a strong correlation was observed between an individual's social media posts expressing feelings of depression and the symptoms they self-reported on a depression screening tool[22]. X, formerly Twitter, is one of the examples of social media that provide microblogging services. User's posts, called "tweets," are limited to 280 characters each [23]. By December 2022, there is a large amount of X monthly users, which is around 368 million people globally [24],leading to the creation of vast and large-scale tweet data.

Several studies have explored the use of machine learning techniques for detecting depression through Twitter data[25–27]. In 2022, Nadeem, et al proposed a method of tweet-based depression detection using TF-IDF and the n-gram technique along with several machine learning algorithms. They discovered that SVM delivers the best performance with the uni-bi-gram technique on binary labeled data, achieving a 96.8% accuracy and a 96.7% F1-score [26]. Reseena Mol et al. study the usage of word2vec (Skip Gram and CBOW), GloVe, and fasttext along with several machine learning algorithms to detect depression from tweets. GloVe method consistently returns the best accuracy and is then used along with Ensemble Method (Random Forest, Bagging Tree, XGBoosting, Gradient Boosting, AdaBoost, and Extra Tree) and gives 97% accuracy [25]. A study conducted by Wongkoblap, A., et al. in 2021 uses modified multiple instance learning (MIL), namely MILA-SocNet and MIL-SocNet. The maximum accuracy obtained is 92%, higher than deep learning, user2vec, LIWC, and other algorithms [27]. texts using deep learning models. Their approach fused two deep learning algorithms: a CNN paired with a BBiLSTM, and yielded an accuracy of 94.28% after optimization [28].

As mentioned in the previous paragraph, deep learning has shown great potential in detecting depression using different types of data, such as text, audio, images, and many more. However, there is a new approach that has emerged in recent years, known as transformer methods. The transformer model, first introduced by Vaswani et al. in 2017, is the pioneering transduction model that uses self-attention to compute depictions of both its inputs and outputs rather than employing sequence-aligned RNNs or convolution [29]. Following its introduction, the transformer method has been adopted and expanded upon by other researchers, leading to the development of newer methods such as BERT, RoBERTa, DistilBERT, ALBERT, and others [30–33].

BERT (Bidirectional Encoder Representations from Transformers) is built upon the transformer architecture with multiple encoder layers. It uses an attention mechanism to understand word relationships in both directions, generating representations for deep learning models. Pre-trained on large text corpora, BERT uses tasks like masked language modeling and next sentence prediction.For specific NLP tasks, an extra output layer and transfer learning allow for cutting-edge performance. BERT's fine-tuning capabilities make it versatile for single texts or text pairs, simplifying tasks by using self-attention on combined text pairs for automatic comparison [30].

BERT and other transformer models have been applied to depression detection. For instance, Filip Nilsson and György Kovács (2022) employed a fine-tuned multiclass BERT transformer on the LT-EDI-ACL2022 dataset, achieving an f1-score of 0.52 using Google's BERT with 12 layers and 110 million parameters [34]. Lu et al. (2022) carried out an extensive study using the DAIC dataset with BERT, resulting in an f1-score of

0.76 [35]. Kwee and Zahra (2022) developed a depression detection system on Reddit posts, finding that BERT using a learning rate of 1×10^{-6} achieved a high accuracy of 97.45% [36]. Devaguptam et al. (2022) used BERT and DeBERTa for depression detection, with DeBERTa without augmentation achieving the highest recall (0.806) and DeBERTa with augmentation achieving the highest f1-score (0.248), despite being limited to one epoch due to computing constraints [37]. Despite the promising results, further research is needed to improve their application.

In previous research, detection was carried out by manually tuning the parameters and settings. However, with the development of techniques called AutoML, which stands for automated machine learning, there is an emerging trend towards automating the detection process. These techniques work by automatically determining hyperparameters to optimize the performance of machine learning systems. Modern deep neural networks, including transformer methods, heavily rely on various hyperparameter choices related to architecture, regularization, and optimization. Automated hyperparameter optimization (HPO) serves several crucial purposes, including reducing human effort in machine learning applications [38]. The automation of ML parameter optimization can be done using meta-heuristic algorithms like Grey Wolf Optimization (GWO). GWO can address the auto-tuning issue by efficiently exploring the hyperparameter space and identifying optimal settings for machine learning systems. This technique emulates the hunting and social structure of grey wolves in the wild, classifying them into alpha, beta, delta, and omega roles, and incorporating the three phases of hunting: locating prey, encircling it, and attacking [39].

GWO has been used for parameter tuning. Mustafa and Husein showed that the DBSCAN-GWO model achieves 98% accuracy in detecting botnet data, outperforming traditional methods [40]. Zhang's 2021 study found that the GWO-SVM model improves accuracy and operating rate in soil liquefaction prediction [41]. In 2024, Xu demonstrated that GWO enhances coal consumption prediction accuracy, surpassing traditional models [42].

The objective of this study is to implement baseline BERT model in detecting depression from tweets. Additionally, it investigates the effectiveness of auto-tuning BERT using the GWO. Finally, the study assesses the performance of the BERT model optimized with GWO in the context of depression detection.

1.2 Theoretical Framework

The theoretical framework for detecting depression through text analysis using BERT can be improved by adding an optimization algorithm known as Grey Wolf Optimization (GWO). This algorithm, implemented with NiaPy tools, helps improve the model's performance by finding the best settings for the model. Using GWO makes the model work better by finding the best configurations that give the most accurate predictions.

1. Depression Detection Theory
   This theory focuses on identifying signs of depression from textual data. It involves understanding the features and patterns within text that are indicative of depressive symptoms. Research in this area includes developing methods to accurately capture these features and designing models that can effectively differentiate between depressive and non-depressive texts.

2. BERT Theory
   BERT (Bidirectional Encoder Representations from Transformers) is an advanced model for processing and understanding natural language. This theory involves using BERT to analyze and classify text data based on its contextual representations. Research includes training BERT on relevant datasets, fine-tuning it for depression detection, and improving its classification accuracy through various techniques.

3. Optimization Algorithm Theory
   Optimization theory is crucial for enhancing the performance of BERT models. The Grey Wolf Optimization (GWO) algorithm is particularly effective in fine-tuning model parameters to attain the highest possible predictive performance. This theory includes understanding the parameter space, optimizing hyperparameters, and evaluating the performance of optimized models.

1.3 Conceptual Framework/Paradigm

The conceptual framework helps to understand the key factors related to this research clearly and shows how they interact with each other using a diagram. This study, which focuses on detecting depression through text analysis and optimization, identifies and explains the following key elements:

1. Text Data
   Text data refers to the text inputs collected from X (formerly Twitter) posts then stored and shared on the Kaggle website. These texts serve as the primary input for the predictive model.

2. Dataset Relabelling
   Relabelling involves the process of verifying and updating the labels of the text data with the help of a psychologist. A psychologist reviews the texts to ensure the labels accurately reflect whether the text indicates depression (”yes”) or not (”no”). This step improves the quality and reliability of the dataset.

3. BERT Model
   This represents the Bidirectional Encoder Representations from Transformers (BERT) model used for text analysis. BERT is designed to understand the context and nuances of language, making it suitable for detecting depression in text. The model processes the text data and classifies the text as either indicating depression (”yes”) or not indicating depression (”no”).

4. Optimization Algorithm (GWO)
   The GWO algorithm introduces optimization into the research framework. It aims to find optimal configurations of the BERT model's parameters that maximize predictive accuracy. The GWO algorithm iteratively adjusts the model's parameters, evaluating their performance using the validation f1-score. The best parameters are selected based on the highest validation f1-score achieved.

The conceptual paradigm illustrates how these variables are interconnected, as depicted in Fig. 1

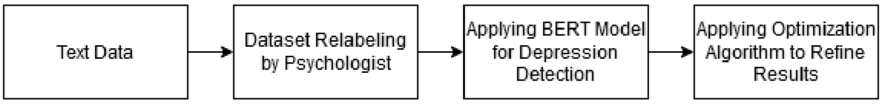

Figure 1: Conceptual Paradigm

The identified components are as follows:

a. Text Data: They are collected from open-source datasets on Kaggle, which were originally scraped from X (formerly Twitter) by the dataset creator.
b. Dataset Relabeling by Psychologist: These datasets are verified and updated labels by a psychologist to ensure accuracy.
c. Applying BERT Model for Depression Detection: This study uses BERT to analyze text and identify depression patterns.
d. Applying Optimization Algorithm to Refine Results: This study uses GWO to find the best model parameters based on validation f1-score.

This paradigm showcases the flow of information and actions within the research process, where text data is collected from open-source datasets on Kaggle, which were originally scraped from Twitter by the dataset creator. The dataset is then relabeled by a psychologist to ensure accuracy. A BERT model is applied to detect depression in the text, and an optimization algorithm (GWO) is used to refine the results and enhance the model's accuracy through parameter tuning.

1.4 Research Problem

The overall problem statement is outlined as follows.

1. How is the performance of the baseline BERT method in tweet-based depression detection?
2. How effective is GWO in enhancing the tweet-based depression detection?
3. How is the performance of the BERT model after optimization using GWO in the context of depression detection?

1.5 Objective and Hypotheses

1.5.1 Objectives

a. To implement baseline BERT model in detecting depression from tweets.
b. To investigate the effectiveness of auto-tuning BERT using the GWO.
c. To compare the performance of BERT baseline with BERT Optimized using GWO.

1.5.2 Hypotheses

Premise 1: Research indicates that while BERT is widely used for natural language processing tasks, including depression detection, many studies do not fully optimize its hyperparameters, which may limit the model's potential performance.
Premise 2: Previous studies [40–42] have shown that the Grey Wolf Optimizer (GWO) enhances the effectiveness of different machine learning approaches, including SVM and DBSCAN, by effectively exploring the hyperparameter space and identifying optimal configurations.
Premise 3: By applying GWO to optimize the hyperparameters of BERT, the model's f1-score, can be significantly improved, resulting in enhancing the overall effectiveness of detecting depression in tweets compared to using a baseline BERT model with standard or default hyperparameter settings.

To conclude, combining BERT and GWO is expected to address the research gap and outperform the baseline BERT model, making it a more robust tool for accurately classifying tweets to detect depression.

1.6 Assumption

This research, makes several key assumptions. First, it is assumed that the BERT model can effectively analyze text data to detect depression due to its advanced language skills. It is also believed that the datasets from Kaggle, originally collected from Twitter, are suitable for training and testing the BERT model for depression detection. Additionally, that the dataset is trusted, relabeled by a psychologist, and is accurate and reliable. It is expected that the GWO algorithm will successfully tune the BERT model's parameters, improve its performance in detecting depression from tweets. Lastly, it is assumed that GWO will significantly enhance the model's accuracy, make it applicable to various text data.

1.7 Scope and Delimitation

A. **Principal Variables**

1. **Independent Variables**
   The independent variables include the text data collected from open-source datasets on Kaggle, originally scraped from Twitter by the dataset creator, and the BERT model architecture used for text analysis.
2. **Dependent Variables**
   The dependent variable is the f1-Score of predicting depression from tweets.

B. **Locale**
The research does not have a specific locale as it focuses on the computational analysis of text data for depression detection, which is not limited to a geographical location.

C. **Timeframe**
The study is conducted over a span of 11 months, from September 2023 to July 2024.

D. **Justification**

1. The chosen timeframe allows for an in-depth exploration of the proposed approach's effectiveness in enhancing depression detection accuracy.
2. The study duration accounts for data collection, data relabeling, model development, optimization processes, comparison with existing methods, and result analysis.

E. **Limitations**
Language Constraint: The dataset is in English, which may restrict the findings to English text data.

1.8 Significance of the Study

The objective of this research is to enhance the detection of depression in tweets data using BERT model and its optimization using the GWO algorithm. This study seeks to enhance the accuracy and reliability of identifying depressive symptoms in text. The GWO algorithm is employed to fine-tune the BERT model's parameters, and enhancing its performance. This research aims to demonstrate that the optimized model significantly outperforms a non-optimized model in terms of evaluation metrics. Ultimately, this study contributes to mental health and natural language processing by presenting effective methods for improving depression detection using machine learning techniques.