

# Fine-tuning Bert Transformers for Detecting Depression from Arabic Social Media

Somaia M. Elimam

Sudan University of Science and Technology/College of Graduate Studies  
Kosti/Sudan

## ABSTRACT

Depression is a real challenge for those who are interested in public health, especially among adolescents and young people. As a result of the tremendous development in the field of technology and the spread of the culture of social networking through the Internet, it became necessary to take advantage of these means in the detection of depression among users of these sites. In this research, we explore the possibility of using social media data to detect and predict depression. In this paper, two Bert transformers were fine-tuned and trained to predict depression in Arabic social media. The proposed models presented a promising performance in comparison with the previous study on the same dataset.

## Keywords

Bert transformers, depression detection, Arabic social media.

## 1. INTRODUCTION

Mental illness is largely undiagnosed and untreated, especially in Arab countries, because people feel ashamed, unaware of this disorder, or don't want to be classified as mentally ill. According to the World Health Organization, depression is one of the major causes of disability worldwide, posing a serious threat to virtually every aspect of human life, including public and private health, academic, communication, and social issues. There are usually considerable changes in feelings, routines, or thoughts associated with this mental disorder, which can lead to suicide. Some developments have been made in the field of depression detection because early diagnosis of this illness would be critical to effective treatment. Observing aspects of a user's social media behavior provides strong evidence that depression can be detected, so we can use social media as a tool to measure depression, where people frequently use social media to vent about their difficulties, so, they tend to talk about their problems through these media more than talking to a psychiatrist which may not only give psychologists and/or psychiatrists more information before making decision but also creates opportunities for early diagnosis [4] [19] [3] [27] [13].

Language is a way for you to express your thoughts and feelings, and it can provide information about the psychological condition you're in right now. We frequently see that people exhibiting depressive symptoms have a propensity to use absolutist language and to constantly bring up the same phrases. Understanding the language of depression allows you to try to help others by recog-

nizing their words and phrases [7]. With the recent advancement in natural language processing and deep learning, many studies have implemented these methods to discuss mental illness and detect depression from social media, especially in English, but few of these studies discuss this in Arabic. In this study, Bert transformers will be investigated to detect and predict depression from Arabic social media.

The rest of this is organized as follows: Section 2 discusses the works related to this work, and Section 3 elaborates on Bert transformers. The Evaluation Dataset, proposed models and the results obtained are discussed in sections 4, 5, and 6 respectively. Finally, section 7 concludes this paper.

## 2. RELATED WORK

Most of the work done in the field of detecting depression from text in social media belongs to the English language, and little research done on this task in the Arabic language. In the following subsections, we investigate the literature for the works that studied depression detection for all languages, and then those that dealt with it in the Arabic language only.

The studies in [23] [22] [25] [17] [18] [6] proposed depression detection models based on Electroencephalogram (EEG), which integrate and adapt convolutional neural network (CNN) and long-short term memory (LSTM) architectures to detect depression using (EEG) signals.

In [24] the authors Built machine learning model to detect depression signs from text extracted from Russian social media VKontakte. They proposed psycholinguistic and stylistic markers of depression in users' posts.

The work [8] explores the ability to detect major depressive disorder from social media by crowdsourcing a set of depression pre-diagnosed users, the researchers capture their social media behavior over a year and monitor their social engagement, emotions, and the language they use. The study built a statistical classifier to estimate the risk of depression.

Kumar et al [15] Built a machine learning framework to predict depression from twitter using SVM, Naïve Bayes, Random Forest, Logistic regression classifiers.

The study [21] investigated a deep learning approach to detect depression from social media by proposing a new weighted activation function. (Neha et al., 2020) analyzed the emotion of users

using real-time videos, the study used Convolutional Neural Networks (CNN), and it designed a chatbot to boost the user's mood if they are depressed. Whereas [14] analyzed depression on Facebook posts by applying traditional machine learning approaches using a set of various psycholinguistic features.

The paper [26] presented a Long Short-Term Memory (LSTM)-based Recurrent Neural Network (RNN) approach to predict symptoms of depression from text, the proposed model was applied on a dataset from an online information channel for young people in Norway.

The study in [10] did a comparative study by investigating CNN-based and RNN-based models across different parameter settings to select the best one to detect and predict signs of depression from Twitter platform.

The authors of [16] developed a depression detection and suicidal ideation system by exploiting machine learning techniques; the proposed system was tested on a dataset extracted from Twitter, and [7] built an Arabic lexicon, ArDep, which contains words and phrases mostly used by depressed people. They collect tweets of Arabic users to build this lexicon, Whereas the study [11] implemented CairoDep, a Bidirectional Encoded from Representations Transformers (BERT) model to detect depressed users from text in Arabic social media. They trained a pre-trained ARABERT and MARBERT Bert models to build their model.

Alghamdi et al [2] implemented rule-based and machine learning approaches to predict depressed users from Arabic social media. They built an Arabic lexicon called ArabDep and used it to train a rule-based algorithm and annotated data with the help of a psychologist to train machine learning classifiers. And [1] proposed a deep learning model to predict depressed Arabic women during the COVID-19 pandemic, they trained a Recurrent Neural Network (RNN) architecture with 10000 tweets collected from 200 users.

### 3. BERT TRANSFORMERS

BERTs represent bidirectional encoder representations derived from transformers. With BERT, deep bidirectional representations are pre-trained from unlabeled text by conditioning both left and right context together at all layers. Bert is based on the fine-tuning approach, which introduces only minimal task-specific parameters and trains on downstream tasks by simply fine-tuning the pre-trained parameters. Bert improves the fine-tuning-based approaches by using a "masked language model" (MLM) pre-training objective [9]. Using the masked language model, some tokens are randomly masked from the input. The objective is to predict the original vocabulary id of the masked word based on only the context in which it appears.

Also, Bert presents a "next sentence prediction" task, which pre-trains text-pair representations jointly. By receiving pairs of sentences as input, the model learns to predict whether the second sentence in the pair corresponds to the subsequent sentence in the original document. During training, 50% of the inputs are pairs in which the second sentence is the next one in the original text, and in the remaining 50%, the second sentence is a randomly selected sentence from the corpus, with the underlying presumption that the random sentence will not be connected to the first [9].

Bert has two stages: pre-training and fine-tuning. The model is trained on unlabeled data across various pre-training tasks during pre-training.

The BERT model is first initialized with the pre-trained parameters, and then all of the parameters are fine-tuned using labeled data from downstream tasks. Even though they start with the same pre-trained parameters, each downstream task has its own fine-tuned model.

## 4. EVALUATION DATASET

An Evaluation Dataset for Depression Detection in Arabic Social Media is a benchmark dataset designed specifically to detect and anticipate depression among Arabic-speaking social media users, the data was collected by searching Twitter about posts that contain depression keywords/phrases. Then the collected data were normalized, preprocessed, and manually labeled into three labels; (depressed, none-depressed, and natural). The final dataset was saved in a csv file to be used for training process [12]. Figure 1 represent a screenshot of the constructed dataset.

## 5. PROPOSED MODELS

To detect depression on Arabic social media, the proposed work fine-tuned two Bert-based models, these two models take into account the linguistic accuracy and computational efficiency. The first model is a multilingual DistilBERT Transformer-based language model, which was trained on 104 languages, including Arabic [20]. DistilBERT is lightweight and faster than full BERT. It was built to reduce computational costs while retaining 97% of BERT's performance, approximately. The second model is araBert [5]), a Bert pre-trained model for Arabic. It was specifically pre-trained on a massive corpus of Arabic text, making it highly effective for understanding the language's unique complexities, especially in managing dialectal variances, morphology, and diacritical marks.

To adapt these two models in the presented study and to achieve high accuracy, the author generated five variants for each of the two models by adjusting the model's hyperparameters (Learning rate, number of Epochs, optimizer, batch size, maximum sentence length, loss function, activation function, and dataset splitting ratio) to find their optimal values. In each of the five variants, one of these parameters was tuned to notice its effect on the result until the optimal value was reached. Table 1 and Figure 2 show the optimal value used for each parameter for the two models and the work done on this study respectively.

Using AdamW as optimizer, binary cross-entropy as a loss function, and sigmoid as activation function, the models were implemented using Python 3 with sklearn, pyTorch, numpy, and random libraries. The implementation took 7 hours and 34 minutes for the DistilBERT model and 44 minutes for the araBERT model.

## 6. MODEL EVALUATION

The study evaluated the constructed models using the Twitter depression detection dataset [12], which contains 5773 tweets labeled as depressed (2828 tweets), natural (1209 tweets), and not-depressed (1736 tweets). The obtained results were compared with the results of our previous work [12], which was built on the same dataset using a deep learning approach. F1-score and accuracy were used as performance measures to compare the constructed models. Table 2 shows the best results of the two models in comparison with the results of the previous work.

The above result shows that the BERT-based models outperform the classical machine learning and deep learning models that are trained on the same dataset, and that is because the two Bert-based model were designed to suit Arabic language features. The

Text	createdAt	screenName	Followers	Friends	Statuses	Sentiment
اشعر بالوحده كمدبنة مات جميع أهلها	28 Nov 2018 01:30:24 GMT	ahmdabozaid1	189	287	351	-1
الجميع يحيط بي لكن اشعر بالوحده بينهم بدونك	27 Nov 2018 21:42:48 GMT	hanyyazeinnn1	492	349	1682	-1
لا أدري لكني اشعر بالغربة رغم أحتواء من حولي	28 Nov 2018 09:33:15 GMT	Ghaz_alah	26	48	624	-1
اريد ان أموت بكل الطرق الممكنة	1 Feb 2019 18:06:47 GMT	habibaali1397	107	489	454	-1
انا مكسور وحزين جدا	7 Feb 2019 14:08:40 GMT	Ahmedy_114	1185	442	10565	-1
انا فعلا محددش يعرفني ولا حاس بيا	6 Dec 2018 13:10:23 GMT	mohmed_zaghloul	1056	2464	11789	-1

Fig. 1. The dataset used in training and model evaluation process.

Table 1. The parameters values used to build models.

Model	Learning rate	# Epochs	batch size	maximum sentence length	Splitting ratio
DistilBERT model	2e-5	4	32	512	90:10
araBert model	5e-5	5	16	256	90:10

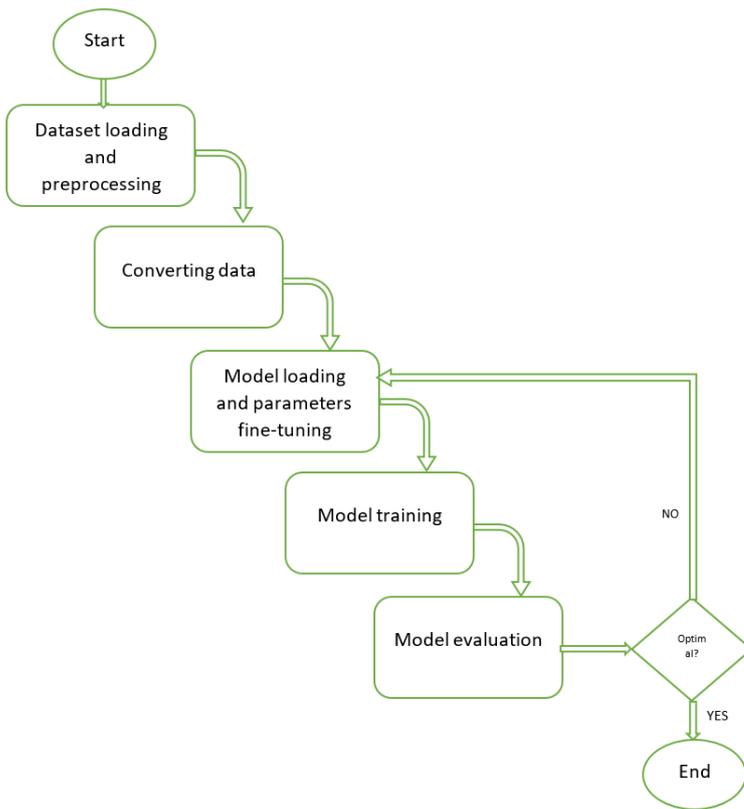


Fig. 2. The proposed model implementation process.

Table 2. The obtained result of the proposed models

Model	F1-score	Accuracy
DistilBERT model	0.64	71.34%
araBert model	0.71	74.00%
S. Elimam	0.62	70.63%

araBERT model performed more accurately than the DistilBERT model as it more specific for the morphological nature of Arabic language.

## 7. CONCLUSION AND FUTURE WORK

In this research, the author showed that BERT sets a new state-of-the-art approach in Arabic sentiment analysis in general and in depression detection specifically. Two BERT-based models were created and tested against the depression detection dataset. The proposed models present promising results outperforming the result of the previous study. These results confirm that Bert-based models can effectively capture complex linguistic cues, oriented sentiment patterns, and emotional expressions related to mental health. and can be used as a base for future work in the field of depression detection from Arabic social media. Future work aims to enhance this research by expanding the dataset and increasing its size, and elaborating more approaches to address mental health.

## 8. REFERENCES

- [1] Eatedal Alabdulkreem. Prediction of depressed arab women using their tweets. *Journal of Decision Systems*, 30:102–117, 2021.
- [2] Norah Saleh Alghamdi, Hanan A. Hosni Mahmoud, Ajith Abraham, Samar Awadh Alanazi, and Laura García-Hernández. Predicting depression symptoms in an arabic psychological forum. *IEEE Access*, 8:57317–57334, 2020.
- [3] Hayda Almeida, Antoine Briand, and Marie Jean Meurs. Detecting early risk of depression from social media user-generated content. In *CEUR Workshop Proceedings*, volume 1866, 2017.
- [4] Salma Almouzini, Maher Khemakhem, and Asem Alageel. Detecting arabic depressed users from twitter data. In *Procedia Computer Science*, volume 163, pages 257–265. Elsevier B.V., 2019.
- [5] Wissam Antoun, Fady Baly, and Hazem Hajj. Arabert: Transformer-based model for arabic language understanding. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools*, pages 9–15, 2020.
- [6] Betul Ay, Ozal Yildirim, Muhammed Talo, Ulas Baran Baloglu, Galip Aydin, Subha D. Puthankattil, and U. Rajendra Acharya. Automated depression detection using deep representation and sequence learning with eeg signals. *Journal of Medical Systems*, 43, 2019.

- [7] Bara Bataineh, Rehab Duwairi, and Malak Abdullah. Ardep: An arabic lexicon for detecting depression. *ACM International Conference Proceeding Series*, pages 146–151, 2019.
- [8] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*, pages 128–137, 2013.
- [9] Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, volume 1, pages 4171–4186, 2019.
- [10] Mahmoud Hussein Orabi Diana Inkpen Ahmed Hussein Orabi, Prasadith Buddhitha. Deep learning for depression detection of twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology*, page 88–97, 2018.
- [11] Mohammad El-Ramly, Hager Abu-Elyazid, Youseef Mo'men, Gameel Alshaer, Nardine Adib, Kareem Alaa Eldeen, and Mariam El-Shazly. Cairodep: Detecting depression in arabic posts using bert transformers. *Proceedings - 2021 IEEE 10th International Conference on Intelligent Computing and Information Systems, ICICIS 2021*, pages 207–212, 2021.
- [12] Somaia Elimam and Mohamed Bougeussa. An evaluation dataset for depression detection in arabic social media. *International Journal of Knowledge Engineering and Data Mining*, 7:113, 2021.
- [13] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. Detecting depression and mental illness on social media: an integrative review, 2017.
- [14] Md Rafiqul Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M Kamal, Hua Wang, and Anwaar Ulhaq. Depression detection from social network data using machine learning techniques. *Health Information Science and Systems*, 6:1–12, 2018.
- [15] Deepthi Kumar, K N Arpitha, G Seema, Dona Sarkar, and Suma Swamy. Predictive analysis of depression via social media mining. pages 565–572, 2020.
- [16] S J Pachouly, Gargee Raut, Kshama Bute, Rushikesh Tambe, and Shruti Bhavsar. Depression detection on social media network (twitter) using sentiment analysis. *International Research Journal of Engineering and Technology*, 08:1834–1839, 2021.
- [17] Abdul Qayyum, Imran Razzak, and Wajid Mumtaz. Hybrid deep shallow network for assessment of depression using electroencephalogram signals. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12534 LNCS:245–257, 2020.
- [18] Abdolkarim Saeedi, Maryam Saeedi, Arash Maghsoudi, and Ahmad Shalhaf. Major depressive disorder diagnosis based on effective connectivity in eeg signals: a convolutional neural network and long short-term memory approach. *Cognitive Neurodynamics*, 15:239–252, 2021.
- [19] Atefeh Safayari and Hamidreza Bolhasani. Depression diagnosis by deep learning using eeg signals: A systematic review. *Medicine in Novel Technology and Devices*, 12:100102, 2021.
- [20] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *ArXiv*, abs/1910.0:2–6, 2019.
- [21] Hrithik Sanyal, Sagar Shukla, and Rajneesh Agrawal. Study of depression detection using deep learning. In *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, volume 2021-Janua, 2021.
- [22] Ayan Seal, Rishabh Bajpai, Jagriti Agnihotri, Anis Yazidi, Enrique Herrera-Viedma, and Ondrej Krejcar. Deprnet: A deep convolution neural network framework for detecting depression using eeg. *IEEE Transactions on Instrumentation and Measurement*, 70, 2021.
- [23] Geetanjali Sharma, Abhishek Parashar, and Amit M. Joshi. Dephnn: A novel hybrid neural network for electroencephalogram (eeg)-based screening of depression. *Biomedical Signal Processing and Control*, 66:102393, 2021.
- [24] Maxim Stankevich, Andrey Latyshev, Evgenia Kuminskaya, Ivan Smirnov, and Oleg Grigoriev. Depression detection from social media texts. *CEUR Workshop Proceedings*, 2523:279–289, 2019.
- [25] Pristy Paul Thoduparambil, Anna Dominic, and Surekha Mariam Varghese. Eeg-based deep learning model for the automatic detection of clinical depression, 2020.
- [26] Zia Uddin. Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications*, 34:721–744, 2022.
- [27] David William and Derwin Suhartono. Text-based depression detection on social media posts: A systematic literature review. In *Procedia Computer Science*, volume 179, pages 582–589. Elsevier B.V., 2021.