International Journal of Computer Science and Information Security (IJCSIS),
Vol. 22, No. 5, October 2024

Depression Detection Using Sentiment Analysis Techniques in Arabic Text

Afaf Hussein Abdelrahman [1], Doaa Elzanfaly [2], Mai El Defrawi [3], Samah Hamed [4]

Department of Information System, Helwan University , Egypt

[1]afafmseleem@gmail.com, [2]doaa.saad@fci.helwan.edu.eg, [3]mai.eldefrawi@gmail.com

***Abstract*—Depression is a mood disorder marked by persistent melancholy and interest loss. It is one of the most dangerous mental health issues that humans may encounter. Machine learning is used to detect depression. Analyzing depression is not a simple task, most people are not usually willing to go to a psychiatrist. or They may not even know that they suffer from depression. Sentiment analysis usually relies on machine learning to analyze human texts. The main goal is to determine if the general sentiment of a piece of text is positive, negative, or neutral. However, there are other factors of consideration, including the frequency and redundancy of the text. There are only a few studies in Arabic, where most of the work is usually done in English. In this paper, Arabic tweets are analyzed to determine depression. This research proposes a new model to identify the level of depression based on several elements while focusing on the timing and frequency of the tweet. A dataset using tweepy API is created to determine the level of depression. This research shows good results as precision is 0.97, recall is 100%, and f1-score is 0.98. Our results outperforms the current techniques by about 30%.**

Index Terms—Depression; sentiment analysis, twitter, supervised learning; machine learning

I. INTRODUCTION

Depression has become a prominent global public health concern, especially in low- and middle-income countries [1]. Major depressive disorder (MDD) is the most prevalent in the world. According to the World Health Organization, about 350 million people in the world are affected by this condition [2, 3, 4].

A person may exhibit any combination of little interest or pleasure in doing things, poor appetite or overeating, feeling bad about yourself, or that you are a failure or have let yourself or your family down. Trouble falling or staying asleep, or sleeping too much, or trouble concentrating on things to be classified as depressed [5].

Social media has become very important. Everyone, whether young or old, uses it to express their thoughts and feelings, such as on Twitter and Facebook.in

In Arab countries, it is found that the idea of visiting a psychiatrist is still feared by people. Social media is resorted by them to express their feelings if a problem is perceived. Due to social media's widespread use, there may be ways to lessen the prevalence of undiagnosed mental illnesses like stress, anxiety, and depression [6].

Users' tweets are analyzed using sentiment analysis and machine learning to identify signs of depression. Sentiment analysis involves the use of natural language processing (NLP) techniques with machine learning to automatically determine the sentiment expressed in a given text, such as positive, negative, or neutral [7].

Machine learning employs two primary techniques: supervised learning, which constructs models from labeled data with predefined input-output pairs for training, and unsupervised learning, which builds models using only input data, uncovering patterns within unlabeled datasets without specified output tags [8], as shown in Fig 1.

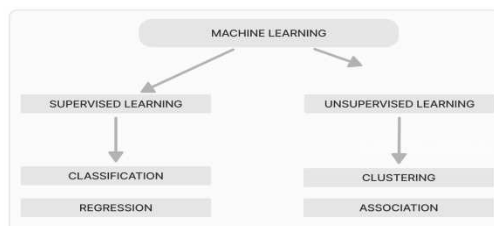

Fig. 1. Machine learning techniques [8]

Supervised learning involves training a model with input/output pairs or labeled data to predict future outcomes accurately based on past data [9].

Because supervised learning produces accurate predictions [10], regression analysis and classification are the two main uses for it.

Classification is a supervised learning approach used to analyze a given data set and to build a model that separates data into a desired and distinct number of classes [11]. Always in classification When the number of classes is less, the efficiency is higher, so most research focuses on a maximum of three classes.

According to [12], Support Vector Machine (SVM), Convolution Neural Network (CNN), Logistic Regression (LR), Random Forests (RF), Decision Trees (DT), are common techniques used for classification.

For unsupervised learning, there are no right or wrong answers, unlike supervised learning, and no teacher. Algorithms are left to their own devices to discover and present an interesting structure in the data [13].

The unsupervised learning approach of Association Rule Learning (ARL) is one of its wide techniques [14].

According to [15], unsupervised learning algorithms such as k-means, Gaussian mixture model (GMM), and fuzzy c-means (FCM) are examples of common clustering techniques.

This research depends on supervised learning with classification.

The data for this research is often collected from psychological websites and Twitter.

The focus of this research is to determine if depression can be detected using user's tweets. The main attributes to consider are the tweet, timing, and frequency.

Detecting depression using sentiment analysis has been studied in several papers [16, 17, 18, 19, 20], but most of them are in English, and there is little research in the Arabic language. These studies are usually limited to two or three classes: positive, negative, and neutral.

The rest of this paper is organized as follows: Section 2 discusses the related work. Section 3 presents the methodology. Section 4 presents the discussion and results. Finally, the conclusion and future work are identified in Section 5.

II. RELATED WORKS

A. Detecting depression in English Text

The goal of a recent study in [21] is to analyze mood disorders such as depression, poor mood, and other symptoms utilizing tweets and emoticons. The experimental results clearly show that the multimodal, multi-view, and multitasking suggested framework achieves an accuracy of 88.29% for the core task of depression identification using the SVM linear kernel function. It was limited to two classes and didn't consider timing and frequency in English.

The authors of study in [22] focused on training text classifiers to detect depression, comparing hybrid and ensemble methods, where ensemble models outperformed hybrids. They used a Reddit dataset with 1841 users (1200 positives and 641 negatives) in two classes, conducted in English. In parallel, a study [23] diagnosed depression by analyzing demographic data and Twitter opinions using five machine learning techniques, with Random Forest showing superior accuracy among three classes, also in English.

The two objectives of the study discussed in this paper [24] are: i) creating a more precise classification function to enhance current lexicon-based techniques for identifying depressing content on social media platforms, and ii) developing a depression diagnostic system embedded with our improved lexicon method, allowing individuals to visualize their depression state instantly via an online interface. In categorizing depression states, our experimental lexicon-based technique attained 77% precision and an F1 score of 74%. It was limited to two classes, and it was in English.

An anxious depression prediction model was presented in [25]. For anxious depression prediction in real-time tweets, it is proposed. The main symptoms of this mixed anxiety-depressive illness are restlessness, insomnia, and erratic thought patterns. Based on the linguistic cues and user posting patterns, the feature set is defined using a 5-tuple vector <word, timing, frequency, sentiment, contrast> where tweets of the first 100 followers of the MS India student forum are analyzed using various linguistic, semantic, and activity features to detect anxious depression disorder using supervised learning (Multinomial Naïve Bayes). This research is similar to our work, but they have a problem as they focus on English and are working on three classes while timing and frequency were considered in this research.

B. Detecting depression in Arabic Text

Both studies in [26] and in [27] focus on predicting depression in Arabic text. Study [26] creates a predictive model using supervised learning focusing on Gulf users and two classes, employing Random Forest, Naive Bayes, AdaBoostM1, and Liblinear. In contrast, a study [27] utilizes natural language processing and machine learning with Arabic text from online forums, comparing lexicon-based and machine-learning-based methodologies on 20,000 stories from Nafsany's forum for two classes and employing a machine-learning-based approach (SVM).

The research in [28] aimed to create a model for analyzing Arabic tweets and diagnosing sadness among Arabic Twitter users. Also, add a new category ("neutral") to increase the diversity of user tweets, resulting in three classifications in comparison to the two previous studies ("depressed," "non-depressed," and "neutral"). Natural language processing methods and machine learning classifiers were used in the model's development. The results showed that the RF classifier performed better than the others, with an accuracy rate of 82.39%. The data was gathered from Arabic Twitter users who responded to the CES-D survey.

A Bi-LSTM model with attention was developed to detect depression in tweets, achieving 83% accuracy on a dataset of 6,000 labeled tweets [29]. By focusing on key features and word weights, it outperformed other models, aiming to understand mental health expressions on Twitter. The attention-based Bi-

LSTM successfully classified Arabic tweets as depressed or not, with WordCloud visualizing common words related to depression.

The study presented in this paper [30] developed an RNN model to analyze Arab women's tweets during the COVID-19 pandemic, predicting depression symptoms effectively. Training on 10,000 tweets from 200 users, the model addressed the stigma around discussing depression in Arab societies, particularly among women. By leveraging social media, it assessed emotions in tweets to diagnose depression. The study was conducted on Arabic tweets from Saudi Arabia to optimize word embedding techniques, evaluate the RNN model's performance, and compare it with existing text. classification methods, with output classes representing positive and negative sentiments.

The authors of the study [31] focus on Arabic tweet sentiment analysis during the COVID-19 pandemic in Saudi Arabia, employing CNN and BiLSTM models to classify sentiments as positive, negative, or neutral. The research compares sentiment analysis before and during the pandemic, noting a rise in negative sentiments during COVID-19. The AraVec-SG-300-3 word2vec model was found to be the most effective word embedding technique. CNN outperformed BiLSTM with 93% accuracy, highlighting the study's success in Arabic sentiment analysis related to COVID-19.

The research in [32] analyzes sentiment in Arabic tweets regarding distance learning in Saudi Arabia during the COVID-19 pandemic. They collected and processed tweets from Saudi Arabia and labeled positive or negative, developing models using N-gram and TF-IDF features with various classifiers. The most accurate model achieved 0.899 accuracy with LR, identifying positive views on distance learning for general education but negative opinions at the university level. The results can guide the Ministry of Education in enhancing distance learning services.

The paper [33] focuses on using Arabic Twitter data to detect mental health issues, proposing a framework for classifying tweets into "Suicide" or "Normal" categories. It emphasizes the importance of early detection through social media analysis and reports high performance using Bidirectional Encoder Representations from Transformers (BERT) and Universal Sentence Encoder (USE) models.

From the related work, it has been concluded that the research on depression analysis is still in need of improvement in both English and Arabic research. Meanwhile, Arabic suffers the most with a lack of interest in detecting depression from social media platforms.

III. METHODOLOGY

The data is cleaned and divided into none, mild, moderate, and severe according to the psychologist. Then, the model was applied to the timing and frequency of tweets.

A. Data Collection

In this research, data was collected from Twitter using tweepy API using some words such as depressed, sad, depressed, anxious, and sad. Data in Egyptian colloquial language.

The data was collected over 6 months (from July to December 2021) by collecting data every week, and duplicate data was deleted by date.

After getting all the tweets, create a Pandas data frame and store the data frame in a dataset.csv file.

B. Data Cleaning and Preprocessing

In the data preparation phase, several crucial steps were undertaken. Initially, rows underwent editing to enhance the quality of descriptions. Subsequently, the data was organized by author for better clarity and relevance. Unnecessary rows, such as instances like (صباح الخير من قسم عزل كورونا), were identified and removed to streamline the dataset. Moreover, each tweet was meticulously processed to eliminate special characters, digits, tags, URLs, and non-Arabic words, ensuring that the text was refined and ready for further analysis. For Example,

'@abdalh180 @tourist_exp فعلاً خساره على توتر وقلق .. وكل شوي قرار يفاجئك.. الله يفرجها علينا وننطلق واحنا مرتاحين بأقرب وقت جد تعبنا من', '#كورونا .. والحمدالله دائماً وابداً'

وحده تقرر حياتنا فيه موت سجلونا_غياب_مستحيل_نداوم حياتنا مو لعبه بكلمه ترا فيها خوف وقلق على الطلاب بسبب كورونا مثال شخص في المدرسه معه كورونا ونقلها لنا ورحت للبيت عند جدي وجدتي كبار في السن و معهم ربو وش يبي يصير لهم https://t.co/JPBCTirQBQ'.

The date and time of publication were stored to distinguish between tweets composed during nighttime and daytime. Emoticons were omitted from the analysis due to their potential for misrepresentation, as an individual might appear jovial while actually feeling sorrowful, leading to inaccuracies in sentiment analysis.

C. Feature Engineering

1) Building the Lexicons: To classify terms and statements into different levels of depression, a psychologist was consulted to categorize all text into the appropriate classes.

Two lexicons were created, one for metonymies and one for words. The metonymy file was divided into the metonymy and its status, such as اضطراب القلق الشديد, and the status is stress according to the expert. The word file was divided into the word and two forms of stemming, the infinitive and the status, such as المروع - مروع - روع - روع and the status is stress.

These lexicons were created using ten symptoms of depression through consultation with a psychologist. They are stress, sadness, suicide, sag, contrition, crying, negative image, sleep, and loss of appetite.

2) *Tweet Status Property:* Using the fuzzy wuzzy method, the percentage of similarity between the metonym in the lexicon and the description from the row in the data set was calculated. If the percentage is greater than or equal to a certain threshold, the status of this metonym is taken and added to the tweet status column. If the percentage is within a certain threshold, the status of this metonym is taken, and it is added to the tweet status property after checking if the description contains all words of the metonym.

The threshold used is 70-85 to find the most suitable similarity between the metonym in the lexicon and the description from the row in the data set.

Then check the word file by taking each word in the description and checking if it is present in the word, two forms of stemming, or the infinitive, and if it exists, its status is added to the tweet status column.

3) *User Status Property:* According to psychologists, if the user's texts contain several symptoms, they are given a certain level (none-mild-moderate-severe) as shown in Table I, which is translated into four equations (dividing the number of statuses for each user by the number of tweets he wrote and obtaining a ratio) as shown in Table II.

TABLE I
USER STATUS CLASSIFICATION BASED ON SYMPTOMS

Number of Symptoms Range	User Status
0 < number of symptoms < 4	none
4 <= number of symptoms < 6	mild
6 <= number of symptoms < 8	moderate
8 <= number of symptoms	severe

TABLE II
USER STATUS CLASSIFICATION BASED ON *RATIO*

Ratio Range	User Status
0 < ratio < 40	none
40 <= ratio < 60	mild
60 <= ratio < 80	moderate
80 <= ratio	severe

This is an example of a user who has four tweets as shown in Fig 2. All of them were analyzed to discover their status. Two of the tweets have no feelings, so their tweet status is set as none, while the other two tweets with feelings were analyzed as follows:

– (قل للرياح تأتي كما شاءت فما عادت سفننا تشتهي شيئا) This tweet is a metaphor for sag according to the Metaphor file, so its tweet status is sag.
– (شعور اني اول مره اشوف ناس كثير بعد كورونا توتر...) This tweet has the word توتر, which expresses stress according to the word file, so its tweet status is stress.

After all tweets for the user were analyzed, the ratio between all tweet statuses and the number of all tweets for the user was calculated, and the user status (none, mild, moderate, or severe) was determined according to the rules mentioned before by psychologists.

	author	description	pubdate	tweet_status	user_status
137	1211Soul	####################...	2021-12-06 10:35:10		1
515	1211Soul	شعور اني اول مره اشوف ناس كثير بعد كورونا توتر...	2021-07-21 14:55:37	stress	1
139	1211Soul	قل للرياح تأتي كما شاءت فما عادت سفننا تشتهي شيئا	2021-11-29 07:02:32	sag	1
141	1211Soul	اغغ من هالبوخه	2021-11-28 20:01:19		1

Fig. 2. Tweet and user status

4) *Target property:* This represents the real condition of the user that the model will try to reach.

first according to the ten symptoms of depression through consultation with a psychologist. They are stress, sadness, suicide, sag, contrition, crying, negative image, sleep, and loss of appetite.

Every tweet is labeled with the number of symptoms appearing in it or marked as negative if it doesn't have any of these symptoms; then, the number of symptoms for every user is summed, and The user status is determined according to these rules with the assistance of a psychiatrist.
(0-3) symptoms give user status none.
(4-5) symptoms give user status mild.
(6-7) Symptoms give user status moderate.
(8-...) symptoms give user status severe.

Then, the column called user_status_real is created, and all tweets for this user are filled out with their status. As shown in Fig 3.

```
def create_target_column(data):
    data.insert(1,'user_status_real',[1,1,1,1,
                                      1,1,1,1,1,1,1,1,
                                      1,1,1,1,1,1,1,1,1,1,1,
                                      0,0,0,0,0,0,0,0,
                                      2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,
                                      0,0,0,
                                      1,1,1,1,
                                      1,1,1,1,
                                      2,2,2,2,2,2,2,2,2,2,2,
                                      0,0,0,0,
                                      3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,
                                      2,2,2,2,2,
                                      1,1,1,1,1,1,1,1,
                                      3,3,3,3,3,3,3,3,3,3,
                                      0,0,0,0,0,0,0,0,
                                      3,3,3,3,3,3,3,3,3,3,3,3,3,3,
                                      0.0.0.
```

Fig. 3. Target property

This is the data frame after feature engineering and creating all columns needed (user_status_real, tweet_status, user_status, tweet_time, user_time_frequency, and user_time) . As shown in Fig 4.

	author	user_status_real	description	pubdate	tweet_status	user_status	tweet_time	user_time_frequency	user_time
137	1211Soul	1	[illegible]	2021-12-06 10:35:10		1	0	0	0
515	1211Soul	1	[illegible]	2021-07-21 14:55:37	stress	1	0	0	0
139	1211Soul	1	[illegible]	2021-11-29 07:02:32	sag	1	0	0	0
141	1211Soul	1	[illegible]	2021-11-28 20:01:19		1	0	0	0
1441	9OdZlLSOa20sj6f	1	[illegible]	2021-09-12 07:52:52	death-stress	1	0	0	1
...	...	...	...	...	...	...	...	...	...
61	yaquob007	0	[illegible]	2022-01-01 13:51:24		0	0	0	0
60	yaquob007	0	[illegible]	2022-01-03 11:02:28		0	0	0	0
59	yaquob007	0	[illegible]	2022-01-03 11:25:21		0	0	0	0

Fig. 4. Feature engineering data frame

As shown in Fig 5.

Fig. 5. Psychologist steps

D. Data Preparation for Machine Learning Pipeline

1) *Tweet Time Property:* The time of the tweet is recorded, if it is between 12 am and 6 am, which is the period of the night according to [27]. Users who are active through the midnight hours, i.e., from 12 am to 6 am, evidently show psychological disturbance in sleep patterns with increased restlessness and overthinking than people active in the other time of the day, then Tweet time is given a certain value.

2) *User Time Property:* Gathering all the timing of all tweets for a certain user (tweet time, which has value 1), it is divided by the number of tweets he wrote and obtaining a ratio.

For the timing of the tweets, they are assigned according to a certain ratio into none, mild, moderate, or severe. The ratio used in this step is less than 25, between 25 and 50, between 50 and 75, and more than 75. The reason for choosing this ratio is based on the psychologist's advice as the most appropriate.

3) *User Time Frequency Property:* Based on the tweet timestamps categorized as either 1 (between 12 am and 6 am) or 0 (other times), the user's time-frequency value is determined. If the count of 1s exceeds that of 0s, the user is considered to tweet frequently from 12 am to 6 am, receiving a time-frequency value of 1. Conversely, if the count of 0s surpasses that of 1s, the user is regarded as tweeting infrequently during that time frame, receiving a time-frequency value of 0.

As shown in Fig 6.

	author	description	pubdate	tweet_status	user_status	tweet_time	user_time_frequency	user_time
137	1211Soul	[illegible]	2021-12-06 10:35:10		1	0	0	0
515	1211Soul	[illegible]	2021-07-21 14:55:37	stress	1	0	0	0
139	1211Soul	[illegible]	2021-11-29 07:02:32	sag	1	0	0	0
141	1211Soul	[illegible]	2021-11-28 20:01:19		1	0	0	0
1441	9OdZILSOa20sj6f	[illegible]	2021-09-12 07:52:52	death stress	1	0	0	1
...	...	...	...	...	...	...	...	...
61	yaquob007	[illegible]	2022-01-01 13:51:24		0	0	0	0
60	yaquob007	[illegible]	2022-01-03 11:02:28		0	0	0	0
59	yaquob007	[illegible]	2022-01-03 11:25:21		0	0	0	0
1166	yaquob007	[illegible]	2021-07-30 18:50:02	stress	0	0	0	0
62	yaquob007	[illegible]	2022-01-01 10:55:52		0	0	0	0

Fig. 6. Tweet time and user time and user time frequency property

4) *Splitting Target and Features Columns:* To prepare the dataset for classification modeling, the target column, which represents the variable to be predicted, was removed from the dataset. The remaining columns, referred to as features, were stored in a variable named 'classification_features_cols.' These features encapsulate the information used for predicting the target variable.

5) *Handling Categorical Data:* One-hot encoding is performed for categorical columns, and boolean columns are converted into integer format. These preprocessing steps are commonly used to transform the data into a suitable format for classification models that expect numerical inputs.

6) *Feature Selection:* Perform feature selection using RFE with logistic regression. It aims to use logistic regression to identify the most relevant features in the classification_features_cols dataset for predicting the classification_target_col.

7) *Split train and Test Data:* Randomly divide the classification_features_cols dataset and classification_target_col target variable into training and testing sets. This allows for evaluating the performance of a classification model on unseen

data by training the model on the training set and evaluating it on the testing set.

8) *Scale Data:* Perform feature scaling using MinMaxScaler on the numeric columns of the training and testing datasets. This scaling ensures that the numeric features are on a similar scale, which can be beneficial for certain machine-learning algorithms that are sensitive to the scale of the features.

IV. DISCUSSION AND RESULTS

A. *Choose Model*

Train and evaluate multiple classification models using different algorithms (LogisticRegression, GaussianNB, KNeighborsClassifier, DecisionTreeClassifier, SVC, RandomForestClassifier). Table III displays the accuracy scores for each model on both the training and testing datasets.

TABLE III

ACCURACY FOR BOTH THE TRAINING AND TESTING DATASETS FOR EACH MODEL

Model	Test Accuracy	Train Accuracy
LogisticRegression	98.04%	100.0%
GaussianNB	67.65%	100.0%
KNeighborsClassifier	90.20%	93.77%
DecisionTreeClassifier	98.04%	100.0%
SVC	96.08%	99.34%
RandomForestClassifier	99.02%	100.0%

B. *Random Forest Classifier Model*

Train a Random Forest Classifier model on the training dataset and use it to predict the target variable for the testing dataset. It allows for evaluating the model's performance on unseen data and obtaining the expected values for further analysis or evaluation. The confusion matrix shown in Table IV, recall, precision, and f1_score are calculated as shown below.
f1_score between y_test and y_pred is 0.98.

TABLE IV

CONFUSION MATRIX

	Predicted Positive	Predicted Negative
Actual positive	86 (TP)	2 (FP)
Actual negative	0 (FN)	14 (TN)

Recall = TP / (TP + FN) = 86/86+0=1
precision = TP / (TP + FP) = 86/86+2 = 0.977

As shown in Fig 7.

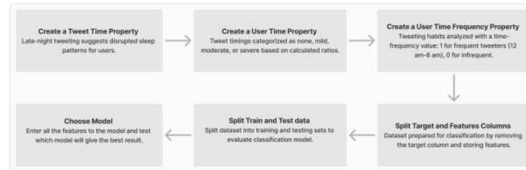

Fig. 7. Proposed model

1) *Results discussion:* We collected the data and analyzed it through the psychologist, and we were able to determine the degree of depression. The model used some new techniques that no one has used before, such as timing and frequency of tweets, and this gives a good indication and we got the best results compared to other research.

In Table V, the accuracy, f1-score, precision, and recall are calculated for this work and are compared against papers [25] and [27].

TABLE V

COMPARISON BETWEEN THE SEARCH RESULTS AND OTHER RESEARCH

Paper	Accuracy	F1-score	Precision	Recall
This Research	99.02%	0.98%	0.97%	100%
[25]	85.09%	79.68%		
[27]	73%	74%	74%	71%

V. CONCLUSION

Depression is one of the most affecting diseases that must be treated quickly because it can lead to suicide in some cases.Therefore, this research discovered depression through analyzing user tweets. The model used is divided into two steps: the first is to create a tweet and user status using the psychologist's rules, and the second is to prepare data for the pipeline by creating the time and time frequency property to determine the depression level. This research has shown promising results while scoring f1_score is 0.98. Recall is 1. Precision is 0.97. In future work, a bigger dataset is going to be used, including more varieties of dialects, including more psychiatric diseases, and a depression dataset automatically generated using the model.

REFERENCES

[1] G. Limenih, A. MacDougall, M. Wedlake, and E. Nouvet, "Depression and global mental health in the global south: A critical analysis of policy and discourse," Int. J. Soc. Determinants Health and Health Serv., vol. 54, no. 2, pp. 95-107, Apr. 2024.

[2] G. Jadhav, S. Babar, and P. Mahalle, "A survey: Performance-aware depression detection," in 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), Mar. 2023, pp. 1242-1249.

[3] L. Orsolini, R. Latini, M. Pompili, G. Serafini, U. Volpe, F. Vellante, M. Fornaro, A. Valchera, C. Tomasetti, S. Fraticelli, and M. Alessandrini, "Understanding the complex of suicide in depression: From research to clinics," Psychiatry Investig., vol. 17, no. 3, p. 207, Mar. 2020.

[4] Y. Asmare, A. Ali, and A. Belachew, "Magnitude and associated factors of depression among people with hypertension in Addis Ababa, Ethiopia: A hospital-based cross-sectional study," BMC Psychiatry, vol. 22, no. 1, p. 327, May 2022.

[5] M. A. Villarroel and E. P. Terlizzi, "Symptoms of depression among adults: United States, 2019," 2019.

[6] J. Kim, J. Lee, E. Park, and J. Han, "A deep learning model for detecting mental illness from user content on social media," Scientific Reports, vol. 10, no. 1, p. 11846, Jul. 2020.

[7] R. Krishna and C. M. Prashanth, "Machine learning based Twitter sentiment analysis and user influence," Int. J. Recent Innov. Trends Comput. Commun., vol. 11, pp. 215-221, 2023.

[8] R. Sharma, K. Sharma, and A. Khanna, "Study of supervised learning and unsupervised learning," Int. J. Res. Appl. Sci. Eng. Technol., vol. 8, no. 6, pp. 588-593, 2020.

[9] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: A survey and review," in Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018, Springer Singapore, 2020, pp. 99-111.

[10] O. A. Alimi, K. Ouahada, A. M. Abu-Mahfouz, S. Rimer, and K. O. Alimi, "A review of research works on supervised learning algorithms for SCADA intrusion detection and classification," Sustainability, vol. 13, no. 17, p. 9597, Aug. 2021.

[11] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," Neurocomputing, vol. 408, pp. 189-215, Sep. 2020.

[12] Z. A. Khan, M. Adil, N. Javaid, M. N. Saqib, M. Shafiq, and J. G. Choi, "Electricity theft detection using supervised learning techniques on smart meter data," Sustainability, vol. 12, p. 8023, 2020.

[13] B. Mahesh, "Machine learning algorithms-a review," Int. J. Sci. Res. [Internet], vol. 9, no. 1, pp. 381-386, Jan. 2020..

[14] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed, "An unsupervised machine learning algorithms: Comprehensive review," Int. J. Comput. Digit. Syst., Mar. 2023.

[15] Y. Lin and S. Chen, "A centroid auto-fused hierarchical fuzzy c-means clustering," IEEE Trans. Fuzzy Syst., vol. 29, no. 7, pp. 2006-2017, Apr. 2020.

[16] P. Arora and P. Arora, "Mining Twitter data for depression detection," in 2019 Int. Conf. Signal Process. Commun. (ICSC), Mar. 2019, pp. 186-189.

[17] M. Mounika, N. S. Gupta, and B. Valarmathi, "Detection of depression-related posts in tweets using classification methods–A comparative analysis," in Proc. Int. Conf. Comput. Netw. Big Data IoT (ICCBI-2019), Springer International Publishing, 2020, pp. 620-630.

[18] R. U. Mustafa, N. Ashraf, F. S. Ahmed, J. Ferzund, B. Shahzad, and A. Gelbukh, "A multiclass depression detection in social media based on sentiment analysis," in 17th Int. Conf. Inf. Technol.-New Gener. (ITNG 2020), Springer International Publishing, 2020, pp. 659-662.

[19] A. Esposito, G. Raimo, M. Maldonato, C. Vogel, M. Conson, and G. Cordasco, "Behavioral sentiment analysis of depressive states," in 2020 11th IEEE Int. Conf. Cogn. Infocommun. (CogInfoCom), Sep. 2020, pp. 000209-000214.

[20] E. Alabdulkreem, "Prediction of depressed Arab women using their tweets," J. Decis. Syst., vol. 30, no. 2-3, pp. 102-117, Jul. 2021.

[21] S. Gupta, A. Singh, and J. Ranjan, "Multimodal, multiview and multitasking depression detection framework endorsed with auxiliary sentiment polarity and emotion detection," Int. J. Syst. Assur. Eng. Manag., vol. 14, suppl. 1, pp. 337-352, Mar. 2023.

[22] L. Ansari, S. Ji, Q. Chen, and E. Cambria, "Ensemble hybrid learning methods for automated depression detection," IEEE Trans. Comput. Soc. Syst., vol. 10, no. 1, pp. 211-219, Mar. 2022.

[23] J. Angskun, S. Tipprasert, and T. Angskun, "Big data analytics on social networks for real-time depression detection," J. Big Data, vol. 9, p. 69, 2022.

[24] B. Z. Yeow and H. N. Chua, "A depression diagnostic system using lexicon-based text sentiment analysis," Int. J. Perceptive Cogn. Comput., vol. 8, no. 1, pp. 29-39, Jan. 2022.

[25] A. Kumar, A. Sharma, and A. Arora, "Anxious depression prediction in real-time social data," in Proc. Int. Conf. Adv. Eng. Sci. Manag. Technol. (ICAESMT19), 2023, pp. 1-7.

[26] S. Almouzini and A. Alageel, "Detecting Arabic depressed users from Twitter data," Procedia Comput. Sci., vol. 163, pp. 257-265, 2019.

[27] N. S. Alghamdi, H. A. Mahmoud, A. Abraham, S. A. Alanazi, and L. García-Hernández, "Predicting depression symptoms in an Arabic psychological forum," IEEE Access, vol. 8, pp. 57317-57334, Mar. 2020.

[28] D. A. Musleh et al., "Twitter Arabic sentiment analysis to detect depression using machine learning," Comput. Mater. Continua, vol. 71, no. 2, pp. 2799-2813, May 2022.

[29] A. M. Almars, "Attention-based Bi-LSTM model for Arabic depression classification," Comput. Mater. Continua, vol. 71, no. 2, pp. 2999-3010, May 2022.

[30] E. Alabdulkreem, "Prediction of depressed Arab women using their tweets," J. Decis. Syst., vol. 30, no. 2-3, pp. 102-117, Jul. 2021.

[31] A. Alqarni and A. Rahman, "Arabic tweets-based sentiment analysis to investigate the impact of COVID-19 in KSA: A deep learning approach," Big Data Cogn. Comput., vol. 7, no. 1, p. 16, 2023.

[32] M. Aljabri et al., "Sentiment analysis of Arabic tweets regarding distance learning in Saudi Arabia during the COVID-19 pandemic," Sensors, vol. 21, no. 16, p. 5431, Aug. 2021.

[33] N. A. Baghdadi et al., "An optimized deep learning approach for suicide detection through Arabic tweets," PeerJ Comput. Sci., vol. 8, p. e1070, Aug. 2022.

Call for Papers: International Journal of Computer Science and Information Security (IJCSIS)

Scope and Topics: The International Journal of Computer Science and Information Security (IJCSIS) invites researchers, practitioners, and academicians to submit original, unpublished contributions covering a wide range of topics in the field of computer science and information security. We welcome submissions that include but are not limited to:

- Computer and Network Security
- Cryptography and Data Security
- Information Assurance
- Artificial Intelligence and Machine Learning in Security
- Cybersecurity Policies and Standards
- Security in Cloud Computing
- Internet of Things (IoT) Security
- Blockchain and Distributed Ledger Technologies
- Secure Software Development
- Privacy-Enhancing Technologies

Submission Guidelines:

- Manuscripts must be original and not currently under consideration for publication elsewhere. All papers should be submitted in English.
- The manuscript should follow the IJCSIS formatting guidelines, available on our website: https://sites.google.com/site/ijcsis/ijcsis
- Submissions must include the title of the paper, abstract, keywords, and full contact information for all authors.
- Papers should be submitted via our submission system here: https://sites.google.com/site/ijcsis/submit-paper

Important Dates:

- **Paper Submission Deadline:** Monthly, 2024
- **Notification of Acceptance:** Within TWO weeks, 2024
- **Final Manuscript Due:** Monthly, 2024
- **Publication Date:** Monthly Issue

Review Process: All submitted papers will undergo a rigorous peer-review process by the IJCSIS editorial board and selected external reviewers. Authors will be notified of the review results by the notification of acceptance date.a

Contact Information: For further inquiries, please contact the editorial office at:

Email: ijcsiseditor@gmail.com

We look forward to receiving your submissions!

Website: https://sites.google.com/site/ijcsis/ijcsis