Classification of Obsessive-Compulsive Disorder Symptoms in Arabic Tweets Using Machine Learning and Word Embedding Techniques

Malak Fahad Al-Haider [1,*], Ali Mustafa Qamar [2], Hasan Shojaa Alkahtani [1], and Hafiz Farooq Ahmad [1]

[1] Computer Science Department, College of Computer Sciences and Information Technology (CCSIT), King Faisal University, Al-Ahsa, Saudi Arabia
[2] Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia
Email: Malak.alhaider@outlook.com (M.F.A.); al.khan@qu.edu.sa (A.M.Q.); hsalkahtani@kfu.edu.sa (H.S.A.), hfahmad@kfu.edu.sa (H.F.A.);
*Corresponding author

***Abstract*—Obsessive-Compulsive Disorder (OCD) is a mental health condition that is characterized by persistent and intrusive thoughts, images, or impulses (obsessions), as well as the presence of repetitive behaviors or mental acts (compulsions) that are aimed at reducing anxiety. These behaviors are typically rigid and performed according to specific rules; furthermore, they can be time-consuming and cause significant distress or impairment in daily functioning. Detecting the symptoms of OCD could help individuals become aware of them and seek a medical diagnosis. People increasingly rely on social media such as Twitter to express and disclose their feelings and thoughts. Although researchers have researched OCD in English, there is no substantial work in this domain concerning Arabic tweets. Therefore, this research proposes investigating and detecting OCD in Arabic tweets using Machine Learning (ML) and word embedding techniques. First, we obtained tweets via web scraping and manually annotated the data by involving medical professionals. Secondly, we conducted exploratory data analysis on the textual data and emojis used to find their correlation. Furthermore, we focused on the quality of word representation. Efficient word representation approaches (word embeddings) combined with recent ML models have shown reasonable progress on text classification tasks. We trained our classification model using the Arabic version of fastText. The proposed models were also tested on our dataset. The analysis indicated that utilizing fastText as a word embedding technique is a particularly promising approach.**

***Keywords*—obsessive-compulsive-disorder, machine learning models, text classification, word embedding, fastText**

I. Introduction

Obsessive-Compulsive Disorder (OCD) is characterized by the occurrence of persistent obsessions, compulsions, or both [1]. These obsessions and compulsions interfere with daily activities and cause severe distress. Individuals with OCD will try to ignore or stop the obsessions, but this will only exacerbate them. Despite efforts to ignore or eliminate these disturbing thoughts or urges, they keep thinking about them. For obsessions, they will have fears related to contamination, pollution/dirt, aggression/harm, sexuality, religion, and perfection [2]. In addition, they will also feel the need for things to be orderly and consistent.

Furthermore, those with OCD will suffer aggressive or horrific thoughts about losing control and harming themselves. Individuals with compulsive symptoms will not only wash their hands until the skin becomes sore from the intensity of rubbing their hands, but they will also perform actions such as frequently checking the doors to ensure they are locked, counting in specific patterns, and exaggerating in reassurance. They may or may not realize their obsessions or compulsions are excessive or unreasonable. However, they are time-consuming and interfere with their daily routines, social functioning, and work [3].

OCD is sometimes referred to as a "vicious" cycle since the obsessions and compulsions occur in a loop that can be extremely difficult to interrupt [4]. The longer one stays in the cycle, the more momentum and strength it accumulates, making it even more difficult to escape. Therefore, it is essential to detect OCD symptoms in the early stages.
According to the WHO, OCD is one of the top ten disabling disorders and is associated with dysfunction and a lower quality of life [5]. Moreover, OCD is regarded as one of the most prevalent neuropsychiatric disorders and is rated as one of the most exhausting and harmful diseases.

People communicate their favorable and unfavorable comments on social media platforms. Rich bodies of work have confirmed that social media is a safe space for many people to vent their feelings by posting information that reflects their emotions [6]. In related research [7], several benefits and uses of social media were discovered: Social engagement, information seeking, leisure, amusement, relaxation, communicative utility, convenience utility, the expression of opinions, information sharing, and

Manuscript received November 25, 2023; revised January 5, 2024; accepted April 11, 2023, published July 8, 2024.

surveillance/knowledge of others. All these applications have made social media platforms powerful data sources.

Twitter is one of the most popular social media platforms, with 217 million daily active users who post approximately 500 million tweets daily [8]. Twitter has been demonstrated to be a rich source of data collection because of its open-source nature, alongside its developer-friendly API that offers superior features in terms of timeliness, speed, and connectivity [9]. Additionally, the 280-character limit on Twitter user posts, such as tweets and comments, facilitates easier analysis compared to platforms allowing longer submissions [8]. Compared to other social media platforms, Twitter users tend to use direct messaging services less frequently, simplifying the process of mining tweets owing to the prevalence of retweets and comments. This results in a dataset containing more sensitive and personal information to be studied to analyze user behavior [8, 9].

The significance of social networks, especially Twitter, in times of crises or pandemics has sparked considerable attention among the scientific community. The content of data collected can offer valuable perspectives on mental disorders, especially during health emergencies like the Ebola outbreak [10] and seasonal flu epidemics [11]. During the COVID-19 outbreak, a survey was carried out among 2,909 individuals in Saudi Arabia to gather information on sociodemographic details and to evaluate responses using the Brief Obsessive–Compulsive Scale (BOCS) and the Perceived Stress Scale (PSS) [12]. Most respondents were women (73.9%) with at least a university degree (81%). The survey found that 57.8% reported newly developed obsessions, 45.9% reported compulsions, and 72.4% experienced moderate to high levels of perceived stress. Notably, new concerns about cleanliness, germs, and viruses were predominantly more common in participants aged 40–49, as well as among workers, students, those enforcing quarantine, and individuals who had spent 20 days or more in quarantine. The emergence of new compulsive behaviors related to handwashing was particularly notable in the age group of 30–49. After the outbreak of COVID-19, the emphasis on hygiene and sterilization advice on Twitter has increased, from washing hands to disinfecting surfaces and wearing masks until all vaccines are completed [13]. As much as this advice confuses the community, people discuss their concerns and opinions on social media. Therefore, studying and comprehending the COVID-19-related content on social media platforms can aid in the classification of tweets.

Mining Twitter content to create datasets, particularly regarding mental disorders, has piqued the interest of scientists in a variety of domains [14–18], including data mining, Machine Learning (ML), Natural Language Processing (NLP), and big data analysis. The use of Twitter for research in mental health has expanded; hence, generating Twitter datasets to be used by academics will expand the research in this area.

Previous works have contributed to detecting various mental disorders through social networking. They used classification techniques to predict Post-Partum Depression (PPT) [19], depression [20], suicide detection [21], and anxiety, bipolar disorder, addiction, and dementia [22]. Even though OCD is one of the most serious disorders that require early detection, there are only a few studies on it. Furthermore, there have been few studies that have created OCD datasets, particularly concerning OCD symptom detection from Twitter content in Arabic. To address this issue, we constructed an Arabic tweet dataset to detect the presence of OCD symptoms. The dataset was created during the pandemic. The significant contributions of this research are threefold:

- To the best of our knowledge, no Arabic dataset is available for OCD symptom detection on the Twitter platform. We built an Arabic dataset to detect OCD symptoms under the supervision of the Erada Complex and Mental Health Hospital in Saudi Arabia, covering the period from March to September 2022.
- We performed a preliminary analysis of the text and emojis in tweets and showed their correlation utilizing the visualization tools.
- Word representation is another area of interest in our research. We investigated whether fastText is a valuable tool for Arabic text in terms of word embedding by employing the Arabic version of fastText. Furthermore, various ML classifiers were applied to the dataset using text classification techniques, including TF-IDF and fastText, and we evaluated the performance of the classification techniques using evaluation metrics.

The remaining paper is organized as follows: Section II details the related work. Section III presents the proposed work, and Section IV contains the results and discussion. In Section V, the limitations of our work are mentioned. Finally, the conclusion with the main findings and remarks is given in Section VI.

II. Literature Review

Social media channels provide fast and dependable data that is valuable for mining and detecting people's disorders. Consequently, multiple studies have created Twitter posts with datasets related to mental disorders in various languages.

Shatte *et al.* [19] investigated whether passive social media cues might be used to identify fathers at risk of Postpartum Depression (PPD). Their dataset was taken from social media posts for 365 days. The evaluation process depended on changing keywords that were used by fathers. They used different variables, such as behavior, emotion, linguistic style, and discussion topics, as features to build support vector machine classifiers. They found that the danger of PPD for fathers can be predicted based on pre-partum data. Additionally, they found that the best algorithm was built based on discussion topic variables only, with a score of 0.82. In addition, SVM also produced good scaling.

Another study by Islam *et al.* [20] described a Facebook dataset for detecting depression. They examined the capability to apply Facebook information and implemented the K-Nearest Neighbors (KNN) classification method for catching depressive emotions.

They analyzed four kinds of features (emotional process, temporal process, linguistic style, and all features) and trained an algorithm to apply each type separately and collectively. The results revealed that the accuracy for various kinds of KNN techniques ranged from 60–70% in different parts of the levels.

Baghdadi *et al.* [21] presented a tweet dataset for detecting mental illnesses, such as Major Depressive Disorder (MDD), on Twitter in Arabic tweets. They proposed a thorough methodology for categorizing tweet inputs into two categories—normal or suicide. Performance metrics were reported via Bidirectional Encoder Representations from Transformers (BERT) and Universal Sentence Encoder (USE) models. They found that the best-weighted sum metric (WSM) was 80.2%, and the best WSM for Arabic BERT models was 95.26%.

Advancing further, an intriguing study by Hinduja *et al.* [22] delves into research to decipher the rationale behind users engaging with social media to express their emotions, viewpoints, and health-related concerns. They focused on the Twitter platform and proposed a generic framework to support proactive mental health monitoring. To achieve accurate sentiment analysis results, they performed detailed data cleaning and preprocessing of tweets using regular expressions based on observed patterns. They developed an ML model, Long Short-Term Memory (LSTM), which is used for the early detection of at-risk social sensors and is based on bespoke event definitions.

Babu and Kanaga *et al.* [23] suggested a data analytic model for identifying depression. The data for this proposed model was collected from user posts on Twitter and Facebook. The patient's User-Generated Content (UGC) for the previous two weeks was collected using the Graph API on Facebook and the REST API on Twitter. The ML model was then trained to classify depression criteria into six ranges (considered normal, mild, moderate, borderline, severe, and extreme). The verdict was deemed negative when the percentage was above the borderline (over 55%).

Another study by Leis *et al.* [24] attempted to detect the linguistic elements of tweets in Spanish as well as the behavioral patterns of Twitter users who compose them, which could indicate signs of depression. Concerning their strengths, their paper aimed for a valid topic to focus on that draws analysis for information in Spanish because many users are ignored or skipped due to specific lexical content-oriented data that are language-based. The problem with this work was the absence of a scientifically and medically validated model to confirm the generalizability of the depression prediction outcomes derived from their employed model.

Spates *et al.* [25] used content analysis to study 4,524 Twitter posts to see how Twitter can be a significant indicator of how suicidal thoughts develop and propagate. They found that, although there were repeated expressions of "wanting to die", most tweets looked non-threatening. Additionally, most talks concerning suicide on Twitter focused on sharing suicide prevention information with other users. This finding implies that Twitter could be a valuable tool for disseminating suicide prevention information across regional borders.

Usman *et al.* [26] built a large-scale sentiment data collection that comprised 90,000 COVID-19-related tweets gathered during the pandemic's early phase, i.e., from February to March 2020. Positive, negative, and neutral emotion classes were assigned to the tweets. They examined the sentiment classification of the gathered tweets by applying various feature sets and classification methods. They discovered that "negative opinion" was a critical factor influencing public sentiment. Initially, there was a public preference for lockdowns, but this sentiment had changed by mid-March. Their research underscores the importance of establishing a responsive and dynamic public health engagement to address and mitigate the proliferation of negative sentiment on social media during a pandemic.

In addition, Bhatia *et al.* [27] introduced the Arabic COVID-19 tweets dataset. This work is intended to extract Arabic tweets from various pandemic periods and carry out various preprocessing operations on them. In addition, a range of state-of-the-art Deep Learning (DL) and ML classifiers were applied to the dataset, and their accuracy was evaluated using several visualization techniques. This study concentrated on the predictive modeling of tweets to ascertain how public opinion altered over time: before, during, and after the pandemic. Additionally, that study performed sentiment analysis on the various pieces of information revealed following the pandemic, both before and after the lockdown. In both periods, DL algorithms outperformed ML ones in terms of accuracy.

Irene *et al.* [28] created the Emotion-COVID-19-Tweet (EmoCT) dataset and used NLP tools to assess tweets in terms of mental health. They trained deep models to categorize each tweet as anger, anticipation, disgust, fear, joy, sadness, surprise, or trust. They manually labeled 1,000 English tweets to create the data set. They found that the two sentiments, "sad" and "fear", correlated more with severe negative emotions. They suggested a method for determining the causes of sadness and fear and studying the emotion trend at the keyword and topic levels. However, the size of the dataset was modest.

Ali and Alenezi [29] described a dataset for a citizen-concern monitoring system based on emotion recognition in Twitter data. They monitored public sentiments and correlated them with COVID-19 symptoms. A hybrid system leveraging rule-based and neural network models annotated the gathered tweets. The rule-based method utilized Arabic emotion and COVID-19 symptom lexicons to classify 0.3 M tweets. At the same time, the neural network was applied to increase the number of sample tweets annotated through the rule-based system. They employed a DL approach using LSTM networks to classify tweets into six distinct emotional categories, as well as symptomatic and non-symptomatic types. Their analysis revealed that the emotions of anger and fear dominated tweet frequency and user engagement, with expressions of happiness coming in subsequently. The study's tracking of user interactions indicated that likes and replies were predominantly used in response to non-symptomatic

tweets. At the same time, retweets were the preferred method for disseminating tweets that mentioned any COVID-19 symptoms.

Lastly, Yang *et al.* [30] focused on the high chronic stress among Americans, especially during the COVID-19 pandemic. Recognizing the limitations of traditional stress measurement methods, the researchers developed an automatic system on Twitter to identify individuals self-reporting chronic stress.

TABLE I. RECENT RESEARCH WORKS REGARDING COVID-19

Study	Mental disorders	Data collection	Social media	Language	Dataset size
Shatte *et al.* [19]	PPD	PRAW (Python Reddit API Wrapper)	Reddit	English	67,796 posts
Islam *et al.* [20]	Depression	NCapture	Facebook	English	7,145 comments
Baghdadi *et al.* [21]	MDD	Twitter scraping (Twint)	Twitter	Arabic	14,576 tweets
Hinduja *et al.* [22]	Bipolar disorder, dementia, and schizophrenia	Twitter streaming API	Twitter	Arabic	2,138 tweets
Babu and Kanaga [23]	Depression	API on Facebook, REST API on Twitter	Twitter and Facebook	English	8,000 posts
Leis *et al.* [24]	Depression symptoms	Twitter API	Twitter	Spanish	140,946 tweets
Spates *et al.* [25]	Suicide thoughts	Twitter streaming API	Twitter	English	4,524 tweets
Usman *et al.* [26]	Sentiment analysis	Twitter streaming API	Twitter	English	90,000 tweets
Bhatia *et al.* [27]	Sentiment analysis	Web scraping snscrape	Twitter	Arabic	2,617 Arabic tweets
Li *et al.* [28]	Mental disorders detection tweets	Twitter streaming API	Twitter	English	1,000 tweets
Ali and Alenezi [29]	Mental disorders detection tweets	Twint	Twitter	Arabic	5.5 M tweets
Yang *et al.* [30]	Stress	Twitter streaming API	Twitter	English	4,159 tweets

Using stress-related keywords, they collected and filtered tweets, achieving an 83.6% accuracy with the Bidirectional Encoder Representation from Transformers. The study demonstrates the feasibility of large-scale chronic stress surveillance and intervention through Twitter.

In contrast to prior studies that compiled datasets using different crawling tools (as outlined in Table I), our approach distinguishes itself in key aspects. While our literature review extensively covers studies on mental health detection in social media—spanning depression, postpartum depression, suicide, anxiety, bipolar disorder, addiction, and dementia—we acknowledge the importance of explicitly addressing the relevance, parallels, or distinctions between these studies and our work on OCD detection, particularly in the context of the Arabic language.

One notable distinction lies in annotation validation. Previous studies often lacked specialist validation for their annotations, whereas, in our research, we rigorously validated annotations with domain specialists, ensuring cultural and linguistic relevance, especially in an Arabic context.

A significant gap in prior works was the limited attention given to the quality of words in Arabic tweets. In contrast, our methodology prioritizes this aspect by employing word embedding techniques, ultimately enhancing the accuracy of our ML classifiers.

In summary, our methodology stands out regarding dataset curation and annotation validation.

It places a unique emphasis on the quality of language in Arabic tweets. This explicit connection enriches the narrative, contributing to the advancement of mental disorder detection on social media within the Arabic-speaking community.

III. MATERIALS AND METHODS

The proposed framework is presented in Fig. 1. This work aims to create an Arabic dataset for OCD detection to determine whether a tweet has OCD symptoms. The first phase was to develop an annotated dataset based on OCD symptoms. After collecting and validating the OCD dataset, five preprocessing steps were performed, and then text analysis was conducted. In the second phase, we built ML classifiers: random forest, decision tree, and KNN. We used TF-IDF with ML classifiers because it proved to be an effective approach in Arabic text classification [31]. Simultaneously, we used the word embedding technique (fastText) with ML to investigate its performance in Arabic and to thus compare the result with TF-IDF. For performance, we used grid search to obtain the best hyperparameters; then, we applied precision, recall, and the F1-Score in k-fold cross-validation techniques.

A. Data Collection

This work presents a collection of 8,711 Arabic posts from Twitter. These Arabic tweets represent the discussions in the period after taking the COVID-19 vaccine. The model aims to classify tweets as "yes" if symptoms are present and "no" if no symptoms are present. The tweets were streamed for around seven months, from March 2022 to September 2022.

This period was selected because most Arab countries had already taken the COVID-19 vaccine by this point. We collected tweets through a crawling procedure via the Tweepy API. Here, an account on the Tweepy API linked to our Twitter account was created. The Tweepy API can accept inputs and return information for the Twitter account to retrieve the tweets. The entire process is depicted in Figs. 2 and 3.

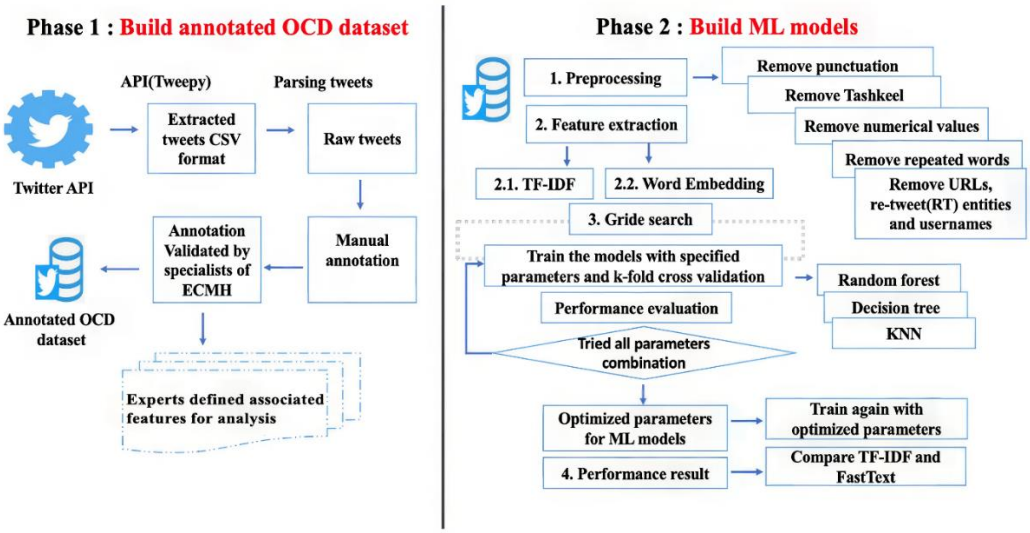

Fig. 1. Proposed framework for detecting OCD symptoms using various approaches.

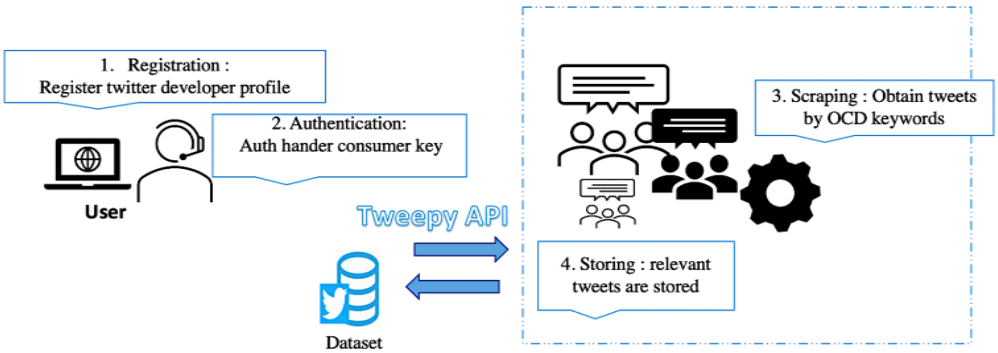

Fig. 2. Data gathering using Tweepy API.

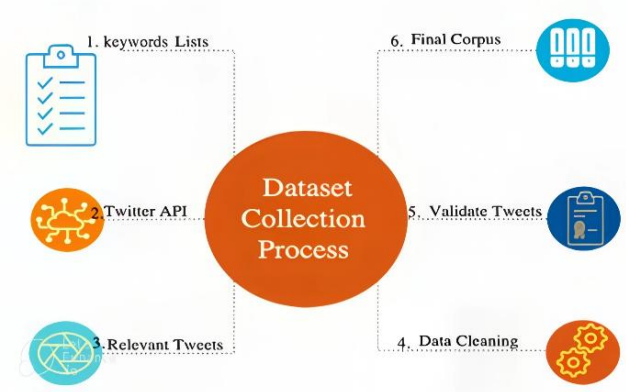

Fig. 3. The OCD dataset collection process.

B. *Data Annotation*

Before training a model in supervised learning, a labeled dataset is necessary [30]. In our study, tweets were systematically selected based on stringent criteria to ensure relevance to Obsessive-Compulsive Disorder (OCD) within the Arabic-speaking community. The selection process involved a combination of targeted keywords derived from the Yale-Brown Obsessive Compulsive Scale (Y-BOCS), culturally relevant terms associated with OCD in Arabic, and topics related to mental health, capturing a diverse range of expressions while maintaining specificity to the context of Eradah Hospital. Table II shows these related keywords.

The annotation process was enriched by the active participation of five mental health specialists exclusively from Eradah Hospital.

These licensed professionals brought a specialized understanding of OCD within the Arabic cultural context, contributing to the development of annotation guidelines and ensuring a clinically informed approach to the identification of OCD-related expressionism tweets. Challenges faced during annotation, including nuances in

language and variations in cultural expressions, were collectively addressed through regular meetings among annotators, specifically focusing on the insights provided by the five Eradah Hospital specialists. The annotation process adhered to a systematic workflow, where annotators, particularly from Eradah Hospital, underwent comprehensive training sessions emphasizing the specific nuances of OCD expressions within the Arabic language and cultural context. Guidelines provided examples tailored to Eradah Hospital specialists' clinical perspectives, fostering a shared understanding of the annotation criteria. Ongoing feedback sessions and regular communication maintained consistency among annotators from Eradah Hospital.

Stringent quality control measures were implemented to uphold the dataset's quality, focusing on Eradah Hospital specialists. Calibration exercises aligned interpretations, and periodic reviews by the five specialists ensured the clinical accuracy of annotations. The subset of data independently reviewed by this team added an extra layer of validation, enhancing the reliability of the annotated dataset. This multi-faceted approach in the annotation process, involving a team of five Eradah Hospital specialists, was implemented to ensure the robustness and clinical accuracy of our labeled dataset for detecting OCD in Arabic tweets.

C. Data Preprocessing

After preparing the dataset, a fundamental step was applied: text preprocessing and cleaning the input tweets to eliminate noise and extraneous data [30].

D. Data Cleaning

Dealing with the data's loud nature is an essential step in cleaning all texts [32, 33].

This procedure removes the contents of tweets from those that are meaningless. We applied the following steps:

- Remove re-tweet entities.
- Remove usernames.
- Remove URLs.
- Remove digits and single Arabic letters.
- Remove non-letter characters (e.g., +, =, %, $).
- Remove punctuation marks.
- Furthermore, in the Arabic language, we have Tashkeel signs, (such as "وغيرها", "والشده", "والضمه", "الفتحه"), which have no meaning and were eliminated.

E. Removing Stop Words

Stop words (such as من, في, على, الى) and their corresponding English terms (to, in, on, from) are frequently used words that are repeated in the dataset and do not provide any valuable information for detecting OCD; therefore, we removed them.

The list of Arabic words used in informal dialects includes words such as "كذا،عشان", and their corresponding English terms (because, like) and others were also removed. As a result, the dimension shrank, and attention was focused on the most crucial information that was associated with OCD symptoms.

TABLE II. OCD KEYWORDS

Obsessive	Translation	Compulsive behaviors	Translation
Anxiety	قلق	Washing	غسل
Obsession	هوس	Strict routine	روتين صارم
Doubt	شك	Suicide	انتحار
Fear	خوف	Coordination	تنسيق
-	-	Counting in patterns	عد بنمط
-	-	Pursuit of perfection	مثالي
-	-	Repeated actions	تكرار
-	-	Cleaning	نظفت
-	-	Checking	تأكد من

IV. EXPLORATORY DATA ANALYSIS

It is essential to demonstrate that the tweets in the OCD dataset constitute an excellent representative sample of the Arabic tweets posted during the COVID-19 vaccine period. People were still skeptical about the vaccine's efficacy and the end of the COVID-19 outbreak. To test this hypothesis, we examined the textual content of the tweets to demonstrate that people are still concerned about taking the vaccine. We illustrate an initial analysis of the dataset, highlighting the most frequent keywords and the emoji interactions that reflect the psychology of users in the dataset. We also examined the tweets to see if people still had interests and concerns about COVID-19 after the vaccines. We created objects called ("All tweet", "TextEmoji", and "TextOnly"). Table III shows the functions of each object.

TABLE III. OBJECTS FOR ANALYSIS

Object	Function
"All tweet"	The tweet has only words without filtering
"TextOnly"	It has only Arabic words in the tweet
"TextEmoji"	It has only Arabic words and emojis in the tweet
"Emoji only"	The tweet has Emojis only

A. Textual Analysis

We focused on word frequency in "TextOnly", which has only Arabic text. Furthermore, we presented the most repeated words over 28 weeks in our dataset.

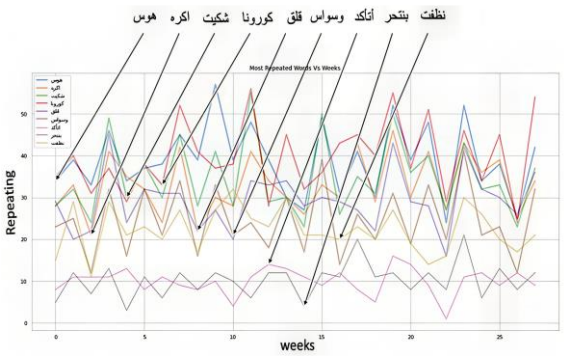

Fig. 4. Distribution of tweets by keywords.

Fig. 4 illustrates the relevant OCD keyword distribution over 28 weeks. We can observe that the terms "كورونا", "هوس", and "اكره", as well as their corresponding English terms "corona", "obsession", and "hate", were the most dominant in our dataset in general. We further looked at the prevalence of the words over the past few weeks, as shown in Fig. 4. In August (week 21), the words "corona",

"obsession", and "doubt" were highly repeated, which is not surprising since it was in summer. People tended to doubt personal hygiene and sterilization. Additionally, we noticed that most of the repeated words were related to hygiene advice, feelings of fear, and corona. These words indicated that people, even after taking vaccines, were still interested in COVID-19.

B. Analysis of Annotation

After carrying out the aforementioned steps, we defined the OCD words, the most relevant words in our dataset that can indicate the presence of OCD symptoms. In terms of validity, the OCD words passed through a pipeline of processing and analysis with HMHC experts, and the experts provided the degree of association to OCD symptoms. Table IV shows these words, and we only consider moderate and high degrees as features. This step helps reduce the amount of redundant data in the dataset and build the model with less machine effort. The degree of association is also shown in Table V.

TABLE IV. OCD WORDS

OCD words	Translation
OCD words = ['هوس', 'اكره', 'كورونا', 'شكيت', 'قلق', 'وسواس', 'نظفت', 'دمرنننننني', 'لازم', 'وربط', 'التدقيق', 'خفت', 'بنتحر', 'أكره', 'اتأكد', 'احس', 'اتاكد', 'الوضوح', 'والحقيقة', 'افكار', 'ظن', 'الكورونا', 'يقفل', 'بالكاد', 'بنتأكد', 'دايم', 'خوف', 'اخاف', 'طبيعي', 'ظنه', 'أتأكد', 'وخيبت', 'كررت', 'نسقت', 'اعقم', 'شعور', 'إصابة', 'قهري', 'افكاري', 'مضاعف', 'أتقدر', 'مجهود', 'والشكوك', 'مرات', 'شك', 'مرتين', 'مثالي', 'كاملة', 'الاحتمالات', 'أشعر', 'كثير', 'مصاب', 'الجراثيم', 'المؤلمة', 'يعكر', 'خايف', 'الأمان', 'لقاح', 'النظافه', 'التعقيم', 'مسيطره', 'فيروس', 'منتطمن', 'بموت', 'الفايروسات', 'الصحة', 'الامراض', 'والميكروبات', 'الخوف']	OCD words = ['obsessive,' 'hate,' 'corona,' 'suspect,' 'worry,' 'obsessive,' 'cleaned,' 'feared,' 'review,' 'bind,' 'necessary,' 'destroyed me,' 'Want to suicide,' 'I hate,' 'I want to check,' 'I feel,' 'I am 'to be sure,' 'thoughts,' 'the truth,' 'thought,' 'think,' 'fear,' 'always,' 'to be sure,' Barely,' 'Closing,' 'Corona,' 'feeling,' 'sterilized,' 'coordinated,' 'repeated,' 'disappointed,' 'certain,' 'thought,' 'natural,' 'afraid,' 'effort,' 'to be appreciated,' 'double,' 'Thoughts,' 'compulsions,' 'illness,' 'feel,' 'possibilities,' 'complete,' 'perfect,' 'twice,' 'doubt,' 'times,' 'doubts,' 'afraid,' 'sour,' 'painful,' 'germs,' 'injured,' 'a lot,' 'safety,' 'vaccine,' 'cleanliness,' 'death,' 'reassured,' 'virus,' 'controlling,' 'sterilization,' 'microbes,' 'diseases,' 'health,' 'viruses,' 'the fear']

Fig. 5. OCD words in the *Yes* category.

Fig. 6. OCD words in the *No* category.

Next, we checked for the presence of OCD words in both labels (Yes and No). The word cloud presents the frequency of the words (see Figs. 5 and 6). We can see that OCD words have a higher frequency with the Yes label than with the No label, which makes sense. There were more positive words in the negative (No) label, such as لطف، سكون، بشائر، تطمئن; their corresponding English terms are contentment, good news, stillness, and kindness.

This analysis indicated that people without OCD symptoms are more optimistic and stable. On the other hand, in the positive (Yes) label, there were more associated terms with OCD and, thus, more negative terms, such as خفت ، بنتحر، مراوغه، خيبه، تمويه; the corresponding English terms are fear, suicide, evasiveness, disappointment, and camouflage.

C. Analysis of Emojis

In "TextEmoji", we focused on both the presence of OCD words and their dynamic interactions with emojis. We created an object called "Tweet Case"; when tweets showed the presence of OCD symptoms, it returned "OCDWords". Similarly, when tweets did not contain OCD symptoms, it returned "No OCDWords". Remarkably, OCDWords have more negative symbols when compared with tweets without OCDWords (see Figs. 7 and 8).

Additionally, people with OCD symptoms used fewer emojis, thus indicating that OCD people tend to be more negative and ambiguous, which is likely due to the nature of their anxious thoughts and their hesitant fears.

Overall, our analysis shows that people continue to experience anxiety and OCD symptoms as a result of COVID-19. To show the importance of our dataset, we compared it with the EmoCT (Emotion-COVID19-Tweet) dataset [30]. EmoCT used Twitter posts as the data source and focused on 1,000 English tweets written during the pandemic.

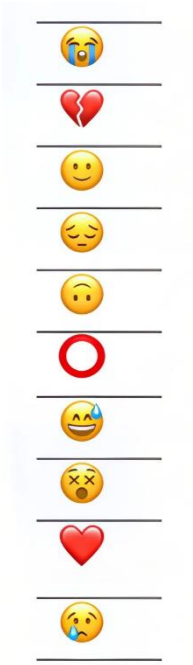

Fig 7. OCD words in the *Yes* category.

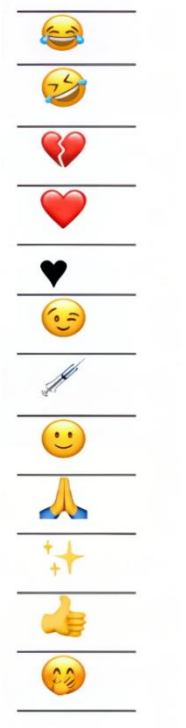

Fig 8. OCD words in the *No* category.

They offered annotated datasets with various annotations concerning mental health, such as anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. However, the validity of labeling was not examined by an expert, and no attention was paid to OCD. Additionally, the period reflects the English tweets during the pandemic, not the period following the vaccine.

In comparison, our work develops the first Arabic tweet dataset that can be used to detect the symptoms of OCD in general. The researchers can use it to detect the symptoms of obsession and compulsive behaviors. Additionally, this research can facilitate the design of experiments to investigate OCD. The dataset is versatile and can be applied to various machine-learning strategies for cluster and classification analyses.

TABLE V. THE DEGREE OF ASSOCIATION TO OCD SYMPTOMS

The degree of association	Words
Moderate	دمرتني /هوس
Moderate	بنتحر/ظن
Moderate	الكورونا/يقفل
High	اعقم
High	النظافة/نظفت
High	كلمات الخوف من الميكروبات
High	كلمات الخوف من الامراض
High	وسواس/خفت/قلق/شكيت
Not associated	اكره/ربط/الحقيقة/احس/يكره/لازم
High	أفكار/الحقيقة/اتاكد/يكره/يقفل/التعقيم/كررت
High	/جراثيم/مرتين/مرات/التدقيق

From a medical point of view, our work is helpful for rapidly indicating the potential presence of OCD; then, medical intervention could confirm the presence. Also, many individuals with OCD found that telemedicine and social media provided a safe place to express their emotions without the stigmatization that would be typically present in a psychiatric clinic.

V. FEATURE EXTRACTION

Feature extraction is one of the critical stages in NLP to better understand tweets. In our work, the initial tweets were cleaned and normalized, then converted into features to be utilized for modeling [30]. We used a strategy to apply weights to specific terms in our dataset before modeling them. We used numerical representation for individual words since numbers are easier for computers to process.

A. Term Frequency-Inverse Document Frequency (TF-IDF)

It is a numerical statistic meant to reflect the importance of a word in a collection or corpus of documents [34]. Term frequency will favor the words with the highest frequency and risk the loss of the significant information that may lie in the less frequent words. It is a way to give a better value for each word. It considers the importance of words in the document, which resulted in a higher value for the most frequent word that appears in a few documents and a lower value for more frequent words in many documents or less frequent words in a document. The *TF-IDF* is represented in Eq. (1).

$$tf\ idf\ (t,d) = tf(t,d) \cdot idf\ (t) \quad (1)$$

where *tf (t, d)* denotes the term frequency function of term *t* in a document *d* and *idf (t)* represents the inverse document frequency of term *t*. The equation of inverse document frequency is described in Eq. (2).

$$idf\ (t) = \log\left(\frac{1+N}{1+df(d,t)}\right)+1) \quad (2)$$

where *N* represents the total number of documents and *df (d, t)* is the document frequency function of the number of documents *d* that contain the term *t*.

In this study, the Twitter API was used to collect tweets. Then, tweets were labeled and cleaned. TF-IDF analysis was conducted on the collected tweets to ascertain their most significant words. Identifying these keywords helps to provide context and significance to tweets related to OCD. The importance of the words within the dataset was determined by their frequency of occurrence in the collected texts. Following this, symptom detection was carried out using the list of essential words identified to ascertain the prevalent symptoms described in the texts related to our dataset. The symptoms of the tweet were then labeled as "Yes" or "No".

B. *Word Embedding (fastText)*

Word embedding is a method to convert textual information to numeric form, which can be used as input to ML classifiers [28]. It can be deduced from the surrounding words or even from those nearby. Modeling words by displaying them in a vector space, known as vector semantics, is the focus of the NLP discipline. As words are embedded into a vector space, the vector representation of a word is called word embedding [29]. fastText is a library developed by Facebook's AI research (FAIR) team for word embedding and text classification learning [30]. The model enables the development of supervised and unsupervised word vector representation learning algorithms. Facebook makes pre-trained models available for 294 languages, including Arabic [31, 32]. In this work, we used fastText to produce the word vectors and then applied ML classifiers. A collection of character n-grams was used to create the word embedding, with each word represented by the sum of its n-gram vectors. In our work, the word "وسواس" can be represented by <الوسواس, اضطراب, شك, شيطان>, which is translated as obsessive, disorder, doubt, and demon, respectively. Figs. 9 and 10 illustrate some of the word similarities.

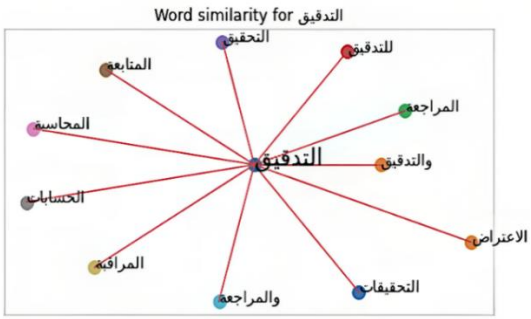

Fig. 9. Word similarity for (التدقيق).

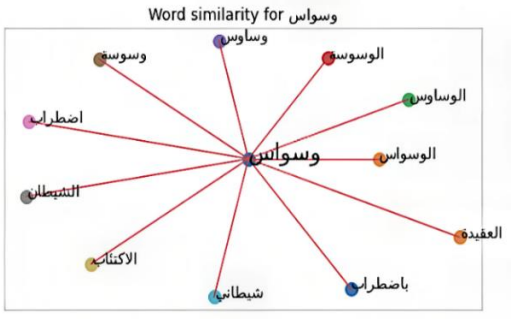

Fig. 10. Word similarity for (وسواس).

TABLE VI. VALUES OF WORD SIMILARITY.

Corona\كورونا	Obsessive\وسواس	Sterilized\عقمت
(كورونا, 0.0)	(وسواس, 0.0)	(عقمت, 0.0)
(الكورونا,10.87)	(الوسواس, 13.12)	(تخشبت,16.61)
(والكورونا,11.30)	(وساوس, 16.42)	(والميكروبات,16.72)
(بالمستشفيات,14.06)	(اضطراب, 17.55)	(الميكروبات,16.72)
(المستشفيات,13.74)	(الشيطان, 17.82)	(والحقيقه,16.98)
(مستشفيات,14.14)	(ينتابني, 18.64)	(المستشفيات, 17.00)
(الفيروسات,14.14)	(اكتئاب,18.79)	(الملقحين,17.04)
(والميكروبات,14.18)	(القلق,18.82)	(الجزائر,17.11)
(الحمدالله,14.32)	(الخوف,18.82)	(والاوبئه,17.14)
(والجراثيم,14.35)	(بفايروس,18.83)	(الجراثيم,17.17)

The values in Table VI indicate how similar the word is to other words with similar meanings. There are three OCD words, which are "وسواس", "كورونا", and "عقمت"; their English counterparts are sterilize, OCD, and corona. For each column, there is a set of words. When the value is 0, it is the same word; when it becomes higher (i.e., second raw), the following closest meaning is the next along. Similarly, the meaning is farther from the word when the value has become very high.

VI. MACHINE LEARNING CLASSIFIERS

After performing feature extraction, a feature vector is fed to classifiers to determine the class of tweets and whether they have OCD symptoms. The aim is to find the class to which new data will be assigned. We used three classifiers and a grid search to obtain the best parameters. RF, KNN, and DT classifiers were used for classification purposes. The extracted features were fed to each classifier.

Our selection of classifiers—Decision Trees (DT), Random Forest (RF), and K-Nearest Neighbors (KNN)—was based on their proven effectiveness in addressing similar problems in Arabic tweets, especially in sentiment analysis and supervised learning [33, 35]. This choice is supported by relevant literature, including the works of

Bouchlaghem *et al.* [33] and Bolbol *et al.* [35], where these classifiers demonstrated notable performance in classifying sentiments in Arabic tweets.

Additionally, these classifiers can integrate with both TF-IDF and fastText. Our methodology incorporates TF-IDF, as demonstrated by Alzanin *et al.* [36], and word embeddings (fastText), as explored by Elfaik and Nfaoui [37]. This strategic integration allows for a comprehensive exploration of feature spaces, maximizing the classifiers' effectiveness in capturing nuances within Arabic social media tweets.

A. *Random Forest*

The random forest method consists of numerous decision trees working together to achieve a single objective [18]. It works as an ensemble learning algorithm. Ensemble learning uses many algorithms to obtain better classification results than any single algorithm. RF builds a variety of DTs based on a variety of algorithms during the training process. The root node discovery and feature node splitting occur randomly in an RF.

The output from RF is a class of patterns representing the average of all the outputs from several classifiers. The rationale behind this endeavor is that the trees defend one another against their unique mistakes. While the output of

some trees may be incorrect and the output of other trees will be correct, the trees can move in the right direction as a group. RF is commonly used as an efficient classifier in Arabic text classification. Therefore, this classifier was used in this study for OCD symptom detection. A typical RF is shown in Fig. 11. In our model, the maximum depth is 7, and the number of estimators is 500.

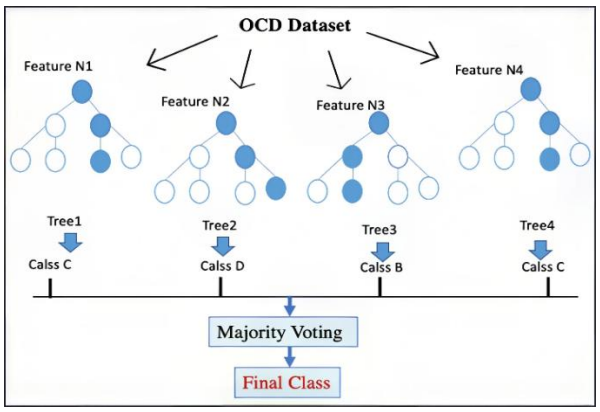

Fig. 11. Random forest.

B. Decision Tree

In supervised learning, the decision tree is an effective and straightforward classifier [33]. It is mainly used to identify a pattern's class, whether OCD or not, using a set of questions. A tree can be built by splitting the source set into subsets depending on an attribute value test. A question is created on each node of the tree-like graph that the decision tree explores. The classifier uses a series of queries designed on various decision tree nodes to try and match the pattern with the feature vector it receives as input. It repeatedly divides the problem into sub-problems until it can determine if it is OCD or not. We adjusted the hyperparameters so that the maximum depth is 17, and the Gini index is used for attribute selection.

C. K-Nearest Neighbor (KNN)

KNN is one of the simplest supervised ML algorithms. This technique finds the similarity between the new instance and available examples and then puts the new instance into the most similar class [27]. It saves all the existing data and categorizes new inputs based on similarity. As new data are generated, they may be instantly classified into a suitable category. Since KNN is a non-parametric technique, it makes no assumptions about the underlying data. The most commonly occurring label around a particular data point is assigned. This process is technically known as "plurality voting", although it is often called "majority voting" in scholarly work. It is important to note that "majority voting" formally requires a result exceeding 50%, so we chose K=5 to avoid a tie in the voting outcome.

VII. Results and Discussion

After constructing the OCD dataset, making a comprehensive analysis, and preprocessing each tweet, we used two techniques separately—TF-IDF and fastText—before applying ML classifiers. In both methods, we converted the processed tweets into a numerical form of real numbers to make them compatible with the necessities of ML classifiers. In TF-IDF, we take the cleaned text, such that there are no stop words, and then apply TF-IDF. We obtained 21,592 attributes, which is the vector of values that is ready for the ML classifiers. In word embedding, fastText transforms the tweets into a vector of real numbers. These vectors contain meaningful information between words. The vectors are closer to each other if the words have similar meanings [31]. In the OCD dataset, we used the Arabic version of fastText, which was computed in two steps. We take the most frequent 1,000 words in the OCD dataset for embedding purposes and then make a dictionary with all these words.

Furthermore, we have 300 values for each one. Later, every word in the dataset is embedded, and we find its meaning through each dimension. As a result, we obtain vectors of 300 real values as a feature. We used the most frequent 1,000 words in the OCD dataset for word embedding; 983 words were embedded, with 300 values for each, and only 17 words were not found in the dictionary.

The hyperparameter tuning process systematically explored the key parameters for each classifier. In the case of k-Nearest Neighbors (KNN), we used grid search to optimize hyperparameters. Specifically, we set k=5 to avoid obtaining an equal voting result, thus enhancing the discriminative power of the classifier. For Decision Trees (DT), hyperparameters were adjusted using grid search, with the maximum depth set to 17 and the Gini index employed for attribute selection. Regarding Random Forest (RF), grid search was applied with a maximum depth of 7 and the number of estimators set to 500, ensuring a balance between model complexity and generalization. We referred to papers addressing similar problems, such as the work by Azizi [38], to explore combinations of hyperparameters.

Table VII discusses the OCD detection performance results with each classifier. After applying ML classifiers to our dataset, the results are depicted in Fig. 12 using the F1-Score. The best results are obtained using fastText with RF. Also, more than twenty features were extracted and implemented, as mentioned in Table IV. The three classifiers used for this purpose were RF, KNN, and DT. The classification was performed using k-fold cross-validation, and the results were analyzed using precision, recall, and the F1-Score.

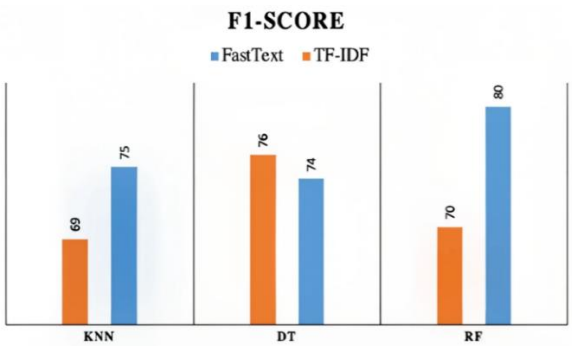

Fig. 12. Comparison of classifiers using F1-Score.

In this study, we utilized a 10-fold cross-validation technique to assess the effectiveness of different classifiers employing TF-IDF and fastText methods. The dataset was divided into ten equal parts, nine for training and one for testing in each iteration. Precision, recall, and F1-Score were calculated for every classifier and method combination across all ten folds. The resulting mean values of these metrics were then analyzed to identify discernible trends and patterns. Table VII presents the mean precision, recall, and F1-Score values for each classifier and method combination across all 10 folds.

TABLE VII. MEAN PERFORMANCE METRICS ACROSS FOLDS FOR DIFFERENT CLASSIFIERS

Metric	Classifier	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	AVG
Precision	RF (fastText)	0.80	0.78	0.76	0.79	0.81	0.75	0.77	0.80	0.76	0.75	0.80
	RF(TF-IDF)	0.82	0.79	0.85	0.81	0.83	0.80	0.84	0.79	0.82	0.85	**0.81**
	DT (fastTex)	0.76	0.74	0.72	0.77	0.75	0.77	0.73	0.76	0.74	0.72	0.74
	DT(TF-IDF)	0.79	0.77	0.81	0.78	0.76	0.82	0.75	0.77	0.79	0.80	0.79
	KNN (fastText)	0.76	0.74	0.72	0.75	0.73	0.75	0.71	0.73	0.72	0.74	0.74
	KNN(TF-IDF)	0.78	0.76	0.79	0.77	0.75	0.80	0.74	0.77	0.79	0.80	0.78
Recall	RF (fastText)	0.76	0.74	0.74	0.77	0.75	0.78	0.77	0.76	0.74	0.72	**0.80**
	RF(TF-IDF)	0.73	0.70	0.75	0.72	0.74	0.71	0.76	0.70	0.73	0.75	0.72
	DT (fastTex)	0.74	0.72	0.70	0.76	0.74	0.77	0.72	0.75	0.73	0.70	0.74
	DT (TF-IDF)	0.77	0.74	0.72	0.77	0.75	0.78	0.73	0.76	0.74	0.72	0.77
	KNN (fastText)	0.75	0.73	0.75	0.71	0.70	0.74	0.75	0.73	0.72	0.75	0.75
	KNN (TF-IDF)	0.72	0.70	0.74	0.72	0.68	0.75	0.70	0.73	0.72	0.74	0.72
F1-Score	RF (fastText)	0.80	0.77	0.78	0.76	0.83	0.77	0.79	0.75	0.82	0.80	**0.80**
	RF (TF-IDF)	0.70	0.71	0.70	0.70	0.69	0.70	0.68	0.71	0.69	0.73	0.70
	DT (fastTex)	0.74	0.72	0.70	0.75	0.73	0.76	0.72	0.75	0.73	0.70	0.74
	DT (TF-IDF)	0.76	0.73	0.72	0.76	0.74	0.77	0.72	0.75	0.73	0.70	0.76
	KNN (fastText)	0.75	0.73	0.71	0.75	0.73	0.75	0.77	0.77	0.72	0.75	0.75
	KNN (TF-IDF)	0.71	0.68	0.72	0.70	0.66	0.73	0.66	0.70	0.71	0.72	0.69

Table VIII illustrates the results produced by the classifiers while using 10-fold cross-validation in terms of mean values of precision, recall, and F-measure. Generally, RF provides the highest values for the F1-Score, reaching 0.80 with fastText. However, the lowest score was in KNN using TF-IDF compared to all classifiers. In fastText, RF has a higher F1-Score than TF-IDF by 10%, and in KNN, it is 0.75 with fastText. In DT, we obtained close results while using both methods. As a result, the performance of the F1-Score improved in RF and KNN using the word embedding technique.

TABLE VIII. COMPARISON OF CLASSIFIERS BASED ON THE MEAN VALUES OF PRECISION, RECALL, AND F1-SCORE

Classifiers	Method	Precision	Recall	F1-Score
RF	TF-IDF	0.81	0.72	0.70
	fastText	0.80	0.80	**0.80**
DT	TF-IDF	0.79	0.77	0.76
	fastText	0.74	0.74	0.74
KNN	TF-IDF	0.78	0.71	0.69
	fastText	0.76	0.75	0.75

The mean precision and recall in TF-IDF are higher when using RF, with 0.81 for precision and lower for recall 0.72, compared with fastText, which obtained values of 0.80 for both methods. Similarly, DT achieved higher results using TF-IDF with 0.79 for precision and 0.77 for recall, as it is 0.74 for precision and recall when using fastText. However, the recall result was higher in KNN when using fastText (0.75) than in TF-IDF, which obtained 0.71. Lastly, KNN received a higher score in precision with 0.78 in TF-IDF and a lower score using fastText with 0.76.

Overall, our study contributes to mental health detection through social media by effectively addressing a recognized gap. The deficiency identified pertains to the absence of annotations tailored explicitly for the detection of mental disorders, particularly OCD, within many datasets. We have meticulously curated a dataset from Twitter, focusing on OCD symptoms, and subjected it to rigorous validation by experts, thereby significantly augmenting its reliability.

In contrast to prior research, which may have overlooked the criticality of validated annotations by specialists, our methodology places paramount importance on ensuring the meticulous validation of annotations associated with OCD symptoms. This emphasis elevates the credibility of our dataset. Furthermore, we surpass conventional feature engineering by integrating advanced word embedding techniques, thereby refining the accuracy of our machine learning classifiers.

Our study not only fills the identified gap but also makes a scholarly contribution to the existing body of research in mental health detection through social media. The introduction of a novel methodology aligns precisely with the precision required for detecting OCD, thereby providing a more nuanced and comprehensive understanding of social media's role in the assessment of mental health.

Recent research has used word embedding in the Arabic language [39], and it has proved that it improves performance. Our study also supports the fact that word embedding, such as fastText, enhances the performance in Arabic texts.

Our study breaks new ground by being the first to employ artificial intelligence to detect Obsessive-Compulsive Disorder (OCD) in Arabic tweets, with a distinct focus on social media. While existing studies have explored OCD, the novelty of our work lies in unveiling and understanding OCD expressions within the dynamic landscape of social media platforms.

(1) Our methodology draws inspiration from related studies, yet the crux of our contribution lies in addressing a gap in existing literature. We are the first to explore and detect OCD within Arabic tweets, making a significant contribution by introducing this novel concept to mental health research on social media using ML.
(2) Uncovering COVID-19's Influence on OCD Expressions: Focused on Arabic tweets, our study delves into the cultural intricacies of OCD expressions on social media, especially considering the context of the COVID-19 pandemic. We analyze how the ongoing health crisis has impacted the manifestation of OCD symptoms, offering unique insights into the intersection of OCD and the pandemic within the Arabic-speaking community.
(3) FastText's Role in Social Media Language Processing: Our analysis highlights fastText's critical role as a tailored word embedding technique for the Arabic language in social media, showcasing its efficiency and potential for future mental health studies in this context.

Ultimately, it is essential to mention that while our study specifically analyzes tweets for the presence of OCD symptoms and does not classify or confirm individuals as having OCD, we acknowledge the importance of addressing potential biases in the dataset and the implications of using social media content for mental health analysis.

- Biases Present in the Dataset: We recognize that our dataset is derived from social media, so it may not fully represent the broader population. Social media users may exhibit certain biases regarding age, gender, socioeconomic status, and cultural background. Acknowledging these limitations is crucial, as they can impact the generalizability of our findings.
- Ethical Reflections: Given that our study analyzes publicly available social media content, we will discuss the ethical considerations surrounding privacy and consent. While tweets are anonymized, we understand the importance of addressing ethical concerns related to potentially sensitive information.
- Expert Labeling Process: To ensure the reliability of our results, we engaged experts to label tweets indicating OCD symptoms manually. However, it is essential to acknowledge that the interpretation of mental health-related content can be subjective, and different experts may have varying opinions. To ensure the reliability and consistency of our results, we maintained a rigorous process, including discussions among the experts to establish clear criteria for labeling. This process helps minimize potential biases arising from the subjective nature of interpreting tweets for mental health symptoms.

VIII. Limitations

In this paper, we present an Arabic dataset for detecting OCD symptoms. However, there are certain limitations. One limitation is that our dataset only covers some Arabic countries, impacting word embedding due to the absence of some words. Since our dataset is randomly chosen based on relevant keywords, it does not encompass all dialects. While utilizing fastText for word embedding with the most frequent 1,000 words in the OCD dataset, a limitation surfaced. Of the selected words, 983 were successfully embedded, leaving 17 absent from the dictionary. We propose developing a pre-trained model tailored to the mental health domain to address this issue. This model could encompass a broader vocabulary, ensuring comprehensive coverage for future studies. We will delve deeper into this limitation once we discuss future work, emphasizing the potential benefits of a specialized pre-trained model for improved word embedding in mental health contexts. Notably, our work does not aim to replace visits to specialized hospitals. It is designed solely to detect symptoms suggestive of possible OCD and cannot aid in diagnosing the disease.

IX. Ethical Considerations

A. Anonymous Tweet Analysis

Our study exclusively focuses on the analysis of anonymous tweets related to Obsessive-Compulsive Disorder (OCD) symptoms. To construct our tweet-based dataset, we employed a symptom-specific retrieval approach. This dataset underwent rigorous NLP text preprocessing for anonymization. This process systematically removed identifiable information such as user accounts, User Mentions and Tags, URLs and Hyperlinks, and Hashtags. The result is a dataset consisting solely of anonymized text, ensuring the confidentiality and privacy of individuals sharing their experiences.

B. Publication Report from ERADA Mental Health Hospital

ERADA Mental Health Hospital issued a supportive publication report following a meticulous review, affirming the ethical validity of our research. This endorsement enhances the credibility of our findings, attesting to our commitment to high ethical standards in mental health research.

X. Conclusion and Future work

OCD is a widespread, chronic, and long-term disorder in which a person experiences uncontrollable, recurring thoughts ("obsessions") and/or activities ("compulsions") that they feel compelled to repeat. Identifying OCD symptoms may lead to consultation for medical diagnosis and treatment. Therefore, this work created a new Arabic Twitter dataset for OCD detection to determine whether a tweet indicates OCD symptoms. We annotated the dataset for the COVID-19 vaccination period with the help of the Erada mental health specialists. The Twitter dataset was annotated by extracting OCD-related features using Tweepy. Then, a comprehensive analysis was conducted on our dataset, starting with regular tweets, texts only, tweets with emoji, and tweets that were exclusively composed of emoji, along with preprocessing texts and

using the latest technique for visualization. This paper also showed how word embedding using fastText improved the F1-score after applying ML classifiers such as KNN and RF compared with TF-IDF. However, the decision tree algorithm received the same F1-score with both methods.

In future work, exploring the development and utilization of a domain-specific pre-trained model in mental health can offer enhanced benefits. Beyond addressing the limitation in word embedding coverage, such a model can improve accuracy, sensitivity, and contextual relevance in detecting mental health-related discussions on social media. This can lead to more robust and nuanced analyses, contributing to advancements in mental health research and aiding in the identification of subtle linguistic cues indicative of various mental health conditions. Furthermore, we plan to use more data along with deep learning algorithms. Moreover, the researchers can extend this work to other related mental health issues.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

M. F. A. (Malak Fahad Alhaider), A. M. Q. (Ali Mustafa Qamar), H. S. A. (Hasan Shojaa Alkahtani), and H. F. A. (Hafiz Farooq Ahmad) carried out Conceptualization, Literature review, and methodology; M. F. A., A. M. Q., and H. F. A. performed exploratory data analysis; M. F. A. and A. M. Q. conducted feature extraction; M. F. A., A. M. Q., H. S. A., and H. F. A. performed machine learning classifiers' implementation, results and discussion; all authors had approved the final version.

REFERENCES

[1] A. G. Guzick, A. Candelari, A. D. Wiese *et al.*, "Obsessive–compulsive disorder during the COVID-19 pandemic: A systematic review," *Current Psychiatry Reports*, vol. 23, no. 11, 2021. https://doi.org/10.1007/s11920-021-01284-2

[2] C. F. Sharpley, "Clinical handbook of psychological disorders: A step-by-step treatment manual," *Behaviour Change*, vol. 5, no. 3, pp. 139–140, 1988.

[3] K. A. Kobak, J. H. Greist, J. W. Jefferson *et al.*, "Behavioral versus pharmacological treatments of obsessive-compulsive disorder: A meta-analysis," *Psychopharmacology*, vol. 136, no. 3, pp. 205–216, 1998.

[4] A. Okasha, "Diagnosis of obsessive-compulsive disorder: A review," *Obsessive-Compulsive Disorder*, pp. 1–41, 2001.

[5] A. Abba-Aji, D. Li, M. Hrabok *et al.*, "COVID-19 pandemic and mental health: Prevalence and correlates of new-onset obsessive-compulsive symptoms in a Canadian province," *Int. J. Environ. Res. Public Health*, vol. 17, 6986, 2020.

[6] J. H. Kim, "Social media use and well-being," *Subjective Well-Being and Life Satisfaction*, pp. 253–271, 2017.

[7] A. Whiting and D. Williams, "Why people use social media: A uses and gratifications approach," *Qualitative Market Research: An International Journal*, vol. 16, no. 4, pp. 362–369, 2013.

[8] G. K. Shahi, A. Dirkson, and T. A. Majchrzak, "An exploratory study of COVID-19 misinformation on Twitter," *Online Social Networks and Media*, vol. 22, 100104, 2021.

[9] J. Shuja, E. Alanazi, W. Alasmary *et al.*, "COVID-19 open-source data sets: A comprehensive survey," *Applied Intelligence*, vol. 51, no. 3, pp. 1296–1325, 2020.

[10] M. Roy, N. Moreau, C. Rousseau *et al.*, "Ebola and localized blame on social media: Analysis of Twitter and Facebook conversations during the 2014–2015 Ebola epidemic," *Culture, Medicine, and Psychiatry*, vol. 44, no. 1, pp. 56–79, 2020.

[11] I. Kagashe, Z. Yan, and I. Suheryani, "Enhancing seasonal influenza surveillance: Topic analysis of widely used medicinal drugs using Twitter data," *Journal of Medical Internet Research*, vol. 19, no. 9, 2017.

[12] D. A. Alateeq, H. N. Almughera, T. N. Almughera *et al.*, "The impact of the coronavirus (COVID-19) pandemic on the development of obsessive-compulsive symptoms in Saudi Arabia," *Saudi Medical Journal*, vol. 42, no. 7, pp. 750–760, 2021. doi: 10.15537/smj.2021.42.7.20210181

[13] L. Pan, J. Wang, X. Wang *et al.*, "Prevention and control of Coronavirus Disease 2019 (COVID-19) in public places," *Environmental Pollution*, vol. 292, 118273, 2022.

[14] M. M. Tadesse, H. Lin, B. Xu *et al.*, "Detection of depression-related posts in reddit social media forum," *IEEE Access*, vol. 7, pp. 44883–44893, 2019.

[15] N. P. Shetty, B. Muniyal, A. Anand *et al.*, "Predicting depression using deep learning and ensemble algorithms on raw twitter data," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, 3751, 2020.

[16] Y. Zhang, S. Wang, A. Hermann *et al.*, "Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women," *Journal of Affective Disorders*, vol. 279, pp. 1–8, 2021.

[17] M. Qorib, T. Oladunni, M. Denis *et al.*, "COVID-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Systems with Applications*, vol. 212, 118715, 2023.

[18] G. Gkotsis, A. Oellrich, T. Hubbard *et al.*, "The language of mental health problems in social media," in *Proc. Third Workshop on Computational Linguistics and Clinical Psychology*, 2016.

[19] A. B. R. Shatte, D. M. Hutchinson, M. Fuller-Tyszkiewicz *et al.*, "Social media markers to identify fathers at risk of postpartum depression: A machine learning approach," *Cyberpsychology, Behavior, and Social Networking*, vol. 23, no. 9, pp. 611–618, 2020.

[20] M. R. Islam, A. R. M. Kamal, N. Sultana *et al.*, "Detecting depression using K-Nearest Neighbors (KNN) classification technique," in *Proc. International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, 2018.

[21] N. A. Baghdadi, A. Malki, H. M. Balaha *et al.*, "An optimized deep learning approach for suicide detection through Arabic tweets," *PeerJ Computer Science*, vol. 8, 2022.

[22] S. Hindujam, M. Afrin, S. Mistry *et al.*, "Machine learning-based proactive social-sensor service for mental health monitoring using Twitter data," *International Journal of Information Management Data Insights*, vol. 2, no. 2, 100113, 2022.

[23] N. V. Babu and E. G. Kanaga, "Sentiment analysis in social media data for depression detection using artificial intelligence: A review," *SN Computer Science*, vol. 3, no. 1, 2021.

[24] A. Leis, F. Ronzano, M. A. Mayer *et al.*, "Detecting signs of depression in tweets in Spanish: Behavioral and linguistic analysis," *Journal of Medical Internet Research*, vol. 21, no. 6, 2019.

[25] K. Spates, X. Ye, and A. Johnson, "'I just might kill myself': Suicide expressions on Twitter," *Death Studies*, vol. 44, no. 3, pp. 189–194, 2018.

[26] U. Naseem, I. Razzak, M. Khushi *et al.*, "Covidsenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 1003–1015, 2021.

[27] S. Bhatia, M. Alhaider, and M. Alarjani, "Sentiment analysis for Arabic tweets on COVID-19 using computational techniques," in *Proc. 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2022.

[28] I. Li, Y. Li, T. Li *et al.*, "What are we depressed about when we talk about COVID-19: Mental health analysis on tweets using natural language processing," *Lecture Notes in Computer Science*, pp. 358–370, 2020.

[29] A. Al-Laith and M. Alenezi, "Monitoring people's emotions and symptoms from Arabic tweets during the COVID-19 pandemic," *Information*, vol. 12, no. 2, 86, 2021.

[30] Y. C. Yang, A. Xie, S. Kim *et al.*, "Automatic detection of Twitter users who express chronic stress experiences via supervised machine learning and natural language processing," *CIN: Computers, Informatics, Nursing*, vol. 41, no. 9, pp. 717–724, 2023.

[31] M. Aljabri, S. M. B. Chrouf, N. A. Alzahrani *et al.*, "Sentiment analysis of Arabic tweets regarding distance learning in Saudi Arabia during the COVID-19 pandemic," *Sensors*, vol. 21, no. 16, 5431, 2021.
[32] A. Al Badawi, L. Huang, C. F. Mun *et al.*, "Privft: Private and fast text classification with Homomorphic encryption," *IEEE Access*, vol. 8, pp. 226544–226556, 2020.
[33] R. Bouchlaghem, A. Elkhelifi, and R. Faiz, "A machine learning approach for classifying sentiments in Arabic tweets," in *Proc. 6th International Conference on Web Intelligence, Mining and Semantics*, June 2016, pp. 1–6.
[34] A. I. Kadhim, "Term weighting for feature extraction on Twitter: A comparison between BM25 and TF-IDF," in *Proc. International Conference on Advanced Science and Engineering (ICOASE)*, 2019.
[35] N. K. Bolbol and A. Y. Maghari, "Sentiment analysis of Arabic tweets using supervised machine learning," in *Proc. International Conference on Promising Electronic Technologies (ICPET)*, December 2020, pp. 89–93.
[36] S. M. Alzanin, A. M. Azmi, H. A. Aboalsamh *et al.*, "Short text classification for Arabic social media tweets," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 6595–6604, 2022.
[37] H. Elfaik and E. H. Nfaoui, "A comparative evaluation of classification algorithms for sentiment analysis using word embeddings," in *Proc. Advanced Intelligent Systems for Sustainable Development (AI2SD'2019) Volume 4 - Advanced Intelligent Systems for Applied Computing Sciences*, *Springer*, 2020, pp. 1–11.
[38] M. Azizi, "A comparison of machine learning techniques to classify tweets relevant to people impacted by dementia and COVID-19," M.S. thesis, Dept. Comp. Sci., University of Saskatchewan, Saskatoon, Canada, 2022.
[39] A. Alwehaibi and K. Roy, "Comparison of pre-trained word vectors for Arabic text classification using deep learning approach," in *Proc. 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018.