IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 24, 2020, accepted March 15, 2020, date of publication March 18, 2020, date of current version April 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2981834

Predicting Depression Symptoms in an Arabic Psychological Forum

NORAH SALEH ALGHAMDI[1], HANAN A. HOSNI MAHMOUD[1,2], AJITH ABRAHAM[3], (Senior Member, IEEE), SAMAR AWADH ALANAZI[4], AND LAURA GARCÍA-HERNÁNDEZ[5]

[1]College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[2]Faculty of Engineering, Department of Computer and Systems Engineering, University of Alexandria, Alexandria 215216, Egypt
[3]Machine Intelligence Research Labs (MIR Labs), Washington, WA 98071, USA
[4]Autism Centre, Prince Sultan Military Medical City, Riyadh 11159, Saudi Arabia
[5]Area of Project Engineering, University of Córdoba, 14071 Córdoba, Spain

Corresponding author: Norah Saleh Alghamdi (nosalghamdi@pnu.edu.sa)

This work was supported by the Deanship of Scientific Research, Princess Nourah bint Abdulrahman University, through the Program Project Funding After Publication, under Grant 41 – PRFA-P-3.

ABSTRACT Recently, social media platforms have been widely used as a communication tool on social networks. Many users have utilized these platforms to reflect their personal lives. These users differ in terms of background, language, age, and educational level. The close relationship between these platforms and their users has created rich information that is related to these users and can be exploited by researchers. Their posts can be analysed using natural language processing (NLP) to predict psychological traits such as depression. However, to the best of our knowledge, no study has utilized social media to predict mental health disorders in Arabic posts, especially depression. Therefore, in this study, we investigate the application of natural language processing and machine learning on Arabic text for the prediction of depression, and we evaluate and compare the performance. Our research method is based on the collection of Arabic text from online forums and the application of either a lexicon-based approach or a machine-learning-based approach. In the former approach, the ArabDep lexicon is created, and a rule-based algorithm is used to predict depression symptoms using the created lexicon; however, in the latter approach, the data are annotated with the help of a psychologist, text features are extracted from Arabic posts, and machine learning algorithms are ultimately applied to predict depression symptoms. We demonstrate that our applied approaches exhibit promising performance in predicting whether a post corresponds to depression symptoms, with an accuracy of more than 80%, a recall of 82% and a precision of 79%.

INDEX TERMS Supervised learning, semi-supervised learning, machine learning, predictive models, depression, lexicon, text analysis.

I. INTRODUCTION

Globally, humans have been supported by merging the power of information technology with their lifestyles. The digital transformation in the health sector has contributed to the enhancement of people's lifestyles and health conditions. Recently, many information technologies have been developed in the health sector for curing illnesses and improving the quality of life. One aspect of the development of information technology in the health sector is the utilization of social media platforms for the realization of a better and healthier life. Mental health monitoring is critical for improving people's lifestyles and increasing human productivity. We consider the standard definition of mental health disorders: disorders that affect the mood, thinking and behaviours [1]. Depression is a mental health disorder that changes a person's mood, thoughts and behaviours. It also affects his body. This disorder causes various physical problems, such as fluctuation of energy, appetite, and sleep patterns. It has been described as a disease with various diagnostic criteria and typically co-exists with other physical or psychological disorders [2]. According to the WHO, depression affects more are 350 million people around the world, which is equivalent to 5% of the global population.

The associate editor coordinating the review of this manuscript and approving it for publication was Nilanjan Dey.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

"The Diagnostic and Statistical Manual of Mental Disorders (DSM) suggests that clinical depression can be diagnosed through the presence of a set of symptoms over a fixed period of time" [3].

The use of social media to predict mental disorders will yield significant improvements in the people's lifestyles and health conditions. Substantial amounts of health information and practices are sought or shared online. This available online information is effectively helping people coping with disorders. Studies have shown that online forums and support groups offer beneficial environments in which people can become linked with others who have similar disorders, pains, or conditions. Additionally, the main advantageous features of these health communities are honesty and self-disclosure. Their users can uninhibitedly discuss feelings or socially unacceptable thoughts. These features are essential therapeutic ingredients [4]. Such online platforms quickly generate data, which must be investigated to discover valuable insights. However, these data require advance analysis tools such as natural language processing (NLP), data mining, and machine learning for the extraction of valuable information, e.g., for detecting or predicting depression. There are many approaches for predicting mental disorders in English-language platforms, which effectively contribute to making therapy more efficient [5]–[8]. References [5], [7], [8] used electroencephalogram (EEG) data to discriminate between cases with versus without depression and applied various processing approaches. Reference [5] classified the selected features, which include the Bhattacharya distance and the t-test result, using logistic regression (LR), support vector machine (SVM) and naïve Bayes (NB) classifiers with 10-fold cross-validation. Reference [7] utilized forehead EEG in conjunction with sensors and algorithms to process the generated signals. Reference [8] collected EEG data from a resting state paradigm and applied SVM, NB, decision tree (DT) and K-nearest neighbour (KNN). Reference [6] utilized NLP to detect depression from social media by creating a lexicon of terms that are commonly used by depressed users. They used the Linguistic Inquiry and Word Count dictionary (LIWC) [25] for feature extraction and latent Dirichlet allocation (LDA) for topic analysis. N-gram modelling is used for word analysis. Various ML classifiers are applied, which include LR, SVM, RF, MLP, and AdaBoost. However, to the best of our knowledge, no research has been conducted on Arabic predictive platforms for mental disorders, especially for depression. Therefore, this study aims at closing this research gap by proposing analytical approaches for the prediction of depression from online Arabic text.

The manipulation of the Arabic language is challenging because it has rich morphological structures in which each word may have many forms and meanings. In addition, the Arabic language has a formal version, which varies from the informal version that is used in every-day speaking. These differences are morphological and syntactical [9]. Most of the contents of social media are written in the informal language, which does not have a standard structure or orthographies. Most available natural language processing solutions for Arabic content are designed for the formal language [10] and exhibit low performance when processing the informal language [11]. In addition to these challenges of Arabic text, no single natural language processing solution can process all the variations of Arabic text problems [12]. Some of the solutions focus on sentiment analysis [13]–[16], while others analyse the Arabic text to understand emotions [17], [18], [11]. No studies have investigated the area of mental health disorders, especially depression.

Our main contributions in this paper are as follows:

1) This paper proposes the use of an Arabic online forum for the identification and prediction of depression symptoms from its posts. We focus mainly on an online platform on which users discuss several mental health disorders and that has a community for supporting depressed users.
2) We collected the ArabDep corpus and labelled the instances based on the category of the forum and via manual annotation by a psychologist who identified the depression symptoms based on the guidelines that were derived from assessment manual DSM-5 [19]. To the best of our knowledge, no one has attempted to build a such a domain-specific dataset.
3) We create the ArabDep lexicon, which is related to Arabic and depression-related words. Three approaches have been used to create the lexicon based on a scoring system, which is used to explore the identification of the depression-related words that are shared by online forum users in their posts. The challenge of identifying correlations between words within a post and depression leads us to exploit the rule-based algorithm, which utilizes the built lexicon to derive valuable insights for the prediction of depression cases from posts. The performance of the lexicon-based method is evaluated by considering the results of the confusion matrix.
4) We evaluated various machine learning models and feature extraction methods, such as n-gram and word embeddings in a real-world textual context, and we compared their performances on predicting depression symptoms.
5) We conducted an experimental evaluation of lexicon-based and machine-learning-based approaches and reported their performances. These experiments offer an initial set of prediction solutions of depression, which may serve as baseline results for further studies of the Arabic language. We demonstrate that our proposed approaches show promise in predicting whether a post corresponds to depression symptoms or not, with accuracy that exceeds 80%, a recall of 82% and a precision of 79% for the lexicon-based approach and with an accuracy that exceeds 73%, a recall of 72% and a precision of 74% for the machine-learning-based approach.

The remainder of this paper is organized as follows: Section II reviews previous works, Section III presents the proposed methodology and describes how the data are collected, labelled and processed using either the lexicon-based approach or the machine-learning-based approach, Section IV evaluates the performances of the proposed methods. Finally, the conclusions of this study are presented in Section V.

II. RELATED WORK

Various studies have been conducted on the prediction and monitoring of mental disorders via natural language analysis and on the search for various features and clues of the languages for the identification of the sentiments of posts. One of the main mental disorders is depression, and various related data have been collected in the literature for the prediction of depression via various approaches, such as using self-assessment questionnaires, self-declaration of diagnosis, membership to specialized online forums and manually/automatically annotated posts.

A. PREDICTING DEPRESSION BASED ON SELF-ASSESSMENT QUESTIONNAIRES

Researchers have collected data from Facebook or Twitter from recruited participants who answer self-assessment questionnaires. Survey is regarded as the second approach after clinical interviews in psychological research. Common self-assessments include the Patient Health Questionnaire (PHQ) [20], the Center for Epidemiologic Studies Depression Scale-Revised (CES-D) [21] and Beck's Depression Inventory (BDI) [22]. Reference [23] collected answers from participants to a depression self-scale assessment, in addition to their posts on Twitter. Depression was detected when participants reported its occurrence. The authors also collected the scores on CES-D and BDI. Behavioural characteristics are associated with social actions, such sharing emotions, involvement, speaking and mentioning antidepressant pills. They build a statistical classifier that offers early approximations of the risk of depression. The study found numerous differences in the collection activities by depressed participants, such as more negative feelings, less interaction, higher degree of self-focus, and the writing of depressing terms. The authors in [24] also utilized the behavioural characteristics on Facebook. They aim at predicting postpartum depression (PPD) in mothers who self-reported their status and conducted a self-assessment survey. The emotions and linguistic cues were measured using [25]. They used statistics and t-tests to characterize and distinguish the mothers' behaviours in the prenatal and postpartum period. Reference [26] conducted a study on the prediction of depression and post-traumatic stress disorder (PTSD) status from tweets. The authors aggregate the data into weeks and days. Various tools for extracting sentiment features, such as labMT, LIWC 2007, and ANEW, were used to measure the happiness status from linguistics of tweets. They trained random forest classifiers to distinguish between users with mental illness and healthy users. Reference [27] used data from Twitter to predict depression from Japanese text. Participants used CES-D to self-assess their depression status. The authors showed that a short period (6 to 16 weeks) of collecting the participants' tweets prior to conducting the CES-D survey was adequate for increasing the accuracy of depression identification. Various features were extracted from the users' activities and the word frequencies in the tweets. MeCab [28] was utilized as morphological stemmer of the Japanese language to accurately acquire frequencies of words. The text analysis was based on the tweet topics per user and was conducted using the topic model LDA [29]. Reference [30] predicted the continuous depression degree of participants on Facebook. Personal surveys of the participants were conducted. The authors estimated the depression degree of each user as the average of the replies to seven facet items of depression within neuroticism [31]. Features are measured including n-grams, topics, lexica and word numbers. A list of words, phrases and topics that are related to depression was provided. The seasonal variations of depression were studied, and it was found that participants tend to have high depression degrees during winter. To acquire a continuous result, various experiments were conducted using regression models during the training phase.

B. PREDICTING DEPRESSION BASED ON SELF-DECLARATION OF DIAGNOSIS

Other studies use public data for users who have "self-declared" their mental disorder diagnoses on public platforms such as Twitter. An example of a self-declaration statement is "I was diagnosed with depression". The objective of the CLPsych 2015 workshop [32] was to investigate and evaluate various solutions to the specified problem for disorders such as depression and PTSD using the same data. Reference [33] generated decision lists based on the presence of N-grams within tweets of users who had revealed their diagnosis of either PTSD or depression. The likelihoods of the test users suffering from PTSD or depression are computed and used to rank the users. The ranking is based on the number of N-grams in each user's tweets. This approach used only the text of tweets and ignored other incorporated data, such as location data. Reference [34] predicted anxiety with consideration of gender, in addition to ten common conditions. This study utilized multi-task learning to predict anxiety from Twitter and compared the performance with that of single-task learning, which is a deep learning network with two hidden layers for independent training for each prediction task. In contrast, in multi-tasking learning, one of the hidden layers is shared among all processed tasks. It's modelling has a flexible design and it can change from the task-agnostic format to a task-specific format. A non-linear logistic function is utilized for the output layer due to the manipulation of binary prediction tasks. Reference [35] acquired data from Twitter that were based on user self-reports regarding their health conditions. This data collection approach was evaluated in various studies, such as [36], and it yielded strong predictive results for practical cases. The Twitter users declared their diagnoses by stating "I have been diagnosed with X", where X denotes one of ten mental conditions. The authors of [35] utilized ''Linguistic Inquiry and Word Count'', which is a psychological dictionary [25], to measure the differences in the characteristics of mental conditions. They calculated the word tokens that fall into a specified category of the dictionary for comparison across all the selected mental conditions.

C. PREDICTING DEPRESSION BASED ON FORUM SUBSCRIPTION

Subscription to online mental health discussion forums is regarded as an important source of public data that are associated with mental health conditions. These forums provide spaces for discussing, asking for advice, and receiving emotional support for all topics that are related to mental health issues. Reference [37] utilized online content that was generated by users to detect depression and to identify the severity degree. The user-generated posts and their mood tags are used to distinguish the linguistic styles and the post sentiments. The authors collected data from LiveJournal using ''search communities by interest''. LIWC features [25] were utilized to differentiate between communities. A combination of statistical analysis and machine learning approaches are used to differentiate the depressive content from the non-depressive content. To distinguish the posts that correspond to depression from the remaining data, various algorithms are used, such as RELIEFF [38], [39] and RFs [40]. The degree of depression is classified as severe, moderate or mild. The severity of depression is identified by examining the valence values of the mood tags by using the ANEW lexicon [41] to represent the tags of the mood as the depression degree. To realize this objective, a hierarchical hidden Markov model (HMM) [42], [43] is implemented on the generated values and classification labels. Another study collected data from reddit forum posts [43], [45] and used the data to study mental health issues. In [44], the authors narrowed their study to students of U.S. universities. The training data were collected from a reddit community for supporting mental health. The prediction model was used to evaluate the estimated degree of distress of a hundred and nine university subreddits. Their approach is based on inductive transfer learning [46]. In a binary classification framework, they trained and tested various classifiers, such as RF, Ada Boost, SVM and LR. As a result, the author found that the percentage of posts about mental health concerns increased over courses, especially for quarter-based academic year schedules compared to semester-based academic year schedules. In [45], the author analysed reddit posts of a user group that posted about mental health issues and changed to posting about suicidal ideation. A statistical approach was applied to predict which user will change from talking about mental health concerns to discussing suicidal ideation. Various sets of features were extracted and used to predict this transition based on three measures: ''linguistic structure'', ''interpersonal awareness'' and ''interaction''. They created five prediction models. Several observations for users, who shifted from mental health posts to suicidal ideation posts, were derived as a result of this study: self-focus, poor linguistic structure, less social engagement, anxiety, and loneliness.

D. PREDICTION USING ANNOTATED ONLINE PUBLIC POSTS

The annotation of publicly accessible posts, such as Twitter posts, is conducted manually to examine and annotate posts that contain keywords that are related to mental health issues. Based on pre-defined theories, the collected public posts are coded by annotators or by using a bottom-up classification method from the post language [47], [48]. In [47], the authors used the annotation approach to differentiate between the language of posts that include mental illness keywords for stigmatization and the language of posts that include mental illness keywords for supporting those who are suffering from mental health issues by sharing valuable information about mental health. The annotation for depression-related studies emphasizes the recognition of posts of users who share their depression experiences, such in [49], [50]. The guidelines that the annotators followed in identifying the depression symptoms are derived from known assessment manuals such as DSM-5 [19]. The study in [51] reduced the number of symptoms by conducting feature experiments that were based on ''feature ablation'' and ''feature elimination''. They aimed at identifying the top-ranked features that contribute to the performance optimization of the classification models. Various machine learning approaches have been proposed for classification for the discovery of information rapidly, accurately and comprehensively [52]–[56].

Although the annotation of online posts is labour-intensive, hidden factors that are related to mental health issues can be inferred, such as weather, which could not be captured by known diagnostic criteria [50], [19]. In addition, additional costs are incurred with the self-assessment questionnaire method, even though it facilitates the collection of the most reliable data for the prediction models. To avoid these costs, additional methods have been proposed for using publicly available data to predict mental health issues. Furthermore, by reviewing the literature, we found that this type of problem has not been studied for Arabic text. In addition, the manipulation of Arabic text differs from that of English text and is more complicated. We will conduct feature extractions with various machine learning methods and a lexicon-based method for the prediction of depression symptoms. We will create our own corpus and predict via a hybrid approach on data from an online forum that have been annotated manually and automatically. Table 1 compares previous studies.

III. PROPOSED METHODOLOGY

In this section, we will introduce our proposed methodology. First, the collected dataset is described in detail. Then, the created corpus is presented, which is followed by a description of the data pre-processing stage. Second, we will introduce our method for predicting depression from Arabic posts using the lexicon-based approach or the machine-learning-based

TABLE 1. Summary of previous studies.

R	Data Source	Language	Prediction based on	Prediction method
[23]	Twitter	English	Questionnaires	Statistics
[24]	Facebook	English	Questionnaires	Statistics: t-tests
[26]	Twitter	English	Questionnaires	Random forest
[27]	Twitter	Japanese	Questionnaires	LDA
[33]	CLPsych 2015	English	Self-declaration	Machine Learning
[34]	Shared Task	English	Self-declaration	Deep Learning
[37]	Forum	English	Membership	RELIEFF, RFs, Markov Model

approach, and we will evaluate the performances of the applied approaches.

A. PREPARATION OF THE ARABDEP CORPUS

Studies have been conducted on the identification of various mental health disorders from public social media contents that are based on understanding the language differences between posts of users who suffer from a specified mental health disorder and posts of a control group of unaffected users in various languages, such as English, Chinese and Japanese [23], [27]. To the best of our knowledge, no attempt has been made to create a corpus for the analysis of depression from Arabic posts. Although several Arabic corpora have been created for NLP, such as in [10], [52], none target depression issues. In this study, we collect data to support the prediction of depression symptoms.

1) DATA COLLECTION AND ANNOTATION

We have sufficiently searched various types of publicly available online content as our sources to create a corpus for research on the use of the Arabic language in depression cases. Within this stage, our main criteria were (i) the quality of the online sources; (ii) the lengths of the posts, and (iii) the performance in differentiating depressed posts from non-depressed posts. Among the variety of publicly available online sources, we considered an online forum, namely, ‘‘Nafsany’’ [53]. It is an online platform that is aimed at encouraging people from Arab countries to publicly share their stories and to obtain feedback, support and advice from other users. It also has private sections that act as clinics for consultations, which are monitored by psychologists. On our study, to ensure privacy, we focused mainly on public posts and all users are anonymized. The Nafsany platform covers substantive subjects regarding diverse medical circumstances. We refer to them as categories, and one of these categories is depression. Nafsany's terms and conditions do not prevent anybody from using its public contents for research, and permission and agreement are obtained for the collection of deidentified data from Nafsany posts. Fig. 1 illustrates how the data were collected from the website. A set of 20000 posted stories were collected from various

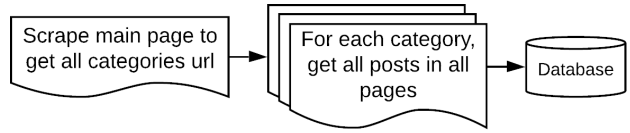


FIGURE 1. Data collection from the Nafsany Forum.

TABLE 2. Statistics on the arabdep corpus.

Characteristic	ARABDEP Corpus
Total posts	20000
Total words	4873544
Average words per post	467
Total characters	20919136

categories of Nafsany's forum. These posts differ in terms of various quantitative characteristics, such as the post length AND the number of words or characters, among other characteristics. TABLE 2 presents the statistical characteristics of the collected corpus. The corpus contains the following:

1) **User ID:** identifier for each user, which is used to collect his/her posts regardless of his/her personal information
2) **ID** for each post.
3) **Date:** the date one which a post was written.
4) **Text** of the post in Arabic.
5) **Class**: the depression label of each post (*depression* —— *non-depression*), which is described in detail in the data annotation subsection.

An important concern in research of this type is the identification of posts that correspond to depression symptoms. We reviewed several of the approaches that are discussed in the related work section: The use of self-assessment questionnaires is more reliable, but it is a tedious process because it demands individual contact with each participant. Another method is searching for expressions that include self-declarations about depression within the posts and retrieving all the posts of the self-declared users. The main problem with this approach is that it collects all users' posts with or without depression symptoms, which may increase the noise of the data, especially if the data are collected from social media such Twitter or Facebook. In our case, we collected data from a specialized forum on mental health disorders. Most of the users on this platform who are suffering from a mental disorder do not declare their diagnosis. Therefore, we cannot rely mainly on the self-declaration approach because many important posts will not be collected. Consequently, we realized the required differentiation between posts with and without depression symptoms via two phases. As we discussed earlier, the Nafsany forum consists of various categories; thus, in the first phase, all posts under the category of depression are collected. We expect that the initial set of collected posts from the first phase will contain many posts that do not correspond to depression symptoms or were posted by non-depressed users who may be active on the depression category forum due to their individual interests, such as having

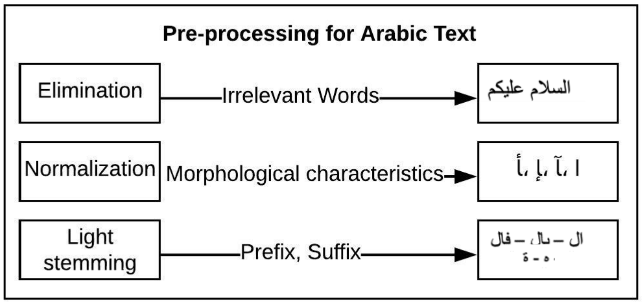


FIGURE 2. Dataset cleaning and pre-processing.

depressed relatives. In the second phase, manual annotation was conducted with the help of a psychologist to distinguish between posts with or without depression symptoms. The objective of this stage is to ensure that the annotations or labels are genuine. There is high confidence in this identification approach for several reasons: Most Nafsany posts are not short and are explicit with anonymous nicknames, and users tend to write without fear. Since we target the depression category of the forum and this category is a supportive area for users who are struggling with depression, we highly expected to collect depression-related posts with little noise from posts that are unrelated to depression. Finally, we support our approach with manual intervention by annotating data according to strict rules that are based on depression symptom classes [50]. These classes were identified from various evaluation sources, such as ''Diagnostic and Statistical Manual of Mental Disorders'' (DMS-5) [19], ''Quick Inventory of Depressive Symptomatology'' (QIDS-SR) [54] and ''Patient Health Questionnaire'' (PHQ9) [20]. Table 3 presents the labels that were used in the annotation process and examples of each.

2) DATA CLEANING AND PRE-PROCESSING

The gathered Arabic text data are pre-processed (see Fig. 2). First, words that will affect the analysis will be eliminated, such as ''السلام عليكم,'' since they are typically used in both depression and non-depression cases. A tool is utilized to automatically eliminate such words. Second, the morphological characteristics of the Arabic language, such as the letter ''alef'' having multiple forms ('ا', 'أ', 'إ', 'آ'), require normalization. Thus, we optimize the performance by reducing the time and the memory that are needed for searching forms that do not affect the meaning of words, by normalizing letters such as 'ة' to 'ه'. Third, the stemming process is conducted. Because the text in the forum is not in the formal Arabic language, the light stemmer performs well by removing only the known prefixes and suffixes without varying the bases of the words [55].

In the following sections, we will present two approaches that we investigate for predicting depression from Arabic text: (i) a lexicon-based approach and (ii) a machinelearning-based approach. Then, we will evaluate the performance of each approach.

TABLE 3. Labels and examples of the arabdep corpus with their translations.

Label: Depression: if there are depression symptoms in the posts
أحتاج للعلاج ... فمن يأخذ بيدي تحية طيبة للقائمين على هذه المساحة التي نسأل الله أن تمنحنا العلاج الذي نبحث عنه مشكلتي لن أتمكن من شرحها في سطور ولكن قد ألخص لكم حالتي بما يلي: أعاني من التوتر والإجهاد النفسي بالإضافة إلى الوسواس القهري والاكتئاب غير راضية عن نفسي ولا أشعر بالسعادة ولا التفاؤل يوما عن يوم أفقد الثقة بنفسي. أشعر بالظلم دائما وابحث عن اهتمام الناس. أشعر باليأس وبداخلي معاناة ١٣ عاما. أشعر بالانكسار وأرى كره الناس لي ولو كان ذلك وهم.. المهم أني أراه. أشعر بالحزن كنت أملك طموحا أكبر من سني واليوم ليس لدى طموح ولا هدف بالحياة أشعر أنني بحاجة للعلاج ولكن لا يوجد بمنطقي مركز نفسي يقف بجانبي فماذا عسان أفعل؟
I need treatment ... someone holds my hand Greetings to those in charge of this space Asking Allah to give us the remedy that we are looking for I will not be able to explain my problem in brief I am suffering from stress and psychological stress in addition to obsessive-compulsive disorder and depression I am not happy with myself and do not feel happy or optimistic Day by day, I am losing my confidence. I always feel wronged and am looking for care from people. I feel despair within me after suffering for 13 years. I feel hate from people even though it is a refraction, and I see it and I feel sad I had more ambition than was typical for my age, whereas today I have no ambition and no purpose in life I feel I need treatment but there is no psychological centre to support me in my city What should I do?
No-Depression: if there are no depression symptoms in the posts
هل تعدد الشخصيات مرض؟ السلام عليكم ودي أسأل عن شيء لما يكون الشخص له أكثر من شخصية مع الناس مثل إنه شخصيته في البيت غير عن اللي في العمل غير مع أصحابه غير مع أقاربه. نقدر نقول عن هذا الشخص مريض نفسي؟ أو متقلب الشخصيات بحسب المواقف اللي يعيشها. تكفون جاوبوني لأني أعاني من هذا الشيء فخفت أنه يكون مرض نفسي فحبيت أقطع الشك باليقين وأسأل أهل الاختصاص
Do I have multiple personality disorder? Peace be upon you I would like to ask for something when a person has more personality with people His personality in the house is not the same as at work, and it is not the same as with his friends or with relatives. Can we call this person mentally ill? Or changing characters depending on the situations he is in. Please answer me because I am suffering from this thing, I fear that it is a mental illness and I want to make sure and to ask people who are in charge.

B. PREDICTION MODELS: LEXICON-BASED APPROACH

1) CONSTRUCTION OF LEXICONS

In this approach, we generate a depression lexicon, namely, the ArabDep Lexicon, for the prediction of depression cases. The ArabDep Lexicon is a collection depression-related terms that are more likely to be found in online posts that are written by individuals who are struggling with depression. In the literature, the lexicon can be created using either a dictionary-based approach or a corpus-based approach. In the former, manual approaches are used to collect depression-related words [37]. The latter approach starts from a list of seeds of depression-related words, and other related words will be collected using statistical and semantic methods [3], [23]. In this study, we investigate two methods for creating the ArabDep Lexicon: (i) category-based lexicon

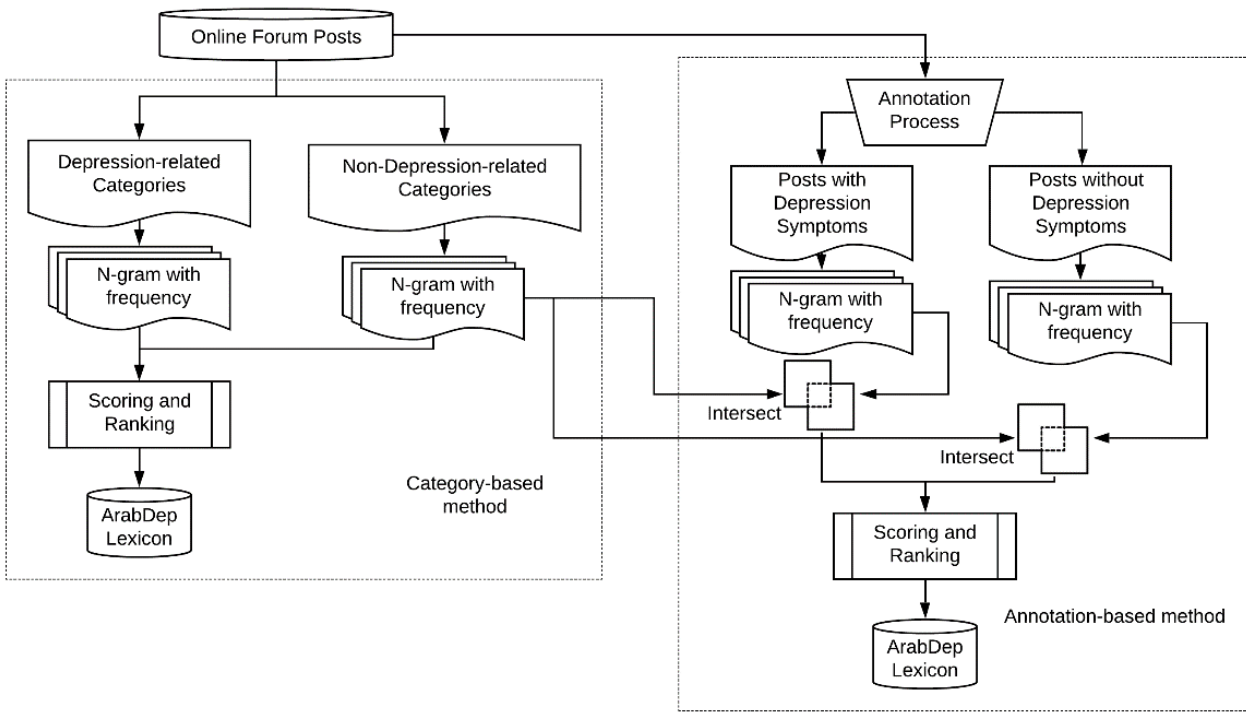


FIGURE 3. Arabdep lexicon construction.

generation and (ii) annotation-based lexicon generation. Fig. 3 illustrates the construction of the ArabDep Lexicon.

In the first method, we divide the posts based on their categories into two corpora: one collection for depression-related categories and another collection for non-depression-related categories. Then, we apply N-gram, which is an important model that is used in NLP. It separates the sentences of the forum posts into words based on the number of grams. We apply variations of N-gram: *unigram*, which divides the post sentences at the level of single words and repeats the process for subsequent words; *bigram,* which collects each pair of neighbouring words in the post sentences into a single group; and *trigram,* which collects each triplet of neighbouring words in the post sentences into a single group. We build three n-gram collections, and we calculate the frequency of each n-gram within a collection. Then, depression n-grams are scored according to (1), where ω_d and ω_{nd} are the frequencies of the depression and non-depression grams, respectively. Then, the depression grams are ordered based on their scores, and the top 1000 grams are collected in the lexicon. The same method is applied for the data before stemming and after stemming.

$$\delta = \omega_d - \omega_{nd} \tag{1}$$

The second method hybridizes the former method with an annotation process, which facilitates the more accurate identification of depression-related words because the psychologist judges the labels based on [50]. The main advantage of this approach is its high degree of granularity, which is due to the due extraction of depression-related words from posts.

TABLE 4. Statistics on the annotated ArabDep corpus.

	Training	Testing	Total
Depression	215	215	430
Not Depression	218	218	346

It starts by dividing the manually annotated posts based on their annotations into two corpora: one for posts that correspond to depression symptoms and another for posts that do not correspond to depression symptoms. Table 4 presents statistics of the corpus after annotation and balancing of the data.

The manual annotations by the psychologist were "clear", "low depression", "strong depression", "no depression" and "could not decide". We consider the first three annotations as depression cases with label = "1" and the last two cases as the non-depression cases with label = "0".

We generate three n-gram collections for n=1, 2, and 3, and we calculate the frequency of each n-gram within each collection. To conduct the double filtration of the generated n-gram collections, we intersect the n-gram collection of the non-depression-related category and the two n-gram collections that are generated by this method and eliminate the intersected part to clean these collections by excluding words that are not related to depression. We apply intersection formula (2), where ω_{nd} is the frequency of n-grams in the non-depression-related category, and ω_i is the frequency of the n-gram collection of posts that correspond to depression symptoms (ds) and posts with non-depression symptoms (nds). As a result, unrelated words, common words and stop

a. Non-stemmed Category ArabDep Lexicon (NSCL)

b. Stemmed Category ArabDep Lexicon (SCL)

c. Annotated ArabDep Lexicon (AL)

FIGURE 4. Distributions of arabdep lexicons (in cloud visualization) that are used in the content of forum posts of Nafsany.

words will be removed from both n-gram collections.

$$\lambda_{\cap} = \omega_{nd} - \omega_i, \quad where\ i \in \{ds, nds\} \tag{2}$$

Then, the depression n-grams are scored according to (1), where ω_d and ω_{nd} are the frequencies of depression and non-depression grams respectively. Next, the depression grams are ranked based on the generated scores, and the top 1000 grams are selected into the lexicon. Fig. 4 presents the distributions of the ArabDep Lexicons (as a cloud visualization) that are used for the content of forum posts of Nafsany. The Non-stemmed Category ArabDep Lexicon (NSCL) and the Stemmed Category ArabDep Lexicon (SCL) are lexicons that were generated via the category-based method, where NSCL was constructed from data before a light stemming process and SCL was constructed from lightly stemmed data.

The Annotated ArabDep Lexicon (AL) is a lexicon that was generated via the annotation-based method. All lexicons are generated with the same generating method; however, according to Fig. 4, the semantic representations vary, and their evaluation is varied accordingly.

2) RULE-BASED ALGORITHM

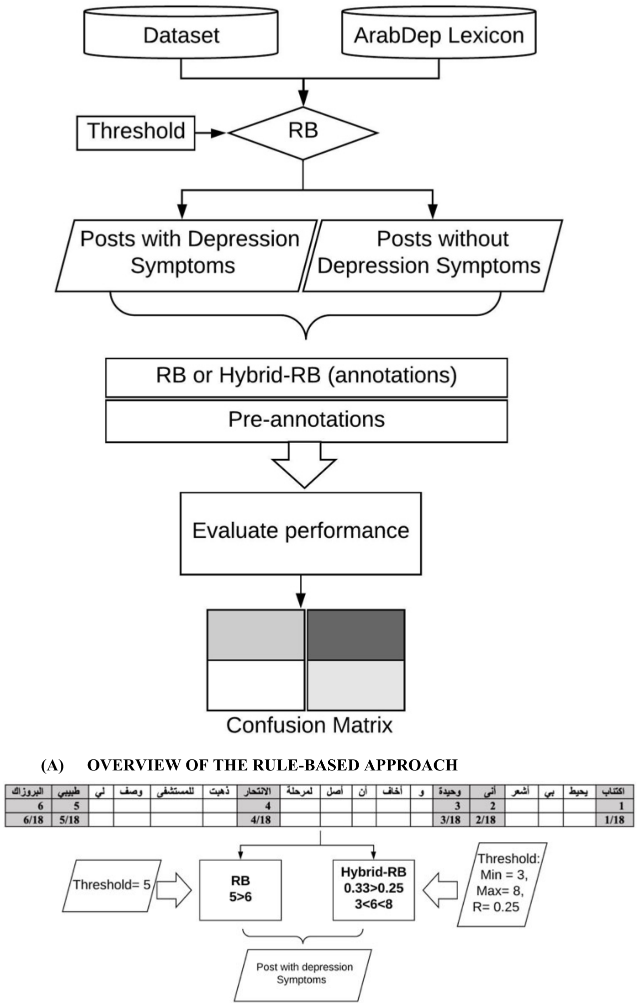
The rule-based algorithm utilizes the generated ArabDep lexicon to predict the depression symptoms from posts. It exploits the strength of the n-gram models that were used to create the ArabDep lexicon and identifies the best combination of grams for the prediction of depression symptoms. The major advantages of this approach are its simplicity and applicability to the Arabic language. The task of predicting depression symptoms using the rule-based algorithm (RB) over the ArabDep lexicon from the computational perspective can be defined as follows: Given an input text T and a lexicon $L = [g_1 \cdots, g_k]$ of depression-related words (grams), identify according to its content whether the text T exceeds the threshold θ of depression-related words g_i, for $i = 1 \cdots k$.

Fig. 5 (a) illustrates the overall process of the RB algorithm with examples. The datasets of depression posts and non-depression posts, along with the corresponding lexicon, are the input parameters of the algorithm. The algorithm will predict whether the sentence corresponds to depression symptoms or not based on a threshold. Fig. 6, 7 and 8 present the pseudo-code of the RB algorithm.

In Fig. 6, Line 1, the algorithm will iterate through a set of thresholds θs that contains either θ or $\theta_{\rm h}$, where θ is the threshold that is defined as the number of depression-related words and $\theta_{\rm h}$ is the hybrid threshold that is defined based on the total number of depression-related words and the ratio of depression-related words. Line 3 maintains the variation of the datasets. If the threshold is $\{\theta_{\min}, \theta_{\max}, \theta_{\Re}\}$, where the number of depression words in a post is within a range of $\theta = \{\theta_{\min}, \theta_{\max}\}$ and the ratio number of depression words in a post is calculated via Equation 3, Line 5 will be executed by calling the hybrid_RB function (see Fig. 6 and 8). Otherwise, Line 7 will be executed, and the RB function will be called (see Fig. 6 and 7). For both functions, if the threshold is met, the post corresponds to depression symptoms; otherwise, the post does not correspond to depression symptoms.

$$\Re = \frac{Depression\ Word\ Count}{Total\ words\ of\ a\ post} \tag{3}$$

Fig. 5 (b) and (c) present examples that demonstrate the rule-based approach for the prediction of depression symptoms. Assume that we have two sentences, as shown in the figure. Fig. 5 (b) shows a few sentences that contain words from the ArabDep lexicon, which are highlighted in grey in the figure. When using the RB function (see Fig. 7) and assuming that the threshold $= \{\theta = 5\}$, the number of depression-related words exceeds the threshold; hence, this post corresponds to depression symptoms. When applying the Hybrid-RB function (see Fig. 8) and assuming that the threshold $= \{\theta_{\min} = 3, \theta_{\max} = 8, \theta_{\Re} = 0.25\}$, the number of depression-related words is 6, which exceeds $\theta_{\min}$ and is less



(A) OVERVIEW OF THE RULE-BASED APPROACH

(B) EXAMPLE OF PREDICTING A POST WITH DEPRESSION SYMPTOMS

(C) EXAMPLE OF PREDICTING A POST WITHOUT DEPRESSION SYMPTOMS

FIGURE 5. Rule-based algorithm demonstration.

than $\theta_{\max}$. The ratio of these words is 0.33, which exceeds $\theta_{\Re}$. Therefore, these values indicate that this post corresponds to depression symptoms.

Fig. 5 (c) presents a few sentences that contain words from the ArabDep lexicon, which are highlighted in grey in the figure. When using the RB function and with same assumption as above, the number of depression-related words is less than the threshold; hence, this post does not correspond to depression symptoms. In contrast, when applying the Hybrid-RB function and with the same assumption as above, the number of depression-related words is 2, which is

```
Rule-based Algorithm
INPUT: Text Data (DT, NDT), where DT denotes depression-related text and NDT non-depression-related text; θs:{θ,θh}, where θ denotes thresholds and θh:{θmin , θmax, θℜ}
OUTPUT: accuracy, precision, recall, F1
INITIALIZATION:
real_depression = coded with 1 for all data from the depression category
real_not_depression = coded with 0 for all data from other categories
real_data = real_not_depression + real_depression

Begin:

for θ in θs:
    depression = []
    for D in DT:
        if θh :
            depression.append(hybrid_RB(D, L, θh.θmin, θh.θmax, θh.θℜ))
        elif θ:
            depression.append(RB(D, L, θ))
    non_depression = []
    for ND in NDD:
        if θh:
            non_depression.append(hybrid_RB(D, L, θh.θmin, θh.θmax, θh.θℜ))
        elif θ:
            non_depression.append(RB(D, L, θ))
    data = depression + non_depression
    performance = performance_score(real_data, data)
End
```

FIGURE 6. Rule-based algorithm pseudo code.

```
INPUT: P: Post, L: Lexicon, θ: Threshold
OUTPUT:1: depression, 0: non-depression
Begin:
    ws = P.split()
    dws = 0
    for w in ws:
        if w in L: dws = dws + 1
        if dws >= θ: return 1
        else: return 0
End
```

FIGURE 7. RB function pseudo code.

```
hybrid_RB Function
INPUT: P: Post, L: Lexicon, θmin , θmax, and θℜ, where θ: Threshold
OUTPUT:1: depression, 0: non-depression
Begin:
    ws = P.split()
    tws = len(ws)
    dws = 0
    for w in ws:
        if w in L: dws = dws + 1
            ℜ= dws / tws
        if dws < θmin: return 0
        elif dws >= θmax: return 1
        else:
            if ℜ > θℜ: return 1
            else: return 0
End
```

FIGURE 8. Hybrid_RB function pseudo code.

less than both $\theta_{\min}$ and $\theta_{\max}$. In addition, the ratio of these words is 0.12, which is less than $\theta_{\Re}$. Consequently, these values are used to predict that this post does not correspond to depression symptoms.

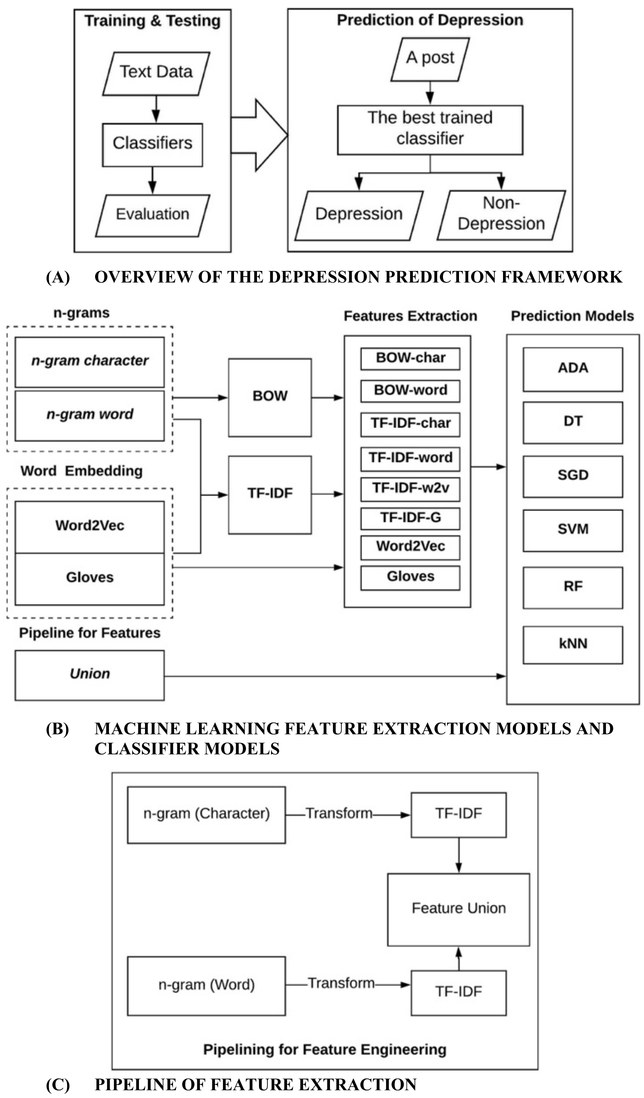


FIGURE 9. **Prediction models: the machine-learning-based approach.**

C. PREDICTION MODELS: MACHINE-LEARNING-BASED APPROACH

Another approach for predicting the presence of depression symptoms is the use of machine learning classification algorithms. The proposed framework is developed by using ADA boost, decision trees, k-nearest neighbour, random forest, support vector machine, and stochastic gradient descent. Various feature extraction models have been used to analyse the use of language in the forum posts. Fig. 9 presents the applied framework. It illustrates the framework of the depression prediction process and the models that have been evaluated for creating the best-trained classifiers.

1) FEATURE EXTRACTION

N-grams and word embedding are used to encode the characters or words to be fed into classifiers.

- **N-gram modelling:** is widely used in text mining and natural language processing to extract features from text. The text of each forum post is divided into characters or words based on a number N, which can be determined via several experiments. Our method is based on calculating the probabilities of incidence of each input post as a unigram, bigram and trigram. For n-gram modelling, bag of words and the term frequency inverse document frequency (TF-IDF), which is a weight-based approach that is used in text mining, are applied as numerical statistics to highlight the importance of each word to each post in the dataset. The essential reason for utilizing these approaches is to scale down the effect of using experientially less informative tokens. When applying bag of words, most structural elements of the text are discarded, such as paragraphs and formatting. The model counts how often each gram occurs in each text. Discarding the structure and conducting the count calculation leads to the concept of a "bag" for representing the text. The computation of this model representation consists of three major steps: (i) tokenization, where each post is split into tokens; (ii) ranking vocabularies; and (iii) encoding the tokens based on their number of occurrences. When applying TF-IDF, the tokenization process is conducted and the token frequency is calculated to provide insight regarding the most informative grams using a small fraction. Equation (4) presents the TF-IDF formula, which is used to compute the weight for each tokenized gram in each post. The term frequency (TF) is the number of occurrences of a gram "g" in a post "p". The inverse document frequency (IDF) is the number of occurrences of gram "g" in the whole corpus, and it is defined in (5), where DF denotes the document frequency of a gram "g" and "N" is the total number of documents/posts [60].

$$TF-IDF_{gp} = TF_{gp} * IDF_{g^1} \tag{4}$$

$$IDF_g = \log\frac{N}{DF_g} \tag{5}$$

- **Word embedding modelling:** is a deep learning method that is used for text representation for the recognition of comparatively important words. This modelling approach is based on the conversion of each word to a corresponding mathematical vector, where each word is represented as a positive or negative decimal number [61]. One of the word embeddings is a global vector for word representation (Glove), which is an unsupervised learning algorithm for generating vectors of words [62]. Glove searches for similar words in the whole context. Another word embedding is word2vec, which uses many documents or a huge corpus to generate a huge number of vectors for representing each word. Our approach is to sum the vector representation using word2vec. Next, using TF-IDF weighting with word2vec, we measure the weights. Then, TF-IDF is concatenated with the weighted word2vec to merge the vectors.

- **Pipeline Union Modelling:** We apply TF-IDF for n-grams to assemble numerous steps to be cross-validated together using various parameters. In our case, we use n-grams for numbers of characters that range from two to five and n-grams for numbers of words that range from one to three. Then, ''Feature Union'' is used to concatenate the outputs of multiple objects and is applied to the pipelines of TF-IDF to concatenate the output features. Fig. 9 (c) illustrates the pipeline components.

2) CLASSIFIERS

- The **AdaBoost (ADA)** algorithm builds a collection of classifiers over training samples by manipulating a set of weights. These weights are adaptively adjusted after each stage of boosting, namely, if these weights are misclassified by the current classifier, they will be increased, whereas if these weights are suitably classified, they will be reduced [63].
- **Decision trees (DTs)** are created via an algorithmic method that determines approaches for splitting a dataset with regards to various conditions. To determine the most suitable condition, the degree of the parent's impurity of the tree before splitting must be compared with the degree of the child's impurity of the tree after splitting. The larger the difference between the parent node and its child node, the more suitable the condition. Various functions can be used to measure the node's impurity, which include the Gini Index, the entropy and the misclassification error. In addition, various algorithms have been proposed for decision trees, such as ID3 [64] and C4.5 [65].
- The **k-nearest neighbours (kNN)** algorithm builds a model by storing the training dataset. To predict a new data point, kNN considers the nearest data points in the same training dataset, which are called neighbours, in calculating the distances between the points of data. Selecting the proper distance value (k) by adjusting this parameter improves the performance of the model. The construction of a model of nearest neighbours is typically very fast; however, the prediction process may be slow when the training set is too large [66].
- The model of **random forests (RF)** operates based on the construction of multiple DTs during training. Its output is the model class of the other classes when solving a classification problem or the mean class of the other classes when solving a prediction problem of the individual trees. Reference [67] presents the generalization error of RF. In the case of the kth tree, a random vector v will be generated. The generalization error depends on two factors: the correlations between the random trees and the strengths of the individual classifiers in the forest. As the number of trees increases, the generalization error will increase.
- **Support vector machine (SVM)** is a non-probabilistic classifier that can find the optimal boundary for each instance. An SVM model represents each post/document as a vector in the space. Then, it computes the separation between the points and a hyperplane [60]. SVM aims at maximizing the distances between the separating hyperplane and the classes. The optimal separation has been realized when a hyperplane that separates the two classes with the largest distance to the closest data points has been identified. This is the linear version of SVM. However, kernels can be utilized, which change the behaviour of the SVM algorithm [68].
- **Stochastic gradient descent (SGD)** is applied to improve the performance of SVM. The input value of the SGD classifier is the sample before the prediction of the next value, and this sample is compared with the actual value. In addition, its algorithm uses a loss function to calculate the distance between the predicted and actual values; if it is large, SGD adjusts the weights of all features and compares the result against each iteration to maximize the similarity between the predicted actual values. During the stage in which the weights of the features are adjusted, over-fitting issues may arise. If the values in the training phase increase and the values in the development phase decrease, an early stop function will terminate the adjustment of the weights [69].

IV. RESULTS AND DISCUSSION

After processing the data via the lexicon-based method or a machine-learning-based method to predict depression symptoms from Arabic text, fair evaluations must be conducted to assess the performance. For the former approach, a range of cases have been evaluated, which include annotation that is based on the forum categories, annotation that is based on manual intervention by a psychologist, light stemming of posts, and no stemming of posts. For the second approach, a range of selected features and designed classifiers with suitable coefficient values have been experimentally evaluated. We evaluated the classification performance when the dataset was divided according to an 80:10:10 ratio of training:validation:testing subsets. Ten-fold cross-validation was conducted for the training stage, which provides a reasonable test of validation. Various performance metrics have been considered: Accuracy is defined as the ratio of correct/true classification; precision calculates how many positive cases have been correctly identified; recall estimates the ratio of positive cases that have been correctly recognized; and F1-score is a harmonic average of the precision and recall, where the closer their values, the higher the F1-score will be. The evaluation metrics are based on the numbers of true-positive predictions (TP), false-positive predictions (FP), true-negative predictions (TN), and false-negative predictions (FN) [70].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F - Score = \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

Additional evaluation metrics are the area under the curve (AUC) and the receiver operating characteristic curve (ROC). ROC corresponds to the true positive rate (TPR) versus the false positive rate (FPR), which is used to evaluate the performance of a model with binary classes under various thresholds of classification. AUC is an aggregate measurement of the classification performance, and it ranges between 0 and 1.

$$TPR = \frac{TP}{TP + FN} \tag{10}$$

$$FPR = \frac{FP}{FP + TN} \tag{11}$$

A. EVALUATION OF THE LEXICON-BASED APPROACH

To visualize the performance of this approach, a confusion matrix was constructed. It facilitates understanding of the relationships between individual cases. A confusion matrix is a contingency table that presents the variances between the true classes (category- or manual-based annotation) and the predicted classes (labels after utilizing lexicons). This presentation approach facilitates the determination of whether the classifier confuses the two classes or labels.

The input for the confusion matrix in Fig. 10 (a) represents the performance on the testing dataset, which contains posts that have been annotated based on depression categories and posts that have been annotated based on non-depression categories. After applying the rule-based algorithm, the confusion matrix shows that the NSCL predicted 73% of the posts that are under depression categories on the Nafsany website as posts that correspond to depression symptoms; these are the true-positive cases. In contrast, 27% of the posts were predicted as posts that do not correspond to depression symptoms, which represent false-positive cases. In addition, the NSCL predicted 75% of the posts that are under non-depression categories on the Nafsany website as posts that do not correspond to depression symptoms, which are the true-negative cases. In contrast, 25% of the posts are incorrectly predicted as posts that correspond to depression symptoms, which represent false-negative cases.

The input for the confusion matrix in Fig. 10 (b) is the same testing dataset as was considered previously but evaluated using SCL. After applying the rule-based algorithm, the confusion matrix in Fig. 10 (b) shows that the SCL predicted 77% of the posts that are under depression categories on the Nafsany website as posts that correspond to depression symptoms, which are the true-positive cases. In contrast, 23% of the posts were incorrectly predicted as posts that do not correspond to depression symptoms, which represent the false-positive cases. In addition, the SCL predicted 75% of the posts that are under non-depression categories on the Nafsany website as posts that do not correspond to depression symptoms, which are the true-negative cases; in contrast, 25% of the posts are falsely predicted as posts that correspond to depression symptoms, which represent the false-negative cases.

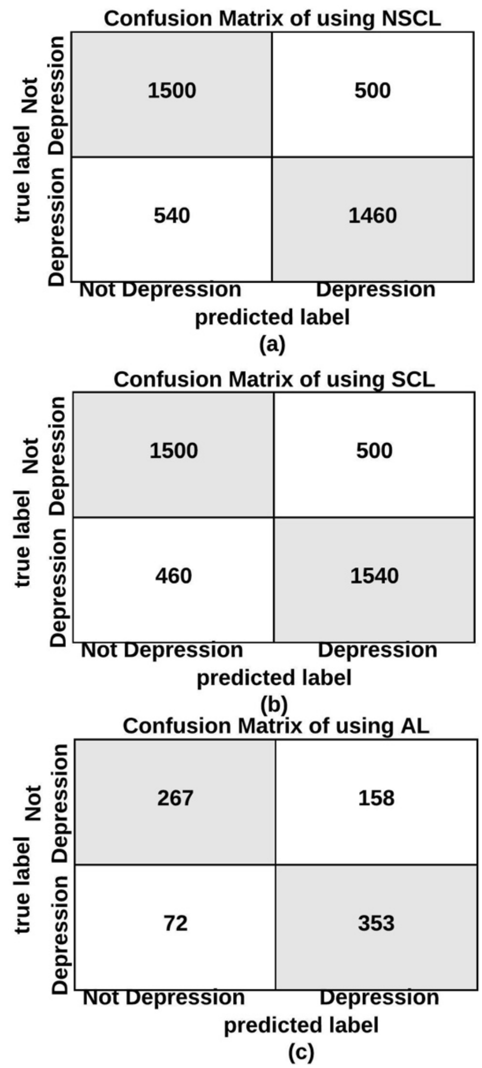


FIGURE 10. Confusion matrices of the lexicon-based approach.

The input for the confusion matrix in Fig. 10 (c) is the testing dataset that contains posts that are annotated with the help of a psychologist and labelled as depression "1" or non-depression "0". After applying the rule-based algorithm, the confusion matrix in Fig. 10 (c) shows that AL outperforms the other lexicon-based approaches. The AL correctly predicted 80% of the posts as posts that correspond to depression symptoms, which are the true-positive cases; in contrast, 20% of the posts were incorrectly predicted as posts that do not correspond to depression symptoms, which represent the false-positive cases. In addition, the AL correctly predicted 82% of the posts as posts that do not correspond to depression symptoms, which are the true-negative cases; in contrast, 18% of the posts were incorrectly predicted as posts that correspond to depression symptoms, which represent the false-negative cases.

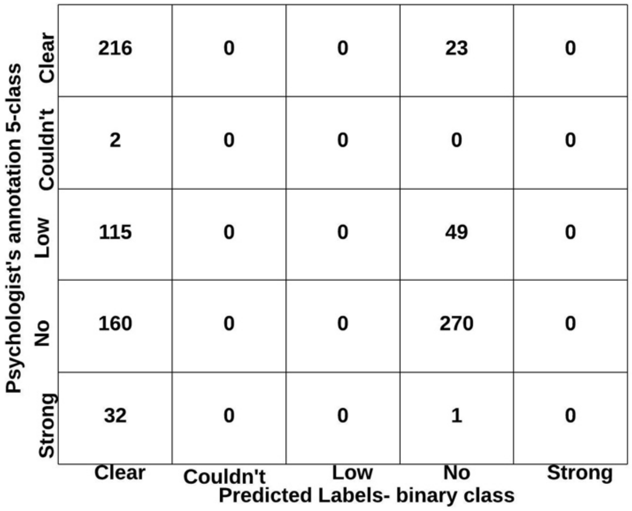


FIGURE 11. Confusion matrix of multiclass annotations by a psychologist vs. binary class labels that were predicted via the lexicon-based approach.

The objective of the experiment in Fig. 11 is to compare the true cases to the predicted cases. In this confusion matrix, as explained previously, the psychologist annotated the data with five classes, which correspond to "0" and "1" cases for the classifier. According to the results for the strong depression cases, the lexicon-based approach predicted depression symptoms for 32 out of 33 cases. For the clear depression cases, the system predicted the occurrences of depression symptoms in 216 cases out of 239. For low depression cases, the system predicted the occurrences of depression symptoms in 115 out of 164 cases. The system yielded more accurate results when the psychologist diagnosed clearer depression cases. For example, on cases that the psychologist labelled as very strong depression cases, the accuracy of our prediction model exceeds 95%. However, as the diagnostic certainty of the psychologist decreased, the accuracy decreased. For example, the accuracy became approximately 75% on cases that the psychologist diagnosed as low depression cases.

TABLE 5 presents the performance of the lexicon-based approach with the variation of the threshold value. The best-performing threshold is $\theta 3$, where the accuracy is 80.45%.

B. EVALUATION OF THE MACHINE-LEARNING-BASED APPROACH

TABLE 7 presents the accuracy results of the six constructed classification models with various NLP-extracted features. The highest accuracy is realized by SGD with TF-IDF and either word-based or character-based models, namely, 73%, which followed by ADA and SVM, which realize 72% accuracy. This performance is realized by ADA when utilizing BOW of characters as its NLP feature extraction approach. In contrast, SVM realized its best performance when applying the TF-IDF model for both words and characters. The word embedding approaches, namely, Gloves and word2vec, failed to perform well since these models require huge data for deep processing and our dataset is not sufficiently large for the use of deep learning to train the classification models.

TABLE 5. Performance results of the lexicon-based approach.

	Accuracy	Precision	Recall	F1
θ1	66.87%	60.67%	95.88	74.32
θ2	74.28%	68.91%	88.48	77.48
θ3	80.45%	79.36%	82.30	80.81
θ4	78.60%	84.57%	69.96	76.58
θ5	76.34%	86.78%	62.14	72.42
θ6	75.31%	90.72%	56.38	69.54
θ7	72.02%	92.13%	48.15	63.24
θ8	70.58%	95.45%	43.21	59.49
θ9	68.31%	96.84%	37.86	54.44

TABLE 6 presents the best performances of the six constructed classification models with the use of various feature extraction approaches. Each performance is represented by various metrics, which include the accuracy, F-score, precision and recall values. The accuracy is used as a measure in various studies that are related to depression detection and prediction. In this study, we present also the precision, recall and F-score values, which enable us to more deeply analyse the outputs. In the application of single features (N-gram, TF-IDF, and BoW) with the classifiers, performance improvements were observed. The best performance is realized with TF-IDF with the SGD learning algorithm, which resulted in 73% accuracy, 74% precision, 71% recall and 74% F1-score.

The detailed experimental results for the word embedding feature extraction have not been reported in TABLE 6 since it performed the poorest among the feature extraction approaches (see TABLE 7). The reason is that this type of feature modelling requires a huge amount of data for deep learning and cannot perform well on small-scale data. In addition, fewer resources are available that contain Arabic text that is related to depression over the internet from an online forum compared to English text.

We conclude that both the lexicon-based and machine-learning-based classification models for predicting depression symptoms perform well even on small datasets. Moreover, based on our results, the predictive power of the lexicon-based approach depends on the selection of a suitable threshold, whereas the predictive power of the machine-learning-based approach depends on the selection of suitable features.

Comparing to related works, references [23], [24], [26], [27] predict depression based on questionnaires.

TABLE 6. Performance results of the machine learning models.

Classifier	Feature Extraction	Acc %	F1 %	Pre %	Rec %
ADA	BoW-Char	72	73	71	75
	BoW-Char (Stem)	69	70	69	72
	BoW-Word	68	70	66	73
	BoW-Word (Stem)	69	71	68	73
	TF-IDF- Char	70	71	71	71
	TF-IDF- Char (Stem)	67	66	68	65
	Pipeline Union	67	67	68	66
	Pipeline Union (Stem)	70	70	71	69
	TF-IDF	65	66	66	66
	TF-IDF (Stem)	65	66	64	68
DT	BoW-Char	66	63	69	58
	BoW-Word	67	60	76	50
	TF-IDF- Char	65	56	76	44
	Pipeline Union	65	56	77	44
	TF-IDF (Stem)	65	62	69	56
KNN	TF-IDF	67	67	69	64
	TF-IDF (Stem)	69	67	74	61
RF	BoW-Char	69	67	71	64
	BoW-Char (Stem)	67	64	72	57
	BoW-Word	66	66	66	66
	BoW-Char (Stem)	70	68	75	62
	TF-IDF- Char	69	67	73	61
	TF-IDF- Char (Stem)	67	65	70	61
	Pipeline Union	66	64	69	59
	Pipeline Union (Stem)	68	66	71	62
	TF-IDF	68	66	72	61
	TF-IDF (Stem)	68	65	72	60
SGD	BoW-Char	70	68	72	65
	BoW-Char (Stem)	65	66	65	66
	BoW-Word	70	71	69	74
	BoW-Char (Stem)	69	71	67	76
	TF-IDF- Char	73	73	74	71
	TF-IDF- Char (Stem)	70	70	70	70
	Pipeline Union	73	73	75	70
	Pipeline Union (Stem)	70	70	70	69
	TF-IDF	73	72	75	70
	TF-IDF (Stem)	73	73	74	72
SVM	BoW-Char	71	73	70	77
	BoW-Char (Stem)	66	68	64	72
	BoW-Word	67	69	66	73
	BoW-Char (Stem)	67	70	65	76
	TF-IDF- Char	72	71	75	68
	TF-IDF- Char (Stem)	71	71	71	71
	Pipeline Union	72	72	73	70
	Pipeline Union (Stem)	69	69	69	69
	TF-IDF	72	72	74	70
	TF-IDF (Stem)	70	71	70	73

The prediction approaches of references [23], [24] were statistical, whereas references [26], [27] used only machine learning. The approaches in references [33] and [34] predict depression based on self-declaration and apply machine learning and deep learning, respectively. Fatima *et al.* [37] used data from forum subscriptions to predict depression. The results of these studies are not comparable, as they were reported based on processing other languages. All these approaches proposed solutions for the English language, except [27] focused on the Japanese language. The manipulation of the Arabic language differs from the manipulation of English and other languages. One of the reasons is the rich morphological structures of the Arabic language, where each word may have many forms and meanings. The formal and informal structures of the Arabic language must be treated differently on either the morphological or syntactical scale. Therefore, we will compare our results with results on manipulating the Arabic language.

TABLE 7. Accuracy results of the machine learning models.

Feature Extraction	ADA	DT	KNN	RF	SGD	SVM
BoW-Char	72	66	64	69	70	71
BoW-Char (Stem)	69	64	63	67	65	66
BoW-Word	68	67	55	66	70	67
BoW-Word (Stem)	69	62	58	70	69	67
Gloves	51	51	51	49	49	51
Gloves (Stem)	51	51	51	49	51	51
Gloves-TFIDF	51	51	51	51	51	51
Gloves-TF-IDF (Stem)	51	51	51	51	51	51
TF-IDF-Char	70	65	62	69	73	72
TF-IDF-Char (Stem)	67	62	62	67	70	71
TF-IDF-Union	67	65	64	66	73	72
TF-IDF-Union (Stem)	70	61	64	68	70	69
TF-IDF-Word	65	61	67	68	73	72
TF-IDF-Word (Stem)	65	65	69	68	73	70
word2vec	57	61	61	58	63	63
word2vec (Stem)	57	56	57	59	64	60
word2vec-TF-IDF	57	54	62	55	61	63
word2vec-TF-IDF (Stem)	54	59	60	56	62	59

The author of [71] investigated how the posts in a social network can be used to classify users via the machine-learning approach based on their mental health levels. They reported their results on data that were collected from Saudi Arabia, but they did not specify the language in which the posts were written. Their reported result is 57% accuracy, 67% precision, and 56% recall. Reference [13] built a lexicon and a sentiment analysis tool for the Arabic language. For their experiments, they reported an accuracy of 70% on Twitter data and an accuracy of 64% on Yahoo-Maktoob data. We report higher performance for the both approaches, namely, the lexicon-based and machine learning approaches, on the Nafsany dataset for the Arabic language compared to [13] and [71]. Fig. 12 compares the results of the proposed methods with those of these available methods.

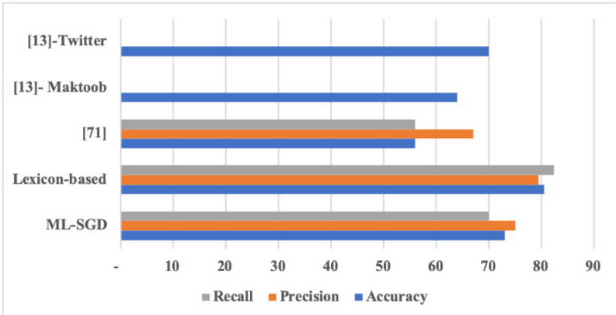


FIGURE 12. Accuracy, precision, and recall results of the proposed methods and the available methods.

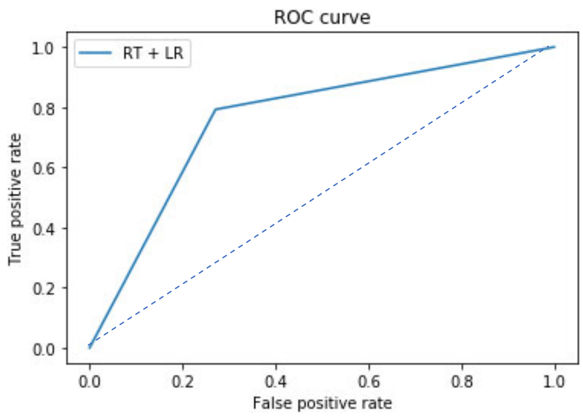


FIGURE 13. ROC curve and AUC = 0.78 for the lexicon-based approach.

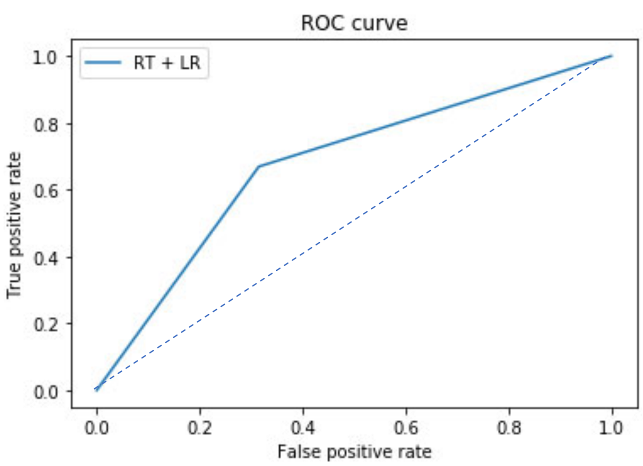


FIGURE 14. ROC curve and AUC= 72% for the machine-learning approach (SGD- TF-IDF).

Fig. 13 and Fig. 14 show the performances of the lexicon-based approach and the machine-learning-based approach for predicting depression symptoms from Arabic text using ROC/AUC. TPR was plotted on the Y-axis, whereas FPR was plotted on the X-axis. The values of AUC for both models are larger than 0.70 and, thus, exceed a random negative value (0.50). The AUC of the lexicon-based approach is 78%, while the AUC of the best classifier of the machine-learning-based approach is 72%.

V. CONCLUSION

This study is likely to enable intelligent instruments to identify and predict depression symptoms from Arabic text based on depression-related words. We initiated research on innovative depression prediction solutions that serve Arabic communities, and we desire to stimulate the research and development of various possible approaches that computationally manipulate Arabic language as expressive guidance for the identification of depression. This paper proposed computational approaches for the utilization of an Arabic online forum on which users discuss and seek advice regarding several mental health disorders. We gathered data from this forum and labelled them either automatically or manually. We proposed a semi-supervised approach (lexicon-based approach) and a supervised approach (machine-learning-based approach) for the prediction of depression from posts. The performances of the proposed approaches were evaluated. The former approach outperformed the latter approach and approaches from related studies by realizing an accuracy that exceeds 80%. However, the machine-learning-based approach was evaluated with various feature extraction models and classifiers. The best performance was realized with TF-IDF as the feature extraction model and SGD as the classifier, which resulted in 73% accuracy.

Although our results demonstrate that the performances of the applied approaches are reasonable, this task is challenging and merits further investigation. Our results could be used as a baseline for the applications of new mechanisms to the prediction of depression from Arabic text. Nevertheless, there is substantial room for improvement, and it is our expectation that the results of this study will motivate others to design and develop innovative and more effective prediction solutions with respect to various variables, such as locations, users and genders. The relationship among these variables, the personalities of the online users and their behaviours in response to depression require further examination. This study did not differentiate between depression categories and regarded them as a single class for the prediction of depression symptoms. We suggest in the future work, various categorizations of depression, such as seasonal affective disorder, bipolar disorder and postpartum depression, should be explored in addition to the degrees or levels of depression. Moreover, regarding the dataset that was collected in this study, additional data will need to be collected and annotated to improve the accuracy of the deep learning models, such Gloves and Word2Vec, since they required large amounts of data for training. An additional improvement to the deep learning approach could be realized by implementing a language model such as BERT as an input for a neural network.

REFERENCES

[1] *World Health Organization*. Accessed: Sep. 21, 2019. [Online]. Available: https://www.who.int/mental health/advocacy/en/

[2] K. Smith, "Mental health: A world of depression," *Nature*, vol. 515, no. 7526, pp. 180–181, Nov. 2014.

[3] A. H. Yazdavar, H. S. Al-Olimat, M. Ebrahimi, G. Bajaj, T. Banerjee, K. Thirunarayan, J. Pathak, and A. Sheth, "Semi-supervised approach to monitoring clinical depressive symptoms in social media," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Jul. 2017, pp. 1191–1198.

[4] G. J. Johnson and P. J. Ambrose, "Neo-tribes: The power and potential of online communities in health care," *Commun. ACM*, vol. 49, no. 1, pp. 107–113, Jan. 2006.

[5] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, "Machine learning framework for the detection of mental stress at multiple levels," *IEEE Access*, vol. 5, pp. 13545–13556, 2017.

[6] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of depression-related posts in Reddit social media forum," *IEEE Access*, vol. 7, pp. 44883–44893, 2019.

[7] C.-T. Lin, C.-H. Chuang, Z. Cao, A. K. Singh, C.-S. Hung, Y.-H. Yu, M. Nascimben, Y.-T. Liu, J.-T. King, T.-P. Su, and S.-J. Wang, "Forehead EEG in support of future feasible personal healthcare solutions: Sleep management, headache prevention, and depression treatment," *IEEE Access*, vol. 5, pp. 10612–10621, 2017.

[8] H. Peng, C. Xia, Z. Wang, J. Zhu, X. Zhang, S. Sun, J. Li, X. Huo, and X. Li, "Multivariate pattern analysis of EEG-based functional connectivity: A study on the identification of depression," *IEEE Access*, vol. 7, pp. 92630–92641, 2019.

[9] N. Habash, *Introduction to Arabic Natural Language Processing* (Synthesis Lectures on Human Language Technologies), U. O. T. G. Hirst, Ed. San Rafael, CA, USA: Morgan & Claypool, 2010.

[10] M. Abdul-Mageed, M. T. Diab, and M. Korayem, "Subjectivity and sentiment analysis of modern standard Arabic," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 2, 2011, pp. 587–591.

[11] A. F. El Gohary, T. I. Sultan, M. A. Hana, and M. El Dosoky, "A computational approach for analyzing and detecting emotions in Arabic text," *Int. J. Eng. Res. Appl.*, vol. 3, pp. 100–107, May/Jun. 2013.

[12] A. Farghaly and K. Shaalan, "Arabic natural language processing: Challenges and solutions," *ACM Trans. Asian Lang. Inf. Process.*, vol. 8, no. 4, pp. 1–22, Dec. 2009.

[13] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, M. Al-Ayyoub, M. N. Al-Kabi, and S. Al-rifai, "Towards improving the lexicon-based approach for arabic sentiment analysis," *Int. J. Inf. Technol. Web Eng.*, vol. 9, no. 3, pp. 55–71, Jul. 2014.

[14] A. Shoukry and A. Rafea, "Sentence-level arabic sentiment analysis," in *Proc. Int. Conf. Collaboration Technol. Syst. (CTS)*, May 2012, pp. 546–550.

[15] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub, "Arabic sentiment analysis: Lexicon-based and corpus-based," in *Proc. IEEE Jordan Conf. Appl. Electr. Eng. Comput. Technol. (AEECT)*, Dec. 2013, pp. 1–6.

[16] M. Abdul-Mageed, M. Diab, and S. Kübler, "SAMAR: Subjectivity and sentiment analysis for arabic social media," *Comput. Speech Lang.*, vol. 28, no. 1, pp. 20–37, Jan. 2014.

[17] M. Al-A'abed and M. Al-Ayyoub, "A lexicon-based approach for emotion analysis of Arabic social media content," in *Proc. Int. Comput. Sci. Inform. Conf. (ICSIC)*, 2016, pp. 343–351.

[18] M. Abdullah, M. AlMasawa, I. Makki, M. Alsolmi, and S. Mahrous, "Emotions extraction from Arabic tweets," *Int. J. Comput. Appl.*, vol. 40, pp. 1–15, 2018.

[19] A. P. Association *et al.*, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, DC, USA: American Psychiatric Association Publisher, 2013.

[20] B. Löwe, K. Kroenke, W. Herzog, and K. Gräfe, "Measuring depression outcome with a brief self-report instrument: Sensitivity to change of the patient health questionnaire (PHQ-9)," *J. Affect. Disorders*, vol. 81, no. 1, pp. 61–66, Jul. 2004.

[21] L. S. Radloff, "The CES-D scale: A self-report depression scale for research in the general population," *Appl. Psychol. Meas.*, vol. 1, no. 3, pp. 385–401, 1977.

[22] A. Beck, R. Steer, and G. Brown, *Beck Depression Inventory-II*. San Antonio, TX, USA: Psychcorp, 1996.

[23] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Proc. 7th Int. AAAI Conf. Weblogs Social Media*, 2013, pp. 128–137.

[24] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, "Characterizing and predicting postpartum depression from shared facebook data," in *Proc. 17th ACM Conf. Comput. Supported Cooperat. Work Social Comput. (CSCW)*, 2014, pp. 626–638.

[25] J. W. Pennebaker, R. J. Booth, and M. E. Francis, *Liwc2007: Linguistic Inquiry and Word Count*. Austin, TX, USA: Liwc.net, 2007.

[26] A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with Twitter data," *Sci. Rep.*, vol. 7, no. 1, p. 13006, Dec. 2017.

[27] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, "Recognizing depression from Twitter activity," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. (CHI)*, 2015, pp. 3187–3196.

[28] T. Kudo. (2006). *Mecab: Yet Another Part-of-Speech and Morphological Analyzer*. [Online]. Available: http://mecab.sourceforge.jp

[29] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[30] H. A. Schwartz, J. Eichstaedt, M. L. Kern, G. Park, M. Sap, D. Stillwell, M. Kosinski, and L. Ungar, "Towards assessing changes in degree of depression through facebook," in *Proc. Workshop Comput. Linguistics Clin. Psychol., From Linguistic Signal to Clin. Reality*, 2014, pp. 118–125.

[31] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *J. Personality*, vol. 60, no. 2, pp. 175–215, Jun. 1992.

[32] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, "CLPsych 2015 shared task: Depression and PTSD on Twitter," in *Proc. 2nd Workshop Comput. Linguistics Clin. Psychol., From Linguistic Signal to Clin. Reality*, 2015, pp. 31–39.

[33] T. Pedersen, "Screening Twitter users for depression and PTSD with lexical decision lists," in *Proc. 2nd Workshop Comput. Linguistics Clin. Psychol., From Linguistic Signal Clin. Reality*, 2015, pp. 46–53.

[34] A. Benton, M. Mitchell, and D. Hovy, "Multi-task learning for mental health using social media text," 2017, *arXiv:1712.03538*. [Online]. Available: http://arxiv.org/abs/1712.03538

[35] G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, "From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses," in *Proc. 2nd Workshop Comput. Linguistics Clin. Psychol., From Linguistic Signal to Clin. Reality*, 2015, pp. 1–10.

[36] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in Twitter," in *Proc. Workshop Comput. Linguistics Clin. Psychol., From Linguistic Signal Clin. Reality*, 2014, pp. 51–60.

[37] I. Fatima, H. Mukhtar, H. F. Ahmad, and K. Rajpoot, "Analysis of user-generated content from online social communities to characterise and predict depression degree," *J. Inf. Sci.*, vol. 44, no. 5, pp. 683–695, Oct. 2018.

[38] M. Robnik-Šikonja and I. Kononenko, "An adaptation of relief for attribute estimation in regression," in *Proc. 14th Int. Conf. Mach. Learn. (ICML)*, vol. 5, 1997, pp. 296–304.

[39] Y. Sun and D. Wu, "A RELIEF based feature extraction algorithm," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2008, pp. 188–195.

[40] T. Kam Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, Aug. 1998.

[41] M. M. Bradley and P. J. Lang, "Affective norms for English words (ANEW): Instruction manual and affective ratings," Citeseer, Center Res. Psychophysiol., Univ. Florida, Gainesville, FL, USA, Tech. Rep. C-1, 1999.

[42] M. Mohamed and P. Gader, "Handwritten word recognition using segmentation-free hidden Markov modeling and segmentation-based dynamic programming techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 5, pp. 548–554, May 1996.

[43] T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse, "An activity monitoring system for elderly care using generative and discriminative models," *Pers. Ubiquitous Comput.*, vol. 14, no. 6, pp. 489–498, Sep. 2010.

[44] S. Bagroy, P. Kumaraguru, and M. De Choudhury, "A social media based index of mental well-being in college campuses," in *Proc. CHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2017, pp. 1634–1646.

[45] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "Discovering shifts to suicidal ideation from mental health content in social media," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2016, pp. 2098–2110.

[46] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[47] J. D. Hwang and K. Hollingshead, "Crazy mad nutters: The language of mental health," in *Proc. 3rd Workshop Comput. Lingusitics Clin. Psychol.*, 2016, pp. 52–62.

[48] M. L. Kern, G. Park, J. C. Eichstaedt, H. A. Schwartz, M. Sap, L. K. Smith, and L. H. Ungar, "Gaining insights from social media language: Methodologies and challenges.," *Psychol. Methods*, vol. 21, no. 4, pp. 507–525, 2016.

[49] P. A. Cavazos-Rehg, M. J. Krauss, S. Sowles, S. Connolly, C. Rosas, M. Bharadwaj, and L. J. Bierut, "A content analysis of depression-related tweets," *Comput. Hum. Behav.*, vol. 54, pp. 351–357, Jan. 2016.

[50] D. Mowery, C. Bryan, and M. Conway, "Towards developing an annotation scheme for depressive disorder symptoms: A preliminary study using Twitter data," in *Proc. 2nd Workshop Comput. Linguistics Clin. Psychol., From Linguistic Signal Clin. Reality*, 2015, pp. 89–98.
[51] D. Mowery, C. Bryan, and M. Conway, "Feature studies to inform the classification of depressive symptoms from Twitter data for population health," 2017, *arXiv:1701.08229*. [Online]. Available: http://arxiv.org/abs/1701.08229
[52] E. Refaee and V. Rieser, "An Arabic Twitter corpus for subjectivity and sentiment analysis," in *Proc. LREC*, 2014, pp. 2268–2273.
[53] *Nafsany Forum*. Accessed: Sep. 24, 2019. [Online]. Available: https://www.nafsany.cc/vb/forumdisplay.php?f=57
[54] L. Baer, D. G. Jacobs, J. Meszler-Reizes, M. Blais, M. Fava, R. Kessler, K. Magruder, J. Murphy, B. Kopans, P. Cukor, L. Leahy, and J. O'Laughlen, "Development of a brief screening instrument: The HANDS," *Psychotherapy Psychosomatics*, vol. 69, no. 1, pp. 35–41, 2000.
[55] M. K. Saad, "The impact of text pre-processing and term weighting on Arabic text classification," M.S. thesis, Islamic Univ., Gaza, Palestine, 2010.
[56] B. Lakshmi and S. Parthasarathy, "Human action recognition using median background and max pool convolution with nearest neighbor," *Int. J. Ambient Comput. Intell.*, vol. 10, no. 2, pp. 34–47, Apr. 2019.
[57] R. Wang and G. Wang, "Web text categorization based on statistical merging algorithm in big data environment," *Int. J. Ambient Comput. Intell.*, vol. 10, no. 3, pp. 17–32, 2019.
[58] M. N. Y. Ali, M. G. Sarowar, M. L. Rahman, J. Chaki, N. Dey, and J. M. R. Tavares, "Adam deep learning with SOM for human sentiment classification," *Int. J. Ambient Comput. Intell.*, vol. 10, no. 3, pp. 92–116, 2019.
[59] X. Yang and X. Jiang, "A hybrid active contour model based on new edge-stop functions for image segmentation," *Int. J. Ambient Comput. Intell.*, vol. 11, no. 1, pp. 87–98, 2020.
[60] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentimental reviews using machine learning techniques," *Procedia Comput. Sci.*, vol. 57, pp. 821–829, Jan. 2015.
[61] A. B. Soliman, K. Eissa, and S. R. El-Beltagy, "AraVec: A set of arabic word embedding models for use in arabic NLP," *Procedia Comput. Sci.*, vol. 117, pp. 256–265, Jan. 2017.
[62] D. Mackay, "Glove," U.S. Patent 852 972, May 7, 1907.
[63] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.
[64] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
[65] S. L. Salzberg, "C4. 5: Programs for machine learning by J. Ross Quinlan. Morgan Kaufmann publishers, inc., 1993," *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, Sep. 1994.
[66] F. Sebastiani, "Machine learning in automated text categorization," *ACM Comput. Surveys*, vol. 34, no. 1, pp. 1–47, Mar. 2002.
[67] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Document Anal. Recognit.*, vol. 1, Aug. 1995, pp. 278–282.
[68] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 2008.
[69] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. COMPSTAT*. Cham, Switzerland: Springer, 2010, pp. 177–186.
[70] T. Basu and C. Murthy, "A feature selection method for improved document classification," in *Advanced Data Mining and Applications. ADMA* (Lecture Notes in Computer Science), vol. 7713, S. Zhou, S. Zhang, and G. Karypis, Eds. Berlin, Germany: Springer, 2012, pp. 296–305.
[71] M. M. Aldarwish and H. F. Ahmad, "Predicting depression levels using social media posts," in *Proc. IEEE 13th Int. Symp. Auto. Decentralized Syst. (ISADS)*, Mar. 2017, pp. 277–280.

NORAH SALEH ALGHAMDI received the bachelor's degree in computer science from Taif University, Taif, Saudi Arabia, in 2004, and the master's degree in computer science and the Ph.D. degree from the Department of Computer Science, La Trobe University, Melbourne, Australia, in 2015. She worked as an Assistant Professor with Taif University, from 2016 to 2017. She has been acting as the Vice Dean of quality assurance with the Deanship of Quality Assurance and Academic Accreditation, PNU, since 2019. She is currently an Assistant Professor with the Department of Computer Science, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests in data mining, machine learning, text analytics, image classification, and deep learning. She participated in organizing the international conference on computing, and in reviewing articles from different journals such as the *International Journal of Web Information Systems*, the *International Journal of Computer Science*, and IEEE ACCESS.

HANAN A. HOSNI MAHMOUD received the Ph.D. degree from the University of Louisiana at Lafayette, in 2001. On leave from the University of Alexandria, Egypt, she established the research unit in the Information Technology Department, College of Computer and Information Sciences, King Saud University. She was the acting Vice Chair of the Computer Science Department, KSU. She is currently an Associate Professor with Princess Nora University, Riyadh, Saudi Arabia. She is very active in the research, where she published more than 40 articles in the field of video processing and VLSI and was awarded five grants in the last four years. She participated in the several international technical committees, Great Lakes Symposium on VLSI in the organizing Committee, and Editorial Review Boards, First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT), Co-Sponsored by IEEE, International Program Committee. She was a Technical Committee Member of the IEEE International Symposium on Systems and Circuits, (ISCASS), a Technical Committee Member and an Organizer of the IEEE Workshop on Signal Processing Systems (SiPS), a Technical Committee Member of the IEEE International Midwest Symposium on Circuits and Systems, and a NITS Reviewer Committee. She was an Editorial Board of the *Journal of Computing and Applications* (JCA) and a Program Committee Member.

AJITH ABRAHAM (Senior Member, IEEE) received the M.Sc. degree from Nanyang Technological University, Singapore, in 1998, and the Ph.D. degree in computer science from Monash University, Melbourne, Australia, in 2001. He is currently the Director of Machine Intelligence Research Labs (MIR Labs), a Not-for-Profit Scientific Network for Innovation and Research Excellence connecting Industry and Academia. The Network with Headquarters in Seattle, USA, he has currently more than 1000 scientific members from more than 100 countries. As an Investigator/Co-Investigator, he received research grants worth more than 100 Million U.S.$ from Australia, USA, EU, Italy, Czech Republic, France, Malaysia, and China. He works in a multidisciplinary environment involving machine intelligence, cyber-physical systems, the Internet of Things, network security, sensor networks, web intelligence, web services, data mining, and applied to various real world problems. In these areas, he has authored or coauthored more than 1300 research publications out of which there are more than 100 books covering various aspects of computer science. One of his books was translated to Japanese and few other articles were translated to Russian and Chinese. About more than 1000 publications are indexed by Scopus and more than 800 are indexed by Thomson ISI Web of Science. Some of the articles are available in the ScienceDirect Top 25 hottest articles. He has more than 700 coauthors originating from more than 40 countries. He has more than 36 000 academic citations (h-index of 89 as per google scholar). He has given more than 100 plenary lectures and conference tutorials (in more than 20 countries). For his research, he received the seven best paper awards at prestigious International conferences held in Belgium, Canada Bahrain, Czech Republic, China, and India. Since 2008, he has been the Chair of the IEEE Systems Man and Cybernetics Society Technical Committee on Soft Computing (which has more than 200 members) and served as a Distinguished Lecturer for the IEEE Computer Society representing Europe from 2011 to 2013. He is currently the Editor-in-Chief of *Engineering Applications of Artificial Intelligence* (EAAI) and serves/served the Editorial Board for more than 15 international journals indexed by Thomson ISI. He is actively involved in the organization of several academic conferences, and some of them are now annual events. More information at: http://www.softcomputing.net/

SAMAR AWADH ALANAZI received the bachelor's degree in clinical psychology from Princess Nourah Bint Abdurrahman University, in 2017. She is currently working as a Clinical Psychologist with Prince Sultan Military Medical City. Her previous research was about Anxiety, Depression, and Stress among mothers of children with and without Cancer in Riyadh.

LAURA GARCÍA-HERNÁNDEZ received the M.Sc. degree in computer science from the Universitat Oberta de Catalunya, Spain, in 2007, and the European Ph.D. degree in engineering from the University of Córdoba, Spain, and also from the Institut Français de Mécanique Avancée, Clermont-Ferrand, France, in 2011.

She has been an Invited Professor during a semester in the Institut Français de Mécanique Avancée, Clermont-Ferrand. She is currently an Associate Professor in project management with the University of Córdoba, Spain. Her primary areas of research are engineering design optimization, intelligent systems, machine learning, user adaptive systems, interactive evolutionary computation, project management, risk prevention in automatic systems, and educational technology. In these fields, she has authored or coauthored more than 60 research publications. She has given several invited talks in different countries. She has realized several postdoctoral internships in different countries with a total duration of more than two years. She received the prestigious National Government Research Grant "José Castillejo" for supporting their post-doc research during six months in the University of Algarve, Portugal. She has been an Investigator Principal in two Spanish research projects and has also been an Investigator Collaborator in some research contracts and projects. She is an Expert Member of ISO/TC 184/SC working team and the National Standards Institute of Spain (UNE). Moreover, she is a member of the Spanish Association of Engineering Projects (IPMA Spain). Considering her research, she received the Young Researcher Award granted by the Spanish Association of Engineering Projects (IPMA), Spain, in 2015. Additionally, she received two times the General Council of Official Colleges Award at prestigious International Conference on Project Management and Engineering both 2017 and 2018 editions. She is the Co-Editor-in-Chief of the *Journal of Information Assurance and Security*. Also, she is an Associate Editor in the following ISI Journals *Applied Soft Computing* and the *Journal of Intelligent Manufacturing*.