IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 18, 2020, accepted September 27, 2020, date of publication October 6, 2020, date of current version October 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3029154

Application of Machine Learning Methods in Mental Health Detection: A Systematic Review

ROHIZAH ABD RAHMAN[1], KHAIRUDDIN OMAR[1], SHAHRUL AZMAN MOHD NOAH[1], (Member, IEEE), MOHD SHAHRUL NIZAM MOHD DANURI[2], (Member, IEEE), AND MOHAMMED ALI AL-GARADI[3]

[1]Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia
[2]Department of Computer Science, Faculty of Science and Information Technology, Kolej Universiti Islam Antarabangsa Selangor, Kajang 43000, Malaysia
[3]Department of Radiology, University of California, San Diego, CA 92093, USA

Corresponding author: Rohizah Abd Rahman (rohizah@ukm.edu.my)

This work was supported by the Professional and Management Incentive Grant, Universiti Kebangsaan Malaysia under Grant GGPP-2019-004.

ABSTRACT This paper presents a critical assessment analysis on mental health detection in Online Social Networks (OSNs) based on the data sources, machine learning techniques, and feature extraction method. The appropriateness of the mental health detection was also investigated by identifying its data analysis method, comparison, challenges, and limitations. This study reviewed articles published in major databases between 2007 and 2018 through keyword searches. The articles were screened base on their titles and abstracts before the full texts were reviewed. The articles were coded in accordance with data set (e.g., data sources, keywords, and geographical locations), method of data analysis, machine learning or deep learning technique, classifier performance, and feature extraction method. 22 articles were selected for review from the total of 2770. As OSNs exhibit high potential as a data source in early detection of mental health problems, most researchers used text analysis on a new data set extracted from different OSNs sources. The extracted data were examined using a statistical analysis or machine learning techniques. Several studies also applied multimethod techniques, which included distributing questionnaires while requesting for the respondents' consent to later access and extract information from his/her OSNs account. Big data in OSNs contribute on mental health problem detection. The presented method is an alternative approach to the early detection of mental health problems rather than using traditional strategies, such as collecting data through questionnaires or devices and sensors, which are time-consuming and costly. However, mental health problem detection through OSNs necessitates a comprehensive adoption, innovative algorithms, and computational linguistics to describe its limitations and challenges. Moreover, referrals from mental health specialists as subject matter experts are also required to help obtain accurate and effective information.

INDEX TERMS Deep learning, feature extraction, machine learning, mental health, online social network.

I. INTRODUCTION

The current changes in the social landscape have contributed significantly to the increase in the rate of mental health problems and psychological disorders. The World Health Organization (WHO) has defined ''mental health'' as the condition of a person who is able to handle his/her stress in life according to his/her ability, but is still able to work normally and productively as well as contribute to the society [1]. Factors that affect mental health probably originate from an individual's way of life, such as work stress, bad financial situation, family issues, relationship problems, and violence, along with environmental factors [2]. These situations can contribute to mental health disorders, such as depression, anxiety, stress, and various psychological disorders that exert an impact on the quality of life and holistic well-being of a person.

Approximately 450 million people worldwide are mentally ill, with the disease accounting for 13% of the global disease burden [3]. WHO estimated that one in four individuals experiences mental disorders in any stage of their lives [4]. In 2018, WHO released a guideline on managing the physical

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

conditions of adults with severe mental health problems. Usually, people will die earlier than the general population if they had severe mental disorders, such as depression, bipolar disorder (BD), psychotic disorder, and schizophrenia [5]. In addition, depression, which can lead to suicidal ideation and suicide attempts, is estimated to affect 350 million people worldwide [6]. Therefore, WHO established a vision wherein people suffering from mental illness are able to recover and live a life like a normal person as outlined in the Comprehensive Mental Health Action Plan (2013–2020) [7].

Mental health problems should be detected and addressed early. Early detection, accurate diagnosis, and effective treatment can alleviate the suffering of people who are dealing with mental health challenges [8]. The effects of mental illness can be severe on the concerned individuals and their families, and on the society as a whole. In general, the traditional methods of mental health detection normally use face-to-face interviews, self-reporting, or questionnaire distribution. However, traditional methods are typically labor-intensive and time-consuming [9]. Thus, previous studies have applied technologies, such as wearable sensors and smartphones in healthcare and mental health detection; however, these technologies are typically used by individuals who have been diagnosed with mental illness and have been monitored over time [10]–[13].

A recent research presented a novel approach of mental health problem detection in online social networks (OSNs) [9], [14]–[34]. OSNs have become popular in recent years and have provided a new medium to communicate and share information. OSNs are used regularly by millions of people worldwide [2]. Through OSNs, users can express their feelings and thoughts by posting different types of data (e.g., text, images, videos, and audios) regarding their daily activities. They can also communicate with friends by commenting on the posts of others [35]. Thus, this new research trend is related to big data research and the increasing availability of resources on the Internet through OSNs (e.g., Facebook, YouTube, Twitter, Instagram, and Sina Weibo). Such increase leads to data overflow, and these huge amount of communication data have become content generators that may be useful for further exploration and analysis [14], [36]–[38]. Therefore, OSNs can generate a massive amount of information that can be used to develop an approach for mental health problem detection [2].

Researchers from the West and the East harvest data from OSNs, such as Twitter, Facebook, and Sina Weibo, and use them as data source for online studies and crowdsourcing. The types of mental health problems detected in OSNs included psychological stress [9], [14], [16]–[18], [22], [24], [29], depression [19], [23], [26]–[28], [30], [31], mental disorders [15], [20], [33], [34], and suicidal ideation [21], [25], [32]. An analysis of the current mental health detection in OSNs is required to comprehend data sets, data analysis methods, feature extraction method, classifier performance (i.e., accuracy and efficiency), challenges, limitations, and future work. The purpose of this systematic review is to conduct a critical assessment analysis of mental health problem detection based on data extracted from OSNs. It intends to explore the competence of mental health problem detection in OSNs, including its challenges, limitations, and future work.

Two common ways to analyze the data from OSNs texts posted by the users are by using dictionary-based and machine learning method. However, there are limitations in using both methods. Thus, nowadays researchers have begun to find alternative ways to improve the performance and efficiency of the analysis. Traditional machine learning faces common training problems, such as overfitting, model interpretation, and generalization [39]. So, researcher moved to deep learning techniques which have emerged in recent years as a powerful tool. This is because machine learning can solve more complex problems, especially in health data [39], [40]. We discuss this in the results section.

Specifically, this systematic review is organized as follows: (1) The first section discusses the objectives of this paper; (2) The second section describes the systematic review methodology with a discussion on the types of mental health problems, the strategies of searching, the criteria of screening and selection, and the quality valuation of the selected articles; (3) The third section presents the results of each study, the data sets (analysis of data source, keywords, duration, and geographical location), feature extraction method, classifier performance, and machine learning techniques used in each study; (4) The fourth section displays the challenges of this study; (5) The fifth section recognizes the limitations of this systematic review; (6) The sixth section offers criticisms for future implications; and (7) The seventh section concludes the study.

II. METHODS

The main purpose of this paper is to explore the adequacy, challenges, and limitations of a mental health problem detection based on OSNs data. The objective of this systematic literature review is to conduct a critical assessment analysis on detection of mental health problems using OSNs. This analysis consists of the data source, the feature extraction method, and classifier performance in machine learning techniques. We also investigated the appropriateness of this pre-mental health detection by identifying its data analysis method, comparison, challenges, and limitations.

A. IDENTIFICATION AND SELECTION OF STUDIES

Most studies that used OSNs as their data source in mental health problem detection were included in the selection. OSNs enable online social interaction among users and facilitate information exchange, as well as diffusion. Therefore, OSNs have created a massive amount of data and offer a distinctive chance for learning and understanding of social communication and interaction among considerably superior populations [41]. Mental health problems refer to psychological disorders, such as stress, depression, anxiety, suicidal ideation, and distress. The current study explains how

TABLE 1. Criteria of inclusion and exclusion.

Criteria of Inclusion	Criteria of Exclusion
Articles in English language (AND)	Studies that unrelated to OSN data (AND)
Articles published between 2007 and 2018 (AND)	Studies that unrelated to mental health problems (AND)
Articles that used user-generated data in OSNs (AND)	Studies that unfulfilled any inclusion criterion
Articles that studied a type of mental health problem (AND)	
Articles that studied user-generated data while keeping users' privacy	

previous researchers used OSNs as their data source in mental health problem detection.

This systematic literature review was conducted to examine the use of social media in mental health detection. The process followed the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) to outline and assess relevant articles [42]. The detailed process of selecting articles was based on the criteria of inclusion and exclusion as described in Section II (C). We used an electronic literature search for suitable articles from the PubMed, IEEE Xplore®, Scopus®, and ScienceDirect databases. The search keywords were as follows: "online social network," "social media," "Twitter," "microblog," "Facebook," "YouTube," "Weibo," "mental health," "mental illness," "mental disorder," "stress," "depression," "anxiety," "suicide," and "PTSD." These researchers used common mental health disorder keywords as defined by the United Kingdom National Institute for Health and Care Excellence, such as "psychological stress", "depression", "anxiety", "suicide", and "posttraumatic stress disorder (PTSD)" [43]. We also referred to the Medical Subject Headings to ensure that the key terms used in mental health are inclusive in the literature [44]. The results indicated that 22 articles were revised as possible suitable studies.

B. CRITERIA OF INCLUSION AND EXCLUSION

After further screening based on the titles and abstracts of articles using the search term in Table 1, the total number of collected articles was 2770. Then, we removed the duplicate articles, hence the total number was reduced to 2735. In the screening stage, two reviewers evaluated the articles individually. 270 competent articles were then distributed to three reviewers for screening and the articles would be proceeded to another stage after meeting the criteria. The criteria of inclusion were as follows: (1) the articles published between 2007 and 2018 in English language, (2) the data were extracted from OSNs, and (3) the articles discuss any type of mental health problems.

After the screening stage, the articles were sent to four reviewers for further evaluation of the articles' eligibility according to Table 1. Then, the reviewers compared and discussed their findings until they reached a consensus and agreement. Lastly, only 22 articles were outlined for the detailed analyses.

C. METHODOLOGICAL QUALITY ASSESSMENT

This systematic review adopted the Critical Appraisal Skills Program (CASP) checklist for evaluating several methodological quality articles comprehensively [45]. The major features and limitations based on the extraction of data such as data source, keywords, duration, and geographical location of data extracted; the quality of data such as data set related to mental health problems; study design such as suitable methodology applied; and the results such as clear study objectives and outcomes, were analyzed and compared to indicate the strengths and weaknesses for each of the studies. The selected articles which have met the criteria of inclusion and exclusion are as provided in Table 1. The articles were selected based on the title, objectives, outcomes, findings, data set, feature extraction method, machine learning techniques, and classifier performance.

III. RESULTS

A. FINDING AND SELECTING STUDIES

In the searching process, 2770 articles were collected. Then, the total number was decreased to 2735 after the duplicates were removed. The irrelevant articles were discarded after the titles and abstracts underwent screening. There were 270 full texts articles that remained for auxiliary screening and were studied in accordance with the criteria outlined in Table 1. Lastly, in this systematic review, we only selected 22 prospective articles to be analyzed intensively before we outlined the comprehensive analytical results and findings.

The 22 articles were published between 2007 and 2018 in English language. A flowchart of citation from the early stage of searching until the final number of articles selected is provided in Figure 1. The figure adopted and followed PRISMA [42]. As presented in Table 2, major characteristics and a summary of the content analysis of the articles are discussed. The analysis of the 22 selected articles was based on the criteria of inclusion and exclusion as provided in Table 1. The summary of the selected articles were in accordance with the data set (data sources, keywords, duration and geographical location), method of data analysis, study objectives, feature extraction method, machine learning techniques, and classifier performances. This paper discusses the results and findings of the selected articles in Section III. The articles also touched on mental health problem detection in OSNs as implemented in various countries. The studies also described the different types of mental health problems.

B. DESCRIPTION OF SELECTED STUDIES

The comprehensive analytical results and findings of the 22 selected articles are discussed in Table 2, Table 3, and Table 4. The summary focused on data extraction (data source, keywords, duration, geographical location, and data

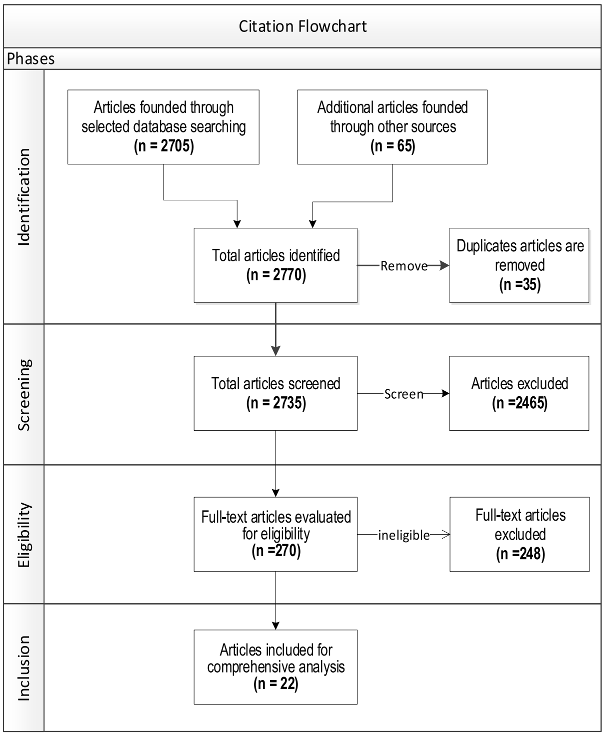

FIGURE 1. Flowchart of Citation from the First Stage of Identification until the Final Stage of Inclusion articles.

set) from OSNs, feature extraction method and objectives, data analysis methods, and classifier performance. The selected studies were categorized distinctly based on the data set, type of mental health problems, data source based on geographical location extraction, and OSNs type. Majority of the studies used new data sets extracted from OSNs mental health problem detection [9], [14], [15], [18]–[27], [29]–[34]. Only one researcher derived the data from another work [28]. The other categories focused on the types of mental health detection. Psychological stress was investigated in eight studies [9], [14], [16]–[18], [22], [24], [29]. Meanwhile, depression was explored in seven studies [19], [23], [26]–[28], [30], [31]. Suicide, on the other hand, was investigated in three studies [21], [25], [32], while mental disorders were examined in four studies [15], [20], [33], [34].

The selected studies also categorized the extracted data on the basis of geographical location. For instance, China was investigated in six studies [9], [16], [22], [25], [28], [29]. Other researchers studied mental illness in the United States [18], [19], [24], [26]. Only one study each was conducted in Japan [23] and Greece [14]. Other studies did not mention the data source based on geographical location extraction. Another category was the type of OSNs data. Several studies used Twitter as data source [17]–[19], [21], [23], [27], [32], [34]. Others used Facebook [14], [31], Sina Weibo [9], [22], [25], [28], [29], and microblog [16]. Lastly, various data were extracted from other types of online sources or crowdsourcing [15], [20], [24], [30], [33].

C. DATA SET

Nearly all the researchers in Table 2 prepared their own data sets, except for one, who used data from another study. The data sets provided were based on the country where the research was conducted, the type of OSNs used, and the duration of data extraction. The advantage of using an original data set is that the information is specific to the objectives of the research depending on the location. Several data were extracted directly from public posts in Twitter or Sina Weibo. Three types of data were extracted. The first type included data extracted from public application programming interface services, such as Twitter [17]–[19], [21], [23], [26], [27], [32], [34] or Sina Weibo [9], [22], [25], [28], [29]. Sina Weibo is similar to Twitter but is available only in China. The second type of data was extracted from Twitter and Facebook user accounts after obtaining their consent. The users were detected of having mental health problems based on their answers to the questionnaires distributed by the researchers [14], [19], [20], [23], [26], [29], [31]. The third type of data was extracted from online sources, such as web fora, microblogs, online communities, crowdsourcing, and Amazon's Mechanical Turk (MTurk) [15], [16], [20], [24], [30], [33].

1) GEOGRAPHICAL LOCATION

Several countries have already begun studying mental diseases within their territories. The extracted data were tagged with their country of origin, such as China, Japan, and the United States. Several studies used their native language and English when preparing their data sets. Table 2 indicates that several researchers extracted data from their respective countries, such as China [9], [16], [22], [25], [28], [29], the United States [18], [19], [24], [26], Japan [23], Greece [14], and other locations that were not provided specifically in the articles.

2) DATA SOURCE

Extracting data from OSNs is the primary purpose of this research on mental health. Determining which OSNs should be used is important. As shown in Table 2, different types of OSNs were used as data sources by previous researchers, including Twitter, Facebook, Sina Weibo, microblogs, and other online sources. Most of the researchers preferred data from Twitter [17]–[19], [21], [23], [26], [27], [32], [34] because it is accessible to the public [36]. Twitter data are rich in information, such as user name, user ID, biography, screen name, URL, account creation data, and tweet texts [36]. Other researchers used Facebook as their data source. Nevertheless, researchers had to request consent from the users who had already been identified as having mental health problems to extract data from their Twitter or Facebook accounts [14], [19], [20], [23], [26], [29], [31]. In China, Sina Weibo has become the choice of many researchers because this OSNs is similar to Twitter [9], [22], [25], [28], [29]. Other researchers used data extracted from microblogs [28], [29].

TABLE 2. Studies related to the detection of mental health in OSNs.

Author/s	Mental Health Types	Data Source (OSNs)	Keywords	Durations	Geo-Location	Data Set
Lin et al. 2017 [9]	Stress	The authors collected 350 million tweets from Sina Weibo.	The authors used tags and comments.	Oct 2009-Oct 2012	China	Four data sets with the first consist of 19,000 tweets (with stress label) and 17,000 tweets (with non-stressed label).
Kandias et al. 2017 [14]	Stress	Data extracted from Facebook.	Not specified	Not specified	Greece	The data set was generated from Facebook users who provided their informed consent (405 fully crawled, 12.346 groups, 98.256 liked objects, 171.054 statuses, and 250.027 comments).
Thelwall 2017 [17]	Stress	3066 English tweets collected from Twitter	Keywords from a variety of sources	1 month (July 2015)	Not specified	The data collected from 3000 stress-related tweets and classified under stress level with scale 1 to 5.
Shuai et al. 2017 [20]	Mental Disorders	Data were collected among 3126 OSNs users via Amazon's MTurk.	Not specified	Not specified	Not specified	3126 OSNs users (1790 males and 1336 females); 389 users were labeled with social network mental disorders, 246 had "Cyber Relationship Addiction," 267 had "Information Overload," and 73 had "Net Compulsion."
O'Dea et al. 2015 [21]	Suicide	14,701 suicide-related tweets collected from Twitter.	Several keywords were derived from [47].	February 18 and April 23, 2014	Not specified	The total is 1820 tweets. Set A consist 829 tweets (training: 746, testing: 83) and Set B consist 991 tweets (Training: 891, and Testing: 100).
Lin et al. 2014 [22]	Stress	600 million tweets collected from Sina Weibo.	Not specified	Oct 2009-Oct2012	China	600 million tweets with five categories of stress such as affection, work, social, physiological, and others.
Tsugawa et al. 2015 [23]	Stress	3200 tweets collected from Twitter.	Not specified	December 4, 2013 to February 8, 2014	Japan	An experiment conducted with 219 participants. Only 214 participants were involved. Data from the participants (males: 121; females: 88) aged 16–55 years were analyzed.
Saleem et al. 2012 [24]	Distress	Data collected from web fora related to psychological issues.	Not specified	Not specified	United States	136 psychological distress labels ranging from PTSD to mild traumatic brain injury and depression symptoms developed via consultation with psychologists.
De Choudhury et al. 2013 [26]	Depression	1583 data of crowd workers (crowdsource: MTurk) who shared their Twitter public profile	Not specified	September 15 to October 31, 2012	United States	637 participants provided access to their Twitter feeds. Subsequently, 476 users (243 males and 233 females) diagnosed with depression.
Deshpande and Rao 2017 [27]	Depression	10,000 tweets collected from Twitter.	Words related to "poor mental well-being"	Not specified	Not specified	10,000 tweets collected to generate training and test datasets with a ratio of 80:20. The training set consists of words that suggest depression tendencies, such as "depressed," "hopeless," and "suicide."
Huang et al. 2014 [25]	Suicide	53 verified suicidal users and over 30,000 posts from Sina Weibo. They obtained 614 true suicidal posts.	Not specified	Not specified	China	614 suicidal posts obtained. To perform a 90 to 10 test, 6140 posts randomly selected from the set of non-suicidal users, for 6754 posts. Finally, 6704 posts obtained.
Wang, Zhang, and Sun 2013[28]	Depression	Data collected from Sina Weibo.	Not specified	August 1–15, 2012	China	The data set was derived from [48].
Xue et al. 2014 [29]	Pressure	The tweets of 459 middle school students (aged 14–20 years) collected from Sina Weibo.	Not specified	July 7, 2013.	China	23 teenagers posted 300 to 1000 tweets and 10,872 tweets. The average number of tweets is 473 per teenager.

TABLE 2. ***(Continued.)*** **Studies related to the detection of mental health in OSNs.**

Saha et al. 2016 [30]	Depression	Data crawled from 620,000 posts made by 80,000 users in 247 online communities.	Not specified	Not specified	Not specified	Data from the Live Journal website contains 620,060 posts from 78,647 users.
Wongkoblap, Vadillo, and Curcin 2018 [31]	Depression	Data of "myPersonality" crawled from Facebook and derived from [49].	Not specified	Not specified	Not specified	Two datasets: 1) self-reported results from Satisfaction with Life Scale, and 2) self-diagnosed results from volunteers who answered the CES-D questionnaires.
Luo et al. 2018 [32]	Suicide	Data collected from 716,899 tweets from Twitter.	Suicide-related terms	January–November 2016	Not specified	191,473 precise suicide-related tweets
Tai et al.2015 [33]	Mental Disorder	Data collected through online social platforms.	Related to depression and being healthy.	Not specified	Not specified	Data from the depressed (keywords: bad, hate, and hard) and healthy (keywords: love, best, and hope) groups based on the extracted keywords.
Saravia et al. 2016 [34]	Mental Illness	Data collected from Twitter.	Related to BPD and BD.	Not specified	Not specified	Data collected from community portals consist of 17 BPD and 12 BD. Each portal has 5000 followers in 145,000 accounts. After filtering, 278 BD and 203 BPD remained.
Chang, Saravia, and Chen 2016 [15]	Mental Disorder	Subconscious crowdsourcing.	Keywords related to BPD and BD	Not specified	Not specified	Data were collected from community portals consist 12 BD and 17 BPD. Then, 5000 followers downloaded from each portal with 145,000 user accounts.
Li et al. 2016 [16]	Stress	Data extracted from a microblog.	Not specified	January 1, 2012 to February 1, 2015	China	**Set 1**: Stressor event (273 from 1 January 2012 to 1 February 2015 related to study; 122 related events, such as examination, contest, and result notification). **Set 2**: Post (124 students who actively used Tencent Weibo. Post from 1 January 2012 to 1 February 2015, 29,232 posts. The average post is 236, the maximum is 1387 and the minimum is 104.
Coppersmith, Harman, and Dredze 2014 [18]	Post-Traumatic Stress Disorder (PTSD)	3200 tweets collected from Twitter.	Related to PTSD	Not specified	United States (military)	260 tweets indicated a diagnosis of PTSD. After filtering, only 244 users were positive samples.
Park et al. 2012 [19]	Depression	65 million tweets collected from Twitter.	Related to depression	June to July 2009	United States	21,103 tweets consist the word "depression" from 14,817 users. 1000 of random tweets (500 tweets from each month) selected for content analysis. 165 participants and 69 participants (male=28, female=41) are active.

In addition, various online resources were used for data extraction, such as Amazon's MTurk, web fora, and online communities. Expectedly, other public resources from different types of OSNs will be possibly used for data extraction in various types of research in the future.

3) DURATION OF DATA EXTRACTION

Only a few researchers provided the duration of data extraction for their original data sets. Several researchers indicated the longest period of data extraction to be approximately three years [9], [16], [22]. Meanwhile, other researchers extracted data for less than one year [17], [19], [21], [23], [26], [32], with the shortest being only one week [28]. None of the researchers stated that the duration of data extraction plays a role in mental health problem detection. Evidently, a conclusion can be drawn that the longer the data extraction duration, the more data can be collected.

4) KEYWORDS

Most of the researchers did not specify the keywords they used in preparing the new data set for their research. Only a few researchers provided their keywords. For example, O'Dea *et al.* (2015) used keywords such as "suicidal," "kill myself," "end my life," "never wake up," "die alone," and "go to sleep forever" in their research on suicide. These keywords were obtained from [21]. Other researchers used keywords from previous studies and reused them in detection research related to suicide attempts. Meanwhile, several researchers indicated that the keywords used in their studies were related to "poor mental well-being" [27]. Overall, the use of keywords was not specified in the selected studies. However, the use of keywords is an important step in conducting research on mental health problem detection. Researchers should seek expert advice to ensure that appropriate keywords related to mental health problems were adopted.

D. FEATURE EXTRACTION

Nearly all the researchers in Table 2 prepared their own data sets, except for one, who used data from another study. Most previous researchers developed mental health detection model based on text content. The text classification process would through basic steps, such as feature extraction, dimension reduction, classifier selection, and evaluation [50]. Feature extraction is a process that reduces the number of resources without losing the syntactic and semantic relations between words. It generates new features from original features and implements a considerable number of variables and complex data. Feature extraction can also reduce the amount of redundant data in a given analysis.

Most common techniques of feature extractions are Term Frequency-Inverse Document Frequency N-Gram, Bag of Word (BoW), (TF-IDF), Word2Vec, and Global Vectors for Word Representation (GloVe). Previous researchers implemented N-Gram features (e.g., unigram, bigram, and trigram) to create word tokens [15], [17], [21], [24], [25], [34]; and TF-IDF features to capture frequent and representative words [14], [15], [21], [24], [34]; bag-of-words [23], [27], and parts of speech [25], [27]. Therefore, Table 3 shows that indicating the feature extraction methods is a crucial step in text classification for mental health problem detection.

Besides the popular feature extraction method above, other researchers implemented Linguistic Inquiry and Word Count (LIWC) and Latent Dirichlet Allocation (LDA) approaches. LIWC is a prevalent sentiment tool through text analysis for feature extraction which is regularly used to capture language characteristics and categories according to psychological perspectives [51]. Several types of linguistic and behavioral features are important in detecting mental health problems. Most studies have used LIWC for feature extraction to detect mental health problems in OSNs by predicting language characteristics [9], [15], [18], [22], [24], [26], [33]. LDA is another technique used to generate topic models in feature extraction by assigning a topic of each word after extracting it in a document [52]. There were several mental illnesses identified using LDA technique after being applied to millions of users posting in OSNs [49]. Tsugawa *et al.* (2015) used LDA in mental health detection in OSNs to create topic models [23].

TABLE 3. Feature extraction methods used in supervised machine learning studies.

Author/s	Feature Extraction Method
Lin et al. 2017 [9]	This study used LIWC2007 in categorizing to positive or negative emotion words.
Kandias et al. 2017 [14]	The feature selection using frequency of TF–IDF and term occurrence.
Thelwall 2017 [17]	All the features used were labeled unigrams, bigrams, and trigrams.
O'Dea et al. 2015 [21]	This research used the basic features of word frequencies or unigrams and TF–IDF instead of simple frequency.
Lin et al. 2014[22]	This study used LIWC2007 for linguistic features.
Tsugawa et al. 2015[23]	This study used bag-of-words for word frequency and LDA for topic models.
Saleem et al. 2012 [24]	The article adopted bag-of-words, unigrams, and TF–IDF. LIWC used for linguistic features.
De Choudhury et al. 2013[26]	This study used LIWC to determine 22 specific linguistic styles.
Deshpande and Rao 2017 [27]	This research used part-of-speech (e.g., adjective, noun, and verb) tags and bag-of-words.
Huang et al. 2014 [25]	This study used N-gram features (unigram, bigram, and trigram) and classified emotional words to positive or negative from a lexicon. Three types of part-of-speech tags (i.e., adjective, noun, and verb) adopted.
Tai et al. 2015 [33]	This research considered the unigram word feature and an LIWC lexicon to determine PTSD user.
Saravia et al. 2016 [34]	This study adopted TF–IDF to model the linguistic features of patients and their pattern of life features (e.g., age and gender). Then, TF–IDF was applied to unigrams and bigrams collected from all the patients' tweets.
Chang et al. 2016 [15]	In this study, TF–IDF applied to unigrams and bigram (calculated the frequency-of-word sequences) and LICW for linguistic features.
Coppersmith et al. 2014 [18]	LICW used to determine the linguistic style of users with PTSD.

E. MACHINE LEARNING TECHNIQUES

Machine learning techniques are currently popular approaches for mental health problem detection. Previous research [9], [14], [17], [21]–[29], [31], [34] used classification techniques in detecting many types of mental health problems, such as stress, suicidal ideation, distress, and depression. Many approaches were developed for the analysis of data used in various types of mental health problem detections. Several researchers created new methods, such as TensiStrength [17] and the Social Network Mental Disorder Detection (SNMDD) model [20]. Other scholars developed

TABLE 4. Classifier performance of studies using supervised machine learning techniques.

Authors	Objectives	Method of Data Analysis	Machine Learning/Deep Learning Technique	Classifier Performance
Lin et al. 2017 [9]	This research proposed a hybrid model for detecting stress by using user content and social interaction in Twitter.	A hybrid model to analyze the data and compared with others machine/deep learning techniques.	• Hybrid (FGM+CNN) • LR • SVM • RF • Gradient-boosted DT • DNN	The **hybrid model** (FGM+CNN) achieved the highest detection performance, i.e., an improvement of 6%–9% in the F_1 score.
Kandias et al. 2017 [14]	This research study the stress level chronicity experienced based on the content posted by the OSNs users.	Content classification was analyzed using machine learning techniques.	• Multinomial NB • SVM • Multinomial LR	**SVM** was selected because it achieved more than 70% (precision, recall, and F score) and better F score values in most categories.
Thelwall 2017 [17]	This research provided a stress and relaxation detection system posted in OSNs.	A new method called TensiStrength compared with machine and deep learning techniques.	• TensiStrength • AdaBoost • SVM • NB • J48 tree • JRip rule • LR • DT	**TensiStrength** can possible to detect expressions of stress and relaxation with accuracy level through user post in Twitter.
Shuai et al. 2017 [20]	This research proposed a framework that can accurately identify potential cases of social network mental disorders.	A new method called SNMDD	• TSVM	**SNMDD** is a new method that possibility to identified a mental disorders' users through OSNs.
O'Dea et al. 2015 [21]	This research examined the suicide level.	Data compared with human coding and machine learning techniques.	• LR • SVM	**SVM** with TF-IDF and without filter was the best performing algorithm.
Lin et al. 2014 [22]	This research proposed a method of stress detection through OSNs.	The primary model, called deep sparse neural network, compared with SVM and ANN while implementing an Auto-Encoder.	• Deep Sparse Neural Network • SVM • ANN	**SVM** achieved good accuracy in linguistic attributes. **Deep Sparse Neural Network** achieved better results in social, linguistic, and visual attributes.
Tsugawa et al. 2015 [23]	This research introduced a method that can recognize depression through users' activities in social media.	Models predict the risk of depression with several features obtained from user activities in Twitter using SVM.	• SVM	**SVM** achieved an accuracy of approximately 70%.
Saleem et al. 2012 [24]	This research proposed a novel technique with a multistage text classification framework to assess psychological status in web fora.	SVM used for distress detection, and MLN was used for noisy distress label detection.	• SVM • MLN	**MLN** provided statistically significant gains over SVM.
De Choudhury et al. 2013 [26]	This research explored the potential of using OSNs to detect depressive user.	SVM used to predict depressed users.	• SVM	**SVM** classifier could predict depression with promising results of 70% classification accuracy.
Deshpande and Rao 2017 [27]	This research applied NLP to Twitter feeds for emotions related to depression.	SVM and NB classifier used for the class prediction process.	• SVM • NB	The result showed that **NB** gained an F_1 score of 83.29, while **SVM** scored 79.73. The precision and recall results were similar. The accuracy of NB was 83% and that of SVM was 79%.

TABLE 4. *(Continued.)* **Classifier performance of studies using supervised machine learning techniques.**

Huang et al. 2014 [25]	This research proposed a real-time system to detect suicidal ideation users.	Data compared with machine learning techniques.	• SVM • RF • J48 tree • LR • Sequential minimal optimization	**SVM** exhibited the best performance, with 68.3% (F-measure), 78.9% (precision) and 60.3% (recall).
Wang, Zhang, and Sun 2013 [28]	This research proposed a model to detect depressive users based on node and linkage features.	DT used as the classifier in a node feature only model.	• DT	The **DT** classifier with node features achieved the highest accuracy was 95% with increased to 15%. The addition of linkage features resulted in a considerably better performance than that with only node features.
Xue et al. 2014 [29]	This research analyzed a psychological pressure experienced by adolescents by collecting data from a microblog.	Data compared with machine learning techniques.	• NB • SVM • ANN • RF • Gaussian process	The **Gaussian process** classifier achieved the highest detection accuracy.
Wongkoblap, Vadillo, and Curcin 2018 [31]	This research explored the relationship between life satisfaction and depression through OSNs.	Data compared with machine learning techniques.	• SVM with RBF • LR • DT • NB	The accuracy of **SVM** with RBF kernel was the best model achieved to 68%.
Saravia et al. 2016 [34]	This research proposed a novel data collection to build a predictive models base on users linguistic and behavioral patterns.	Data analyzed using an RF classifier.	• RF	The precision of **RF** was 96% for BPD and BD.
Chang, Saravia, and Chen 2016 [15]	This research developed a model to determine mental disorders user base on users linguistic and behavioral patterns.	Data analyzed using an RF classifier.	• RF	The performance of stressor event detection using **RF** was approximately 13.72% (precision), 19.18% (recall), and 16.50% (F_1 measure).

hybrid methods that comprised two types of machine learning techniques [9] and implemented deep learning techniques in their research [22]. Majority of past researchers compared different types of available machine learning techniques. The most commonly used machine learning techniques in mental health problem detection included Support Vector Machines (SVM), Naïve Bayes (NB), logistic regression (LR), random forest (RF), Decision Tree (DT), Gaussian Process, K-means, and Artificial Neural Network (ANN). This algorithm is appropriate in performing text classification. Most of the text classification processes will go through for four phases: feature extraction, dimension reduction, selection of machine learning techniques, and evaluation [50]. Generally, once the date is filtered, a formal feature extraction method is implemented. Then, the data will be classified using machine learning technique. Furthermore, all the machine learning techniques are compared to the efficiency of each classifier based on the metrics of Precision, Recall, F-measure, and Accuracy [53]. These few years, deep learning technique has become popular and achieved surpassing results in comparison to previous machine learning techniques in text classification [50]. Previous research implemented deep learning techniques such as Deep Neural Network (DNN) [53] and Sparse Deep Neural Network [22] in mental health problem detection.

Table 4 lists the machine learning techniques used in mental health problem detection. The most commonly used technique is SVM, which was applied by 13 researchers [14], [17], [20]–[27], [31], [32]. SVM has become one of the standard machine learning and data mining tools used to solve linear and nonlinear two-group classification problems [54], [55]. The second most used machine learning or deep learning techniques were LR and RF, which were both adopted by five researchers. NB ranked third with four researchers choosing this technique. By contrast, the least used machine learning or deep learning techniques are AdaBoost, Gaussian process, gradient-boosted DT, the hybrid technique of factor graph model (FGM) with Convolutional Neural Network (CNN), JRip Rule, Markov logic networks (MLNs), Multinomial LR, Multinomial NB, sequential minimal optimization, SVM with Radial Basis Function (RBF), TensiStrength, and Transductive SVM (TSVM) [9], [17], [20], [29].

IV. DISCUSSION AND CHALLENGES

The information available in OSNs provides a huge amount of data with immense potential to be explored in modern research. Millions of data can be extracted from OSNs to understand the phenomenon selected for a study. Therefore, many researchers used data from OSNs in recent studies, such as those related to pandemics and cyberbullying. In this work, the researchers focused on mental health problem detection through OSNs.

Several findings can be referred to by researchers for future studies. First, only a few studies on mental illness found useful information, such as people with mental health problems isolating themselves and not communicating with others. Lin *et al.* (2017) reported that people with mental health problems are less complicated and less connected to other users in their social structures compared with non-stressed users [9]. Furthermore, Park *et al.* (2014) determined that people with depression will only post about himself/herself without interacting with others [19]. Meanwhile, another researcher found that depressed individuals exhibit lower social activities but higher in negative emotions, self-attraction, medical dependency, and religiosity judgments [26].

Second, the researchers found an additional finding related to the use of different languages in mental health problem detection during data analytics in OSNs. Tsugawa *et al.* (2015) detected people with depression through different languages from the data source using the same method introduced by De Choudry *et al.* (2013) [23], [26]. The features obtained from the Japanese language were able to predict depression with 69% accuracy similar to that of the English language [23], [26].

In general, several challenges are involved in mental health problem detection in OSNs due to a number of issues in non-face-to-face communication and human–computer interaction. The major challenge typically arises from language barrier; this situation commonly occurs while determining the exact meaning of mental health behind the words and languages written in OSNs [9], [16], [23], [25]. Several approaches are available to solve this issue. For example, the use of machine learning can help understand and determine the possibility of an existing mental health problem behind the words and languages written in OSNs. Another challenge is the account privacy policy that is currently being enforced by most OSNs service providers; this policy has made extracting data from OSNs difficult for most researchers [14]. Privacy and security policies are among the challenges faced by researchers during data preparation due to the collection of public user data, such as those from Twitter. In this review, we identified several challenges that should be addressed in the future. These challenges are described as follows.

A. QUALITY OF DATA SETS AND MODEL INTERPRETATION

As mentioned previously, nearly all the researchers listed in Table 2 prepared their own data sets, except for one who used from another study. On the one hand, the advantage of using an original data set is that the information will be specific to the objectives of a research depending on the location. On the other hand, creating an original data set can impose several biases. Most data that represent human behavior can be marred by undesirable and unconscious biases. Machine learning-based systems should prudently reflect how biases affect the data that are required to train a model [56], and practices that report and control such biases should be adopted. The power of machine learning, but also one of its weaknesses, is the ability of such algorithms to distinguish features in historical data that humans cannot easily find [57]. Simple models that normally involve simple tasks are interpretable given that the number of features assessed by the model is relatively small and the relationship among them is easy to understand. However, complex tasks fundamentally include several features that may work collaboratively within a model to make a prediction; therefore, interpreting such models is likely difficult because several complex statistical patterns that correlate to minor indicators across various features are involved [57].

B. MENTAL HEALTH PROBLEM DETECTION OVER TIME

One of the interesting and challenging tasks is mental health problem detection over time. In contrast with other text classification tasks, mental health status can vary significantly over time. For example, a mental health case reported in an OSNs website may begin with a simple mental health issue (i.e., a weak signal) and end with a suicide case (i.e., a strong signal). Consequently, different mental health scenarios that change over time should be considered while building machine learning models. A model should be effective in detecting weak signals and in continuously evolving mental health detection cases over time.

C. MULTICATEGORIES OF MENTAL HEALTH PROBLEMS

The researcher also found that most previous studies outlined mental health problems as a general problem without focusing on a specific type of mental health problems. Most previous studies did not consider multicategories of mental health problems except for a few that included this issue in their research focus. Categorizing mental health problems is difficult because numerous feature selection processes should be performed by researchers. Consequently, researchers choose to either generalize or specify the category of a mental health problem.

D. DATA PREPROCESSING

The preparation of a new data set is one of the challenges in mental health problem detection. Researchers implement novel data preparation techniques that include a new set of features to detect mental health problems using OSNs. Researchers should also select a machine learning technique to train data extraction from OSNs.

E. LENGTH OF POSTS

One of the unique features of OSNs is their short post length, which can range from a few words to a few sentences in several social media websites. This feature makes mental health problem detection different from previous research fields that used only short texts such as sentiment analysis to classify whether a post is positive or negative. In mental health problem detection, the analysis of long text blogs will definitely be more informative than the analysis of posts on social media websites, such as tweets. Consequently, understanding mental health problems from a limited post length is one of the important challenges that must be considered when developing machine learning models for mental health problem detection.

F. MULTILINGUAL CONTENT

Posts in OSNs can be simply written using a specific language or a combination of several languages within a single post. The analysis of multilingual texts is an exciting challenge that must be addressed in the future.

G. DATA SPARSITY

The data sparsity problem can reach a considerable extent in OSNs due to the high level of informal textual distinctiveness. Addressing the data sparsity problem is crucial because it may negatively affect the performance of a machine leaning model. Considering the practical significance of minimizing data sparsity, future research should further investigate this challenge.

H. PUBLICLY AVAILABLE DATA SETS

This systematic review shows that only a few studies have used publicly available data sets. However, publicly available data sets are important in creating benchmark data sets to compare the performance of different algorithms proposed in previous studies on mental health problem detection. The creation of publicly available data sets while preserving user privacy is another challenge that must be addressed in the future.

I. DATA QUANTITY AND GENERALIZABILITY

In recent years, deep learning algorithms have become popular approaches with the ability to improve text classification performance. The major advantage of deep learning is automatically extracting features without requiring human experts. However, one of the challenges that limits the application of deep learning models is generalizability. For example, a model trained on the data from specific users may not perform well on data from other users with different writing styles. Moreover, a model trained on user data from the United Kingdom or the United States may not perform well on user data from Asia. Consequently, a large dataset is required to comprehensively cover the patterns of most users to have more generalized models. Collecting extensive and diverse data can make deep learning models more generalizable and less vulnerable to bias [56].

J. ETHICAL CODE

Researchers should fully understand the ethical code of conduct before collecting data from OSNs and should apply good research practice by sending permission requests to OSNs users and providers. Most previous researchers used one type of OSNs as data source. They also performed text analysis in OSNs for mental health problem detection. Interpreting mental health problems is difficult because OSNs users typically use other types of data to express themselves.

V. LIMITATIONS

This study exhibits limitations in the selection of articles because it used only four journal databases (i.e., PubMed, IEEE Xplore®, ScienceDirect, and Scopus®). Moreover, only articles published in English and related to mental health problems were included. The duration of the search for articles started in October 2018, and collected articles were published between 2007 and 2018. Only articles related to data extracted from OSNs were selected, and the articles must be about detecting and mental health problem detection.

VI. FUTURE IMPLICATIONS

The researchers believe that the challenges outlined in the previous section can be addressed in future studies on mental health problem detection. Several aspects, such as methods, geographical locations, data extraction type, language, and multiple data sources, should be considered. Data were collected from specific geographical locations, such as China, Greece, Japan, and the United States [9], [14], [18], [19], [22]–[29]. Meanwhile, several researchers developed new or hybrid algorithms for mental health problem detection [9], [17]. The current researchers believe that implementing a new method and creating a new data set based on countries and localities may improve research on mental health detection in the future.

Another aspect is extracting data from multiple OSNs. This aspect may improve the results of mental health detection. The current researchers will observe any possibility of extracting data from other OSNs apart from those that were commonly used in previous research. Other researchers also raised this concern, and they intend to implement user data logs from multiple OSNs and new techniques for integrating multisource data [27]. The language barrier issue and extracting other languages from OSNs provide a potential future direction for research based on geographical location and native language because many OSNs users use their native language, instead of only English, in texting.

Furthermore, the researchers also believe that using another type of data (e.g., pictures, audios, and videos), instead of texts only in OSNs can be one of the potential areas for exploration in future research. The immensely rich content available in OSNs should not be disregarded because they can provide valuable data to future researchers. In conclusion, the aforementioned aspects may lead researchers to new feature selection processes before implementing machine learning algorithms in mental health problem detection.

VII. CONCLUSION

The purpose of this systematic review is to conduct a critical assessment on mental health problem detection. This analysis consists of the data source, the feature extraction method, and classifier performance in machine learning or deep learning techniques. This systematic review also investigates the appropriateness of pre-mental health detection by identifying its method of data analysis, challenges, and limitations. This study concludes that OSNs exhibit high potential as data sources of mental health problems detection but can never be substituted with traditional mental health detection methods based on face-to-face interviews, self-reporting, or questionnaire distribution. Nevertheless, OSNs can provide complementary data, and the combination of the two approaches in detecting mental health can improve future research. Mental health problem detection can probably decrease the number of people who are suffering from mental health problems or whose conditions perceived to become worse without further treatment. However, this study requires comprehensive adoption, innovative algorithms, and computational linguistics to increase the accuracy and precision of mental health problem detection in the future.

REFERENCES

[1] *Promoting Mental Health: Concepts, Emerging Evidence, Practice*. World Health Org., Geneva, Switzerland, 2004.
[2] R. A. Rahman, K. Omar, S. A. M. Noah, and M. S. N. M. Danuri, "A survey on mental health detection in Online Social Network," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, nos. 2–4, pp. 1431–1436, 2018.
[3] Global Burden of Disease Study 2013 Collaborators, "Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: A systematic analysis for the global burden of disease study 2013," *Lancet*, vol. 386, no. 9995, pp. 743–800, 2015.
[4] *The World Health Report 2001: Mental health: New Understanding, New Hope*, World Health Org., Geneva, Switzerland, 2001.
[5] *Management of Physical Health Conditions in Adults With Severe Mental Disorders: WHO Guidelines*, World Health Org., Geneva, Switzerland, 2018.
[6] M. Marcus, M. T. Yasamy, M. van Ommeren, D. Chisholm, and S. Saxena, "Depression: A global public health concern," World Health Org., Geneva, Switzerland, Paper Depression, 2012, pp. 6–8.
[7] *Mental Health Action Plan 2013-2020*, World Health Org., Geneva, Switzerland, 2013.
[8] C. L. M. Keyes, "Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health," *Amer. Psychologist*, vol. 62, no. 2, pp. 95–108, 2007.
[9] H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T.-S. Chua, "Detecting stress based on social interactions in social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1820–1833, Sep. 2017.
[10] E. Garcia-Ceja, V. Osmani, and O. Mayora, "Automatic stress detection in working environments from Smartphones' accelerometer data: A first step," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 4, pp. 1053–1060, Jul. 2016.
[11] A. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," in *Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact.*, Sep. 2013, pp. 671–676.
[12] T. G. M. Vrijkotte, L. J. P. van Doornen, and E. J. C. de Geus, "Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability," *Hypertension*, vol. 35, no. 4, pp. 880–886, Apr. 2000.
[13] G. Ali, A. Ali, F. Ali, U. Draz, F. Majeed, S. Yasin, T. Ali, and N. Haider, "Artificial neural network based ensemble approach for multicultural facial expressions analysis," *IEEE Access*, vol. 8, pp. 134950–134963, 2020.
[14] M. Kandias, D. Gritzalis, V. Stavrou, and K. Nikoloulis, "Stress level detection via OSN usage pattern and chronicity analysis: An OSINT threat intelligence module," *Comput. Secur.*, vol. 69, pp. 3–17, Aug. 2017.
[15] C.-H. Chang, E. Saravia, and Y.-S. Chen, "Subconscious crowdsourcing: A feasible data collection mechanism for mental disorder detection on social media," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 374–379.
[16] Q. Li, Y. Xue, L. Zhao, J. Jia, and L. Feng, "Analyzing and identifying teens' stressful periods and stressor events from a microblog," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1434–1448, Sep. 2017.
[17] M. Thelwall, "TensiStrength: Stress and relaxation magnitude detection for social media texts," *Inf. Process. Manage.*, vol. 53, no. 1, pp. 106–121, Jan. 2017.
[18] G. A. Coppersmith, C. T. Harman, and M. H. Dredze, "Measuring post traumatic stress disorder in Twitter," in *Proc. 7th Int. Conf. Weblogs Social Media (ICWSM)*, 2014, pp. 1–4.
[19] M. Park, C. Cha, and M. Cha, "Depressive moods of users portrayed in Twitter," in *Proc. ACM SIGKDD Workshop Healthcare Informat. (HI-KDD)*, 2012, pp. 1–8.
[20] H.-H. Shuai, C.-Y. Shen, D.-N. Yang, Y.-F.-C. Lan, W.-C. Lee, P. S. Yu, and M.-S. Chen, "A comprehensive study on social network mental disorders detection via online social media mining," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 7, pp. 1212–1225, Jul. 2018.
[21] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on Twitter," *Internet Intervent.*, vol. 2, no. 2, pp. 183–188, May 2015.
[22] H. Lin, J. Jia, Q. Guo, Y. Xue, J. Huang, L. Cai, and L. Feng, "Psychological stress detection from cross-media microblog data using deep sparse neural network," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.
[23] 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.
[24] S. Saleem, R. Prasad, S. Vitaladevuni, M. Pacula, M. Crystal, B. Marx, D. Sloan, J. Vasterling,and T. Speroff, "Automatic detection of psychological distress indicators and severity assessment from online forum posts," in *Proc. COLING*, 2012, pp. 2375–2388.
[25] X. Huang, L. Zhang, D. Chiu, T. Liu, X. Li, and T. Zhu, "Detecting suicidal ideation in Chinese microblogs with psychological lexicons," in *Proc. IEEE 11th Int. Conf. Ubiquitous Intell. Comput. IEEE 11th Int. Conf. Autonomic Trusted Comput. IEEE 14th Int. Conf. Scalable Comput. Commun. Associated Workshops*, Dec. 2014, pp. 844–849.
[26] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Proc. 7th Int. AAAI Conf. Weblogs Social Media (ICWSM)*, 2013, pp. 128–137.
[27] M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence," in *Proc. Int. Conf. Intell. Sustain. Syst. (ICISS)*, Dec. 2017, pp. 858–862.
[28] X. Wang, C. Zhang, and L. Sun, "An improved model for depression detection in micro-blog social network," in *Proc. IEEE 13th Int. Conf. Data Mining Workshops*, Dec. 2013, pp. 80–87.
[29] Y. Xue, Q. Li, L. Jin, L. Feng, D. A. Clifton, and G. D. Clifford, "Detecting adolescent psychological pressures from micro-blog," in *Proc. Int. Conf. Health Inf. Sci.*, 2014, pp. 83–94.
[30] B. Saha, T. Nguyen, D. Phung, and S. Venkatesh, "A framework for classifying online mental health-related communities with an interest in depression," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 4, pp. 1008–1015, Jul. 2016.
[31] A. Wongkoblap, M. A. Vadillo, and V. Curcin, "A multilevel predictive model for detecting social network users with depression," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Jun. 2018, pp. 130–135.
[32] J. Luo, J. Du, C. Tao, H. Xu, and Y. Zhang, "Exploring temporal patterns of suicidal behavior on Twitter," in *Proc. IEEE Int. Conf. Healthcare Informat. Workshop (ICHI-W)*, Jun. 2018, pp. 55–56.
[33] C.-H. Tai, Z.-H. Tan, Y.-S. Lin, and Y.-S. Chang, "Mental disorder detection and measurement using latent Dirichlet allocation and sentiwordnet," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 1215–1220.
[34] E. Saravia, C.-H. Chang, R. J. De Lorenzo, and Y.-S. Chen, "MIDAS: Mental illness detection and analysis via social media," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 1418–1421.
[35] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of social media," *Bus. Horizons*, vol. 53, no. 1, pp. 59–68, Jan. 2010.
[36] M. A. Al-garadi, K. D. Varathan, and S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network," *Comput. Hum. Behav.*, vol. 63, pp. 433–443, Oct. 2016.
[37] M. K. Othman and M. S. N. M. Danuri, "Proposed conceptual framework of dengue active surveillance system (DASS) in Malaysia," in *Proc. Int. Conf. Inf. Commun. Technol. (ICICTM)*, 2016, pp. 90–96.
[38] M. Raza, M. Awais, K. Ali, N. Aslam, V. V. Paranthaman, M. Imran, and F. Ali, "Establishing effective communications in disaster affected areas and artificial intelligence based detection using social media platform," *Future Gener. Comput. Syst.*, vol. 112, pp. 1057–1069, Nov. 2020.
[39] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

[40] X. Li, Y. Jiang, M. Li, and S. Yin, "Lightweight attention convolutional neural network for retinal vessel segmentation," *IEEE Trans. Ind. Informat.*, early access, May 11, 2020, doi: 10.1109/TII.2020.2993842.
[41] M. A. Al-garadi, M. S. Khan, K. D. Varathan, G. Mujtaba, and A. M. Al-Kabsi, "Using online social networks to track a pandemic: A systematic review," *J. Biomed. Informat.*, vol. 62, pp. 1–11, Aug. 2016.
[42] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. A. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration," *J. Clin. Epidemiol.*, vol. 62, no. 10, pp. e1–e34, 2009.
[43] *National Collaborating Centre for Mental Health, National Institute for Health, and Clinical Excellence, Common Mental Health Disorders: Identification and Pathways to Care*, Brit. Psychol. Soc. Roy. College Psychiatrists, London, U.K., 2011.
[44] National Library of Medicine. (2018). *Medical Subject Headings (MeSH)*. Accessed: Oct. 29, 2018. [Online]. Available: https://meshb.nlm.nih.gov/
[45] J. Singh, "Critical appraisal skills programme," *J. Pharmacol. Pharmacother.*, vol. 4, no. 1, pp. 76–77, 2013.
[46] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social Web 1," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 1, pp. 163–173, 2012.
[47] J. Jashinsky *et al.*, "Tracking suicide risk factors through Twitter in the US," *Crisis*, vol. 35, pp. 51–59, 2014.
[48] X. Wang, C. Zhang, Y. Ji, L. Sun, and L. Wu, "A depression detection model based on sentiment analysis in micro-blog social network," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*, 2013, pp. 201–213.
[49] M. Kosinski, S. Matz, S. D. Gosling, and D. Stillwell, "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines," *Amer. Psychol.*, vol. 70, no. 6, pp. 543–556, 2015.
[50] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, pp. 1–68, 2019.
[51] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *J. Lang. Social Psychol.*, vol. 29, no. 1, pp. 24–54, Mar. 2010.
[52] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
[53] H. Lin, J. Jia, Q. Guo, Y. Xue, Q. Li, J. Huang, L. Cai, and L. Feng, "User-level psychological stress detection from social media using deep neural network," in *Proc. ACM Int. Conf. Multimedia (MM)*, 2014, pp. 507–516.
[54] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[55] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," in *Machine Learning and Its Applications: Advanced Lectures*, G. Paliouras, V. Karkaletsis, and C. D. Spyropoulos, Eds. Berlin, Germany: Springer-Verlag, 2001, pp. 249–257.
[56] M. A. Gianfrancesco, S. Tamang, J. Yazdany, and G. Schmajuk, "Potential biases in machine learning algorithms using electronic health record data," *JAMA Intern. Med.*, vol. 178, no. 11, pp. 1544–1547, 2018.
[57] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England J. Med.*, vol. 380, no. 14, pp. 1347–1358, 2019.

ROHIZAH ABD RAHMAN received the bachelor's degree in computer science from Universiti Sains Malaysia, in 2002, and the master's degree in information technology from Universiti Kebangsaan Malaysia, in 2006, where she is currently pursuing the D.Phil. degree in computer science. Her Ph.D. research is related to artificial intelligence, data science, natural language processing, and machine learning. Her master's degree was also involved in artificial intelligence and implemented neural network, and particle swarm optimization algorithm. Have more than 12-year experiences in teaching and learning in programming (Java, C++, C and Python), database, and data structure.

KHAIRUDDIN OMAR received the bachelor's and master's degrees in computer science from Universiti Kebangsaan Malaysia, in 1986 and 1989, respectively, and the D.Phil. degree from Universiti Putra Malaysia, in 2000. He is currently a Professor with the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia. His research interests include artificial intelligence (pattern recognition in decision making with uncertainty—ayesian reasoning, neural networks, fuzzy logic, fuzzy neural networks, and image processing – 2D and 3D, edge detection, thinning, segmentation, feature extraction, image improvement, texture, resolution, image transforms, e.g., Trace, Fourier, Wavelet) with applications to Jawi/Arabic Manuscripts, biometric authentication.

SHAHRUL AZMAN MOHD NOAH (Member, IEEE) received the B.Sc. degree (Hons.) in mathematics from Universiti Kebangsaan Malaysia, in 1992, and the M.Sc. and Ph.D. degrees in information studies from the University of Sheffield, U.K., in 1994 and 1998, respectively. He is currently a Professor with the Center for Artificial Intelligence Technology, Universiti Kebangsaan Malaysia and currently heads the Knowledge Technology research group. His research interests include information retrieval and ontology with special emphasis on semantic search and recommender systems. He has published more than 200 research articles in these areas. He is currently the Chair of the Persatuan Capaian Maklumat dan Pengurusan Pengetahuan (PECAMP), and a member of the International Association for Ontology and its Applications (IAOA) and IEEE Computer Science Society associations.

MOHD SHAHRUL NIZAM MOHD DANURI (Member, IEEE) received the B.Sc. degree (Hons.) in computer science from Universiti Sains Malaysia, in 2002, the M.S. degree in intellectual property from Universiti Kebangsaan Malaysia, in 2006, and the D.Phil. degree in information management from Universiti Teknologi MARA, Malaysia, in 2017. He is currently a Senior Lecturer with the Faculty of Science and Information Technology, Kolej Universiti Islam Antarabangsa Selangor, Malaysia. He is also a registered Professional Technologist and one of the Technology and Technical Accreditation Council (TTAC) assessment panel under Malaysia Board of Technology (MBOT). His research interests include information systems, the Internet of Things, data science, machine learning, and ICT applications in agriculture.

MOHAMMED ALI AL-GARADI received the Ph.D. degree in computer science from the University of Malaya, Malaysia, in 2017. He is a Postdoctoral Scholar at the University of California San Diego (UCSD), USA. He has published several research articles in refereed journals and conferences. He served as a reviewer for several journals, including the IEEE Communications Magazine, the IEEE Transactions on Knowledge and Data Engineering, IEEE Access, *Future Generation Computer Systems, Computers and Electrical Engineering*, and the *Journal of Network and Computer Applications*. He also received several national and international awards during his Ph.D. research. His research interests include big data analytics, machine learning, deep learning, biomedical informatics, and social media analytics.