Computer Methods and Programs in Biomedicine Update 2 (2022) 100066

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine Update

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine-update

Machine learning models to detect anxiety and depression through social media: A scoping review

Arfan Ahmed [a,*], Sarah Aziz [a], Carla T. Toro [b], Mahmood Alzubaidi [c], Sara Irshaidat [d], Hashem Abu Serhan [d], Alaa A. Abd-alrazaq [a], Mowafa Househ [c,*]

[a] AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar
[b] Institute of Digital Healthcare, WMG University of Warwick, Warwick, UK
[c] College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
[d] Jordan University Hospital, Amman, Jordan

ARTICLE INFO

Keywords:
Anxiety
Depression
Social media
Social networking
Artificial intelligence
Machine learning
COVID-19

ABSTRACT

Despite improvement in detection rates, the prevalence of mental health disorders such as anxiety and depression are on the rise especially since the outbreak of the COVID-19 pandemic. Symptoms of mental health disorders have been noted and observed on social media forums such Facebook. We explored machine learning models used to detect anxiety and depression through social media. Six bibliographic databases were searched for conducting the review following PRISMA-ScR protocol. We included 54 of 2219 retrieved studies. Users suffering from anxiety or depression were identified in the reviewed studies by screening their online presence and their sharing of diagnosis by patterns in their language and online activity. Majority of the studies (70%, 38/54) were conducted at the peak of the COVID-19 pandemic (2019–2020). The studies made use of social media data from a variety of different platforms to develop predictive models for the detection of depression or anxiety. These included Twitter, Facebook, Instagram, Reddit, Sina Weibo, and a combination of different social sites posts. We report the most common Machine Learning models identified. Identification of those suffering from anxiety and depression disorders may be achieved using prediction models to detect user's language on social media and has the potential to complimenting traditional screening. Such analysis could also provide insights into the mental health of the public especially so when access to health professionals can be restricted due to lockdowns and temporary closure of services such as we saw during the peak of the COVID-19 pandemic.

1. Introduction

Depression and anxiety, often referred to as 'common mental illness', are prevalent at a global scale. For depression, the lifetime prevalence varies between cultures and is higher in higher-income countries such as the USA at 20% [1,2]. Not only do depression and anxiety pose a large economic burden on society [3], there is also a large toll on affected individuals with respect to years lost due to ill health: disability adjusted life years (DALYs) for mental illness are similar to cardiovascular and circulatory diseases. Regarding Years Lived with Disability (YLDs), the proportion is high at 32·4% for mental health disorders [4]. Lifespan is shorter for affected individuals [5], furthermore, depression is a strong risk factor for suicide [6]. The current situation is bleak exasperated largely due to the COVID-19 pandemic due to quarantines, self-isolation, lockdowns and a general feeling of fear causing a dramatic rise in cases of stress, anxiety, and depression [7,8].

For those who have access to mental health services, there is a large proportion of individuals with common mental health disorders who have not sought diagnosis or medical help of any kind. There are several reasons behind low help-seeking behavior; the most common include: stigma and concerns of judgment by others including friends, family, employers or health services or insurance, belief that the situation will improve with time, poor understanding of treatments that could be efficacious, amongst others [9].

The past decade has witnessed high interest from computer and data scientists in the issues presented here regarding the limitations of current diagnostic practice for anxiety and depression. Artificial Intelligence (AI)- technology based predictive models use a variety of data types including *in-vivo* structural and functional imaging data and neuropsychological data [10]. An algorithm is used to train a predictive

* Corresponding authors.
E-mail address: mhouseh@hbku.edu.qa (M. Household).

https://doi.org/10.1016/j.cmpbup.2022.100066
Received 1 November 2021; Received in revised form 27 August 2022; Accepted 4 September 2022
Available online 9 September 2022

model on the training set (i.e. part of the dataset) and the test set (the remainder of the dataset) is used for evaluation, this avoids overfitting (cross-validation process). Several possible metrics can be recorded such as those we report in Table 3. In an umbrella review of systematic reviews in this area, there are promising findings for disorders such as Alzheimer's disease and schizophrenia, however, there is less published research based on imaging and neuropsychological data for depression and anxiety [10]. AI-technology base predictive models have also been developed using data from Electronic Health Records (EHRs) by comparing health records with diagnosis of depression with controls and mining data in years before diagnosis to explore predictors of depression before it was confirmed clinically [11]. Speech data has also been used to develop AI-based predictive models of mental health disorders, and certain acoustic features have been found to be different for several mental health disorders compared to controls [12].

The prolific uptake and integration into the daily life of social media offers another and, in many cases, a much larger data repository for the development of predictive models to detect mental health disorders such as depression and anxiety using user-generated data such as written posts, blogs, photos, and videos. The first study of its kind by De Chaudhury and colleagues at Microsoft (2013) [13] examined 69 K Twitter postings shared by individuals identified as having depression using the centre for Epidemiological Studies Depression Scale (CES-D) and developed a support vector machine (SVM) classifier to predict whether a Twitter post could be depression-indicative with an accuracy of >70% and precision of 0.82. This landmark study highlighted to the research community the value of insights into mental health risk factors and population-level disclosure of mental health symptoms on social media platforms, that ordinarily may not be disclosed. This has led to dozens of new studies using AI-technology based prediction models and user-generated data from different social media platforms. These include Twitter, Facebook, and Instagram. Also, other platforms with a more uneven global distribution of users include Reddit, an online forum popular in the USA (where over half of the users are based), Sina Weibo which is the second most popular social media platform in China, and Vkontakte, which is popular in Russia.

1.1. Research problem and aim

As there has been an explosion of social media studies since a study in 2013 [13], and although a number of mental health related Machine Learning (ML) reviews exist [14–16], we found no scoping reviews on ML and social media that focus on anxiety and depression following PRISMA guidelines. There is a need for social media studies to be collated and reviewed. . Based on this rationale/necessity, the aim of this scoping review is to explore ML models used to detect anxiety and depression through social media.

2. Methods

A scoping review intended to fulfill the above-mentioned objective of the study. In this section, the details of the utilized methodology to carry out the review are explained. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [17] were followed as guidelines for this scoping review.

2.1. Search strategy

2.1.1. Search sources

Six databases including PsycINFO, Medline, Embase, ACM Digital Library, IEEE Xplore, and Google scholar. In the case of Google Scholar, the first 10 pages of the search were selected, as Google Scholar delivers the most pertinent studies first. The search process was carried out from 10th to 14th May 2021.

2.1.2. Search terms

Three various collections of search terms were identified for the present study to search databases. The first collection of the search terms included terms related to social media (i.e., Facebook and Twitter). The second collection of the terms was related to the target disease (i.e., anxiety and depression). The last collection of the terms consisted of terms related to AI (e.g., artificial intelligence OR machine learning). The search query used in this review is as follows: (("social media" OR "social network" OR Facebook OR Twitter OR tweet) AND (artificial intelligence" OR "machine learning" OR "Deep Learning" OR "reinforcement learning" OR "Data Mining" OR "Big Data" OR "Supervised learning" OR "unsupervised learning" OR "Text Analysis" OR "Text Mining" OR "Predictive Analytics") AND (anxiety OR depressive OR depression OR anxious))

2.2. Study eligibility criteria

For the study to be selected, it had to propose an AI-based approach or technique primarily to detect anxiety and depression through social media such as Twitter and Facebook. Only studies published in English since 2010 were selected, and only peer-reviewed articles, theses, dissertations, conference proceedings, and reports were considered in this review. Reviews, conference abstracts, proposals, and editorials were removed.

2.2.1. Study selection

The studies filtration process was performed in three stages, starting with identifying studies, screening title and abstract, and full-text screening by two reviewers. The first reviewer identified and removed duplicate studies. At Stage Two, the second reviewer screened titles and abstracts from all the returned studies. Furthermore, the two reviewers independently screened the full text of studies that passed the title and abstract screening. Any disagreements between reviewers were resolved through discussion.

2.2.2. Data extraction and data synthesis

The data extraction form was developed and tested with five included studies shown in Multimedia Appendix B. Two reviewers independently used Excel sheet to extract data related to the characteristics of identified studies, predictive models used, and data source trained upon. The narrative approach was utilized to synthesize the extracted data of the included studies. We recorded the ML predictive model reported in the study (e.g., Linear Regression (LR), SVM, any ML algorithm, deep learning (DL) model), data source trained (e.g. Facebook, Twitter), data collection (e.g., how data is collected i.e. recruited to take a depression survey and share their Facebook or Twitter data or data is collected from existing public online sources), language of data samples (e.g., English, Arabic, etc.) mental illness diagnosed (e.g., self-declared, manually expert annotated, The Center for Epidemiological Studies-Depression (CES-D), patient health questionnaire (PHQ), etc.), number of data sample used (e.g. in case of tweets how many tweets are being analyzed), mental disorder analyzed (e.g. Depression, Anxiety, etc.), how the evaluation of data performed (e.g. Accuracy, F1-score, etc.), and best result attained (how good is their detection results).

3. Results

3.1. Search results

The study selection process for this scoping review is illustrated in Fig. 1. A total of 2219 citations were identified from searching the predefined six bibliographic databases. 112 duplicates were removed from the results. We screened the title and abstract of the remaining 2107 studies and excluded: 671 publications that seemed irrelevant, 975 publications that did not use social media, 204 publications that did not target anxiety and/or depression, and 75 publications that were not

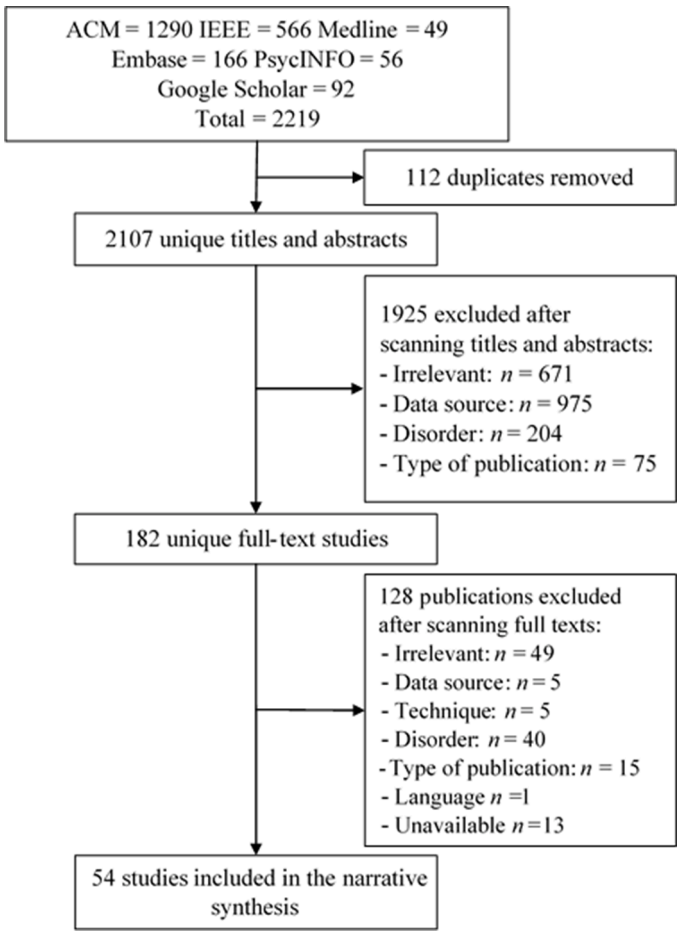

Fig. 1. PRISMA chart.

journal articles, dissertations, or conference papers. The remaining 182 items were screened in full text, and 128 items were removed from them for the following reasons: 49 studies were not relevant, 5 studies did not use social media, 5 studies did not use AI techniques, 40 studies did not target anxiety and/or depression, 15 publications that were not journal articles, dissertations, or conference papers, 1 study was not written in the English language, and 13 studies could be found in full-text. A total of 54 articles were included in this scoping review.

3.2. General features of included papers

The article publication dates ranged from 2010 to 2021. About 70% (38/54) of the studies were published in the last 2 years (i.e., 2019 and 2020) (Table 1). While about 56% (30/54) of the included studies were conference papers, about 44% (24/54) were journal articles. About 66% of the studies ($n = 36/54$) were conducted in Asia with China, India, and Bangladesh being the most common countries (12, 8, and 7 papers, respectively). Nine (16%) of the papers were published by groups based in Europe (France, Germany, Ireland, Spain, UK), seven (13%) from North America (including Canada), and two studies (3%) were included from Australia and New Zealand (Table 2).

3.3. Mental health disorders

Most of the included papers focused on depression ($n = 47/54$, 87%) while only 7 (13%) papers focused on anxiety disorders which included social anxiety ($n = 2$), post-traumatic stress disorder (PTSD; $n = 2$), and obsessive-compulsive disorder (OCD, $n = 1$). A variety of standardized questionnaire-based reference standards were used to identify the target

Table 1
Inclusion and exclusion criteria.

Criteria	Specified criteria
Inclusion	Focus on predicting anxiety and depression through analysis of social media data
	Predictive or classification models that focus on analysis based on users generated text posts
	ML techniques applied on social media dataReported in the English language
	Studies from 2010 to 2021
	Studies reported in peer-reviewed journal articles, thesis, dissertations, and conference proceedings.
Exclusion	Analyzed the correlation between social network data and symptoms of mental illness
	Analyzed textual contents only by human coding or manual annotation
	Examined data from online communities (e.g., LiveJournal)
	Studies that focus on internet addiction
	Examined the influence of cyberbullying on mental health
	Did not explain where the datasets came from (source of social media).

population for model development, among which CES-D scale ($n = 7/54$, 13%) and PHQ of varying versions ($n = 2/54$, 3%) were the most popular ones, whereas only 7% studies ($n = 4/54$) used other questionnaires (Hamilton Depression Rating Scale (HDRS), The Sport Anxiety Scale (SAS) questionnaire, NEO Personality Inventory-Revised (NEO-PI- R) questionnaire) per individual study once. Not all studies included standardized questionnaires to identify the target population and instead identified the study population by selecting data from mental health sub-Reddits (a specific online community, and the posts associated with it) or groups ($n = 6/54$, 11%) or by exclusively selecting posts from users within sub-Reddits that had a post with "I have a diagnosis of depression" self-declaring their mental status ($n = 10/54$, 19%). Many studies were annotated after collecting data in bulk from different social websites these annotations were mostly automated ($n = 11/54$, 20%) by performing a preprocessing step in which the developed algorithm identifies mental health keywords or hashtags within the posts, while 13% studies performed manual annotation ($n = 7/54$) posts being thoroughly examined and identified as depressive or anxious by various mental health experts. Also, as most of the studies were developing algorithms that determine the depressive or anxious level of posts, they did not search specific dataset and rather randomly selected population ($n = 6/54$,11%). The remaining study used a pre-defined depressive dataset for study and one did not specify how they determined the mental health level of the population.

3.4. Social media data source and predictive models

Table 3 outlines the studies included in this scoping review use of social media data from a variety of different platforms to develop predictive models for the detection of depression or anxiety. These included Twitter ($n = 18/54$, 33%), Facebook ($n = 6/54$,11%), Instagram ($n = 2/54$, 3%), Reddit ($n = 13/54$, 24%), Sina Weibo ($n = 6/54$,11%), some studies reported a combination of different social sites posts, a combination of Facebook and Twitter ($n = 4/54$, 7%), while other remaining social sites, like Vkontakte and different combinations (twitter+-Blueapp,Twitter+Reditt, Twitter+Sina Weibo), were used by one study each. AI, ML, and DL models were used as summarized in Table 3. The table is presented in its current format in line with previous literature for presenting such data [18].

The studies reviewed in this scoping review made use of prediction models to detect and classify users according to their mental disorders. Predictive models use a training data set (a set of selected features) in order for a machine learning algorithm to learn patterns from that data.

Table 4 shows the most commonly used models (complete with abbreviation explanations) were Ada-Boost, CNN, GRU, KNN, LR, LSTM, MLP, RF, DT, VGG-Net, and XGboost which were used by different studies for analyzing various social sites data. Other models that were

Table 2
general characteristics of studies.

Characteristics	Number of studies	Studies reference (refer to Appendix A)
Year of publication	2013: 2	S40,S48
	2014: 1	S51
		S39,S46
	2015: 2	S47
	2016: 1	S17,S35,S38
	2017: 3	S11,S16,S20,S28,S33,S36,S54
	2018: 7	S5,S8-10,S12,S13,S15,S19,S21, S23,S25,S29,S34,S41,S42,S44, S45,S50,S53
	2019: 19	S1-S4,S6,S7,S14,S18,S22,S24, S26,S27,S30-S32,S37,S43,S49, S52
	2020: 19	
Country	Australia: 1	S49
	Bangladesh: 7	S12,S13,S15,S16,S20,S24,S32
	Canada:1	S35
	China: 12	S4,S6,S10,S18,S23,S26,S30, S37,S39,S45,S51,S54
	France:1	S48
	Germany: 1	S52
	India: 8	S5,S7,S9,S14,S17,S22,S34,S43
	Indonesia: 2	S36,S50
	Ireland:1	S29
	Japan:1	S46
	Kazakhstan: 1	S31
	Korea:2	S3,S47
	New Zealand:1	S2
	Pakistan: 1	S8
	Saudi Arabia: 2	S38,S44
	Spain:3	S1,S21,S25
	UK: 3	S11,S27,S42
	USA: 6	S19,S28,S33,S40,S41,S53
Type of publication	Conference: 30	S6,S7,S10-S15,S19,S20,S22, S24,S27,S30-S33,S36,S38-S40, S42,S43,S46-S48,S51-S54
	Journal article: 24	S1-S5,S8,S9,S16-S18,S21,S23, S25,S26,S28,S29,S34,S35,S37, S41,S44,S45, S49,S50
Disorder(*)	Anxiety Disorders (including social anxiety, PTSD, OCD): 7	S3,S8,S9,S26,S41,S45,S53
	Depression: 47	S1,S3-25,S27,S28,S31-S37,S39-S46,S48-S50,S52-S54
Mental illness criteria	CES-D: 7	S11,S36,S39,S40,S42,S46,S48
	PHQ (any version): 2	S5,S28
	Self-declared: 10	S1,S2,S18,S21,S23,S24,S33, S50,S52,S54
	Mental health subreddits or groups: 6	S3,S7,S8,S25,S41,S44
	Manually annotated: 7	S12,S13,S19,S20,S29,S32,S35
	Automated annotation (mental health keywords or mental hashtags):11	S4,S9,S14,S31,S37,S38,S43, S45,S47,S49,S51
	Predefined depression dataset: 1	S6
	Random selected: 6	S15,S16,S17,S27,S30,S34
	Other or Unspecified: 4	S10,S22,S26,S53

* the numbers don't add up some of the studies addressed both disorders.

used on only one individual social data source were BRR, NLP techniques, SVR, GPR, LR, Bert, MFFN, models built based on linguistic and behavioral features, GBM, and multiclass tree models. Model performances were evaluated in almost all the studies using various metrics; of which F1-score (n = 27/54, 24%) being the most opted primary measuring metrics, followed by Accuracy (n = 16/54,30%) other metrics like Recall, Precision, AUC, RMSE, Pearson correlation, and ERDE were used by single studies only. One study did not use any measuring metric. The performance of the models ranged from 0.043 to 0.99 with respect to each social site. Different languages were used by different data sources for analysis comprising of English (n = 47/54, 87%) the

Table 3
Predictive models and their primary metrics observed in reviewed studies.

Data type	No.	Predictive model used														Primary Performance metrices	Performance range(for data type)	Language of Data Samples Analyzed	Range of participants	References
		Ada-Boost	CNN	GRU	KNN	LR	LSTM	MLP	NB	Proposed Algo	RF	DT	SVM	XGBOOST	Others					
twitter	18	Y	Y		Y	Y	Y		Y	Y	Y	Y	Y	Y	BRR, NLP,SVR	F1 score: 7 Accuracy: 8 RMSE: 1 Not used:1	0.47 to 0.99	English - 14 Bangla-1 Indonesian- 1 Spanish- 1	55 - 20,000	S2,S9,S12,S13, S17, S21,S22,S31,S34, S35,S36,S42,S45, S46,S47, S48,S50, S52
facebook	6		Y	Y			Y	Y		Y					GPR	F1 score: 3 Accuracy: 3	0.61 to 0.90	English - 6	90 - 5,947	S11,S16,S20,S27, S42,S43
Instagram	2					Y									lr	Accuracy: 1 AUC:1	0.60 to 0.71	English - 2	749 - 1,908	S14,S28
Reditt	13	Y	Y	Y		Y	Y	Y	Y		Y	Y	Y	Y		F1 score:10 Accuracy: 2 ERDE: 1	0.043 to 0.96	English -13	365 - 48,537	S1,S3,S7,S8,S19,S23, S24,S25,S33,S41, S44,S49,S54
Sina Weibo	6		Y				Y						Y		Bert,MFFN, built models based on linguistic and behavioral features	F1 score: 3 Accuracy: 2 Precision:1	0.5-0.97	English - 3 Chinese- 3	1000 - 30,000	S4,S6,S10,S30,S37, S39
facebook + twitter	4				Y	Y			Y			Y	Y		VGG-Net	F1 score: 1 Accuracy: 2 Pearson correlation: 1	0.56 to 0.77	English - 4 Bangla- 1	150 - 3,498	S15,S32,S38,S53
others	5		Y			Y					Y	Y		Y	GBM,Multi-class tree models	F1 score: 3 Accuracy: 1 Recall: 1	0.86 to 0.96	English - 5	619 - 1,000,500	S5,S18,S26,S31,S51

Table 4
Models used within reviewed studies.

Models Name	Abbreviation	number of studies	Study Reference
Adaptive Boosting	AdaBoost	1	S23
Bayesian Ridge Regression	BRR	1	S29
Bidirectional Encoder Representations from Transformers	Bert	1	S10
built classification and regression models based on linguistic and behavioral features		1	S39
Convulational Neural Network	CNN	7	S2,S3,S10,S11,S19, S51,S52
linear regression (elastic-net regularized)	LR	1	S28
Gradient Boosting Machine	GBM	2	S9,S31
Gaussian Process Regression.	GPR	1	S27
Gated Recurrent Neural Network	GRU	4	S11,S13, S33,S42
K- Nearest Neighbour	KNN	3	S16,S20,S32
Logistic Regression (including L1,L2 regulized)	LR	6	S1,S5,S14,S23,S32, S41
Long Short Term Memory (including Attention-Based Bidirectional Long Short-Term Memory (Attention-BiLSTM))	LSTM	8	S7,S10,S12,S24, S30,S37,S42,S54
Tree based (including Multi-Class Trees)		2	S25,S26
Multilayer Perceptrons	MLP	2	S11,S23
Multimodal Feature Fusion Network	MFFN	1	S6
Natural Language Processing techniques (with lexical approach)	NLP	3	S21,S36
Navie bayes (including Multinomial)	NB	8	S8, S9,S15,S17,S32, S34,S38, S47
Propose algorithms (including count the occurance of depressive words)		2	S43,S45
Random Forest	RF	6	S8,S9,S22,S23,S31, S32
Decicison Tree (including Rule based)	DT	3	S5,S16, S50
Support Vector Regression	SVR	1	S34
Support Vector Machine (including deep integrated support vector machine (DISVM))	SVM	15	S4,S8,S10,S15,S16, S17,S23,S32,S33, S35,S38, S40,S46, S48,S49
Visual Geometry Group Network	VGG-Net	1	S53
eXtreme Gradient Boosting	XGBOOST	4	S3,S5,S18,S54

most used one, followed by Chinese ($n = 3/54$, 6%), and Bangla ($n = 2/54$, 4%) whereas Indonesian, Spanish were used once per study. Most studies specified the number of participants used in the study for their social network analysis. The minimum number of participants used was 55 whereas the 1,000,500 was the highest used by any study. Studies focusing on Twitter posts yielded higher performance metrics overall. The minimum performed was obtained by one study that used Instagram data type. Sina Weibo being the Chinese social networking site mostly analyzed the posts that were in Chinese.

4. Discussion

4.1. Principal findings

The objective of this scoping review was to explore ML models used to detect anxiety and depression through social media. We highlighted the most common ML models used including their performance measurement metrics along with their application to popular social media networks with the most common languages being targeted as English. Most of the papers focused on depression and used data from multiple social media platforms including Twitter, Facebook, Instagram, Reddit, Sina Weibo, and Vkontakte yielding predictive models with reported performance metrics. Whereas previous integrative studies [18] reviewed studies aimed at predicting mental health illness using social media and reported success in identifying at-risk or depressed individuals by monitoring social media. To the best of our knowledge no recent scoping review exists, we, therefore, report an up-to-date scoping review following PRISMA-ScR guidelines. The use of PRISMA-ScR gives a greater understanding of relevant terminology, core concepts, and key items to report for scoping reviews.

Wide access to publicly available social media data from a large corpus of users offers the power to predict common mental disorders (or symptoms of these) using AI models. The rationale behind most of the studies in this review was common: to explore novel approaches to diagnose mental disorders for where existing diagnostic process and practice harbor several limitations including the reliance on subjective interpretations and accounts to reach diagnosis, limited availability of mental health services in low- to medium- income countries and low help-seeking behaviors [9,19,20]. The promise of AI-technology-based predictive models could lead to integration with existing practice, for example, to assist mental health practitioners in assessing mental health symptoms more objectively than offered by currently used diagnostic procedures such as manuals and standardized questionnaires. AI-technology-based predictive models could also offer the opportunity to identify symptoms at an earlier stage possibly before psychosocial consequences become problematic. They could also offer the opportunity to personalize treatments based on an individual's symptoms that may not necessarily be picked up by standard currently available diagnostic procedures.

Compared to predictive models using other types of data such as from EHRs, social media data presents researchers with a fundamental challenge on how the target population i.e., those with depression can be identified. Some studies use standardized questionnaires such as CES-D [21] however the time required to contact users and barriers to uptake can lead to limited numbers of participants which can impact the success of a predictive model, where more data is likely to yield better performing models. Other studies [22], based their data mining on posts from sub-Reddit communities focusing on mental health disorders such as depression or anxiety, however, no further selection criteria were used. Superior to this are methods such as those described by Martinez-Castano et al. (2020) [23] where people were identified by extracting self-expressions of diagnoses from participants social media posts submissions by submitting queries to Reddit's search service (queries such as, "I was diagnosed with depression". Expressions such as "I am depressed", "I have depression", or "I think I have depression", were excluded). Furthermore, the control group was obtained by random sampling from the entire community of Reddit users and manual addition of users who were active on the depression threads but had no depression [23]. In summary, the allowing environment and sense of community and trust offered to social media users on some of these platforms combined with robust research design could lead to identification of target populations with depression or anxiety that are as valid and reliable as with the use of standardized questionnaires (for where there will also be limitations regarding noise and inaccuraciesThe approach used by Martinez-Castano and colleagues (2020) [23] could also address the serious problem that currently exists regarding low disclosure of mental health symptoms and low help-seeking behaviors.

This field is fast paced and growing, with current advancements and research in ML techniques we are likely to see continuous improvements. For example, a study in 2019 [22] reported how the models are improving since the 2013 landmark studies where several experiments were conducted including on anxiety and depression utilizing SVM, NB, and RF and found they outperformed with Co-training technique (a type of semi supervised learning approach) as compared to their individual

use in terms of Precision, Recall, and F-measure. Therefore, using co-training technique-based approach seems promising compared with the state-of-the-art classifiers [22].

4.2. Strengths and limitations

4.2.1. Strengths

We were able to produce a high-quality review by conducting and reporting according to the PRISMA Extension for Scoping Reviews. The most popular databases within the healthcare and information technology fields were searched to maximize the number of studies retrieved. Publication bias risk was minimized by searching the first 10 pages of Google scholar. Multiple reviewers independently screened the studies and extracted the data minimizing the risk of selection bias.

4.2.2. Limitations

Due to practical constraints, our search was restricted to English studies published between 2010 and 2021, and we were not able to search interdisciplinary databases (e.g., Web of Science and Scopus). Furthermore, we did not use other terms related to social media (instgram, Reddit, TikTok, Snapchat, etc.) and AI (models such as SVM, LR, CNN, ANN, RNN, RF, etc.). Accordingly, we could possibly have missed some relevant studies.

4.3. Practical and research implications

4.3.1. Practical implications

The volume of studies that have been published since the original studies by De Chaudhury and colleagues in 2013, provides us with exciting insights to the capability of AI technologies to assist and ultimately transform diagnostic practice. Thought-provoking developments during this time are evident. For example, the work by Martinez-Castano et al. [23] (included in this review) summarizes the development of a multi-component platform for real-time processing of social media data for the early detection of depression. Continuous analysis of social media data along the temporal dimension is crucial given that common mental health problems such as anxiety and depression fluctuate with time and there can be short periods of remission and exacerbation. In times of a pandemic, we would expect the use of social media to increase as the public use it to express themselves. This review observed majority (70%) of the studies were conducted during the peak of the COVID-19 pandemic (2019–2020). This highlights interest by researchers to conduct such studies during this period of rise in mental health disorders. It would however be interesting to conduct a full systematic review of these studies to determine their link to COVID-19 before drawing any definitive conclusions which goes beyond the purpose of this scoping review. For this research field to make greater strides and bridge the gap between research to clinical care, reproducibility and replication using external validation of many of the models presented here will be the next step. Only then could many of the studies that are still in the early proof-of-concept stage, progress to the next stage of piloting and integrating into existing diagnostic process to examine feasibility from multiple clinical and practical perspectives. Furthermore, our search did not return any studies in the MENA region (except 2 studies in Saudi Arabia) or any regions that are currently in conflict zones where the rates of anxiety and depression could provide meaningful insight into the mental health of those living through conflicts, we encourage research targeting such regions.

Although the promise of predictive models summarized in this review to transform mental health diagnostics is exciting, there are ethical considerations that should be highlighted. For AI technologies to be integrated into current practice for diagnosing common mental health disorders, there should be availability of appropriate treatment. However, a treatment lag is evident and can be common [24], therefore the reasons behind treatment lag need to be unpicked, and strategies to combat these existing barriers are essential while at the same time ongoing work can continue to further develop predictive models. Only when both research challenges are addressed should there be piloting of integration of new predictive models into existing care models.

4.3.2. Research implications

In a similar vein to AI technology-based predictive models that are being developed for neurological disorders, such as Alzheimer's disease, there are ethical concerns around the accuracy of predictive models and a need for consensus on what predictive power is sufficient in order to justify the risks and consequences of false-negatives and false-positives [25]. Another ethical concern regards mental health literacy and limitations some individuals may have in understanding a diagnosis, or risk of having a mental disorder and what the implications may be regarding the impact on life and work, treatment, and 'cure'. It's also possible that a limited possibility to explain how a 'black box' algorithm works could cause concerns about transparency of predictive models. An ever-growing amount of literature is published in this field of research. Therefore, this scoping review is suitable for a quick orientation for different stakeholders and may inform further scientific investigations. Future studies also need to consider incorporating current text mining models that have been validated and published for emotion classification especially those that focus on anxiety and/or depression. A further systematic review could encourage more work in this field and collaboration with healthcare professionals to incorporate social media account activity in helping diagnosis, furthermore, a systematic review should report on demographics which would highlight the gap in the current literature on social media usage to anxiety and depression among age and gender for example. Finally, there is a need for more efforts to develop robust governance models for user data and social media data security to ensure that users have control of their data and that there are no consequences on health insurance premiums or risk of employment discrimination [25].

5. Conclusion

Identification of those suffering from anxiety and depression disorders may be achieved using prediction models to detect user's language on social media and has the potential to complimenting traditional screening. Continuous analysis of social media data along the temporal dimension is crucial. We saw a dramatic rise in studies we reviewed during the COVID-19 peak (2019–2020). Whereas some ethical considerations are needed, AI-technology based predictive models could offer the opportunity to identify symptoms at an earlier stage possibly before psychosocial consequences become problematic.

Funding

N/A

Institutional review board statement

N/A

Informed consent statement

N/A

Data availability statement

N/A

CRediT authorship contribution statement

Arfan Ahmed: Conceptualization, Validation, Formal analysis, Writing – original draft, Writing – review & editing. **Sarah Aziz:** Conceptualization, Methodology, Validation, Formal analysis, Writing –

original draft, Writing – review & editing. **Carla T. Toro:** Writing – original draft, Writing – review & editing. **Mahmood Alzubaidi:** Methodology, Validation. **Sara Irshaidat:** Data curation. **Hashem Abu Serhan:** Data curation. **Alaa A. Abd-alrazaq:** Conceptualization, Methodology, Supervision. **Mowafa Househ:** Conceptualization, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

N/A.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpbup.2022.100066.

Appendix A. Study reference table

Reference	Study Title
S1	Martínez-Castaño, R., J.C. Pichel, and D.E. Losada, *A big data platform for real time analysis of signs of depression in social media.* International Journal of Environmental Research and Public Health, 2020. **17**(13): p. 4752.
S2	Ismail, N.H., et al., *A deep learning approach for identifying cancer survivors living with post-traumatic stress disorder on Twitter.* BMC Medical Informatics and Decision Making, 2020. **20**(4): p. 1–11.
S3	Kim, J., et al., *A deep learning model for detecting mental illness from user content on social media.* Scientific Reports, 2020. **10**(1): p. 1–6.
S4	Ding, Y., et al., *A depression recognition method for college students using deep integrated support vector algorithm.* IEEE Access, 2020. **8**: p. 75,616–75,629.
S5	Jain, S., et al. *A Machine Learning based Depression Analysis and Suicidal Ideation Detection System using Questionnaires and Twitter*. in *2019 IEEE Students Conference on Engineering and Systems (SCES)*. 2019. IEEE.
S6	Wang, Y., et al. *A Multimodal Feature Fusion-Based Method for Individual Depression Detection on Sina Weibo*. in *2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC)*. 2020. IEEE.
S7	Mahapatra, A., S.R. Naik, and M. Mishra. *A Novel Approach for Identifying Social Media Posts Indicative of Depression*. in *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*. 2020. IEEE.
S8	Tariq, S., et al., *A novel co-training-based approach for the classification of mental illnesses using social media posts.* IEEE Access, 2019. **7**: p. 166,165–166,172.
S9	Kumar, A., A. Sharma, and A. Arora. *Anxious depression prediction in real-time social data*. in *International Conference on Advances in Engineering Science Management & Technology (ICAESMT)−2019, Uttaranchal University, Dehradun, India*. 2019.
S10	Wang, X., et al. *Assessing depression risk in Chinese microblogs: a corpus and machine learning methods*. in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*. 2019. IEEE.
S11	Wongkoblap, A., M.A. Vadillo, and V. Curcin. *Classifying depressed users with multiple instance learning from social network data*. in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. 2018. IEEE.
S12	Depression Analysis from Social Media Data in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique
S13	Uddin, A.H., D. Bapery, and A.S.M. Arif. *Depression Analysis from Social Media Data in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique*. in *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*. 2019. IEEE.
S14	Jain, V., et al. *Depression and Impaired Mental Health Analysis from Social Media Platforms using Predictive Modelling Techniques*. in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. 2020. IEEE.
S15	Al Asad, N., et al. *Depression detection by analyzing social media posts of user*. in *2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON)*. 2019. IEEE.
S16	Islam, M.R., et al., *Depression detection from social network data using machine learning techniques.* Health information science and systems, 2018. **6**(1): p. 1–12.
S17	Deshpande, M. and V. Rao. *Depression detection using emotion artificial intelligence*. in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. 2017. IEEE.
S18	Li, Y., et al., *Depressive Emotion Detection and Behavior Analysis of Men Who Have Sex With Men via Social Media.* Frontiers in Psychiatry, 2020. **11**: p. 830.
S19	Shrestha, A. and F. Spezzano. *Detecting depressed users in online forums*. in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 2019.
S20	Islam, M.R., et al. *Detecting depression using k-nearest neighbors (knn) classification technique*. in *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*. 2018. IEEE.
S21	Leis, A., et al., *Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis.* Journal of medical Internet research, 2019. **21**(6): p. e14199.
S22	Kamite, S.R. and V. Kamble. *Detection of Depression in Social Media via Twitter Using Machine learning Approach*. in *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*. 2020. IEEE.
S23	Tadesse, M.M., et al., *Detection of depression-related posts in reddit social media forum.* IEEE Access, 2019. **7**: p. 44,883–44,893.
S24	Shah, F.M., et al. *Early Depression Detection from Social Network Using Deep Learning Techniques*. in *2020 IEEE Region 10 Symposium (TENSYMP)*. 2020. IEEE.
S25	Cacheda, F., et al., *Early detection of depression: social network analysis and random forest techniques.* Journal of medical Internet research, 2019. **21**(6): p. e12554.
S26	Ta, N., et al., *Evaluating Public Anxiety for Topic-based Communities in Social Networks.* IEEE Transactions on Knowledge and Data Engineering, 2020.
S27	Lushi Chen, L., et al., *Examining the Role of Mood Patterns in Predicting Self-Reported Depressive symptoms.* arXiv e-prints, 2020: p. arXiv: 2006.07887.
S28	Ricard, B.J., et al., *Exploring the utility of community-generated social media content for detecting depression: an analytical study on Instagram.* Journal of medical Internet research, 2018. **20**(12): p. e11817.
S29	Gruda, D. and S. Hasan, *Feeling anxious? Perceiving anxiety in tweets using machine learning.* Computers in Human Behavior, 2019. **98**: p. 245–255.
S30	Wang, X., et al. *Leverage Social Media for Personalized Stress Detection*. in *Proceedings of the 28th ACM International Conference on Multimedia*. 2020.
S31	Narynov, S., et al. *Machine Learning Approach to Identifying Depression Related Posts on Social Media*. in *2020 20th International Conference on Control, Automation and Systems (ICCAS)*. 2020. IEEE.
S32	Victor, D.B., et al. *Machine Learning Techniques for Depression Analysis on Social Media-Case Study on Bengali Community*. in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. 2020. IEEE.
S33	Sadeque, F., D. Xu, and S. Bethard. *Measuring the latency of depression detection in social media*. in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 2018.
S34	Arora, P. and P. Arora. *Mining twitter data for depression detection*. in *2019 International Conference on Signal Processing and Communication (ICSC)*. 2019. IEEE.
S35	Jamil, Z., *Monitoring tweets for depression to detect at-risk users*. 2017, Université d'Ottawa/University of Ottawa.
S36	Oyong, I., E. Utami, and E.T. Luthfi. *Natural language processing and lexical approach for depression symptoms screening of Indonesian twitter user*. in *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*. 2018. IEEE.
S37	Yao, X., et al., *Patterns and Longitudinal Changes in Negative Emotions of People with Depression on Sina Weibo.* Telemedicine and e-Health, 2020. **26**(6): p. 734–743.
S38	Aldarwish, M. M., & Ahmad, H. F. (2017, March). Predicting depression levels using social media posts. In *2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS)* (pp. 277–280). IEEE.

(*continued on next page*)

(*continued*)

S39	Hu, Q., et al., *Predicting Depression of Social Media User on Different Observation Windows*.
S40	De Choudhury, M., et al. *Predicting depression via social media*. in *Proceedings of the International AAAI Conference on Web and Social Media*. 2013.
S41	Thorstad, R. and P. Wolff, *Predicting future mental illness from social media: A big-data approach.* Behavior research methods, 2019. **51**(4): p. 1586–1600.
S42	Wongkoblap, A., M.A. Vadillo, and V. Curcin. *Predicting Social Network Users with Depression from Simulated Temporal Data*. in *IEEE EUROCON 2019–18th International Conference on Smart Technologies*. 2019. IEEE.
S43	Vanlalawmpuia, R. and M. Lalhmingliana. *Prediction of Depression in Social Network Sites Using Data Mining*. in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. 2020. IEEE.
S44	De Choudhury, M., S. Counts, and E. Horvitz. *Predicting postpartum changes in emotion and behavior via social media*. in *Proceedings of the SIGCHI conference on human factors in computing systems*. 2013.
S45	Zhou, T.H., G.L. Hu, and L. Wang, *Psychological disorder identifying method based on emotion perception over social networks.* International journal of environmental research and public health, 2019. **16**(6): p. 953.
S46	Tsugawa, S., et al. *Recognizing depression from twitter activity*. in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2015.
S47	Lee, J.H., J.M. Kim, and Y.S. Choi. *SNS data visualization for analyzing spatial-temporal distribution of social anxiety*. in *Proceedings of the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory*. 2016.
S48	De Choudhury, M., S. Counts, and E. Horvitz. *Social media as a measurement tool of depression in populations*. in *Proceedings of the 5th annual ACM web science conference*. 2013.
S49	Shatte, A.B., et al., *Social media markers to identify fathers at risk of postpartum depression: a machine learning approach.* Cyberpsychology, Behavior, and Social Networking, 2020. **23**(9): p. 611–618.
S50	Syarif, I., N. Ningtias, and T. Badriyah. *Study on Mental Disorder Detection via Social Media Mining*. in *2019 4th International Conference on Computing, Communications and Security (ICCCS)*. 2019. IEEE.
S51	Lin, H., et al. *User-level psychological stress detection from social media using deep neural network*. in *Proceedings of the 22nd ACM international conference on Multimedia*. 2014.
S52	Trotzek, M., S. Koitka, and C.M. Friedrich, *Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences.* IEEE Transactions on Knowledge and Data Engineering, 2018. **32**(3): p. 588–601.
S53	Guntuku, S.C., et al. *What twitter profile and posted images reveal about depression and anxiety*. in *Proceedings of the international AAAI conference on web and social media*. 2019.
S54	Cong, Q., et al. *XA-BiLSTM: A deep learning approach for depression detection in imbalanced data*. in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2018. IEEE.

Appendix B. Data extraction form

Concept	Definition
Study Characteristics	
Author	The first author of the study.
Year Submission	The year in which the study was submitted.
Country of publication	The country where the study was published.
Publication type	The paper type (i.e., peer-reviewed, conference or preprint).
AI technique characteristics	
AI models/ algorithms	The specific AI models or algorithms that were used (e.g., Decision tree, Random forest, Convolutional neural network).
Data source trained upon	what data the predictive models are trained upon I.e. which social media platform facebook,twitter, any social forum,or survey forum
Metric used for results measurement	what metrics used to evaluate the detection process i.e. accuracy, recall, sensitivity or AUC
Dataset Characteristics	
Data types	what data the predictive models are trained upon I.e. which social media platform facebook,twitter, any social forum,or survey forum
Language of Data samples	what language data analysis was performed upon i.e. English tweets, Chinese's tweets or any other language
Dataset size	The total number of data that were used (I.e., in case of tweets how many tweets are being analysed)
Number of Users	The total number of users identified in each study to analyze their social media accounts and extract data from.
Mental illness identification criteria	How the users regarded as having depression or anxiety were identified,i.e. any depression test (PHQ,CES-D etc.), self-declared or randomly selected etc.
Mental disorders identified	what type of mental order patient detected to tweet or use social media most to express themself

References

[1] D.S. Hasin, et al., Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States, JAMA Psychiatry 75 (4) (2018) 336–346.

[2] R.C. Kessler, E.J. Bromet, The epidemiology of depression across cultures, Annu. Rev. Public Health 34 (2013) 119–138.

[3] A. Konnopka, H. König, Economic burden of anxiety disorders: a systematic review and meta-analysis, Pharmacoeconomics 38 (1) (2020) 25–37.

[4] D. Vigo, G. Thornicroft, R. Atun, Estimating the true global burden of mental illness, Lancet Psychiatry 3 (2) (2016) 171–178.

[5] O. Plana-Ripoll, et al., A comprehensive analysis of mortality-related health metrics associated with mental disorders: a nationwide, register-based cohort study, Lancet 394 (10211) (2019) 1827–1835.

[6] K. Hawton, et al., Risk factors for suicide in individuals with depression: a systematic review, J. Affect. Disord. 147 (1–3) (2013) 17–28.

[7] S.K. Brooks, et al., The psychological impact of quarantine and how to reduce it: rapid review of the evidence, Lancet (2020).

[8] Shihabuddin, L. How to manage stress and anxiety from coronavirus (COVID-19). 2020 [cited 2020 13/10/2020]; Available from: https://www.rwjbh.org/blog/2020/march/how-to-manage-stress-and-anxiety-from-coronaviru/.

[9] L.H. Andrade, et al., Barriers to mental health treatment: results from the WHO World Mental Health surveys, Psychol. Med. 44 (6) (2014) 1303–1317.

[10] A.A.-A.J.S.D.A.C.T.T.A.A.M.A.M. Househ, The performance of artificial intelligence-driven technologies in diagnosing mental disorders: an umbrella review, J. Med. Internet Res. (2021).

[11] L. Nichols, et al., Derivation of a prediction model for a diagnosis of depression in young adults: a matched case-control study using electronic primary care records, Early Interv. Psychiatry 12 (3) (2018) 444–455.

[12] D.M. Low, K.H. Bentley, S.S. Ghosh, Automated assessment of psychiatric disorders using speech: a systematic review, Laryngoscope Investig. Otolaryngol. 5 (1) (2020) 96–116.

[13] M. De Choudhury, S. Counts, E. Horvitz, Predicting postpartum changes in emotion and behavior via social media, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, SIGCHI, 2013.

[14] S. Chancellor, M. De Choudhury, Methods in predictive techniques for mental health status on social media: a critical review, NPJ Digital Med. 3 (1) (2020) 43.

[15] A.B.R. Shatte, D.M. Hutchinson, S.J. Teague, Machine learning in mental health: a scoping review of methods and applications, Psychol. Med. 49 (9) (2019) 1426–1448.

[16] R. Skaik, D. Inkpen, Using social media for mental health surveillance: a review, ACM Comput. Surv. 53 (6) (2020). Article 129.

[17] A.C. Tricco, et al., PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation, Ann. Intern. Med. 169 (7) (2018) 467–473.

[18] S.C. Guntuku, et al., Detecting depression and mental illness on social media: an integrative review, Curr. Opin. Behav. Sci. 18 (2017) 43–49.

[19] D.C. Rettew, et al., Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews, Int. J. Methods Psychiatr. Res. 18 (3) (2009) 169–184.
[20] Y. Wang, et al., A Multimodal Feature Fusion-Based Method for Individual Depression Detection on Sina Weibo, in: Proceedings of the 39th International Performance Computing and Communications Conference (IPCCC), IEEE, 2020.
[21] Q. Hu, et al.. Predicting Depression of Social Media User on Different, Observation Windows, 2015. Dec, https://ieeexplore.ieee.org/document/7396831.
[22] S. Tariq, et al., A novel co-training-based approach for the classification of mental illnesses using social media posts, IEEE Access 7 (2019) 166165–166172.
[23] R. Martínez-Castaño, J.C. Pichel, D.E. Losada, A big data platform for real time analysis of signs of depression in social media, Int. J. Environ. Res. Public Health 17 (13) (2020) 4752.
[24] R. Kohn, et al., The treatment gap in mental health care, Bull. World Health Organ. 82 (2004) 858–866.
[25] F. Ursin, C. Timmermann, F. Steger, Ethical implications of Alzheimer's disease prediction in asymptomatic individuals through artificial intelligence, Diagnostics 11 (3) (2021) 440.