International Journal on Semantic Web and Information Systems
Volume 20 • Issue 1

Machine Learning for Depression Detection on Web and Social Media:
A Systematic Review

Lin Gan, University of Sanya, China
https://orcid.org/0000-0002-2263-8836

Yingqi Guo, Department of Social Work, Department of Geography (joint), Smart Society Lab., Hong Kong Baptist University, Hong Kong*

Tao Yang, University of Sanya, China*
https://orcid.org/0000-0001-7061-9835

ABSTRACT

Depression, a significant psychiatric disorder, affects individuals' physical well-being and daily functioning. This focused analysis provides a comprehensive exploration of contemporary research conducted between 2012 and 2023 that delves into the utilization of sophisticated machine learning methodologies aimed at identifying correlates of depression within social media content. Our study meticulously dissects various data sources and performs a comprehensive examination of different machine learning algorithms cited in the researched articles and literature, aiming to pinpoint an approach that can enhance detection accuracy. Furthermore, we have scrutinized the use of varied data from social media platforms and pinpointed emerging trends, notably spotlighting novel applications of artificial neural networks for image processing and classification, along with advanced gait image models. Our results offer essential direction for future research focused on enhancing detection precision, acting as a valuable reference for academic and industry scholars in this field.

KEYWORDS

Depression Detection, Machine Learning, Multimodal, Sentiment Analysis, Social Media, Text Mining

1. INTRODUCTION

Depression, a pervasive mental disorder plaguing nearly 5% of adults worldwide, is characterized by chronic low mood, enduring sadness, and feelings of hopelessness and helplessness. Those afflicted often lose the intrinsic pleasure in activities they once relished, grapple with alterations in sleep and appetite, and consistently struggle with fatigue and concentration difficulties. According to the Blue Book of Depression in China 2022, the lifetime prevalence of depression among adults in China stands staggeringly high at 6.8%, equating to an approximate total of 95 million individuals undergoing this debilitating condition. The advent and progression of the COVID-19 pandemic have seemingly

DOI: 10.4018/IJSWIS.342126
*Corresponding Author

amplified the global burden of mental disorders. The pandemic has escalated the rates of severe depression and anxiety by 28% and 26%, respectively, as per research conducted by the University of Washington and the University of Queensland (Santomauro Mantilla Herrera Shadid Zheng Ashbaugh et al., 2021). Women and younger individuals appear to endure the most brunt (Qian & Yan, 2013).

The backdrop of this research lies in the rising prevalence and swift evolution of social media, which has sparked increasing interest in investigating its implications for mental health. The surge in social media users has motivated scholars to mull over the possibilities of deploying social media information to discern depression and related mental health issues. Previous investigations principally depend on the utilization of conventional paradigms and techniques, such as text analysis, sentiment analysis, and sociometric analysis.

In light of the escalating empirical data highlighting the association between particular linguistic patterns and affective manifestations on digital social networks as putative markers of depressive disorders, the incorporation of machine learning (ML) into the detection of depression has garnered substantial interest. The adeptness of ML methodologies lies in their ability to meticulously appraise the presence of depression through the examination of textual information amassed from social media platforms. A comprehensive meta-analysis was conducted, concentrating on the deployment of ML models to proficiently identify signs of depression within social media content (Liu et al., 2022).

1.1 Machine Learning

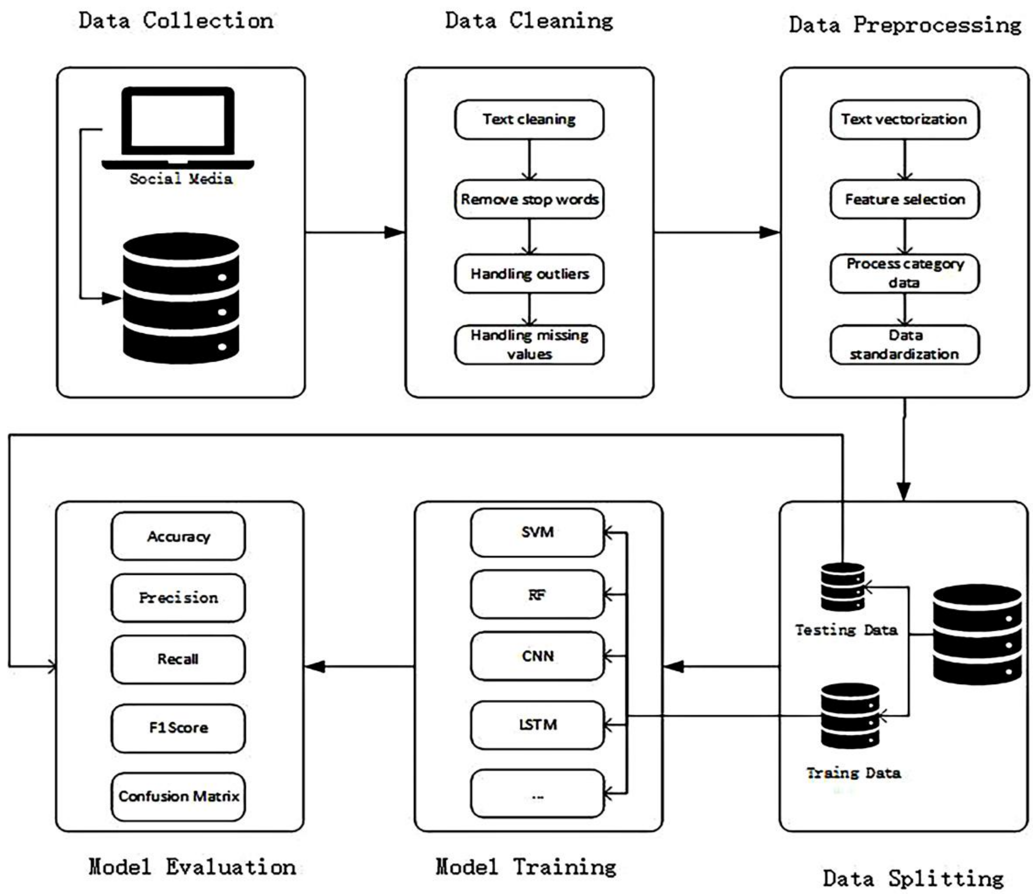
The typical procedure of ML in identifying depression embodies multiple stages, with the primary ones elaborated in Figure 1. In the data collection phase, a large amount of social media textual data, including user-generated content and comments, needs to be gathered. The preprocessing stage involves operations such as cleaning, deduplication, and tokenization to eliminate irrelevant information and noise, making the data more standardized and easy to process. Feature extraction utilizes various algorithms and techniques to extract depression-related features from the textual data, such as word frequency and sentiment tendencies. Finally, by training the model to classify and recognize these features, depression detection is achieved (Figure 1).

In the wake of the advancements in deep learning, an expanding repertoire of algorithms has been extensively integrated for the purpose of detecting depression. Deep learning technology can automatically extract deep features from text, understand the meanings of words and sentences in context, and better identify depression-related text. Compared to traditional methodologies, a plethora of emerging novel models and algorithms in recent years have demonstrated enhanced accuracy in detecting depression and proficiency in addressing time-series issues. These advancements play a crucial role in the prediction and diagnosis of depression, particularly considering the significance of early and sequential information for patients afflicted with depressive disorders. Therefore, deep learning considers early-time series, preventing the algorithm from inaccurately identifying depression patients, and exhibits better generalization performance.

It should be noted that, while ML has made strides in detecting depression on social media, challenges and limitations remain, such as data acquisition and annotation, feature extraction and selection, model accuracy and generalization ability, and clinical validation. Given the need for further research to understand the algorithms related to depression detection on social media platforms, in this study the authors reviewed and summarized the detection models based on ML. The authors aimed to understand the evolution of the most relevant and recent algorithms and summarize previous literature. Specifically, the purposes of this review were to determine:

1. The application of relevant algorithms in detecting depression on social media platforms.
2. The innovations and shortcomings when reviewing and comparing the methods of depression detection in the literature.
3. How to select the best model among multiple algorithms and multimodal evaluations.
4. The methodological limitations in using social media data for mental health research that can be critically distilled from the literature.

Figure 1. Data Preprocessing and Model Training Flowchart



This study focused on existing methods utilizing different artificial intelligence (AI) and ML technologies. The second section describes the inclusion and exclusion criteria along with specific methods employed. The third section presents the literature review. The fourth section offers a comparison and analysis of the technologies and literature. The fifth section delves deeply into future development trends, potential limitations, and ethical considerations. The sixth section is the conclusion.

2. METHODS

Patient and public involvement
No patient involved

2.1 Search Strategy

The authors conducted a comprehensive scoping review to identify relevant keywords. The researchers meticulously carried out an exhaustive survey encompassing all scholarly-reviewed research published prior to August 31, 2023. This review effectively utilized six extensive databases that cover diverse domains: Web of Science Core Collection, PubMed, ProQuest, Ovid MEDLINE, PsycINFO, and Embase. The authors used the keywords identified in the scoping review as the search strategy: (machine learning OR Machine Learning OR ML OR Deep Learning OR deep learning OR DL)

Table 1. Inclusion and Exclusion Criteria

Criteria	Description
Inclusion	All ages, genders, ethnicities, and countries; peer-reviewed English articles; focus on early detection of depression using ML algorithms.
Exclusion	Grey literature, opinion articles, methodology/validity/feasibility studies, literature reviews, case studies, and exclusively medical or physiological studies.

and (Depreesion OR Depreesion Detaction OR mood* OR mental*) and (text AND multimodal OR multimode).

2.2 Inclusion Criteria

The scope of the inquiry was not circumscribed by specific categories of participants, thereby allowing for an examination of demographics across the entire spectrum, encompassing all age groups, genders, ethnicities, and nationalities. The authors deemed admissible for inclusion solely peer-reviewed publications in English. This study focused solely on the early detection of depression using ML algorithms.

2.3 Exclusion Criteria

The authors excluded grey literature, opinion articles, methodology/validity/feasibility studies, literature reviews, case studies, and exclusively medical or physiological studies.

2.4 Screening and Study Selection

The process of paper screening comprised four steps: (1) Using the literature management system Endnote to exclude duplicate papers; (2) screening by title; (3) screening by abstract; (4) screening by full text. The authors reconciled discrepancies within the preliminary selection through a combination of face-to-face and digital conferencing. Upon exhaustive evaluation of the complete manuscripts, the researchers achieved a unanimous accord on the definitive roster of articles to be shortlisted. Out of 10,980 citations identified from title searching, 20 studies met the inclusion criteria following full-text screening.

2.5 Data Extraction and Synthesis

The reviewer extracted data from the shortlisted articles into a summary table, categorizing them by author, publication year, dataset, algorithm model, and performance metric evaluation (Table 2).

3. LITERATURE REVIEW

The development of natural language processing (NLP) can be traced back to the 1950s, when early research focused on rule-based and knowledge-based methods such as writing grammar rules and dictionaries for sentence analysis. In the 1980s, improved computing power and the emergence of large corpora enabled statistical methods to gain dominance in NLP. In particular, in the past decade, the evolution of deep learning technology has catalyzed substantial advancements in the field of NLP. The advent of sophisticated deep neural network architectures, such as recurrent neural networks (RNN), long short-term memory (LSTM) networks, and transformers, has dramatically enhanced the efficacy and precision of NLP tasks.

Chiong et al. (2021) discussed several text preprocessing and text-based feature methods and ML classifiers. They used two publicly available datasets labeled Twitter to train and test ML models. They used three additional non-Twitter depression datasets (from Facebook, Reddit, and electronic diaries) to test the performance of trained models against other social media sources. Using the dynamic sampling experiment method, the over- and undersampling logistic regression (LR) scores were 8.7% and 10.4%, respectively. LR performed best in both sampling methods, and dynamic undersampling improved the accuracy of depression detection. In the integrated classifier experiment, random forest (RF) provided better results than all single and integrated classifiers, and the RF model was superior to other classifiers in using general text for depression detection.

Tong et al. (2023) used the cost-sensitive boosting pruning tree (CBPT) algorithm to detect depression. They proposed a new classifier, the CBPT, demonstrating its classification ability on two publicly accessible Twitter depression detection datasets. When using eight boosting classifiers to classify five public datasets (i.e., TTDD, CLPsych2015, LSVT, Statlog, Glas) [Mean Accuracy/ F1-Score±Standard Deviation], the CBPT showed the best results in four cases, that is, TTDD (Accuracy: 88.39±0.60, F1-Score: 86.90±0.62), CLPsych2015 (Accuracy: 70.69±1.84, F1-Score: 66.54±2.42), showing CBPT's superiority compared to other depression detection frameworks in two Twitter datasets.

According to Chen et al. (2018), perinatal depression can be detected through deep learning's LSTM algorithm in social media posts. Their study extracted and classified the emotional characteristics of WeChat users based on the LSTM network model, applying it to the screening scenario of perinatal depression. The dataset used in this study mainly came from WeChat, and the Edinburgh Postnatal Depression Scale was used to screen emotions for model validation (including 10 test items such as emotion, pleasure, guilt, anxiety, fear, insomnia, coping ability, sadness, crying, and self-imposed). In the article's experimental section, the authors confirmed that the deep learning LSTM network model can be used for perinatal depression screening, achieving similar results to screening using the Edinburgh Postnatal Depression Scale.

Alsagri and Ykhlef (2020) collected over 30 million tweets as the data source and used support vector machines (SVMs), naive Bayes (NB), and decision trees (DTs) as classifier techniques to identify the degree of depression. They used Rstudio IDE (R 3.3) version for data preparation, feature extraction, and classification tasks to determine whether Twitter users were depressed. Second, they trained the classifier using 10-fold cross-validation and tested it on a fixed test set. The experimental results showed that the rich feature discrimination model could achieve better results, with higher accuracy and F-measure scores in detecting depressed users.

According to Kabir et al. (2023), the degree of depression can be viewed by examining subtle signs of depression in microblog text information. The article's experimental section provides a new dataset of 40,191 tweets, labeled "depressed" or "not depressed" by expert annotators. Kabir et al. further classified each tweet into three levels within the "depressed" category: (1) Mild, (2) moderate, and (3) severe. Each label provided a related confidence score to verify the quality of the annotation. They selected several baseline models to evaluate the proposed dataset, including classic ML algorithms, SVM, bidirectional LSTM (BiLSTM), bidirectional encoder representations from transformers (BERT), and DistilBERT models. In the class-wise AUC-ROC (i.e., area under the curve and receiver operating characteristic, respectively) curve, DistilBERT had the highest linear ROC performance, with nondepression at 0.788841, mild at 0.747211, moderate at 0.787959, and severe at 0.866003. The final experimental results showed that even the most advanced models could not understand the contextual semantics of the data, and the keyword division between different degrees of depression was also evident. The same keywords could be expressed in different ways to convey different emotions. Kabir et al. concluded by discussing the need to establish a new vocabulary database with a larger semantic network to expand the seed terms.

Hemanthkumar and Latha (2019) conducted sentiment analysis on tweets, dividing a dataset of 43,000 tweets into training and testing data at a ratio of 70:30. The dataset consisted of two columns of text, the tweets, and the sentiment as positive (1) or negative (0), and was found using the NB classifier and SVM algorithm. In the preprocessing stage, steps such as emoji extraction, hyperlink removal, slang replacement, timestamp deletion, number deletion, symbol deletion, spelling correction, proper noun deletion, lemmatization, and stop word removal were performed. Then, the researchers used the bag-of-words model to train the model, enabling the identification of the frequency of occurrence of a particular word in the text, helping to build the prediction model. Then, they preprocessed and tested 30% of the test data to determine whether the tweet was positive or negative. Subsequently, they evaluated the results through a confusion matrix. The authors concluded that the NB algorithm is better than the SVM algorithm for depression detection.

Liaw and Chua (2022) combined social network analysis and ML to identify and predict the behavior of people with depression on Twitter. Their paper introduces user network and engagement features, which can fully reflect a user's social behavior and mental health status and effectively identify depression based on text. Liaw and Chua used four indicators to evaluate the model: Accuracy, precision, recall, and F1-score. DART (XGB) achieved the best performance in terms of accuracy and F1-score at 82.05%. Another finding was that the number of depression-related keywords in liked tweets in user participation behaviors, such as retweeting, liking, and sharing, was the most important indicator, with a gain of at least twice compared to the other 88% of the features. The topic modeling feature of liked tweets also had a higher gain and was important in detecting depression. In addition, features extracted from original tweets, replies, and liked tweets had higher gains and were more important in detecting depression than retweets and quoted tweets.

Yazdavar et al. (2017) suggested that clinical depression symptoms can be detected from Twitter posts using semisupervised ML. They aimed to reflect the PHQ-9 questionnaire used by clinicians by analyzing the language patterns and topic preferences of users self-reporting depressive symptoms on Twitter. This method stood out for its ability to capture clinical depression indicators, with an accuracy rate of 68% and a sensitivity rate of 72%. The improvements highlighted in the study include an active automatic detection method, a bottom-up approach using user tweet processing and distributed semantics to capture clinical depression, and another method based on combining the first latent Dirichlet allocation (LDA) model and a top-down SStot model to extract depressive symptoms from tweets using a dictionary. This method advances mental health monitoring through social media platforms by utilizing vocabulary usage patterns and topic expressions consistent with clinical findings in the PHQ-9 medical chart. Although this differs from traditional detection methods, it shows potential in promoting early detection and intervention of mental health care.

Fang et al. (2023) introduced an efficient and robust multimodal fusion model (MFM-Att). The experiment included various data types (i.e., audio, visual, and text), to capture a wide range of potential indicators of depression in the data, and a multilevel attention mechanism to extract depression-related features. The MFM-Att utilized intramodal attention mechanisms to learn depression-related features within each modality and abstract relationships between each modality and other modalities, allowing for a more detailed understanding of depression signals in the data. The multilevel attention was designed to minimize information redundancy caused by feature diversity and improve overall model performance.

Ma et al. (2023) used a deep learning approach to predict an individual's depression tendency automatically, based on multiple sources of personal data. The study utilized three models: A BERT-based social network model for text, a three-dimensional convolutional model for gait images, and an LSTM-based model for gait key points. A majority voting method, majority voting, was then used to construct a comprehensive model that could operate normally even when some data were missing, thereby improving the accuracy of detecting depression tendency. Since the study used five-fold cross-validation, there was some fluctuation in the results of each fold test. The final results showed that the accuracy of the social network model alone was 69.74%, the accuracy of the gait image model was 74.74%, and the accuracy of the gait key point model was 69.69%. After integrating these models through ensemble learning, the accuracy rate reached 91.59%, improving the model's stability. It is worth noting that the use of five-fold cross-validation in this study resulted in longer training times for the models, necessitating more advanced methods to improve convergence speed and optimize balance.

Burdisso et al. (2019) provided a new supervised learning model, SS3, for text classification, naturally supporting incremental classification and learning. SS3 was designed as a general framework for handling early risk detection problems. Burdisso et al. evaluated the model on the CLEF eRisk2017 pilot task and found it had lower computational costs and the ability to explain its basic principles. The experimental results showed that the SS3 classifier outperformed other supervised ML models (i.e., SVM, multinomial NB, and neural networks) and standard classifiers. The text classifier SS3 aims to meet the requirements of incremental classification, early classification, and support for interpretability simultaneously. During document classification, SS3 generated a confidence vector by decomposing the input document into different hierarchical structures and applying specific functions. This process can meet the requirement of early classification while gradually generating the entire document's classification results through recursive propagation of information, thus achieving incremental classification. Finally, the actual classification operation was performed based on the final confidence vector.

Zeberga et al. (2022) developed a novel framework for identifying depression and anxiety-related posts on social media using BERT and knowledge distillation techniques. The system employed word2vec, BERT, and BiLSTM to analyze and detect signs of depression and anxiety in social media posts. It leveraged the BERT to capture the contextual and semantic nuances of mental health text data while utilizing a BiLSTM sequence processing model as a classifier that can account for both preceding and succeeding words in a sentence. Additionally, the study incorporated knowledge distillation techniques to transfer the knowledge from a large pretrained model (i.e., BERT) to a smaller model (i.e., DistilBERT) for improved performance and accuracy. The data collection framework was based on Reddit and Twitter, and, after numerous hyperparameter optimizations, the model achieved an accuracy of 98%, surpassing the performance of other related models.

Wani et al. (2023) presented an AI and deep learning-based model for a binary classification problem, specifically identifying patients with depression on social media platforms. In the experimental section, they captured behavioral-biological feature signals using Word2Vec and term frequency-inverse document frequency (i.e., TF-IDF) models incorporating mixed features. Wani et al. trained convolutional neural network (CNN) and LSTM models to utilize these features. They collected six distinct datasets, resulting in 24 experiments. The researchers employed evaluation metrics such as accuracy, recall, precision, and F1-score to assess the performance of the learning models. The experimental findings revealed that both Word2Vec LSTM and Word2Vec (CNN with LSTM) models achieved exceptional accuracy rates of 99.02% and 99.01%, respectively, outperforming all other experiments conducted in this study. Furthermore, Word2Vec features better detected depressive symptoms on Facebook (95.02% accuracy using CNN) and YouTube (98.15% accuracy using CNN with LSTM).

Angskun et al. (2022) presented a novel model capturing depression symptoms effectively by analyzing Twitter users' demographic characteristics and text sentiment. They conducted the experiment in four stages. First, they tested various ML techniques, including SVMs, DTs, NB, RF, and deep learning, using different features to evaluate their performance. The results revealed that the mixed features (i.e., a combination of demographic characteristics and Twitter user information) achieved the highest average F-measure of 0.728 among all ML techniques. Angskun et al. employed three feature selection methods (i.e., RF, variance analysis, and SVM recursive feature elimination) in a second experiment to enhance depression detection performance. They drew three conclusions from this experiment: (1) The optimal number of selected features decreased from 24 to 19 for RF, 15 for variance analysis, and 22 for SVM recursive feature elimination; (2) variance analysis demonstrated significantly faster processing time than RF and SVM recursive feature elimination; (3) the number of features did not impact the processing time of feature selection. In the third experiment, they measured the construction time and usage time of models based on RF, variance analysis, and SVM recursive feature elimination using different numbers of features determined from previous experiments. The experimental results indicated that variance analysis exhibited superior performance in terms of feature selection. Finally, based on evaluation metrics such as F-measure accuracy precision and recall rate, the researchers concluded that RF was the most suitable ML technique for detecting depression.

Tadesse et al. (2019) conducted sentiment analysis on Reddit posts, utilizing Linguistic Inquiry and Word Count (LIWC) categories, N-gram probability, LDA models, and various combinations of these models as baseline features to detect depression among users. The researchers employed four main classifiers and one artificial neural network classifier, while they used 10-fold cross-validation for result verification. By comparing different classification techniques, the experimental findings demonstrated that the most effective sentiment classification model was LIWC+LDA+bigram combined with a multilayer perceptron (MLP) neural network model. This combination yielded superior performance in depression detection with an accuracy rate of 91% and an F1-score of 0.93. Notably, this outcome surpassed models utilizing single features (e.g., LIWC, LDA or N-gram) or only certain feature combinations. Furthermore, a significant finding indicated that appropriate feature selection and multifeature combination can greatly enhance depression prediction performance.

Almars (2022) analyzed Arabic social media content to understand users' emotions and detect signs of depression. Due to the complexity of Arabic and the lack of available resources, the detection of depression from Arabic social media has lagged behind that of other languages. Almars manually classified 6000 tweets as depressive or nondepressive. The experimental element of the study combined a BiLSTM with an attention mechanism. Compared to traditional ML methods, this method better handled complex language phenomena, captured semantic information, and improved the model's expression ability. On the other hand, the attention mechanism allowed the model to focus on key information related to depression when processing input sequences, thereby improving the model's detection accuracy. Through comparative experiments, Almars compared this model with traditional ML methods, demonstrating its superiority in depression detection tasks, with an accuracy rate of 83%.

Ghosh et al. (2022) developed an end-to-end multimodal multitask (MT) system. The framework combined information from both text and image modalities and used bidirectional gated recurrent unit (i.e., BiGRU) and attention mechanisms to model contextual information. The MT system utilized Twitter users' metadata information, such as user descriptions, geographical locations, and avatar URLs, to study the psychological states of depressed and nondepressed users. The experimental data were divided into a benchmark dataset of depressed and nondepressed users, further subdivided into two labeled datasets, namely, D1 (depressed users) and D2 (nondepressed users), and an unlabeled dataset, D3, (depressed candidate dataset). Emotional information was introduced as another feature of the task-specific output layer in the penultimate layer of the proposed MT framework, and finally linearly connected features from different modalities and passed them through a fully connected layer to the output layer to obtain the best system performance. The experimental results revealed that the proposed method achieved 70% accuracy, superior to other model methods.

Guo et al. (2023) developed a Chinese dataset for depression detection based on social media, using an extended depression dictionary (DUT-SL) to extract depression-related lexical features from text and correlation measures to fuse these features. Then, they used ML methods to analyze the relationship between these features and depression. The experimental results showed that the Soft Voting model performed best at each k value (300, 400, 500, 600, 700). When k=500, the accuracy of Soft Voting reached the highest level of 94.36%, with an F1-score of 94.28%. In addition, LR also demonstrated excellent and consistent performance, exceeding Soft Voting at k=700. The experimenters also used the WU 3D dataset for validation, and the second set of experimental results showed that the light gradient boosting machine (LightGBM) model achieved the best performance at each k value. When k=100, LightGBM achieved the highest accuracy of 96.14% and an F1-score of 93.78%. In summary, ensemble learning models performed better than others, with the most stable performance from LR.

Neuman et al. (2012) developed an automatic depression screening tool, pedesis, which extracts relevant concept domains describing depression from the metaphorical relationships hidden in the network. Psychological experts used this information to construct a "depression dictionary." At the same time, the dictionary automatically assessed the severity of depression in text or whether the text involves the topic of depression. In the experiment, Neuman et al. tested pedesis on three corpora. Compared with the expert's prediction, the system's prediction accuracy for depression problems increased by 9%. On the blog corpus, the system's correct classification rate was 84.2% ($p < .001$). By comparing the system's prediction with the psychological expert's judgment, the system achieved an average accuracy of 78% and an average recall rate of 76%.

Zhang H. et al. (2023) developed a neural network hybrid model, the mixed-type data distribution (MTDD), based on social media data, analyzing the content posted by users on social media platforms to help build a postlevel method to detect individual depression tendencies. Compared with existing methods, the MTDD model had low data acquisition costs and was easy to operate. At the same time, it avoided the problem of unpublished and incomplete data in depression detection. Second, the MTDD model was based on the hybrid deep neural network model, combining the advantages of CNN and the BiLSTM network to avoid the problem of poor generalization ability of a single model in identifying depression tendency. Third, the MTDD model learning was based on multimodal features to learn the vector representation of depressive text, including text features, semantic features, and domain knowledge, making the model more robust. The experimental results show that the MTDD model could detect users who may have depression tendencies with an F1-score of 95%.

Anshul et al. (2023) conducted a study on the detection of depressed and nondepressed users in a COVID-19 dataset on Twitter. They proposed an innovative multimodal framework that combines text, user-specific, and image analysis to detect depression in social media users. In order to capture more contextual information about users' emotional states, the authors introduced three methods. The first method captures external features by extracting URLs present in tweets, the second method extracts text content from images posted by users in tweets, utilizing intrinsic features that include a range of topics, emotions, depression-related information, and user-specific information extracted for specific users. Additionally, the authors described a user's information by extracting five groups of features from different modalities in the experiments. The third method incorporates a visual neural network deep learning model, which uses deep learning algorithms to generate embedding vectors for posted images to create visual feature vectors for prediction. In the depression classification stage, the authors trained the model using the feature set of all users. For experimentation, they separately employed LR, XGBoost, and neural network classifiers to detect depressed users in the given dataset. The classification results showed that combining the predictions of the three classifiers yielded better results than using any individual classifier alone. During the experimental evaluation, the researchers tested the proposed multimodal feature-based ensemble learning (MFEL) model using a new dataset and additional benchmark datasets, comparing its performance with other state-of-the-art algorithms. The final experimental results indicated that the MFEL model outperformed LR, XGBoost, and neural network models in terms of accuracy, recall, F1-score, and precision on the Tsinghua database. It achieved an F1-score of 91.9% and an accuracy of 91.7%. Performance on the COVID-19 new dataset surpassed several other models. In terms of accuracy, it reduced the error by 8.3% compared to MDL, and MFEL improved accuracy by 2.7% compared to MMTFIDF. This improvement is attributed to the MFEL method not only extracting multimodal features from user tweets, but also extracting various features including visual features from users. Additionally, MFEL extracted factors such as vocabulary categories and emotional intensity from tweets, contributing to the effective identification of depressive signs during the outbreak of the pandemic.

Bokolo and Liu (2023) proposed a comprehensive approach using ML techniques to detect depression from user tweets. The dataset comprised 600,000 Twitter texts, divided into two subsets, namely, a training set and a test set, with a data allocation ratio of 8:2. The researchers selected LR, Bernoulli NB, RF, DistilBERT, SqueezeBERT, DeBERTA, and RoBERTa models for model training. Experimental results showed that, in the detection of depression from tweets, the RoBERTa model achieved the highest accuracy (0.981) and the highest average precision (0.97). Bokolo and Liu employed a rigorous cross-validation process to evaluate all the ML models used. The dataset was divided into 10 parts, with each split also dividing the data into an 80% training set and a 20% validation set. This approach helped mitigate potential impacts of data partitioning on model performance and effectively assessed the models' generalization capabilities. The results and evaluations of this study indicated that LR, Bernoulli NB, and RF models demonstrated robust performance, exhibiting high accuracy and precise predictive capabilities. These models and algorithms have the ability to analyze social media data for mental health monitoring.

Zogan et al. (2022) introduced an advanced multiaspect depression detection hierarchical attention network (MDHAN) with interpretable multilevel attention networks. This model utilizes multiaspect features and word embedding features to automatically detect depressed users on social media while explaining the model's prediction results. The researchers obtained the experimental data from user posts on Twitter along with additional features. The so-called feature extraction involves encoding user posts using a two-level attention mechanism applied at the tweet and word levels. This calculates the importance of each tweet and word, capturing semantic sequence features from the temporal sequence of user posts. After establishing the hierarchical attention model, the importance of each tweet and word is computed through two levels of attention mechanisms. These attention weights are then used to train a deep neural network. In the experimental evaluation, the MDHAN model performed well due to its utilization of a rich set of features and the construction of a robust model by jointly learning the estimates of these features. Furthermore, MDHAN achieved optimal performance, with an F1-score of 89%, indicating that combining the hierarchical attention network (HAN) with a multiaspect user timeline semantic feature strategy is sufficient for detecting depression on Twitter.

Hussain et al. (2019) utilized language data from social media user posts to detect individual symptoms of depression. The researchers proposed a framework to identify social information as an important predictive factor for depression. Building upon this framework, the researchers developed an application (i.e., SMPP) that detects depression-related markers among Facebook users using data-driven methods and ML classification techniques. They conducted the evaluation on a dataset of 4,350 users assessed for depression using the Center for Epidemiologic Studies Depression Scale (CES-D). By this analysis, they identified a set of features distinguishing individuals with and without depression. Through a thorough assessment of key features in individuals with and without depression on social media, the results revealed the following:

1. **Educational Level:** The level of education may influence the open expression of symptoms and signs of depression, suggesting a need for further research to assess the relationship between education and depressive symptoms.
2. **Age and Relationship Status:** Age and relationship status are associated with depression. Single individuals and those with partners experienced an increased risk of depression with age, while married individuals had a lower incidence of depression.
3. **Gender:** The proportion of depression was lower in males compared to females, indicating a relationship between gender and depression.
4. **Social Network Size:** Users with smaller social circles who posted more depression-related markers were more likely to suffer from depression. Individuals with depression tended not to increase the size of their friend networks and leaned toward social isolation.
5. **Participation in Activities:** The frequency of participation in activities showed a negative correlation with depression. Individuals without depression were more likely to engage in social activities.
6. **Facebook Activity:** The number of Facebook activities was related to Facebook usage, and longer Facebook usage times were associated with more invitations.
7. **Use of Social Features:** Individuals with depression were less likely to use the like function and show interest in friends' activities.

These findings highlight the potential of analyzing social media data for detecting and understanding depression, providing insights into various factors associated with depressive symptoms and behaviors.

Ali et al. (2022) explored the classification of mental health issues in Facebook status updates using ML techniques. The focus of their study was primarily on the detection of depression, gradually expanding to six other common mental health issues: Anxiety, psychosis deviation, paranoia, unrealistic, and mild manic. To accurately determine users' psychological states, the researchers employed four different ML classifiers: RF, SVM, NB, and k-Nearest Neighbors (KNN). They incorporated various feature selection techniques during training and testing. Also, they compared the performance of ML classifiers (including SVM, KNN, and NB) in mental health behavior classification. The results showed that the RF classifier based on the uni-gram feature set outperformed other competing classifiers in classification effectiveness. To determine the optimal solution, the researchers also conducted an applicability analysis using a cost-effectiveness function, further confirming that the RF classifier achieved the best training and testing results on unlabeled status updates.

Ghosh et al. (2023) introduced a BiLSTM-CNN model based on an attention mechanism for detecting depressive moods in Bengali social media text. The proposed framework is lighter and more robust, compared to traditional models, displaying superior performance. To address the issue of scarce data, the researchers created a meticulously curated Bengali text dataset, undergoing various preprocessing stages and utilizing three unique embeddings to enhance accuracy. With the incorporation of an attention mechanism, the model achieved an impressive accuracy of 94.3%, sensitivity of 92.63%, and specificity of 95.12%. Extending its effectiveness to multiple languages including English, the model demonstrated excellent performance in comparative evaluations, surpassing traditional ML models, ensemble methods, transformers, and existing architectures. Given the surge in social media content, achieving automation through advanced models has become crucial for swiftly identifying and intervening in events related to depression. Ghosh et al.'s (2023) attention-based BiLSTM-CNN model, with its outstanding performance in detecting depressive text, brings hope for future research endeavors and underscores the importance of understanding the heterogeneous text forms prevalent in social media interactions. Additionally, researchers consider building models and promoting dataset diversification for low-resource languages such as Bengali as important focal points for future considerations.

Wu and Tang (2022), in response to the inability of NLP and traditional ML techniques to accurately detect multiple categories of cyberbullying across vast social media, proposed a hierarchical squeeze attention network (HSAN) model for the detection and classification of the severity of cyberbullying. This network model mainly includes four structures: Word encoding, squeeze attention mechanism (SAM) for words, sentence encoding, and SAM for sentences. The researchers tested it on data they collected from the "Gossip" section of the PTT forum and four public datasets. Subsequently, they compared their HSAN model with algorithms and network models such as SVM, RF, and TextCNN. The final experimental results showed that the newly developed SAM in Wu and Tang's (2022) article can fully analyze data and determine the severity of cyberbullying. Moreover, the HSAN network model demonstrated superior performance in detecting the severity of cyberbullying incidents compared to other ML and deep learning models, with an accuracy value of up to 62.3% and F1 values also higher than those of other network models.

Due to the current lack of a highly objective indicator for detecting mild depression on a global scale, Shin et al. (2021) researched the potential of using voice as a biomarker to gauge the severity of depression. They derived the experimental data from recordings taken during live interviews using the Mini International Neuropsychiatric Interview (MINI) with 93 volunteers, from which they extracted 21 voice features. Then, they used these 21 voice features to construct a model. The culmination of the experimental findings revealed that the MLP network model outperformed other ML modalities, attaining an AUC score of 65.9%, a sensitivity measure of 65.6%, and a specificity rate of 66.2%. This investigation underscored the voice alterations occurring during depressive episodes, substantiating the capacity of ML to accurately discriminate between nondepressed individuals and those with mild depression, hence suggesting the viability of vocal detection as a tool for identifying mild depressive states.

4. ANALYSIS

In recent years, numerous researchers have conducted meta-analyses or reviews to understand the available technologies for automatic depression detection. The above outline helps understand the latest computational models and advances in ML for depression detection and address the research gaps.

Based on the above review and literature, the authors generated Figure 2 illustrating the sources and distribution of data in depression detection. Currently, data for depression detection primarily originate from major social media platforms such as Facebook, Twitter, Reddit, Weibo, and WeChat. These platforms provide a wealth of real-time and diverse data, including text, images, and videos, facilitating the understanding of users' emotions and mental states. Common algorithms employed include RF, gradient boosting trees, RNN, LSTM, and gated recurrent unit. When implemented, these algorithmic frameworks are adept at capturing long-term dependencies and discerning patterns within temporal sequence data, rendering them apt for chronological forecasting and taxonomic assignments. Social media data, in comparison to traditional psychological assessment scales, more authentically reflect users' lives and emotional states. ML and NLP techniques allow to process and analyze these data, extracting features and patterns related to depression to train efficient depression detection models.

Regarding data types, the authors' literature references indicate a shift in depression detection types (Figures 2 and 3). Figure 2 illustrates the distribution proportions of depression user data across various source platforms. Twitter captures the largest portion with 41.0%, while Weibo and WeChat account for 2.6% and 5.1%, respectively. YouTube, Reddit, Facebook, and Instagram and other smaller social media platforms collectively contribute to 2.6%, 12.8%, 10.3%, and 25.6%. These findings potentially indicate variations in activities and engagement levels among depression users on these platforms.

Furthermore, in terms of the publication quantity and distribution over the years for ML-based depression detection (Figure 3), the authors' analysis of the accumulated scholarly literature reveals a gradual shift towards the utilization of machine algorithms and models in identifying depression

Figure 2. Data Source Chart

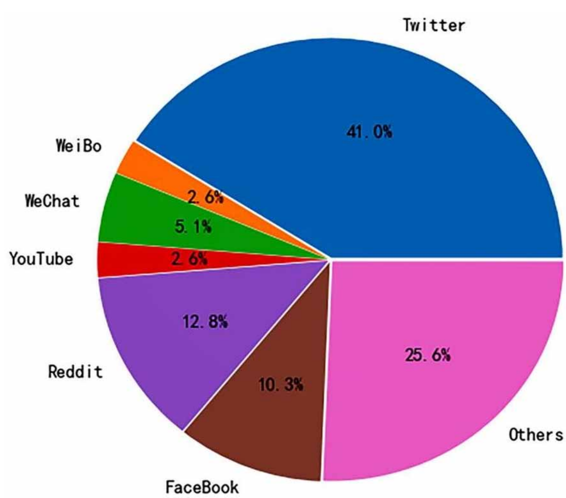


Figure 3. The Article Quantity and Publication Year

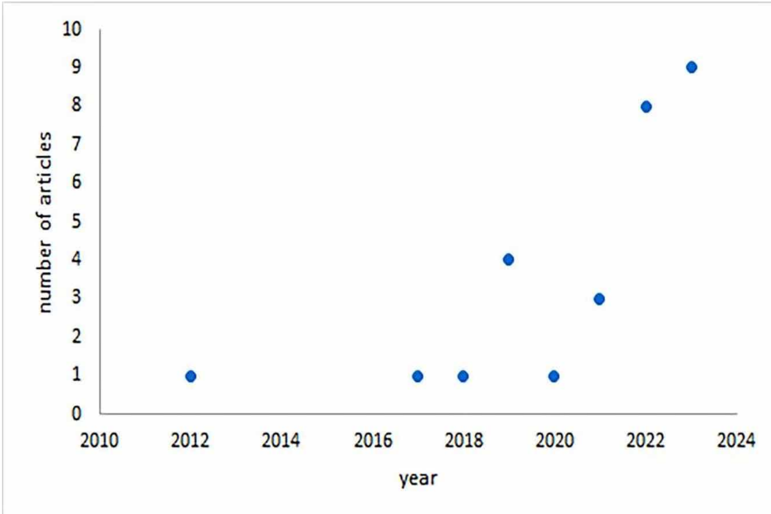
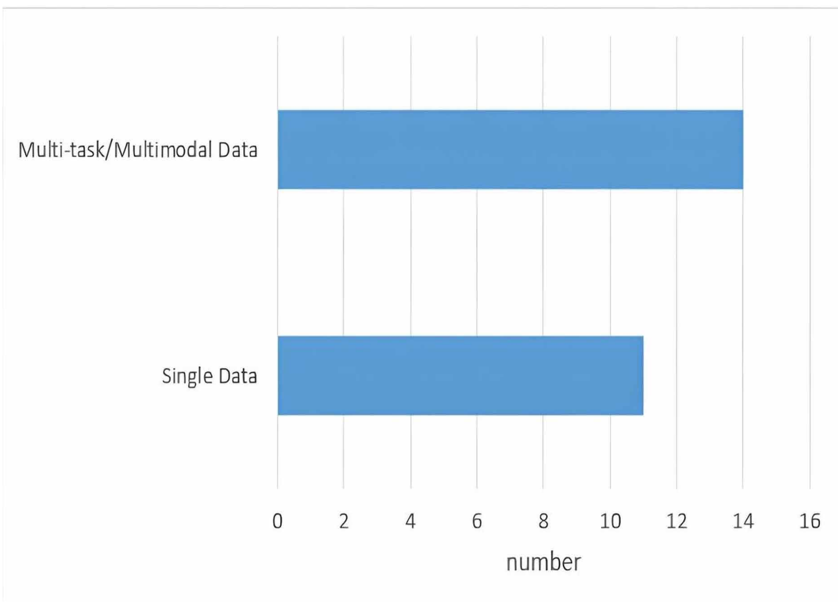


Figure 4. Data Type



over time. Initially, ML techniques for detecting depression primarily focused on traditional statistical methods and feature engineering. Researchers have meticulously extracted attributes associated with depression and employed various ML approaches for classification and prediction purposes. While these approaches have shown some level of success, their effectiveness is limited by the constraints inherent in manual feature selection and model complexity. As deep learning has advanced, there has been a shift towards its application in detecting depression. Deep learning exhibits strong capabilities for automated feature extraction, enabling it to discern more complex and sophisticated feature representations from vast data repositories. In the realm of conventional modalities, deep learning paradigms demonstrate exceptional proficiency in capturing intricate data configurations and nonlinear associations, thereby enhancing the precision and robustness of depression diagnostics. As time progresses, an increasing number of research endeavors are incorporating ML and deep learning strategies, reflecting a growing recognition of their potential to revolutionize the field of depression detection.

Figures 3 and 4 show that 56% of the articles focus on a single modality data type, while 44% explore multimodal approaches, indicating a growing trend in researching depression detection using various data sources. These data types include information from text, images, audio, and physiological signals, offering a more comprehensive and accurate diagnosis of depression. For instance, analyzing text and image content on social media can provide insights into a patient's social activity level, emotional expression, and lifestyle attitudes. Monitoring physiological signals such as electroencephalogram (EEG) or heart rate variability can offer information on a patient's physiological responses and emotional states. The transition from a specific data type (e.g., single modality) to multiple data types in depression detection involves leveraging deep learning methods to handle and understand multimodal information, providing a more holistic understanding of the psychological states and characteristics of depression or mental illness patients, thereby improving the accuracy of early depression detection.

Regarding the primary investigative objective, which aims to determine the utilization of relevant algorithms in detecting depression across social media platforms, Table 2 presents a comprehensive compilation of existing studies and methodologies employed for this purpose based on the literature review. Out of the 28 studies identified in this review, ML constitutes 23.68% of the methodological approaches, while deep learning algorithms account for 44.73%. To extract attributes associated with depression from such textual data, it undergoes preprocessing and subsequent analysis using NLP techniques.

Furthermore, as Tables 3 and 4 show, conventional ML methodologies have been extensively employed for the detection and categorization of depression. These algorithms achieve this by extracting features from textual information to identify potential indicators of depression, and then categorizing them into labels that reflect varying levels of severity or symptomatology.

In recent years, amidst the continuous advancement of deep learning technologies, a growing body of research has been dedicated to exploring depression diagnostic techniques based on deep neural networks. These methodologies often leverage architectures such as CNNs or RNN. By training on extensive datasets containing text, speech or image data, these approaches achieve automatic identification and classification of depression. Not only do these methods enhance detection accuracy, but they also provide more comprehensive and detailed analytical results, offering robust support for clinical diagnosis and treatment. Deep learning models have relatively better feature representation and generalization ability and perform better in depression detection tasks.

For the second study objective, all of the indicators in Table 4 are standard methods of evaluation for depression detection. These criteria can be used to evaluate the performance of the model on the training set and test set and whether the model can adapt to unknown datasets. As Table 4 shows, the literature review revealed nine evaluation criteria and performance evaluation methods, among which accuracy and F1-score are used most frequently. Accuracy is defined as the proportion of correctly classified samples to the total number of samples by the model, and it is one of the most frequently used methods for evaluating the performance of depression detection. Generally speaking, the higher the accuracy, the better the classification performance of the model.

In the context of depression detection, the F1 score is the average of Precision and Recall. When evaluating models, it is typically desired for both Precision and Recall to be at high levels to ensure accurate identification and minimize the error identification rate.

Regarding the third research objective, in solving multiple algorithms and multimodal evaluations, the pursuit of optimal models within multi-algorithm frameworks or multimodal data processing

Table 2. Overview of Relevant Surveys and Reviewed Papers

No.	Authors	Year of publication	Data collection volume	Datasets	Algorithm or model	Performance metrics	Results
1	Chiong et al.	2021	The data are 348,723 in all.	Twitter, Facebook, Reddit, electronic diary.	ML classifier oversampling (over- and undersampling) LR.	Over- and undersampling LR scores 8.7% and 10.4%, respectively.	The RF model performs better than other classifiers in detecting depression using general text.
2	Tong et al.	2022	69,672 tweets	Twitter	Cost-sensitive boosting pruning trees.	TTDD (accuracy: 88.39±0.60; F1-score: 86.900.62), CLPsych201 5 (accuracy: 70.69±1.84; F1-score: 66.54±2.42).	The new classifier, the CBPT, offers advantages over other depression detection frameworks.
3	Chen et al.	2018	Users of WeChat moment is 171702.	Wechat	LSTM	Preliminarily confirmation that the LSTM network model could be used for perinatal depression screening and had good performance, improving the efficiency and accuracy of early screening.	The deep learning LSTM network model can be used for perinatal depression screening.
4	Alsagri & Ykhlef	2020	300,000 tweets	Twitter	SVM, NB, DT.	SVM-Linear (SVM-L) the best model, showing an increase reaching 82% accuracy. F measure in SVM-L also increased, reaching 0.79.	The rich feature discrimination model could achieve better results, with higher accuracy and F-measure scores in detecting depressed users.
5	Kabir et al.	2023	40191 tweets	Twitter	SVM, BiLSTM, BERT, DistilBERT	In the class-wise AUC-ROC curve, DistilBERT has the highest linear ROC performance, with 0.788841 for nondepression, 0.747211 for mild depression, 0.787959 for moderate depression, and 0.866003 for severe depression.	Even the most advanced models could not understand the contextual semantics of the data, and the keyword division between different degrees of depression was also evident.
6	Hemanthkumar & Latha	2019	701 were users with depression, and 594 were nondepression users.	Twitter	NB classifier and SVM.	Evaluation using a confusion matrix concluded that the polynomial Bayes algorithm was superior to the SVM algorithm for depression detection.	The depression detection Web application mentioned in the experiment is capable of predicting depressive trends in tweets in both English and Arabic in advance.
7	Liaw & Chua	2022	23 million tweets	Twitter	User network and engagement features.	The models with proposed features outperformed other ML models without proposed features, and the best model achieved performance of 82.05% in terms of accuracy and F1-score. In addition, features for liked tweets, compared to other features, had a return of at least twice as much.	ML methods can effectively identify depression based on text, and data such as users' replies to tweets and liking behaviors are also important in detecting depression.
8	Yazdavar et al.	2017	4000 users	Twitter	Semisupervised statistical model.	Accuracy rate reached 68%, and precision rate reached 72%.	The authors' research methods showed potential in promoting early detection and intervention in mental health care.
9	Fang et al.	2023	185 samples	DAIC-WOZ	MFM-Att and root mean square error.	The MFM-Att model achieved a better performance than the state-of-the-art model in terms of root mean square error on the DAIC-WOZ dataset.	The authors' network model can understand depression signals in the data in more detail.

continued on following page

Table 2. Continued

No.	Authors	Year of publication	Data collection volume	Datasets	Algorithm or model	Performance metrics	Results
10	Ma et al.	2023	10400 tweets	Wechat, Weibo, mobile phone recording, ans social network data.	BERT model of social networks, three-dimensional convolutional gait image model, gait key points of LSTM, majority voting.	Accuracy of social network model in a single data source was 69.74%, accuracy of the gait image model was 74.74%, and accuracy of the gait key point model was 69.69%. After ensemble learning, accuracy reached 91.59%, improving the model's stability.	The multimodal research method is very important for the detection of depression.
11	Burdisso et al.	2019	9579 samples	Reddit	Supervised learning model SS3 translation.	SS3 classifier superior to other supervised ML models (e.g., SVM, multinomial naive Bayes, and neural network) and standard classifiers.	The authors' classifier outperforms other classifiers and has a lower computational cost.
12	Zeberga et al.	2022	100,000 tweets, 95,000 from Reddit.	Reddit and Twitter.	Word2ve, BERT, and BiLSTM.	Accuracy rate of 98% exceeded the performance of other models.	The authors' new model showed good performance in the study of depression, and the accuracy of the model is high.
13	Wani et al.	2023	The multiplatform data total is 52944 samples.	Facebook, Twitter, and YouTube.	CNN and LSTM.	Accuracy of Word2Vec LSTM and Word2Vec (CNN with LSTM) models reached 99.02% and 99.01%, respectively; accuracy of Word2Vec features on Facebook and YouTube reached 95.02% using CNN and 98.15% using CNN with LSTM, respectively.	The accuracy of Word2Vec LSTM and Word2Vec (CNN with LSTM) models is higher.
14	Angskun et al.	2022	222 tweets, 1522 retweets,and 16 hashtags.	Twitter	SVM, DT, NB, RF	Highest average F-measure was 0.728; the optimal number of features for RF, variance analysis, and SVM recursive feature elimination decreased from 24 to 19, 15, and 22, respectively; feature selection speed of variance analysis was 7x and 917x times faster than RF and SVM recursive feature elimination, respectively.	RF was the most suitable ML technique for detecting depression.
15	Tadesse et al.	2019	1841 posts	Reddit	LIWC, N-gram probability, LDA model, and multifeature combination.	LIWC+LDA+bigram and MLP neural network model best emotional classification model, performing best in depression detection, with accuracy of 91% and F1-score of 0.93.	The most effective sentiment classification model was LIWC+LDA+bigram combined with an MLP neural network model.
16	Almars	2021	6000 tweets	Twitter	BiLSTM	Accuracy rate reached 83%.	Comparative experiments proved that the authors' model has more advantages in depression detection tasks, compared with traditional ML methods for detecting depression.

continued on following page

Table 2. Continued

No.	Authors	Year of publication	Data collection volume	Datasets	Algorithm or model	Performance metrics	Results
17	Ghosh et al.	2022	11879 users	Twitter	Multimodal MT system (MTUD+Img+Cat), standard deviation (ST), F1-score, accuracy, and Pearson score.	After adding image features to the user description, accuracy of the system improved by 2.07% compared to other ST models, and the F1-score improved by 1.11%. Addition of category information also improved the performance of the ST model by 0.87%, with an accuracy improvement of 0.87% and an F1-score improvement of 1.64%. By jointly learning depression detection, emotion recognition tasks, and modeling the problem in an MT setting, overall performance of the system improved by 1.53% in accuracy and 1.36% in F1-score. The best Pearson's r score for the "joy" class was 0.81, followed by "sadness" with 0.66 and "disgust" with 0.63.	The performance of the proposed method is better than other model methods for detecting depression.
18	Guo et al.	2023	2200 users	Weibo and WU 3D dataset.	DUT-SL, soft voting, LR, LightGBM model, integrated learning model.	When k = 500, accuracy of soft voting peaked at 94.36%, and F1-score at 94.28%. LR also showed consistent and excellent performance, surpassing soft voting when k = 700. When validated on the WU 3D dataset, LightGBM performed best at each value of k. When k=100, LightGBM achieved peak accuracy of 96.14% and F1 of 93.78%.	The performance of ensemble learning model is better than other models, and the performance of LR is the most stable on the depression dataset.
19	Neuman et al.	2012	681,288 posts	Website blogs	Pedesis DUT-SL	The system's predictive accuracy for depression improved by 9%. When tested on a blog corpus, the system achieved an accuracy of 84.2% ($p < .001$). By comparing the system's predictions with those of mental health professionals, the system ultimately achieved an average precision of 78% and an average recall rate of 76%.	The automatic depression screening tool, pedesis, demonstrated good performance.
20	Zhang et al.	2023	1,600,000 tweets	Reddit and Twitter.	MTDD	High accuracy, with an F1-score of 95%.	The MTDD model could detect users who may have depression tendencies.

continued on following page

Table 2. Continued

No.	Authors	Year of publication	Data collection volume	Datasets	Algorithm or model	Performance metrics	Results
21	Anshul et al.	2023	2059,326 tweets	Twitter	MFEL, LR, XGBoost, and neural network classifiers.	The MFEL model produces an F1-score of 91.9% and an accuracy of 91.7%. In terms of precision, recall, F1-score, and accuracy, it shows an improvement of 2.4%, 4.3%, 3.4%, and 3.4%, respectively, over LR, with 2.1%, 1.4%, 1.7% and 2.0%, respectively, over XGBoost, with 1.4%, 9.3%, 5.6%, and 5.3%, respectively, over the neural network model.	The MFEL model is superior to LR, XGBoost and neural network models in detecting depression.
22	Bokolo et al.	2023	632,000 tweets	Twitter	LR, Bernoulli NB, RF, DistilBERT, SqueezeBERT, DeBERTA, RoBERTa models	The RoBERTa model achieves the highest accuracy ratio of 0.981 and the highest mean accuracy of 0.97 (across 10 cross-validation folds) in detecting depression from tweets.	LR, Bernoulli NB, and RF models demonstrated robust performance, exhibiting high accuracy and precise predictive capabilities.
23	Zogan et al.	2021	2159 positive users and 2049 negative users.	Twitter	MDHAN	MDHAN achieved an optimal performance of 89% on the F1-score.	Combining the HAN with a multiaspect user timeline semantic feature strategy is sufficient for detecting depression on Twitter.
24	Hussain et al.	2019	6562 users	Facebook	Lexicon-based classifier, SVM, deep canonical correlation analysis, NB, and DT	The Spearman's correlation of the LIWC feature of the users' status updates with CES-D scores ($p< 0.05$). Among the listed features, nine features revealed significant correlations with the CES-D score. Age, friend network size, events, and work were negatively correlated with depression, whereas gender, likes, depression markers, help-seeking markers, and SUD markers had positive correlations ($p=0.004$, $p=0.014$, $p=0.005$, $p=0.005$, and $p=0.006$, respectively). The chi-square test for categorical variables was utilized to test the correlation. It is used to find whether there is a significant correlation between the two variables.	Analyzing social media data helps in detecting and understanding depression, providing insights into a variety of factors associated with depressive symptoms and behaviors.
25	Ali et al.	2022	2100 samples	Facebook	RF, SVM, NB, KNN.	The RF classifier based on uni-gram feature set is superior to other classifiers in classification effect, further confirming that the RF classifier is trained and tested best on unlabeled post updates.	The RF classifier based on unary feature set is superior to other similar classifiers in the detection of depression.

continued on following page

Table 2. Continued

No.	Authors	Year of publication	Data collection volume	Datasets	Algorithm or model	Performance metrics	Results
26	Ghosh et al.	2023	15,031 social media statuses	Facebook, Twitter, and YouTube.	LSTM-CNN	Bangla and social text, with a sensitivity of 92.63% and specificity of 95.12%.	Extending its efficacy to multiple languages including English, the model exhibited outstanding performance in comparative evaluations, surpassing traditional ML models, ensemble methods, transformers, and existing architectures.
27	Wu & Tang	2022	5000 samples	PTT Forum	HSAN, SVM, RF, and text CNN.	The HSAN network model has shown superior performance in detecting the severity of online bullying incidents compared to other ML and deep learning models, with an accuracy value as high as 62.3%, and F1 values also surpassing those of other network models.	The authors' newly developed SAM can thoroughly analyze data and determine the severity of cyberbullying. The HSAN network model demonstrates superior performance in detecting the severity of cyberbullying incidents compared to other ML and deep learning models.
28	Shin et al.	2021	93 volunteers	Record voices from the MINI	Linear regression, NB, SVM,, MLP, and neural network model.	Compared to linear regression, NB, and SVM, the MLP network demonstrates outstanding performance with an AUC of 65.9%, sensitivity of 65.6%, and specificity of 66.2%.	The study illustrated changes in voice during depressive episodes, confirming that ML can accurately distinguish between nondepressed individuals and those with mild depression, suggesting the potential of using voice for detecting mild depression.

involves employing various methods and technologies, including hyperparameter optimization and fine-tuning in transfer learning, among others. The most frequently mentioned method in the literature is multitype voting, an ensemble learning method that makes final decisions by combining the prediction results of various types of models. As Table 3 indicates, different types of models are combined using the multitype voting method to make weighted averages of their prediction results, with the resulting value used as the final prediction result. Second, it also includes interpretability and robustness. Models with strong interpretability can better explain their prediction results, while models with strong robustness can better cope with noise and outliers in the dataset. For example, classification can be performed using models with strong interpretability, such as SVMs, or classification can be performed using models with strong robustness, such as DTs. In addition, cross-validation can be used to evaluate model performance. Cross-validation segregates the dataset into numerous subdivisions, designating a singular subset as the testing cohort while allocating the residual as the training ensemble for model refinement and assessment. Perpetuating the validation process across multiple iterations engenders more precise appraisals of the model's operational efficacy.

Within the context of depression detection efficacy as delineated by the referenced literature and demonstrated in Table 5, the performance of various algorithms is contingent upon data characteristics and task requirements. For text classification tasks, NB and LR typically exhibit commendable performance due to their high learning efficiency and robustness with textual data. DTs are also competent in handling text data; however, they are prone to overfitting and may not

Table 3. Overview of Common Algorithms and Model Classification

Algorithm	The frequency of the cited article
NB	6
DT	4
SVM	7
Linear regression model	4
Soft voting	2
RF	4
LightGBM Model	1
LSTM (including all extension of LSTM architecture)	7
CNN (contains all CNN extensions)	5
KNN (contains all KNN extensions)	9
BERT (contains all extensions)	2
MFM-Att	3
N-gram	1
LDA	2
Cost-sensitive boosting pruning trees	1
Semisupervised statistical model	1
XGBoost	1
MDHAN	1

Table 4. Overview of Evaluation Criteria

Evaluation criteria	Frequency	Total number of frequencies
Accuracy	8	37
F1-score	8	
Precision	8	
Recall	6	
AUC-ROC curve	2	
Confusion matrix	1	
Root-mean-square error	2	
Confidence	1	
Average recall	1	

display the stability of NB and LR. In the case of numerical data, SVM and RF usually possess strong generalization capabilities and robustness but require more computational resources. Gradient boosting tree algorithms such as XGBoost and LightGBM offer rapid processing of numerical data and generally higher prediction accuracy.

Regarding efficiency, NB and KNN algorithms typically have lower computational complexity, making them suitable for large-scale datasets. Nonetheless, the performance of these algorithms

Table 5. Algorithm Comparison

Algorithm	Advantages	Disadvantages	Complexity	Scalability	Robustness	Learning efficiency	Applicable data types
NB	Simple, fast, and easy to implement.	Sensitive to data distribution, assumes independence of features.	Low	High	Medium	High	Text, numerical
DT	Easy to understand and strong interpretability.	Prone to overfitting and poor generalization ability.	Medium	Medium	Medium	High	Text, numerical
SVM	Strong generalization ability and good robustness.	Slow training speed and sensitive to parameters.	High	Low	High	Medium	Numerical
LR	Simple, easy to understand, and strong interpretability.	Weak nonlinear fitting ability.	Low	High	Medium	High	Numerical
RF	Strong generalization ability and good robustness.	High model complexity and slow training speed.	High	High	High	Medium	Text, numerical
LightGBM	Fast training speed, high efficiency, and strong generalization ability.	High model complexity and weak interpretability.	High	High	High	High	Text, numerical
LSTM	Capable of processing time series data.	High model complexity and slow training speed.	High	Low	Medium	Low	Time series
CNN	Capable of processing image data.	High model complexity and slow training speed.	High	Low	High	Low	Images
KNN	Simple and easy to implement.	Large computation and sensitive to noise.	Low	Low	Medium	High	Numerical
BERT	Capable of processing natural language data.	High model complexity and slow training speed.	High	Low	Medium	Low	Text
XGBoost	Strong generalization ability, good robustness, and fast training speed.	High model complexity and weak interpretability.	High	High	High	High	Text, numerical

might degrade in the presence of skewed data distributions and noise when handling substantial datasets. Conversely, deep learning models, such as LSTM networks and CNN, usually entail higher model complexity and training speed, thus necessitating more computational resources for large-scale data processing.

Each algorithm possesses inherent limitations. For instance, NB assumes mutual independence among features, which may not be valid for complex data relationships. DTs are susceptible to overfitting and struggle with generalization to new data. SVMs are sensitive to parameter settings and exhibit slower training speeds. Deep learning models require extensive data and computational resources and lack interpretability.

In summary, ML technologies are paving new paths for the diagnosis and treatment of depression. By analyzing patients' behavior, physiological data, and social media interactions, these techniques can identify potential signs of depression, aiding in early diagnosis and intervention, thereby enhancing the effectiveness of treatments. In the utilization of social media data, researchers have developed algorithms capable of analyzing text content on platforms such as Weibo and Twitter (Angskun et al., 2022; Anshul et al., 2023; Chiong et al., 2021; Dhelim et al., 2023; Ghosh et al., 2023; Guo et al., 2023). These algorithms, by examining users' posts, interaction frequency, and emotional tendencies, can detect inclinations towards depression or emotional distress. Furthermore, ML can also be employed within clinical settings to develop personalized treatment plans. By analyzing how patients respond to different treatment options, ML models can predict which type of treatment is most likely to be effective for a specific patient.

5. DISCUSSION

5.1 Potential Research Areas

Based on a review of literature in the past decade, the authors identified the future research scope of ML in depression detection to include mainly the following aspects: Identifying biological markers of depression, exploring abnormal patterns of brain networks, analyzing social media data, and quantifying psychological symptoms.

5.1.1 Biomarkers

By predicting various Biomarkers (e.g., EEG data, facial expressions, speech features, and genetic mutations), it might be possible to foresee or identify whether a patient has depression or potential risk thereof. Currently, a series of fundamental and clinical studies have been initiated in the field of Biomarker Identification. Some researchers utilize ML techniques for the analysis of EEG to detect depression. Others employ ML to analyze speech and facial expression features, achieving prediction and recognition of patients' depressive symptoms (Burdisso et al., 2019; Ma et al., 2023). Most of these studies are at the exploratory stage, with some having achieved anticipated outcomes, but further clinical research validation is still needed.

5.1.2 Abnormal Brain Network Patterns

By using neuroimaging technologies such as MRI or fMRI, some researchers can explore the abnormal patterns of brain networks in patients with depression, in combination with ML. The conjunction of neuroimaging and ML provides researchers with powerful tools. Researchers typically use MRI or fMRI data, training various ML models (e.g., SVM, RF, and deep neural networks) to identify differences in brain structure and function between patients with depression and healthy individuals (Zhuo et al., 2019).

5.1.3 Social Media Data Analysis

By analyzing texts or images on social media, ML can reveal changes in behavior and emotions of patients, which helps in early detection of depression. Most of these studies, as can be derived from the literature in this paper, are at the application stage. However, due to the involvement of a large amount of personal privacy issues, how to effectively use social media data while protecting personal privacy remains a major issue. The authors will discuss this data and data anonymization problem specifically in the limitations section below.

5.1.4 Quantification of Psychological Symptoms

With data collected by sensors and mobile devices, ML can quantify individuals' behavioral and emotional symptoms, aiding in disease tracking and assessment. Some studies are currently using sensor data from smartphones and wearable devices (e.g., GPS signals and motion sensors), through ML techniques, to quantify the daily behavior and emotional changes of patients with depression. These studies can help doctors and patients for more accurate monitoring and assessment of the disease, usually at the exploratory and preliminary application stage (Müller et al., 2021).

5.2 Research Limitations and Ethical Considerations

A number of articles have mentioned a series of data privacy and ethical considerations involved in the domain of depression detection using ML methods. Despite the massive potential brought by the application of ML and AI in detecting depression, scholars also face a set of challenges and limitations, and the need to balance the relationship between technological development and individual privacy protection. This also addresses the fourth question the authors raised in this study, pertaining to the limitations and ethical considerations of ML in depression monitoring. The following aspects provide a detailed analysis of this issue.

Firstly, ML models rely heavily on large-scale training data. In many instances, obtaining sufficient and accurate depression data can be challenging, especially considering these data need to be annotated and labeled by clinical experts. On the other hand, detection of depression often requires the use of highly sensitive personal health data, including psychological assessments and medical records. Anonymization and deidentification of data during the data collection and processing phases are of paramount significance. At the same time, it is necessary to establish a comprehensive data security system, implementing measures such as encryption and access control to guard against data maliciously acquired by unauthorized sources.

Secondly, some literature experiments have problems with poor generalization. Existing ML models often perform well on training data, but poorly on new and unseen data. This can be attributed to differences in the definition and diagnosis method of depression targeted in training.

In addition, fairness and bias in algorithms call for ethical considerations. ML models can potentially be influenced by bias in training data, leading to underperformance in specific experiments and clinical scenarios. Therefore, ensuring diversity and representativeness of the data in model training is crucial to avoid discrimination against specific groups.

Lastly, the protection of patients' right to information and decision-making is part of ethical considerations. Patients need to be fully informed about how their data will be used, as well as the limitations and uncertainties of the model. Medical professionals should retain a dominant role in decision-making, ensuring the output from ML models serves as a reference rather than a replacement.

In conclusion, when applying ML methods in depression detection, considering data privacy and ethical issues is a necessary measure to safeguard patient rights and propel technological progress.

6. CONCLUSION

This paper provides a comprehensive review of ML methods for depression detection, introducing various types of depression detection methods, including those based on text, images, and multimodal data. These methods have achieved certain results in depression detection. The accuracy of depression detection is the core issue of ML methods. Although some methods can achieve good classification performance, some difficulties and challenges remain.

Through the analysis of existing research, the authors identified current limitations and challenges, providing valuable guidance for future work in this domain. Future research could focus on long-term tracking and longitudinal studies. Based on the literature and screening conducted, the authors observed that past studies primarily concentrated on content generated by English-speaking users from platforms such as Twitter, Reddit, Facebook, and Instagram, as well as regional social networks such as Weibo and WeChat in China. There is currently insufficient dataset coverage for similar research on other regions or regional network platforms. Therefore, the development and investigation of multilingual deep learning models for the most common mental health issues, such as depression, self-harm, and suicide, have not been thoroughly explored and will be a crucial area for future exploration.

The hallmark of this study lies in the exhaustive assessment of both traditional and emerging ML algorithms, with a niche focus on their application for depression detection as represented in the larger body of literature. As part of the authors' innovative exploration, they also delved into the meaningful analysis of how data from social media are utilized, and evaluated the burgeoning role of advanced technologies such as artificial neural networks and gait image models in this field. These explorations were further enhanced as the authors advanced beyond the technological domain to provide an insightful discourse on the future trajectories in this field, carefully weighing the potential limitations and navigating the ethical implications associated with the application of ML for depression detection. The authors' findings and analyses offer a substantial contribution to the understanding and future direction of this rapidly evolving domain.

COMPETING INTERESTS

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

FUNDING

No funding was received for this work.

PROCESS DATES

March, 2024

Received: January 25, 2024, Revision: February 22, 2024, Accepted: February 22, 2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Yingqi Guo, yingqiguo@hkbu.edu.hk and Tao Yang, mailto: 328748306@qq.com.

REFERENCES

Ajmani, L., Chancellor, S., Mehta, B., Fiesler, C., Zimmer, M., & de Choudhury, M. (2023). A systematic review of ethics disclosures in predictive mental health research. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency.* doi:10.1145/3593013.3594082

Ali, M., Baqir, A., Husnain Raza Sherazi, H., Hussain, A., Hassan Alshehri, A., & Ali Imran, M. (2022). Machine learning based psychotic behaviors prediction from Facebook status updates. *Computers, Materials & Continua*, *72*(2), 2411–2427. doi:10.32604/cmc.2022.024704

Almars, A. M. (2022). Attention-based Bi-LSTM model for Arabic depression classification. *Computers, Materials & Continua*, *71*(2), 3091–3106. doi:10.32604/cmc.2022.022609

Alsagri, H. S., & Ykhlef, M. (2020). Machine learning-based approach for depression detection in Twitter using content and activity features. *IEICE Transactions on Information and Systems, E103.D*(8), 1825–1832. 10.1587/transinf.2020EDP7023

Andersson, S., Bathula, D. R., Iliadis, S. I., Walter, M., & Skalkidou, A. (2021). Predicting women with depressive symptoms postpartum with machine learning methods. *Scientific Reports*, *11*(1), 7877. Advance online publication. doi:10.1038/s41598-021-86368-y PMID:33846362

Angskun, J., Tipprasert, S., & Angskun, T. (2022). Big data analytics on social networks for real-time depression detection. *Journal of Big Data*, *9*(1), 69. doi:10.1186/s40537-022-00622-2 PMID:35610999

Anshul, A., Pranav, G. S., Rehman, M. Z., & Kumar, N. (2023). A multimodal framework for depression detection during COVID-19 via harvesting social media. *IEEE Transactions on Computational Social Systems*, 1–17. doi:10.1109/TCSS.2023.3309229

Babu, N. V., & Kanaga, E. G. (2021). Sentiment analysis in social media data for depression detection using artificial intelligence: A review. *SN Computer Science*, *3*(1), 74. Advance online publication. doi:10.1007/s42979-021-00958-1 PMID:34816124

Bokolo, B. G., & Liu, Q. (2023). Deep learning-based depression detection from social media: Comparative evaluation of ML and transformer techniques. *Electronics (Basel)*, *12*(21), 4396. doi:10.3390/electronics12214396

Burdisso, S. G., Errecalde, M., & Montes-y-Gómez, M. (2019). A text classification framework for simple and effective early depression detection over social media streams. *Expert Systems with Applications*, *133*, 182–197. doi:10.1016/j.eswa.2019.05.023

Chen, Y., Zhou, B., Zhang, W., Gong, W., & Sun, G. (2018). Sentiment analysis based on deep learning and its application in screening for perinatal depression. In *Proceedings of the IEEE Third International Conference on Data Science in Cyberspace (DSC)* (pp. 451—456). doi:10.1109/DSC.2018.00073

Chiong, R., Budhi, G. S., Dhakal, S., & Chiong, F. (2021). A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, *135*, 104499. doi:10.1016/j.compbiomed.2021.104499 PMID:34174760

Dhelim, S., Chen, L. L., Das, S. K., Ning, H., Nugent, C. D., Leavey, G., Pesch, D., Bantry-White, E., & Burns, D. M. (2023). Detecting mental distresses using social behavior analysis in the context of COVID-19: A survey. *ACM Computing Surveys*, *55*(14s), 1–30. doi:10.1145/3589784

Fang, M., Peng, S., Liang, Y., Hung, C., & Liu, S. (2023). A multimodal fusion model with multi-level attention mechanism for depression detection. *Biomedical Signal Processing and Control*, *82*, 104561. doi:10.1016/j.bspc.2022.104561

Gan, L. et al.. (2024). Experimental comparison of three topic modeling methods with LDA, Top2Vec and BERTopic. In H. Lu & J. Cai (Eds.), *Artificial intelligence and robotics. ISAIR 2023. Communications in computer and information science* (Vol. 1998, pp. 376–391). Springer., doi:10.1007/978-981-99-9109-9_37

Gausman, J., & Langer, A. (2020). Sex and gender disparities in the COVID-19 pandemic. *Journal of Women's Health*, *29*(4), 465–466. doi:10.1089/jwh.2020.8472 PMID:32320331

Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2022). What does your bio say? Inferring Twitter users' depression status from multimodal profile information using deep learning. *IEEE Transactions on Computational Social Systems*, *9*(5), 1484–1494. doi:10.1109/TCSS.2021.3116242

Ghosh, T., Banna, M., Nahian, M. J. A., Uddin, M. N., Kaiser, M. S., & Mahmud, M. (2023). An attention-based hybrid architecture with explainability for depressive social media text detection in Bangla. *Expert Systems with Applications*, *213*, 119007. doi:10.1016/j.eswa.2022.119007

Guo, Z., Ding, N., Zhai, M., Zhang, Z., & Li, Z. (2023). Leveraging domain knowledge to improve depression detection on Chinese social media. *IEEE Transactions on Computational Social Systems*, *10*(4), 1528–1536. doi:10.1109/TCSS.2023.3267183

Hemanthkumar, M., & Latha, A. (2019). Depression detection with sentiment analysis of tweets. *International Research Journal of Engineering and Technology*, *6*, 3–7.

Hussain, J., Satti, F. A., Afzal, M., Khan, W. A., Bilal, H. S., Ansaar, M. Z., Ahmad, H. F., Hur, T. H., Bang, J. H., Kim, J., Park, G. H., Seung, H., & Lee, S. (2019). Exploring the dominant features of social media for depression detection. *Journal of Information Science*, *46*(6), 739–759. doi:10.1177/0165551519860469

Jones, N. P., Siegle, G. J., & Mandell, D. L. (2015). Motivational and emotional influences on cognitive control in depression: A pupillometry study. *Cognitive, Affective & Behavioral Neuroscience*, *15*(2), 263–275. doi:10.3758/s13415-014-0323-6 PMID:25280561

Kabir, M., Ahmed, T., Hasan, M. B., Laskar, M. T. R., Joarder, T. K., Mahmud, H., & Hasan, K. M. A. (2023). DEPTWEET: A typology for social media texts to detect depression severities. *Computers in Human Behavior*, *139*, 107503. doi:10.1016/j.chb.2022.107503

Liaw, A. S., & Chua, H. N. (2022). Depression detection on social media with user network and engagement features using machine learning methods. In *Proceedings of the IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)* (pp. 1—6). doi:10.1109/IICAIET55139.2022.9936814

Ling, Z., & Hao, Z. J. (2022). An intrusion detection system based on normalized mutual information antibodies feature selection and adaptive quantum artificial immune system. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–25. doi:10.4018/IJSWIS.308469

Ling, Z., & Hao, Z. J. (2022). Intrusion detection using normalized mutual information feature selection and parallel quantum genetic algorithm. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–24. doi:10.4018/IJSWIS.307324

Liu, D., Ahmed, F., Shahid, M., Guo, J., & Feng, X. L. (2022). Detecting and measuring depression on social media using a machine learning approach: Systematic review. *JMIR Mental Health*, *9*(3), e27244. doi:10.2196/27244 PMID:35230252

Ma, W., Qiu, S., Miao, J., Li, M., Tian, Z., Zhang, B., & Dong, W. (2023). Detecting depression tendency based on deep learning and multi-sources data. *Biomedical Signal Processing and Control*, *86*, 105226. doi:10.1016/j.bspc.2023.105226

Malhotra, A., & Jindal, R. (2022). Deep learning techniques for suicide and depression detection from online social media: A scoping review. *Applied Soft Computing*, *130*, 109713. doi:10.1016/j.asoc.2022.109713

Meng, Y., Speier, W., Ong, M. K., & Arnold, C. W. (2021). Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. *IEEE Journal of Biomedical and Health Informatics*, *25*(8), 3121–3129. doi:10.1109/JBHI.2021.3063721 PMID:33661740

Mishra, A., Joshi, B. K., Arya, V., Gupta, A. K., & Chui, K. T. (2022). Detection of distributed denial of service (DDoS) attacks using computational intelligence and majority vote-based ensemble approach. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–10. doi:10.4018/IJSSCI.309707

Motrico, E., Bina, R., Domínguez-Salas, S., Mateus, V., Contreras-García, Y., Carrasco-Portiño, M., Ajaz, E., Apter, G., Christoforou, A., Dikmen-Yildiz, P., Felice, E., Hancheva, C., Vousoura, E., Wilson, C. A., Buhagiar, R., Cadarso-Suárez, C., Costa, R., Devouche, E., Ganho-Ávila, A., & Mesquita, A. et al. (2021). Impact of the COVID-19 pandemic on perinatal mental health (Riseup-PPD-COVID-19): Protocol for an international prospective cohort study. *BMC Public Health*, *21*(1), 368. Advance online publication. doi:10.1186/s12889-021-10330-w PMID:33596889

Müller, S. R., Chen, X., Peters, H., Chaintreau, A., & Matz, S. C. (2021). Depression predictions from GPS-based mobility do not generalize well to large demographically heterogeneous samples. *Scientific Reports*, *11*(1), 11. doi:10.1038/s41598-021-93087-x PMID:34234186

Neuman, Y., Cohen, Y., Assaf, D., & Kedma, G. (2012). Proactive screening for depression through metaphorical and automatic text analysis. *Artificial Intelligence in Medicine*, *56*(1), 19–25. doi:10.1016/j.artmed.2012.06.001 PMID:22771201

Owusu, P. N., Reininghaus, U., Koppe, G., Dankwa-Mullan, I., & Bärnighausen, T. (2021). Artificial intelligence applications in social media for depression screening: A systematic review protocol for content validity processes. *PLoS One*, *16*(11), e0259499. doi:10.1371/journal.pone.0259499 PMID:34748571

Pathoee, K., Rawat, D., Mishra, A., Arya, V., Rafsanjani, M. K., & Gupta, A. K. (2022). A cloud-based predictive model for the detection of breast cancer. *International Journal of Cloud Applications and Computing*, *12*(1), 1–12. doi:10.4018/IJCAC.310041

Qian, Y., & Yan, X. (2013). Prevalence of postpartum depression in China: A systematic analysis. *Chinese Journal of Practical Nursing*, 1—3.

Robertson, C., Carney, J., & Trudell, S. (2023). Language about the future on social media as a novel marker of anxiety and depression: A big-data and experimental analysis. *Current Research in Behavioral Sciences*, *4*, 100104. doi:10.1016/j.crbeha.2023.100104 PMID:37397228

Santomauro Mantilla Herrera Shadid Zheng Ashbaugh, D. F., Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E. F., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., & Ferrari, A. J. et al. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet*, *398*(10312), 1700–1712. doi:10.1016/S0140-6736(21)02143-7 PMID:34634250

Schindler, M., & Domahidi, E. (2023). The computational turn in online mental health research: A systematic review. *New Media & Society*, *25*(10), 2781–2799. doi:10.1177/14614448221122212

Shin, D., Cho, W., Park, C. H., Rhee, S. J., Kim, M. J., Lee, H., Kim, N. S., & Ahn, Y. M. (2021). Detection of minor and major depression through voice as a biomarker using machine learning. *Journal of Clinical Medicine*, *10*(14), 3046. doi:10.3390/jcm10143046 PMID:34300212

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in Reddit social media forum. *IEEE Access : Practical Innovations, Open Solutions*, *7*, 44883–44893. doi:10.1109/ACCESS.2019.2909180

Tembhurne, J. V., Almin, M. M., & Diwan, T. (2022). Mc-DNN: Fake news detection using multi-channel deep neural networks. [IJSWIS]. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–20. doi:10.4018/IJSWIS.295553

Tong, L., Liu, Z., Jiang, Z., Zhou, F., Chen, L., Lyu, J., & Zhou, H. (2023). Cost-sensitive boosting pruning trees for depression detection on Twitter. *IEEE Transactions on Affective Computing*, *14*(3), 1898–1911. doi:10.1109/TAFFC.2022.3145634

Wani, M. A., ElAffendi, M. A., Shakil, K. A., Imran, A. S., & Abd El-Latif, A. A. (2023). Depression screening in humans with AI and deep learning techniques. *IEEE Transactions on Computational Social Systems*, *10*(4), 2074–2089. doi:10.1109/TCSS.2022.3200213

Wu, J. L., & Tang, C. Y. (2022). Classifying the severity of cyberbullying incidents by using a hierarchical squashing-attention network. *Applied Sciences (Basel, Switzerland)*, *12*(7), 3502. doi:10.3390/app12073502

Yazdavar, A. H., Al-Olimat, H. S., Ebrahimi, M., Bajaj, G., Banerjee, T., Thirunarayan, K., Pathak, J., & Sheth, A. (2017). Semi-supervised approach to monitoring clinical depressive symptoms in social media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. doi:10.1145/3110025.3123028

Zeberga, K., Attique, M., Shah, B., Ali, F., Jembre, Y. Z., & Chung, T. S. (2022). A novel text mining approach for mental health prediction using Bi-LSTM and BERT Model. *Computational Intelligence and Neuroscience*, *2022*, 1–18. doi:10.1155/2022/7893775 PMID:35281185

Zhang, H., Wang, H., Han, S., Li, W., & Zhuang, L. (2023). Detecting depression tendency with multimodal features. *Computer Methods and Programs in Biomedicine*, *240*, 107702. doi:10.1016/j.cmpb.2023.107702 PMID:37531689

Zhang, Q., Guo, Z., Zhu, Y., Vijayakumar, P., Castiglione, A., & Gupta, B. B. (2023). A deep learning-based fast fake news detection model for cyber-physical social services. *Pattern Recognition Letters*, *168*, 31–38. doi:10.1016/j.patrec.2023.02.026

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *NPJ Digital Medicine*, *5*(1), 46. Advance online publication. doi:10.1038/s41746-022-00589-7 PMID:35396451

Zhang, T., Yang, K., Ji, S., & Ananiadou, S. (2023). Emotion fusion for mental illness detection from social media: A survey. *Information Fusion*, *92*, 231–246. doi:10.1016/j.inffus.2022.11.031

Zhuo, C., Li, G., Lin, X., Jiang, D., Xu, Y., Tian, H., Wang, W., & Song, X. (2019). The rise and fall of MRI studies in major depressive disorder. *Translational Psychiatry*, *9*(1), 9. doi:10.1038/s41398-019-0680-6 PMID:31819044

Zogan, H., Razzak, I., Wang, X., Jameel, S., & Xu, G. (2022). Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. *World Wide Web (Bussum)*, *25*(1), 281–304. doi:10.1007/s11280-021-00992-2 PMID:35106059

Zou, M. L., Li, M. X., & Cho, V. (2020). Depression and disclosure behavior via social media: A study of university students in China. *Heliyon*, *6*(2), e03368. Advance online publication. doi:10.1016/j.heliyon.2020.e03368 PMID:32099917

Lin Gan, received her M.S. in Hong Kong Polytechnic University, in 2019. Her research interests include computational sociology, text analysis, multimodal medical monitoring.

Yingqi Guo, assistant professor at Hong Kong Baptist University, she received her Ph.D. in University of Hong Kong in 2017. Her research interests including: smart society, environment and well-being, GIS/GPS/RS and Wearable Technologies in Wellbeing Research, mixed methods, gender.

Tao Yang, Professor of University of Sanya. His research interests include Big data and intelligent agriculture.