US-China Education Review A, May 2026, Vol. 16, No. 5, 364-374
doi: 10.17265/2161-623X/2026.05.012

Social Media-Based Depression Prediction and Assessment: Methods, Challenges, and Future Directions

CHEN Meifen
Shenzhen Polytechnic University, Shenzhen, China

Depression is a common mental disorder and a major global public health concern. The widespread use of social media has created new opportunities for the early recognition and supplementary assessment of depression-related risk signals, because users frequently disclose their emotions, behaviors, and interpersonal experiences online. This review synthesizes the literature on social-media-based depression detection, with an emphasis on the major categories of features used in this field, including textual, behavioural, temporal, emotional, visual, and multimodal signals. It further summarizes the principal modeling strategies, ranging from shallow machine-learning classifiers to deep-learning and attention-based architectures. Recent bibliometric, systematic review, and meta-analytic evidence indicates that this field continues to grow and that prediction performance is often promising, although substantial heterogeneity remains across platforms, labeling strategies, feature sets, task definitions, and evaluation metrics. Overall, social-media-based depression assessment has considerable value as a supplementary screening approach, particularly for early risk identification and longitudinal monitoring. However, it should not be regarded as a substitute for clinical diagnosis. Future research should place greater emphasis on label validity, feature validity, model interpretability, cross-platform generalizability, ethical governance, and clinical translation.

Keywords: social network, depression, prediction, assessment method

Introduction

Depression is a prevalent mental disorder characterized by a significant and persistent low mood as its primary clinical feature. Patients frequently exhibit symptoms including low mood, slowed cognition, cognitive dysfunction, sleep disturbances, abnormal behaviors, and severe suicidal ideation. Prolonged depression may also result in behaviors such as alcoholism and violent abuse, which severely impact patients' quality of life and overall physical and mental health (Hu & Liu, 2019; Xin et al., 2015). According to World Health Organization statistics, approximately 332 million individuals are affected globally. The prevalence of depression is higher among women than men, and it can lead to suicide. Given that many individuals express their emotions, stress,

Acknowledgements: This study was financed by: (1) Guangdong Province Philosophy and Social Science Planning Project: A Study on the Reading Behavior Characteristics and Intervention Strategies of Electronic Picture Books for Children with Autism from a Cognitive Neuroscience Perspective (No.: GD25CJY24); (2) Shenzhen Philosophy and Social Science Project: A Study on the Relationship between Interpersonal Relationships and Personality Traits of Vocational College Students in the Context of Artificial Intelligence (No.: SZ2024D010); (3) The Third Batch of School-Enterprise Joint R&D Centers of Shenzhen Polytechnic University and Jimuyida for 3D Digital Innovation Applications (Grant No.602431009PQ); (4) Research Projects of Department of Education of Guangdong Province-2024WCXTD037.

CHEN Meifen, Ph.D., Associate Professor, School of Digital Media, Shenzhen Polytechnic University, Shenzhen, Guangdong, China.

and psychological distress online, social media has emerged as a crucial data source for investigating auxiliary indicators of depression (World Health Organization, 2023). Consequently, early detection, diagnosis, and intervention for patients with depression are of paramount importance.

The current identification and diagnosis of patients with depression primarily depend on structured rating scales and clinical interviews. While this approach can accurately assess patients' depressive states, it tends to engage primarily those with severe depression who are willing to self-evaluate using these scales and seek in-person consultations with professionals. In contrast, many individuals with mild depression or depressive tendencies remain unaware of their condition or lack the motivation to actively pursue medical advice, which hinders their access to timely intervention. Furthermore, the limited frequency of in-person diagnoses restricts the ability to monitor the dynamic changes in patients' symptoms in real time. In recent years, the proliferation of various social networking platforms has enabled individuals to share information, express emotions, and engage in interactions, thereby facilitating the early recognition, risk stratification, and supplementary assessment of depression. Several scholars have analyzed users' depressive states utilizing social network data (Men, Wei, & Wu, 2020). Recent systematic review and meta-analytic evidence continues to indicate that machine-learning-based methods for detecting depression on social media remain promising; however, significant heterogeneity persists across different platforms, labeling strategies, feature sets, and outcome definitions. Therefore, depression assessment based on social network data should currently be regarded as a supplementary tool rather than a substitute for conventional clinical assessment. In light of this, the present paper reviews several prominent methods for analyzing depressive states based on social network data, outlines the advancements in the application of these research methods, summarizes their respective advantages and disadvantages, and discusses potential future research directions.

Literature Retrieval Strategy

This study employed preliminary machine searches followed by secondary manual screening to retrieve and evaluate literature from the CNKI and Web of Science databases. Initially, the keywords "depression," "social network," and "assessment" were selected based on the article's theme. Subsequently, the retrieval scope was expanded to encompass the broader meanings of these keywords and their common application contexts. For Chinese literature published in the most recent 10 years, the search query was 抑郁症 & ((社交网络 or 社交媒体 or 社交网站 or 微博) & (检测 or 识别 or 预测)), which yielded 23 relevant articles. For English literature published in the most recent 10 years, the retrieval strategy was depression & (("social media" or "social network" or "social networking site" or "Facebook" or "Twitter" or "Instagram" or "Forum" or "Weibo") & (detect or identify or predict)), which yielded 485 relevant articles. After querying the relevant literature from the databases, a manual screening was performed to assess the alignment of the literature with the theme of this paper. Three screening criteria were established: first, the research theme must focus on depression; second, the research data must be derived from social networks; and third, the research methods must primarily involve quantitative approaches, such as predictive algorithm analysis, predictive model construction, and the application of prediction methods, while excluding literature that employs only simple qualitative analysis. Following this screening process, more than 50 relevant studies were ultimately identified. Among these, the majority utilized Twitter as the data source, followed by studies using Sina Weibo and Facebook, while those employing data from image-sharing platforms like Instagram were comparatively fewer.

Depression Assessment Methods Based on Social Networks

With the advent of the Internet, social networks have gained immense popularity, prompting an increasing number of individuals to share their lives and emotions on these platforms. Research indicates that patients with depression are more inclined to communicate on social networks while maintaining a certain degree of distance. Platforms such as Facebook, Twitter, and Sina Weibo serve as primary venues for these individuals to express their feelings, share their emotional states, and seek assistance. Several studies have demonstrated significant differences in language use and social behavior between patients with depression and typical users across various social platforms (Chancellor et al., 2016; Jiang, 2019). For instance, users experiencing depression tend to utilize first-person pronouns and past-tense verbs more frequently (Nadeem et al., 2016). These findings establish a foundation for assessing depression through social networks. Existing research has identified that the existing research can be broadly organized into two stages: feature classification and extraction, followed by model construction.

Feature Classification and Extraction

Social network data encompasses a wealth of information. To accurately identify users with depression, it is essential to extract effective features that will facilitate subsequent model development. The features typically employed include text features, behavioral features, emotional features, and image features. The specific classification and extraction methods are illustrated in Figure 1.

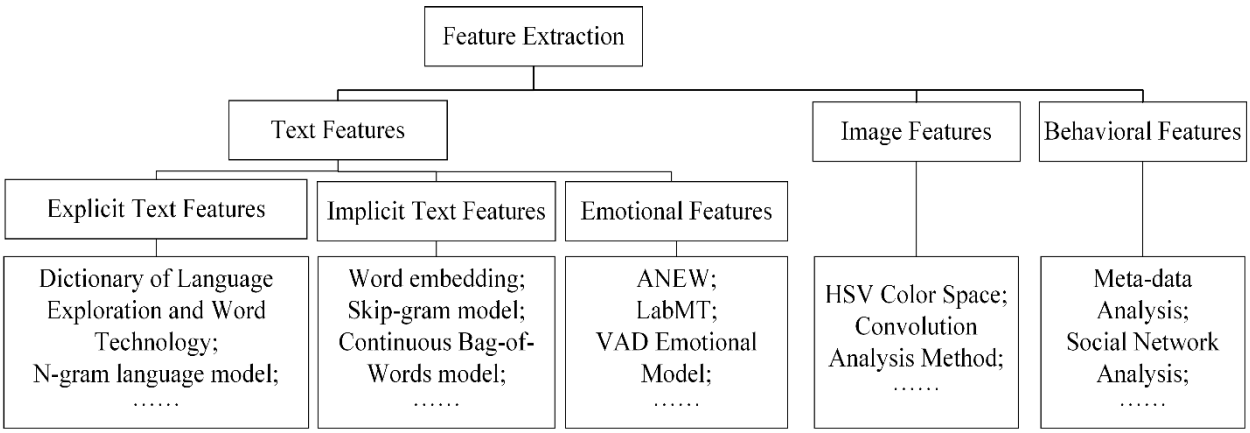

Figure 1. Types and methods of feature extraction for depression in social networks.

Text Features

Explicit text features. In machine learning, extracting and selecting textual features is crucial for improving model performance. Studies that predict patients with depression have employed diverse features to boost model accuracy. Early work emphasized users' textual and syntactic characteristics in social media posts. A widely used tool for text analysis is the Linguistic Inquiry and Word Count (LIWC) dictionary introduced by Pennebaker et al. (2007), which quantifies word categories in text content, particularly those relevant to psychological processes. Many studies have used this approach to extract key textual features for predicting depression (Tsugawa et al., 2015; Reece et al., 2017; Wongkoblap, Vadillo, & Curcin, 2018; Cheng et al., 2017). User posts on social media platforms such as Twitter are often unstructured, because they contain novel terms, exhibit lexical and grammatical errors, and are constrained by maximum character limits. N-gram language models address this lack

of structure effectively. These probabilistic models assume that a word's occurrence depends only on a limited number of preceding words.

The N-gram model is also widely used for text feature extraction in studies of depression (Benton, Mitchell, & Hovy, 2017). Another research direction employs topic modeling, commonly using methods such as the LDA (Latent Dirichlet Allocation) topic model (Blei et al., 2003). The N-gram model accounts for the order of word occurrence, whereas the LDA model treats words as conditionally independent. At the Second Symposium on Computational Linguistics and Clinical Psychology, participants competed in a shared task to identify patients with depression and post-traumatic stress disorder among 1,800 Twitter users. The highest-ranked model combined Supervised LDA, Supervised Anchor (SA), and term frequency–inverse document frequency (TF-IDF) algorithms to extract textual features (Resnik et al., 2015).

Implicit text features. Explicit text feature extraction typically relies on the researcher's subjective judgment to count occurrences of specific words and to analyze text structure and themes. This approach suits early studies with small datasets. With the rise of big data and cloud storage, however, vast volumes of social network data have become available, offering richer information for detecting depression. Applying explicit feature extraction to such large-scale data would require extensive engineering effort and introduce numerous complications. In the era of deep learning, researchers instead use representation learning algorithms to generate implicit text features automatically from datasets. Word embeddings are a central representation learning approach in natural language processing; they perform implicit feature extraction by mapping words to numeric vectors. Concretely, embeddings project the original high-dimensional space—whose dimension equals the vocabulary size—into a much lower-dimensional continuous real-valued space, assigning each word or phrase a vector representation.

In recent years, distributed word-embedding methods have been widely applied to extract textual features from social networks. One approach uses neural-network language models, which afford intrinsic smoothing and avoid the complex frequency-smoothing techniques required by traditional N-gram models; they also handle out-of-vocabulary words and compute their probabilities. Another common approach is Word2vec, which comprises the Skip-Gram and continuous bag-of-words (CBOW) models. Orabi et al. (2018) applied these two models to automatically extract features from tweets and fed them into a neural network to predict patients with depression. Studies show that such automatic feature-extraction pipelines substantially improve data representation quality and downstream prediction performance.

More recent research in the Chinese context indicates that text-only feature extraction remains valuable and does not necessarily require increasingly complex multimodal pipelines. Lyu et al. (2023) collected depression scores and Weibo posts from 789 users and extracted 117 lexical features by combining Simplified Chinese LIWC with culture-related and suicide-related lexicons. Their best-performing linear regression model achieved a Pearson correlation of 0.33 with self-reported depression scores and a split-half reliability of 0.75. This finding suggests that, under Chinese-language conditions, expanding psychologically meaningful lexicons can strengthen text-based depression recognition while preserving relatively good interpretability and cross-platform portability.

Emotional features. Emotional signals in users' text and emojis constitute an important class of features. Textual emotions can be quantified using tools such as LabMT (Language Assessment by Mechanical Turk), ANEW (Affective Norms for English Words), and the LIWC emotion dictionary. Shen et al. (2017) derived sentiment features by counting the positive and negative words in users' posts using the LIWC sentiment dictionary. To address language differences, Cheng et al. (2017) developed a Chinese vocabulary based on the

DAS-21 to extract emotional features from Sina Weibo posts and found that users with depression expressed more sadness, anxiety, and loss than ordinary users. On the social networking platform Weibo, recognizing that existing sentiment lexicons are inadequate for predicting depression, Fang (2017) extracted depression-related terms from the Weibo corpus and used them to construct a depression-specific lexicon for sentiment analysis of users' posts.

Shen et al. (2017) also applied the VAD emotion model (valence–arousal–dominance) to derive users' emotional state, intensity, and perceived control as features. Beyond emotional words, emojis are prevalent in social media and help compensate for the lack of prosody, facial expressions, and body language in written communication. For example, a smiling face typically signals positive affect, whereas a crying face signals negative affect. Emoji polarity can be determined with an emoji dictionary. Ricard et al. (2018) converted emojis appearing in titles and comments into emotion scores using the Emoji Emotion Scale and used those scores as features in a depression prediction model; they found that the sadness score was the most effective feature for identifying users with depression.

Image Features

Images are more intuitive and vivid than text, and they can convey more complex information. Avatars in users' profiles and images in their posts both provide valuable cues for predicting depression. Image features include brightness, saturation, cold-color ratio, and clarity. Some researchers extracted brightness, saturation, and clarity from Instagram posts and found that users prone to depression are less likely to publish brightly colored images and more likely to post bluish or grayish pictures (Reece & Danforth, 2016). Depressed users tend to favor black-and-white filters that render images blurry and faded. Shen et al. (2017) extracted image brightness and warmth and employed a feature comprising five dominant tones in HSV color space. Beyond intrinsic image attributes, depicted people and landscapes constitute meaningful features. Studies examining user-posted images that include faces and facial expressions found that users with depressive tendencies typically posted fewer human faces than non-depressed users. Their photos were often close-ups, and faces of relatives or friends appeared infrequently.

Behavioral Features

In addition to text and images, users' behavioral patterns contain substantial information. Prior work classifies these features into three categories: posting characteristics, interaction characteristics, and communication characteristics. Posting characteristics comprise metrics such as post count, posting frequency, and temporal distribution. Men et al. (2020) selected metadata including total post count and average daily posts as behavioral features, and divided the day into 12 two-hour intervals. They assessed periodicity by counting the number of posts in each interval. Their results indicate that the overall post volume for the control (nondepressed) group is substantially higher than for the depressed group. Conversely, because depressed individuals often experience sleep disturbances such as insomnia or early-morning awakening, they may post more frequently during nighttime hours. The study used a domain-specific activity metric to extract users' nighttime posting patterns as an indicator of depression. Interactive features encompass responses, comments, likes, and shares. Saravia et al. (2016) extracted retweets, quotations, replies, and mentions on Twitter and used them as inputs to a prediction model. They reported that depressed users exhibited fewer interactions and lower engagement with others on social networks. Interaction characteristics characterize metrics such as follower count, followee count, social-circle size, and the strength of social ties. Vedula and Parthasarathy (2017) used social network data to

construct personal network graphs, analyzed their clustering coefficients, densities, and depths, and treated these metrics as users' interaction characteristics. Studies have found that, compared with typical users, patients with depression have fewer followers on average and weaker social ties.

Model Construction

After feature selection, the process advances to model selection. Choosing an appropriate model to assess users with depression is a critical step. As machine learning and artificial intelligence continue to develop, models that detect depressed users from social networks are evolving. Currently, the primary approaches fall into two categories: shallow learning and deep learning, as shown in Figure 2.

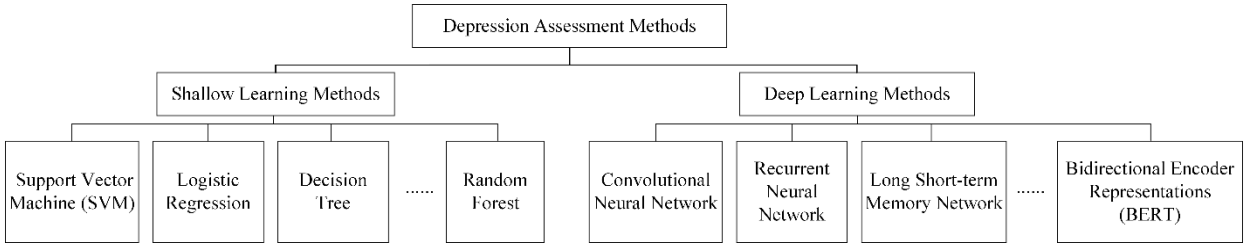

Figure 2. Depression assessment methods.

Early studies relied on small datasets and shallow learning methods to identify depressed users in social networks. Tsugawa et al. (2015) extracted textual and behavioral features from the activity records of 209 Twitter users and applied support vector machines to predict depression, achieving a relatively high accuracy. Eichstaedt et al. (2018) collected Facebook usage records from 683 users, extracted features such as content, length, and posting frequency, and used a penalized logistic regression model to identify patients with depression. Decision trees and random forests are also widely used to predict depressive users. A decision tree is a tree-like model that derives a final decision through a sequence of sub-decisions. Vedula and Parthasarathy (2017) developed a decision tree model based on Twitter usage, employing features such as textual content, emotional indicators, and interpersonal network metrics to identify individuals with depression.Random forest integrates multiple decision trees through ensemble learning, offering low computational overhead and strong generalization. Reece et al. (2017) employed a random forest to analyze color features of images posted by users in an Instagram dataset and thereby identify users with depression. The model's simple structure facilitates modeling, enables visual representation of feature importance, and provides excellent interpretability.

With the advent of the big data era, massive datasets offer richer information for research, and predictive models have shifted toward more powerful deep learning approaches. Compared with shallow methods, deep architectures better capture complex data through hierarchical feature representations and thus improve model accuracy. Orabi et al. (2018) used word embedding to extract features automatically from the dataset and applied convolutional neural networks (CNNs) to predict patients with depression; their approach outperformed earlier shallow models. Rohan Kshirsagar et al. (2017) employed a hierarchical recurrent neural network (RNN) with an attention mechanism as a predictive model. The attention mechanism isolates salient information from extensive inputs and suppresses irrelevant data, thereby substantially improving the model's efficiency and accuracy. Zogan et al. (2022) applied four models—Support Vector Machine (SVM), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and the Transformer-based Bidirectional Encoder Representation (BERT)—to the same dataset to predict user depression. They reported that the three deep-learning approaches outperformed SVM by a substantial margin.

In recent years, researchers have combined multiple features and applied multimodal learning techniques to identify depressed users on social networks. A modality denotes a source or form of information. Multimodal learning integrates diverse modalities (for example, text, images, audio, and video) to construct models. By exploiting complementarity among modalities and reducing redundancy, this approach yields improved feature representations. On social media, users express thoughts and emotions through text, images, and short videos (Duong, Lebret, & Aberer, 2017). Thus, multimodal learning is well suited to social network analysis. Shen et al. (2017) constructed a multimodal dictionary that combined text topics, social network metrics, image features, and emotion indicators to predict depression in 1,402 Twitter users. Cheng and Chen (2022) and Liu et al. (2023) used deep-learning multimodal fusion to integrate text and image features and predict college students' depression status. Both approaches outperformed unimodal baselines.

In a related multimodal direction, Cheng and Chen combined text, image, posting time, and time intervals through a multimodal time-aware attention network, reporting F1-scores of 95.6% on Instagram and 90.8% on Twitter. More recently, Liu et al. (2024) applied a simplified multi-head attention mechanism to the WU3D Weibo dataset of approximately 1.15 million posts from around 21,000 users, achieving an F1-score of 0.9473 while also offering individual-level interpretability through attention visualization. Alongside the improvement in predictive performance, recent studies have placed increasing emphasis on model explainability. Zogan et al. (2022) proposed an explainable hybrid deep learning framework that jointly incorporated textual, behavioural, temporal, and semantic features, arguing that explainability is important for user trust and for understanding why a model reaches a given prediction. These studies do not overturn the basic framework of "feature extraction + model construction," but they do indicate that recent progress lies mainly in better multimodal fusion, more refined temporal modeling, and stronger interpretability.

Summary of Relevant Research

Table 1 summarizes recent research on depression assessment using social network data, listing each study's publication year, data sources, selected features, and prediction methods, and reporting the models' performance metrics.

Most studies combine multiple feature types for feature selection. Early prediction work used simple classifiers such as linear regression, support vector machines, and decision trees. These models are straightforward, easy to interpret, and simple to fit, but they have limited capacity and handle few parameters, which constrains their ability to learn complex tasks. With cloud computing and big data, researchers have turned to deeper neural architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These models stack multiple hidden layers, with each layer transforming the previous layer's output so that initial "low-level" features become "high-level" representations more closely related to the prediction target, enabling more complex classification. Recently, multimodal learning models that fuse text, video, and image modalities have gained attention and have been applied to depression assessment.

Early models predominantly used the area under the curve (AUC) as the evaluation metric, whereas neural network-based models consistently employed the F1 score. Across studies, these models achieved relatively high reported performance varies with platform, labels, task definition, and unit of analysis on the evaluated datasets. Models that incorporate a wide range of feature types and greater complexity likewise report comparatively high prediction accuracy.

Table 1

Studies on Depression Assessment Based on Social Networks

Ref.	Year	Platform	Features / modalities	Model type	Performance metrics
Vedula & Parthasarathy	2017	Twitter	Text,behavioural,emotion	Decision Tree	ACC=0.9
Eichstaedt et al.	2018	Facebook	Text	Logistic Regression	AUC=0.72
Ricard	2018	Instagram	Text + Behavioral + Emotional + Image Features	Logistic Regression	AUC=0.72
Orabi et al.	2018	Twitter	Word Embedding + Demographic Features	Convolutional Neural Network (CNN)	F1=0.870
Yazdavar et al.	2020	Twitter	Text + Demographic + Image Features	Multimodal Learning	F1=0.900
Zogan et al.	2022	Twitter	Textual, behavioural,temporal,s temporal, semantic	Explainable hybrid deep learning (MDHAN)	F1=0.893
Liu et al.	2024	Weibo	Text ,user features with attention-based fusion	Hybrid model with simplified multi-head attention	F1 = 0.947

Recent bibliometric, systematic review, and meta-analytic evidence further suggests that this research area has continued to grow and that social-media-based detection remains promising, although substantial heterogeneity persists across platforms, labeling strategies, feature sets, and outcome definitions (Kim, Lee, & Park, 2021; Liu et al., 2022; Phiri et al., 2025). Kim, Lee, and Park (2021) identified 565 relevant publications from 2015 to 2020 and reported continuous growth in publication counts. Subsequent systematic review and meta-analytic evidence likewise suggests that social-media-based depression prediction remains effective, but results vary substantially according to platform type, labeling strategy, outcome measure, and model type. Therefore, when comparing findings across studies, attention should be paid not only to performance indices such as AUC or F1, but also to differences in sample construction, annotation procedures, and task definition.

Issues and Prospects

Research using social network data and machine learning to analyze depressive states has achieved initial success, but notable limitations and shortcomings remain.

Feature Validity

In feature engineering, researchers typically select relevant features based on experience and subjective judgment, but such selections often lack guaranteed validity. Studies commonly rely on simple correlation analysis to explore feature-label relationships (Etkin, 2018). Employing deeper causal analysis can establish stronger connections and thus improve feature validity (Boes et al., 2018; Dond éet al., 2017). Because model performance depends directly on feature quality, unverified features can produce substantially erroneous predictions. In addition to feature validity, recent studies suggest that label validity and annotation quality are equally critical. Kabir et al. (2023) proposed DEPTWEET, a dataset of 40,191 tweets annotated under a DSM-5- and PHQ-9-informed typology with four labels—non-depressed, mild, moderate, and severe—and an associated confidence score for each label. Compared with simple binary labels or weak self-disclosure labels, this line of work provides stronger support for discussing severity recognition, annotation reliability, and the quality of training data in depression detection on social media.

Limited Model Interpretability

Trained models often yield parameters with ambiguous theoretical meaning, allowing different researchers to interpret the same parameter in divergent ways. When model parameters correspond to psychiatric or psychological constructs, no assessment index exists to quantify the degree of correspondence (Huys, Moutoussis, & Williams, 2011). Deep learning models in particular tend to lack feature-level interpretability and thus do not clarify specific social-network behaviors or performance patterns of patients with depression. Although higher data dimensionality and multilayer network architectures can improve predictive accuracy, they typically reduce model interpretability.

Limited Model Generalizability

The objective of machine learning is to produce a model that performs well on new samples rather than only on the training set. This capacity is termed the model's generalization ability. Larger training sets increase the likelihood of learning a model with strong generalization. Conversely, small sample sizes tend to produce large measurement errors and greater model variability (Schnack & Kahn, 2016). Excessively homogeneous samples can cause the model's predictions to reflect sample idiosyncrasies instead of the underlying disease (Mendelson et al., 2017). In our cross-validation, subsamples drawn from the same dataset yielded relatively high accuracy, but the model's generalization and reliability must still be tested on new samples from different sources, including varying gender, age, education, and cultural backgrounds.

Limited Clinical Applicability

Most studies have focused on theoretical model development and have not produced runnable platforms or applications, resulting in a low rate of clinical adoption. User-friendly applications and platforms could provide early warnings, support clinicians' diagnosis of depression, and help ensure timely treatment. Currently, some smartphone and wearable applications monitor heart rate, sleep quality, daily activity, and phone usage in people with depression (Mendelson et al., 2016; Paula et al., 2016). We argue that social-network–based prediction software holds substantial potential, but its translation into clinical practice must be accelerated.

For future research, we recommend grounding studies in established clinical psychology and psychiatry theories, strengthening interdisciplinary links, and improving methodological standardization to enhance practical impact. We also encourage collaboration with related experts—clinical psychiatrists, medical researchers, and social workers who have experience assisting people with mental illness—because they interact frequently with patients with depression and thus have a clearer understanding of patients' presentations. The practical knowledge these professionals gain from patient interaction can offer highly valuable guidance for theoretical research.

References

Benton, A., Mitchell, M., & Hovy, D. (2017). Multi-task learning for mental health using social media text. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 1*, 152-162.

Blei, D. M., Ng, A. Y., Jordan, M. I. et al. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022.

Boes, A. D., Uitermarkt, B. D., Albazron, F. M. et al. (2018). Rostral anterior cingulate cortex is a structural correlate of repetitive TMS treatment response in depression. *Brain Stimulation, 11*(3), 575-582. https://doi.org/10.1016/j.brs.2018.01.029

Chancellor, S., Lin, Z. Y. et al. (2016). Quantifying and predicting mental illness severity in online pro-eating disorder communities. *Acm Conference on Computer-supported Cooperative Work & Social Computing. ACM*.

Cheng, J.-C., & Chen, A. L. P. (2022). Multimodal time-aware attention networks for depression detection. *Journal of Intelligent*

Information Systems, 59, 319-339. https://doi.org/10.1007/s10844-022-00704-w

Cheng, Q., Li, T. M. H., Kwok, C.-L. L., Zhu, T., & Yip, P. S. F. (2017). Assessing suicide risk and emotional distress in Chinese social media: A text mining and machine learning study. *Journal of Medical Internet Research, 19*(7), e243. https://doi.org/10.2196/jmir.7276

Dondé, C., Amad, A., Nieto, I. et al. (2017). Transcranial direct-current stimulation (tDCS) for bipolar depression: A systematic review and meta-analysis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 78*, 123-131. https://doi.org/10.1016/j.pnpbp.2017.05.014

Duong, C. T., Lebret, R., & Aberer, K. (2017). Multimodal classification for analysing social media. arXiv. https://arxiv.org/abs/1708.02099

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences, 115*(44), 11203-11208. https://doi.org/10.1073/pnas.1802331115

Etkin, A. (2018). Addressing the causality gap in human psychiatric neuroscience. *JAMA Psychiatry, 75*(1), 3-4. https://doi.org/10.1001/jamapsychiatry.2017.3610

Fang, Z. Y. (2017). Depression prediction of Weibo users based on word vector method. *Electronic Technology & Software Engineering*, (7), 199-200.

Hu, Y. Q., & Liu, Z. H. (2019). Intervention study on mental health of depressed college students: The differential impact of school support across institution types. *Journal of Educational Science of Hunan Normal University, 18*(5), 120-125.

Huys, Q. J. M., Moutoussis, M., & Williams, J. (2011). Are computational models of any use to psychiatry? *Neural Networks, 24*(6), 544-551. https://doi.org/10.1016/j.neunet.2011.03.001

Jiang, W. Y., Jiang, G. R., Xu, S. et al. (2019). Trends in national mental health literacy: Based on big data analysis of depression-themed Weibo. In *Abstracts of the 22nd National Academic Conference on Psychology.*

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

Kim, J., Lee, D., & Park, E. (2021). Machine learning for mental health in social media: Bibliometric study. *Journal of Medical Internet Research, 23*(3), e24870. https://doi.org/10.2196/24870

Kshirsagar, R., Morris, R., & Bowman, S. (2017). Detecting and explaining crisis. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality* (pp. 66-73). Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-3108

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

Liu, J., Chen, W., Wang, L. et al. (2024). A hybrid depression detection model and correlation analysis for social media based on attention mechanism. *International Journal of Machine Learning and Cybernetics, 15*, 2631-2642. https://doi.org/10.1007/s13042-023-02053-8

Lyu, S., Ren, X., Du, Y., & Zhao, N. (2023). Detecting depression of Chinese microblog users via text analysis: Combining Linguistic Inquiry Word Count (LIWC) with culture and suicide related lexicons. *Frontiers in Psychiatry, 14*, 1121583. https://doi.org/10.3389/fpsyt.2023.1121583

Men, X. P., Wei, R. B., & Wu, X. L. (2020). Analysis and detection of language and behavioral characteristics of depressed users in social networks. *Journal of Modern Information, 40*(6), 76-87.

Mendelson, A. F., Zuluaga, M. A., Lorenzi, M. et al. (2017). Selection bias in the reported performances of AD classification pipelines. *NeuroImage: Clinical, 14*, 400-412. https://doi.org/10.1016/j.nicl.2016.12.018

Nadeem, M., Horn, M., Coppersmith, G. et al. (2016). Identifying depression on Twitter[EB/OL]. https://arxiv.org/ftp/arxiv/papers/1607/1607.07384.pdf

Orabi, A. H., Buddhitha, P., Orabi, M. H., & Inkpen, D. (2018). Deep learning for depression detection of Twitter users. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic* (pp. 88-97). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-0609

Paula, G., Ferrer, L. et al. (2016). Psychologist in a pocket: Lexicon development and content validation of a mobile-based app for depression screening. *JMIR mHealth and uHealth, 4*(3), e102. https://doi.org/10.2196/mhealth.5284

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007*. Austin (TX): LIWC.net.

Phiri, D., Makowa, F., Amelia, V. L., Phiri, Y. V. A., Dlamini, L. P., & Chung, M. H. (2025). Text-based depression prediction on social media using machine learning: Systematic review and meta-analysis. *Journal of Medical Internet Research, 27*, e59002. https://doi.org/10.2196/59002

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, *6*(1), 15. https://doi.org/10.1140/epjds/s13688-017-0110-z

Reece, A. G., Reagan, A. J., Lix, K. L. M., Dodds, P. S., Danforth, C. M., & Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports, 7*, 13006. DOI:10.1038/s41598-017-12961-9

Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V.-A., & Boyd-Graber, J. (2015). The University of Maryland CLPsych 2015 shared task system. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (pp. 54-60). Association for Computational Linguistics. https://doi.org/10.3115/v1/W15-1207.

Ricard, B. J., Marsch, L. A., Crosier, B. S., & Hassanpour, S. (2018). Exploring the utility of community-generated social media content for detecting depression: An analytical study on Instagram. *Journal of Medical Internet Research, 20*(12), e11817. https://doi.org/10.2196/11817

Saravia, E., Chang, C.-H., Wu, H.-C., & Chen, Y.-S. (2016). MIDAS: Mental illness detection and analysis via social media. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 896-903). IEEE. https://doi.org/10.1109/ASONAM.2016.7752348

Schnack, H. G., & Kahn, R. S. (2016). Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters. *Frontiers in Psychiatry, 7*, 50. https://doi.org/10.3389/fpsyt.2016.00050

Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.-S., & Zhu, W. (2017). Depression detection via harvesting social media: A multimodal dictionary learning solution. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)* (pp. 3838-3844). https://doi.org/10.24963/ijcai.2017/536

Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., & Ohsaki, H. (2015). Recognizing depression from Twitter activity. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3187-3196). DOI:10.1145/2702123. 2702280

Vedula, N., & Parthasarathy, S. (2017). Emotional and linguistic cues of depression from social media. In *Proceedings of the 2017 International Conference on Digital Health* (pp. 127-136). ACM. https://doi.org/10.1145/3079452.3079465.

Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2018). A multilevel predictive model for detecting social network users with depression. *2018 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 130-135). https://doi.org/10.1109/ICHI.2018.00022

World Health Organization. Depressive disorder (depression) [EB/OL]. (2023-03-31) [2025-12-08]. https://www.who.int/news-room/fact-sheets/detail/depression

Xin, L. M., Chen, L., Yang, F. D. et al. (2015). Risk factors for suicide attempts in patients with depression accompanied by anxiety symptoms. *Chinese Journal of Nervous and Mental Diseases*, *41*(10), 613-617.

Yazdavar, A. H., Mahdavinejad, M. S., Bajaj, G., Romine, W., Sheth, A., Monadjemi, A. H. et al. (2020). Multimodal mental health analysis in social media. *PLOS ONE, 15*(4), e0226248. https://doi.org/10.1371/journal.pone.0226248

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, 25*, 281-304. https://doi.org/10.1007/s11280-021-00992-2