A Systematic Review of Machine Learning Algorithms for Mental Health Detection Using Social Media Data

Madina Amiri[1,a], Sundresan Perumal[1,b], Norita MD Norwawi[1,c], Nurazah Ismail[2,d]

[1]Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM)
Bandar Baru Nilai, Malaysia
[2]Psychiatry Unit, Faculty of Medicine and Health Sciences, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, Malaysia
[a]madina.amiri1001@raudah.usim.edu.my, [b]Sundresan.p@usim.my, [c]norita@usim.edu.my, [d]azah.ismail@usim.edu.my

Abstract
Mental health plays a key role in our daily lives, affecting our thoughts, behaviors, and relationships. However, mental health issues like depression, anxiety, and stress are becoming more common due to multiple factors such as family problems, failed relationships, and social demands. Early detection and timely treatment are crucial to prevent harm to individuals and society. This systematic review looks at how machine learning can help identify mental health issues by analyzing data from social media. We review methods used in recent studies, such as natural language processing, sentiment analysis, and boosting algorithms, based on 100 peer-reviewed articles. These methods are grouped into supervised machine learning and unsupervised machine learning algorithms. Finally, we identify gaps such as a lack of clinical assessment and psychological acknowledgment in current studies. Overall, we suggest integrating psychological acknowledgment with machine learning to improve accuracy and responsibility in using machine learning for mental health assessment on social media.

Keywords: mental disorders, machine learning, social media

1. Introduction
Mental health is one of the most important factors in today's life; all our emotions, thoughts, and behaviors are related to mental health. Mental disorders are abnormalities in individuals' behaviors and actions, disrupting their daily routines. People are resorting to suicide and criminal activities because of the worsening of mental health disorders (Murray et al., 2023). These disorders can manifest in various forms, such as mood disorders (i.e., depression, suicidal thoughts), anxiety disorders, and psychotic disorders. All these types of disorders are significant risks to families and society and even to the patient's own life (Fucci et al., 2022).

As a result, people are uncomfortable sharing their mental health problems with friends or family, they do not seek help at the correct time to aggravate the disease. However, emergent algorithms

from Machine Learning (ML), Deep Learning (DL), Statically techniques, Natural Language Processing (NLP), and other techniques help to early detect mental disorders from social media data (Kolliakou et al., 2020). Machine learning (ML) is a branch of artificial intelligence that focuses on the creation and analysis of statistical algorithms that can effectively generalize and accomplish jobs without explicit instructions. People are suffering from mental disorders, crimes, and suicide are spreading widely nowadays, youth are attacking their family members and youths are committing suicide, and depressed and anxiety disorder patients are feeling embers to share their mental status. Because of limited awareness, older family members often overlook the mental health of their children. These problems are leading to the society's low happiness index (Kim et al., 2023)). Researchers and health policymakers should put their efforts to use the best technologies and techniques to prevent the worsening of mental disorders among individuals. They should also identify gaps in the existing literature that may need to be addressed over time through further experience with machine learning algorithms and research. With these objectives in mind, the current review was designed to summarize the existing literature on detection of mental disorders.

2. Methodology

In this section, we will describe in detail how we have collected and selected the articles for this systematic literature review. In addition, we will explain how we have analyzed the articles collected for this systematic literature review.

2.1 Search methodology and article selection

The current article is a narrative review of the existing literature on the detection of mental disorders by collecting data from various social media platforms such as X/Twitter and Facebook and collecting data from various pre-defined datasets such as eRisk 2018/Reddit and CLPSych 2015. A search of the IEEE, ELSEVIER, ScienceDirect and ResearchGate databases was undertaken using the search terms "mental disorders," "detection," "machine learning algorithms," "mental health," "psychiatry," "psychology," "anxiety," "depression" and "social media" in various permutations and combinations. A total of 100 citations were retrieved using this method. On reviewing the above citations, 80 articles were excluded because they dealt with other techniques of detection of mental disorders, such as statistical methods and natural language processing. A careful review of 20 articles is revealed in this study.

2.2 Methodological and thematic analysis of selected articles

A total of 20 articles were included in this review. Of these, only five articles could be genuinely labeled as "original research." All five studies were cross-sectional and observational in design. The remaining 15 articles were consistent with mental disorders detection using statistical methods and boosting algorithms as it was not possible to conduct a formal systematic review or meta-analysis given the nature of the above publications. Instead, a narrative review was undertaken, giving priority to the few observational studies available and briefly summarizing the salient themes from the other types of publication.

Three broad themes were identified across the 20 publications and were used to structure the review: (a) systematic literature reviews on mental disorder detection, (b) commentary and correspondence broadly addressing the use of machine learning algorithms in mental disorder

detection, (c) commentary and correspondence addressing the impact and prevalence of mental disorders.

3. Results

Five studies from different countries examined supervised machine learning for mental disorders detection (Devi & Gopi, 2024; Title-Tlatelpa et al., 2021; Yang et al., 2020; Shatte et al., 2020.; 2020; Allan H et al., 2019). Their results are summarized in the table below (Table 1).

Table 1: Supervised machine learning algorithms used for mental illness detection

Reference & Year	Technique	Dataset		Classification Type (Multi/Binary)	Mental Illness Class
		Platform	Size		
Devi & Gopi (2024)	Support Victore Machine	Facebook, Twitter and Instagram	-	Multiple	Depression, Anxiety, Suicidal thoughts, Phobia
(H et al., 2024)	Support Victore Machine	Youtube	-	Binary	Homophobia
Titla-Tlatelpa et al (2021)	Support Victore Machine, Random Forest, Decision Tree	eRisk 2018/Reddit Twitter	1.33M posts + (Shen et al., 2017) dataset size tweets	Binary	Depression
Yang et al (2020)	Support Victore Machine, Naive Bayes, Random Forest, Left-to-right parsing, K-Nearest Neighbors, Neighbors	Facebook	22M status updates from 153K users	Binary	Depression
Shatte et al (2020)	Support Victore Machine	Reddit	68K posts	Multi	Depression

As seen in the results above, all the papers focus on depression, with most of them emphasizing on SVM machine learning algorithms (Devi & Gopi, 2023; Shatte et al., 2020). Some studies combined other algorithms with SVM to achieve better performance (Yang et al., 2020; H et al., 2024; Title-Tlatelpa et al., 2021). As in the results, researchers used social media data for mental disorders detection studies.

Five of these papers (5/20) showed the use of unsupervised machine learning algorithms with rules for detecting mental disorders from social media.

Table 2: Unsupervised machine learning algorithms are used for detecting mental illnesses

Reference & Year	Technique	Dataset		Classification Type (Binary/Multi)	Mental illness Class
		Platform	Size		
Aragon et al. (2021)	Extension K-means (AP)	e-risk forum	--	Multi	Anorexia, Depression
Bae et al., 2021	Density-Based Spatial Clustering of Applications with Noise, Linear Discriminant Analysis	Reddit	425K posts	Binary	Schizophrenia
Alharthi, 2020	K+DENCLUE, AHDC	Twitter, Reddit	140K tweets + posts	Binary	Depression
Lopez-castroman et al., 2019	K-means	Questionnaire	2254 patients	Multi	Anorexia, Depression
Joshi & Patwardhan, 2020	Fuzzy C-means, Extended k-means	Twitter	80.2K tweets.	Multi	Mental illness

K-means unsupervised clustering algorithm is mostly used for the detection of mental disorders (Lopez-castroman et al., 2021; Narynov et al., 2020). Hence, researchers also used social media such as Twitter, Reddit, VKontakte, and other platforms to collect a wide range of data. In addition, Fuzzy C-means, PAM Hierarchical, DBSCAN, and LDA unsupervised algorithms also showed a significant performance in various types of mental disorders such as schizophrenia, suicidal thoughts and depression (Joshi & Patwardhan, 2020; Bae et al., 2022; Alharthi, 2020).

Boosting algorithms helps the researcher to increase the performance and accuracy of their work, many researchers are using boosting algorithms in mental disorders detection to boost accuracy for a better result. In this article, we collected five papers in which researchers used boosting algorithms for better performance of mental disorders detection. (Musleh et al., 2022) achieved an accuracy of 82.39% by using the Adaboost algorithm to detect depression from 4542 tweets. In addition, (Hidayatullah & Maharani, 2022; P. Kumar, 2022) used XGBoost and Gradient Boost algorithms to detect depression and anxiety using Twitter social media platforms, which achieved an accuracy of 81.25% and 78%. Hence, Catboost is also used in depression detection studies, which achieved a high accuracy of 91%.

After reviewing the papers carefully, there are some considerable gaps that need to be improved, such as having a small dataset. In addition, some studies have only worked for detecting depression among one gender, male or female. The overall gap that needs to be more focused on is the lack of psychological acknowledgment and lack of clinical assessment.

4. Conclusion

This study emphasized the critical role that mental health plays in modern society, particularly the concerning rise in mental illnesses such as depression and anxiety, which usually go untreated owing to stigma and a lack of understanding. Following a thorough study of the current literature, we uncovered the exciting potential of advanced methodologies such as Machine Learning (ML) and statistical methods for early diagnosis of these diseases.

The findings show a high dependence on social media data for research, indicating its usefulness in capturing real-time emotional states and behavioral trends. Several research has demonstrated effective techniques, particularly the use of Support Vector Machines (SVM) and various boosting algorithms. However, discipline still has significant gaps. More thorough longitudinal research is required, as the prevalence of cross-sectional studies limits our ability to fully comprehend causal relationships and long-term effects. Additionally, the current focus on illnesses and depression highlights the need for more thorough research that addresses a variety of mental health conditions.

Scholars and health officials must collaborate to take advantage of technological advancements as mental health becomes a more urgent global concern. This necessitates increasing public awareness and educating people about mental health issues in addition to fortifying the current detection mechanisms. Addressing the constraints found in this research would pave the way for more effective detection models, resulting in improved mental health outcomes and reduced social hazards associated with untreated mental diseases. Future research should attempt to fill these gaps, promoting a comprehensive approach to mental health that values innovation while putting patients' well-being first.

References

Alharthi, S. A. (2020). Advances in Computer Sciences Empirical Study of Features and Unsupervised Sentiment Analysis Techniques for Depression Detection in Social Media. 1–10.

Aragon, M. E., Lopez-Monroy, A. P., Gonzalez-Gurrola, L. C., & Montes-Y-Gomez, M. (2023). Detecting Mental Disorders in Social Media Through Emotional Patterns - The Case of Anorexia and Depression. IEEE Transactions on Affective Computing, 14(1), 211–222. https://doi.org/10.1109/TAFFC.2021.3075638

Bae, Y. J., Shim, M., & Lee, W. H. (2021). Schizophrenia Detection Using Machine Learning Approach from Social Media Content. Sensors, 21(17), 5924. https://doi.org/10.3390/s21175924

Devi, T. J., & Gopi, A. (2024). An Efficient Novel Approach for Early Detection of Mental Health Disorders Through Distributed Machine Learning Paradigms from Public Societal Communication. International Journal of Intelligent Systems and Applications in Engineering, 12(2), 767–778. https://ijisae.org/index.php/IJISAE/article/view/4319

Fucci, D., Romano, S., Baldassarre, M., Caivano, D., Scanniello, G., Thuran, B., & Juristo, N. (2022). A Longitudinal Cohort Study on the Retainment of Test-Driven Development. ArXiv, 1(1). https://arxiv.org/pdf/1807.02971.pdf

H, S., Sivakumar, S., R, R., Jayaguptha, N., & Thenmozhi, D. (2024, March 1). Quartet@LT-EDI 2024: Support Vector Machine Based Approach For Homophobia/Transphobia Detection In Social Media Comments (B. R. Chakravarthi, B. B, P. Buitelaar, T. Durairaj, G. Kovács, & M. Á. García Cumbreras, Eds.). ACLWeb; Association for Computational Linguistics.

https://aclanthology.org/2024.ltedi-1.28/

Hidayatullah, M. R., & Warih Maharani. (2022). Depression Detection on Twitter Social Media Using Decision Tree. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 6(4), 677–683. https://doi.org/10.29207/resti.v6i4.4275

Joshi, D., & Patwardhan, Dr. Manasi. (2020). An analysis of mental health of social media users using unsupervised approach. Computers in Human Behavior Reports, 2, 100036. https://doi.org/10.1016/j.chbr.2020.100036

Kolliakou, A., Bakolis, I., Chandran, D., Derczynski, L., Werbeloff, N., Osborn, D. P. J., Bontcheva, K., & Stewart, R. (2020). Mental health-related conversations on social media and crisis episodes: a time-series regression analysis. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-57835-9

Kim, A. Y., Jang, E. H., Lee, S.-H., Choi, K.-Y., Park, J. G., & Shin, H.-C. (2023). Automatic Depression Detection Using Smartphone-Based Text-Dependent Speech Signals: Deep Convolutional Neural Network Approach. Journal of Medical Internet Research, 25(1), e34474. https://doi.org/10.2196/34474

Kumar, P., Samanta, P., Dutta, S., Chatterjee, M., & Sarkar, D. (2022). Feature-Based Depression Detection from Twitter Data Using Machine Learning Techniques. Journal of Scientific Research, 66(02), 220–228. https://doi.org/10.37398/jsr.2022.660229

Lopez-Castroman, J., Abad-Tortosa, D., Cobo, A., Courtet, P., Barrigón, M. L., Artés-Rodríguez, A., & Baca-García, E. (2019). PSYCHIATRIC PROFILES OF E-HEALTH USERS: WHAT CAN WE LEARN USING DATA MINING TECHNIQUES? (Preprint). JMIR Mental Health. https://doi.org/10.2196/17116

Musleh, D. A., Alkhales, T. A., Almakki, R. A., Alnajim, S. E., Almarshad, S. K., Alhasaniah, R. S., Aljameel, S. S., & Almuqhim, A. A. (2022). Twitter Arabic sentiment analysis to detect depression using machine learning. Computers, Materials and Continua, 71(2), 3463–3477. https://doi.org/10.32604/cmc.2022.022508

Murray, A. L., Brown, R., Zhu, X., Speyer, L. G., Yang, Y., Xiao, Z., Ribeaud, D., & Eisner, M. (2023). Prompt-level predictors of compliance in an ecological momentary assessment study of young adults' mental health. Journal of Affective Disorders, 322(May 2022), 125–131. https://doi.org/10.1016/j.jad.2022.11.014

Shatte, A. B. R., Hutchinson, D. M., Fuller-Tyszkiewicz, M., & Teague, S. J. (2020). Social Media Markers to Identify Fathers at Risk of Postpartum Depression: A Machine Learning Approach. Cyberpsychology, Behavior, and Social Networking, 23(9), 611–618. https://doi.org/10.1089/cyber.2019.0746

Sinha, P. P. (2019). # suicidal - A Multipronged Approach to Identify and Explore Suicidal Ideation in Twitter. 941–950.

Titla-Tlatelpa, J. de J., Ortega-Mendoza, R. M., Montes-y-Gómez, M., & Villaseñor-Pineda, L. (2021). A profile-based sentiment-aware approach for depression detection in social media. EPJ Data Science, 10(1). https://doi.org/10.1140/epjds/s13688-021-00309-3

Yang, Y., Yang, Y., Pan, A., Xu, Z., Wang, L., & Zhang, Y. (2022). Identifying Depression in Parkinson' s Disease by Using Combined Diffusion Tensor Imaging and Support Vector Machine. 13(June), 1–8. https://doi.org/10.3389/fneur.2022.878691

Biodata

Madina Amiri
Msc Master candidate

Ts.Dr. Sundresan Perumal
sundresan.p@usim,edu.my
Ph.D in Science and Technology from University Sains Islam Malaysia
Deputy Director (Innovation, Product Development and Commercialization) at Universiti Sains Islam Malaysia

Prof. Ts. Dr. Norita Md. Norwawi
norita@usim.edu.my
PhD in Computational Recognition-Printed Decision Model Based on Temporal Data Mining Approach in a Multiagent Environment for Reservoir Flood Control Decision
Head of Cybersecurity & System Research Unit at Universiti Sains Islam Malaysia

Dr Nurazah Ismail
azah.ismail@usim,edu.my
MBBS (USIM), Doctor of Psychiatry (DrPsych) Universiti Kebangsaan Malaysia
Medical lecturer & PsychiatristUniversiti Sains Islam Malaysia