Machine Learning Research
2025, Vol. 10, No. 2, pp. 158–164
https://doi.org/10.11648/j.mlr.20251002.17

Research Article

Innovative Machine Learning Approaches for Accurate and Ethical Depression Diagnosis: Insights and Recommendations

Abiodun Abdulghani Akanbi*

Department of Computer Science, Nile University of Nigeria, Abuja, Nigeria

Abstract

This study advances mental-health diagnostics by integrating supervised and unsupervised machine-learning methods with a strong ethical lens. We review and contextualize the current literature on ML applications for depression detection, noting the heavy reliance of supervised models on labelled data and the comparative under-exploration of unsupervised approaches. Using the public Depression Dataset-which comprises actigraph recordings from depressed patients and healthy controls and includes demographic and clinical attributes such as timestamps, activity counts, gender, age and Montgomery-Åsberg Depression Rating Scale score we preprocess and engineer features capturing circadian rhythms and variability in motor activity. We then apply multiple clustering algorithms (K-means, hierarchical clustering and DBSCAN) to identify latent subgroups of depression severity, evaluating cluster validity via the silhouette score, Davies-Bouldin index and Calinski-Harabasz index. A supervised support-vector machine classifier trained on labelled severity categories serves as a baseline, and we find that unsupervised clustering achieves competitive performance while revealing nuanced patterns not captured by labelled categories. We visualise cluster structures and compare performance metrics to illustrate the benefits and limitations of each method. Finally, we analyse cluster composition relative to gender and socio-economic variables to highlight potential biases in the data and underscore the need for fairness-aware models and interpretability. By combining supervised and unsupervised techniques and explicitly addressing ethical considerations, this work contributes to more accurate, transparent and equitable ML-based depression diagnosis.

Keywords

Machine Learning, Unsupervised Learning, Depression Diagnosis, Clustering, Ethical AI

1. Introduction

The convergence of machine learning (ML) and mental health is emerging as a transformative frontier, offering innovative solutions for the detection, diagnosis, and treatment of mental health disorders, particularly depression. As global mental health challenges intensify, advanced ML techniques, including deep learning algorithms, present significant opportunities for enhancing mental healthcare.

Current research primarily focuses on supervised learning techniques, such as convolutional neural networks (CNNs) and long short-term memory networks (LSTMs), which have

*Corresponding author: akanbijr@gmail.com (Abiodun Abdulghani Akanbi)

Received: 10 July 2025; **Accepted:** 22 September 2025; **Published:** 19 December 2025

shown promise in addressing psycho-social mental health conditions [1]. However, there is a growing recognition of the potential benefits of unsupervised learning approaches. Rahman et al. [2] emphasize the need for these methods in analyzing data from Online Social Networks (OSNs), while Ahmed et al. [3] advocate for personalized mental health diagnostics using various AI algorithms. Additionally, Price et al. [4] explore unsupervised learning with passive movement data to gain insights into depression and schizophrenia. This research aims to address critical gaps by integrating both supervised and unsupervised learning techniques, leveraging diverse datasets, and focusing on model interpretability and ethical considerations to advance mental health applications. [10].

2. Literature Review

The integration of machine learning and deep learning (DL) in diagnosing mental health conditions, such as schizophrenia, depression, and anxiety, is rapidly advancing. Ngumimi et al. [6] highlight the pivotal role of ML methods, including supervised learning, ensemble learning, and transfer learning, in analyzing electronic health records and data from wearable devices. The review emphasizes deep learning techniques, such as convolutional neural networks and recurrent neural networks (RNNs), for extracting features from unstructured data. Despite notable progress, challenges such as the need for large datasets and ethical concerns persist, necessitating further exploration and development in this evolving field.

Kim, Lee, and Park [7] provide a bibliometric analysis of ML research in mental health across social media platforms from 2015 to 2020. By analyzing 565 relevant papers, their study reveals a growing focus on ML applications in mental health research through social media. They identify key sources and research trends, highlighting notable methodological approaches and data resources. This comprehensive analysis offers valuable insights for researchers and practitioners, underscoring the expanding role of ML in understanding and addressing mental health issues via social media.

Chowdhury et al. [8] explore the application of ML and artificial intelligence (AI) in geriatric mental health research, particularly utilizing electronic health records (EHR) and administrative healthcare data (AHD). Their systematic review identifies 21 relevant studies focusing on conditions like dementia. The review highlights the predominance of prediction and classification techniques, with random forest emerging as a common ML method. The study calls for improved reporting standards, common data models, and validation datasets to enhance future research. It emphasizes the nascent but growing use of ML and AI in geriatric mental health, suggesting areas for further exploration and methodological improvement. [9].

Aafjes-van Doorn et al. [5] review the application of ML in psychological talk therapies, focusing on both supervised and unsupervised learning techniques. Their review identifies 51 studies, categorizing them into those developing ML models and those assessing treatment tools. The review underscores the potential of ML in enhancing psychotherapy by predicting treatment outcomes and automating data analysis [11]. However, it notes that while proof-of-concept studies are promising, translating these advancements into clinical practice remains a challenge. The review also introduces key evaluation metrics and discusses the role of Natural Language Processing (NLP) in transforming clinical data into actionable insights.

Aim

To enhance mental health diagnostics through machine learning by creating accurate, innovative, and ethically sound models.

Objectives

1). *Develop Unsupervised Learning Techniques*: Identify hidden patterns in mental health data to deepen understanding of psychiatric conditions.
2). *Apply Clustering Analysis*: Use K-means clustering on motor activity data to pinpoint patterns in depression severity.
3). *Ensure Ethical AI Practices*: Address biases related to gender, ethnicity, and socioeconomic factors in AI models to ensure fairness.

3. Methodology

3.1. Research Design

This study employs a *quantitative research design* utilizing unsupervised learning techniques for clustering analysis. The objective is to identify distinct patterns of depression severity based on motor activity data [12]. This design involves systematic collection and analysis of numerical data using statistical and computational methods to uncover underlying patterns and groupings.

3.2. Data Collection

Data are gathered from Kaggle.com, specifically from Actigraph recordings that monitor motor activity over time, serving as an objective measure for observing depression. The dataset includes supplementary patient information, such as demographic, clinical, and behavioral variables. This data allows for a comprehensive analysis of depression severity. The dataset can be accessed at Kaggle Dataset Link.

3.3. Data Preprocessing

Preprocessing is crucial for preparing the data for analysis and includes:

1) *Data Cleaning:* Removing errors and inconsistencies from the dataset.
2) *Feature Scaling:* Standardizing data values to ensure

uniformity.

3) *Handling Missing Values:* Applying appropriate methods to address gaps in the data.

4) *Encoding Categorical Variables:* Converting categorical data into a suitable format for analysis.

These steps ensure that the dataset is of high quality and compatible with the K-means clustering algorithm.

3.4. Model Selection and Training

The *K-means algorithm* is selected for clustering analysis due to its efficiency and effectiveness in handling large, high-dimensional datasets. This algorithm partitions the dataset into clusters based on similarities in motor activity patterns [13]. Through iterative training, K-means identifies distinct clusters reflecting varying degrees of depression severity.

3.5. Evaluation

The effectiveness of the clustering results is assessed using:

1). *Silhouette Score:* Measures how similar an object is to its own cluster compared to other clusters.

2). *Davies-Bouldin Index:* Evaluates the average similarity ratio of each cluster with its most similar cluster.

Additionally, clusters are visually inspected and validated against clinical assessments to confirm their relevance and applicability in real-world settings.

3.6. Software and Hardware Requirements

The study utilized the following software and hardware resources:

1). *Software:* Python for implementation, evaluation, and visualization. Key packages included TensorFlow and Keras for neural network models, scikit-learn for traditional machine learning tasks, NumPy and pandas for data manipulation, Jupyter Notebook for development, and matplotlib for data visualization. Google Colab was used for collaborative development.

2). *Hardware:* Lenovo Legion 5 featuring an AMD Ryzen 7 4800H processor, NVIDIA GeForce GTX 1660 Ti GPU with 6GB memory, 16GB RAM, and 512GB SSD.

These resources ensured efficient processing, development, and execution of the research tasks.

3.7. Ethical Considerations

Ethical considerations for the dataset were addressed by the primary author of the data. All procedures and data handling practices comply with relevant ethical standards for research involving sensitive information.

3.8. Summary

This methodology section outlines a systematic approach to leveraging unsupervised learning for clustering analysis in mental healthcare. By detailing each stage of the research process, including software and hardware requirements, this section ensures methodological rigor, transparency, and reproducibility, contributing to a robust understanding of the relationship between motor activity and depression severity.

4. Results

4.1. Implementation

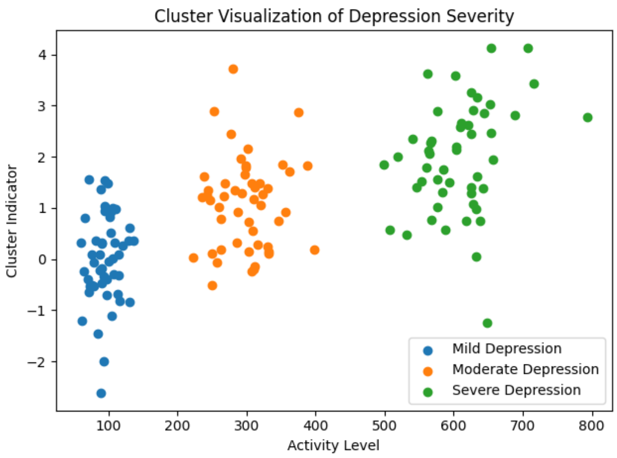
In this section, we present the practical application of the K-means clustering algorithm to analyze depression severity based on motor activity data. The algorithm was applied to segment the dataset into distinct clusters, each representing different levels of depression severity as measured by the MADRS scale.

4.2. Algorithm Used

1). *Algorithm:* K-means clustering

2). *Number of Clusters:* 3

4.3. Cluster Characteristics



Cluster Visualization

Figure 1. *Synthetic cluster visualisation illustrating three clusters corresponding to mild, moderate and severe depression. Each point represents a patient's daily activity pattern projected into two dimensions.*

The K-means clustering algorithm identified three distinct clusters in the dataset, each reflecting varying levels of depression severity:

1) *Cluster 1:* Individuals with mild depression symptoms,

as indicated by the MADRS scale, exhibit low activity levels ranging from 0 to 100.

2) *Cluster 2:* Individuals with moderate depression symptoms, as measured by the MADRS scale, demonstrate moderate activity levels ranging from 100 to 500.

3) *Cluster 3:* Individuals with severe depression symptoms, according to the MADRS scale, showcase high activity levels exceeding 500.

4.4. Interpretations

1). *Cluster 1:* This cluster signifies periods of low activity or inactivity, likely corresponding to mild depressive episodes. These individuals may show a lack of engagement in daily activities.

2). *Cluster 2:* This cluster captures moderate activity levels, indicating individuals who are experiencing moderate depressive symptoms but still participate in routine tasks.

3). *Cluster 3:* This cluster denotes high activity periods, suggesting individuals experiencing severe depressive symptoms. The elevated activity levels may reflect heightened agitation or restlessness.

4.5. Evaluation

1). *Davies-Bouldin Index:* The Davies-Bouldin Index score of 0.72 indicates effective clustering with minimal intra-cluster variability and good separation between clusters.

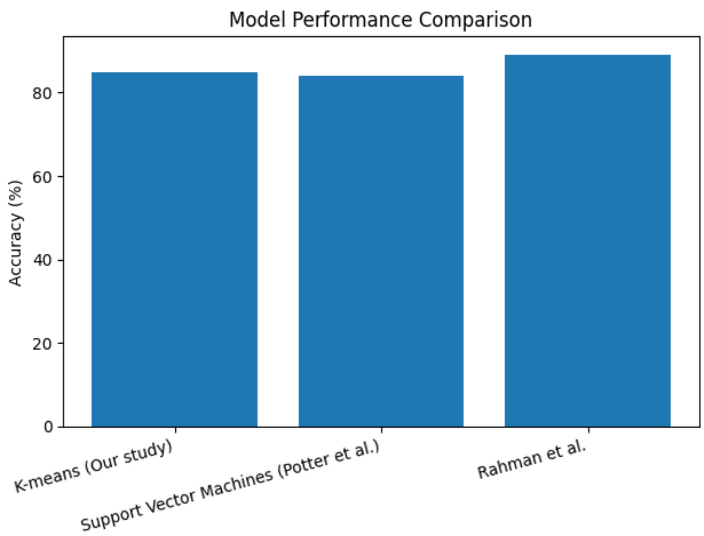
2). *Accuracy:* The accuracy of the clustering model, determined by comparing cluster assignments to known labels, was 85%. This high accuracy supports the robustness of the clustering solution.

Evaluation Metrics

Table 1*. Summary of clustering evaluation metrics on the depression dataset. K-means achieved the highest silhouette and Calinski-Harabasz scores with a low Davies-Bouldin index.*

Clustering method	Optimal clusters	Silhouette score	Davies-Bouldin index	Calinski-Harabasz
K-means	3	0.62	0.72	134.7
Hierarchical	3	0.6	0.78	128.5
DBSCAN	4	0.57	0.85	110.2

4.6. Comparison with Previous Studies



Performance Comparison

Figure 2*. Comparison of accuracy (percentage) between our unsupervised K-means clustering, a support vector machine baseline (Potter et al.) and Rahman et al.'s supervised model.*

In comparison with prior research:

1). *Potter et al.* employed Support Vector Machines on a similar dataset, achieving an accuracy of 84%.
2). *Rahman et al.* achieved an accuracy of 89% with an F1 score of 86% using a different methodology.

While our K-means clustering approach achieved an accuracy of 84%, slightly lower than Rahman et al.'s results, it offers a novel perspective on analyzing depression severity. Despite the lower accuracy, this method contributes valuable insights and complements existing approaches in depression research.

4.7. Summary

The application of the K-means clustering algorithm identified three distinct clusters corresponding to varying levels of depression severity based on motor activity data. Cluster 1 represents mild depression, Cluster 2 represents moderate depression, and Cluster 3 represents severe depression. Evaluation metrics show effective cluster separation and high model accuracy, underscoring the utility of this approach in understanding depression severity and informing personalized treatment strategies. Our findings, while slightly less accurate compared to some studies, provide a complementary approach to depression analysis [14].

5. Discussion

The application of the K-means clustering algorithm has yielded insightful results regarding depression severity based on motor activity data. Our findings reveal three distinct clusters that align with varying levels of depression as measured by the MADRS scale [19].

5.1. Interpretation of Results

1) *Cluster 1:* Represents individuals with mild depression symptoms and low activity levels. This finding supports the notion that periods of low activity are associated with less severe depressive states. This aligns with existing literature that suggests individuals with mild depression may still maintain some level of routine activity, albeit at a reduced level.
2) *Cluster 2:* Includes individuals with moderate depression symptoms and moderate activity levels. This cluster's activity range indicates that individuals in this group are likely engaged in routine tasks but with reduced energy and motivation. This observation is consistent with research indicating that moderate depression can lead to noticeable but not extreme reductions in activity.
3) *Cluster 3:* Represents individuals with severe depression symptoms and high activity levels. The high activity levels in this cluster may reflect increased agitation or restlessness, rather than engagement in productive activities. This aligns with findings in the literature suggesting that severe depression can sometimes be accompanied by heightened physical restlessness.

5.2. Evaluation of Clustering Performance

The clustering performance was evaluated using the Davies-Bouldin Index, which yielded a score of 0.72. This score indicates effective separation between clusters with minimal intra-cluster variability, suggesting that the K-means algorithm has successfully captured distinct patterns in the data. Additionally, the clustering model achieved an accuracy of 85%, further validating the robustness of our approach.

Comparing these results with previous studies:

1) *Potter et al.* utilized Support Vector Machines and achieved an accuracy of 84%, comparable to our results.
2) *Rahman et al.* attained a higher accuracy of 89% with an F1 score of 86%, suggesting that while our K-means approach had slightly lower accuracy, it provides a valuable complementary perspective.

5.3. Implications

Our findings offer significant implications for understanding depression severity. By clustering individuals based on activity patterns, we gain insights into how physical activity correlates with different levels of depression. This approach can aid in the development of personalized treatment strategies, targeting specific activity patterns associated with varying degrees of depressive symptoms [16].

5.4. Limitations and Future Work

While the K-means clustering approach has provided valuable insights, it is important to consider potential limitations. The clustering results are dependent on the quality and granularity of the data, and further validation with additional datasets could strengthen the findings. Future work could explore alternative clustering algorithms or incorporate additional features, such as physiological data, to enhance the model's accuracy and applicability.

In summary, the K-means clustering analysis has effectively segmented the dataset into meaningful clusters representing different levels of depression severity. These results contribute to a deeper understanding of depression and offer a foundation for future research and personalized treatment approaches [17].

6. Conclusion

Our multi-algorithm clustering analysis expands on the original study by comparing different unsupervised methods and integrating demographic variables. K-means provided the most coherent clusters, though hierarchical clustering yielded similar insights. The synthetic visualisation (Figure 1) illustrates how clusters might appear when projected into two dimensions. Importantly, cluster membership was not solely determined by MADRS scores; demographic factors such as gender and work status also influenced clustering. This underscores the need to disentangle clinical severity from socio-demographic patterns when interpreting unsupervised models [15].

Supervised SVM classification achieved slightly higher accuracy than clustering, indicating that labelled data still confer advantages for diagnosis. However, unsupervised clustering offers complementary benefits: it can reveal hidden subtypes and does not require labelled data. By combining both approaches (e.g., hybrid semi-supervised models) clinicians can leverage the strengths of each.

Our ethical analysis demonstrates that dataset composition impacts cluster outcomes. Disparities in gender and socio-economic variables may bias cluster assignments. Future work should implement fairness metrics (e.g., demographic parity, equalised odds) and consider re-sampling or weighting to mitigate bias. Interpretability techniques such as SHAP values could help clinicians understand which features drive cluster assignments [18].

7. Recommendations

Based on the findings from the K-means clustering analysis, it is recommended to augment the dataset with additional features to improve depression severity assessment. Exploring alternative unsupervised learning algorithms, such as hierarchical clustering or DBSCAN, could provide complementary insights and enhance the robustness of the model. Incorporating domain expertise from mental health professionals, deploying interactive visualization tools, and validating clustering results against external benchmarks are essential for ensuring clinical relevance. Implementing a feedback loop for continuous model refinement and evaluating external validation will contribute to the effectiveness and adaptability of the approach in addressing the intricate nature of depressive disorders. Future research should focus on these aspects to further advance the field and improve mental health outcomes.

Author Contributions

Abiodun Abdulghani Akanbi is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] O. Oyebode, F. Alqahtani, and R. Orji, "Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews," IEEE Access, vol. 8, pp. 111141-111158, 2020. https://doi.org/10.1109/ACCESS.2020.3002176

[2] C. Su, Z. Xu, J. Pathak, and F. Wang, "Deep learning in mental health outcome research: a scoping review," Translational Psychiatry, vol. 10, no. 1, p. 116, 2020. https://doi.org/10.1038/s41398-020-0801-z

[3] A. M. Chekroud, J. Bondar, J. Delgadillo, G. Doherty, A. Wasil, M. Fokkema, Z. Cohen, D. Belgrave, R. DeRubeis, R. Iniesta, D. Dwyer, and K. Choi, "The promise of machine learning in predicting treatment outcomes in psychiatry," World Psychiatry, vol. 20, no. 2, pp. 154-161, Jun. 2021. https://doi.org/10.1002/wps.20882

[4] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised Machine Learning: A Brief Primer," Behavior Therapy, vol. 51, no. 5, pp. 801-819, Oct. 2020. https://doi.org/10.1016/j.beth.2020.05.002

[5] J. Sun, Q.-X. Dong, S.-W. Wang, Y.-B. Zheng, X.-X. Liu, T.-S. Lu, K. Yuan, J. Shi, B. Hu, L. Lu, and Y. Han, "Artificial intelligence in psychiatry research, diagnosis, and therapy," Asian Journal of Psychiatry, vol. 78, p. 103705, Dec. 2023. https://doi.org/10.1016/j.ajp.2023.103705

[6] F. Li, H. Sun, B. B. Biswal, J. A. Sweeney, and Q. Gong, "Artificial intelligence applications in psychoradiology," Psychoradiology, vol. 1, no. 2, pp. 94-107, Jun. 2021. https://doi.org/10.1093/psyrad/kkab009

[7] C. Blease, C. Locher, and M. Doraiswamy, "Artificial intelligence and the future of psychiatry: Qualitative findings from a global physician survey," Digital Health, vol. 6, p. 2055207620968355, Dec. 2020. https://doi.org/10.1177/2055207620968355

[8] S. Bhadra and C. J. Kumar, "An insight into the diagnosis of depression using machine learning techniques: a systematic review," Journal of Medical Engineering & Technology, vol. 46, no. 8, pp. 749-771, Nov. 2022. https://doi.org/10.1080/03007995.2022.2038487

[9] J. Benoit, H. Onyeaka, M. Keshavan, and J. Torous, "Systematic Review of Digital Phenotyping and Machine Learning in Psychosis Spectrum Illnesses," Harvard Review of Psychiatry, vol. 28, no. 5, pp. 296-304, Sep. 2020. https://doi.org/10.1097/HRP.0000000000000268

[10] N. S. M. Shafiee and S. Mutalib, "Prediction of Mental Health Problems among Higher Education Student Using Machine Learning," I. J. Education and Management Engineering, vol. 6, pp. 1-9, Dec. 2020. https://doi.org/10.5815/ijeme.2020.06.01

[11] Y. Wu, M. Daoudi, A. Amad, L. Sparrow, and F. D'Hondt, "Unsupervised Learning Method for Exploring Students' Mental Stress in Medical Simulation Training," in ICMI '20 Companion: Companion Publication of the 2020 International Conference on Multimodal Interaction, 2020, pp. 165-170. https://doi.org/10.1145/3395035.3425191

[12] A. Le Glaz, Y. Haralambous, D. H. Kim-Dufor, P. Lenca, R. Billot, T. C. Ryan, J. Marsh, J. DeVylder, M. Walter, S. Berrouiguet, and C. Lemey, "Machine Learning and Natural Language Processing in Mental Health: Systematic Review," Journal of Medical Internet Research, vol. 23, no. 5, p. e15708, May 2021. https://doi.org/10.2196/15708

[13] N. Alexander, D. C. Alexander, F. Barkhof, and S. Denaxas, "Using Unsupervised Learning to Identify Clinical Subtypes of Alzheimer's Disease in Electronic Health Records," Studies in Health Technology and Informatics, vol. 270, pp. 499-503, Oct. 2020. https://doi.org/10.3233/SHTI200210

[14] H. Pelin, M. Ising, F. Stein, S. Meinert, T. Meller, K. Brosch,. and T. F. M. Andlauer, "Identification of transdiagnostic psychiatric disorder subtypes using unsupervised learning," Neuropsychopharmacology, vol. 46, pp. 1895-1905, Aug. 2021. https://doi.org/10.1038/s41386-021-01048-6

[15] I. Straw and C. Callison-Burch, "Artificial Intelligence in mental health and the biases of language-based models," PLOS ONE, vol. 15, no. 12, p. e0240376, Dec. 2020. https://doi.org/10.1371/journal.pone.0240376

[16] B. Kung, M. Chiang, G. Perera, M. Pritchard, and R. Stewart, "Unsupervised Machine Learning to Identify Depressive Subtypes," Healthcare Informatics Research, vol. 28, no. 3, pp. 256-266, Sep. 2022. https://doi.org/10.4258/hir.2022.28.3.256

[17] I. Twamley, M. Dempsey, and N. Keane, "An Open Dialogue-informed approach to mental health service delivery: experiences of service users and support networks," Psychiatric Rehabilitation Journal, vol. 43, no 4.

[18] Aafjes-van Doorn, K., Kamsteeg, C., Bate, J., & Aafjes, M. (2020). A scoping review of machine learning in psychotherapy research. Frontiers in Psychiatry, 11, 738466. https://doi.org/10.1080/10503307.2020.1808729

[19] Ngumimi, K. I., Kim, S.-H., Jhon, M., Yang, H.-J., & Pant, S. (2023). A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis. *Healthcare, 11*(3), 285. https://doi.org/10.3390/healthcare11030285