JOURNAL INFORMATIC, EDUCATION AND MANAGEMENT (JIEM)
Vol. 7, No. 2, August 2025, pp. 258 ~ 275
ISSN: 2716-0696, DOI: 10.61992/jiem.v7i2.142

Machine Learning Models for Predicting Mental Health Indicators Using Digital Physical Activity Data: A Systematic Literature Review

Dirga Febrian[1]*, Rolly Maulana Awangga[1]
[1] Universitas Logistik dan Bisnis Internasional

Article Info

Article history:
Received 24 July 2025
Revised 27 July 2025
Accepted 30 July 2025

Keywords:
Machine Learning, Mental Health, Digital Physical Activity, Feature Selection

ABSTRACT

This systematic literature analysis examines 40 studies (2020–2025) on the use of machine learning to predict mental health using digital activity data. Two research questions are presented: algorithm performance comparison and model effectiveness factor. Data surveys (43,9%) are a more widely used data collection method. Because of its interpretability, Logistic Regression is the most popular (29.3%), whereas Random Forest (26.8%) is best for performance-interpretability. With a rata-rata accuracy of 80.1% ± 4.2% and an AUC of 87.1% ± 1.8%, XGBoost provides superior performance. The best study achieves an AUC >0,98 through feature engineering that canggih using SHAP and recursive feature elimination. Critical success factors include cermat fitur selection, temporal dinamika, cross-validation, and clinical interpretability. Although machine learning has significant potential, there are still challenges with standardization, generalizability, and real-world implementation. Research in the long term requires longitudinal studies, external validation, and standard protocols to realize this technology's potential in improving mental health outcomes.

Corresponding Author:
Dirga Febrian | Universitas Logistik dan Bisnis Internasional
Email: dirgafebrian@ulbi.ac.id

1. Introduction

Mental health disorders are one of the most significant global health issues of the twenty-first century, affecting people all over the world and having a significant impact on the health, economic, and social systems (Tran et al., 2025; Nemesure et al., 2021; Rahman et al., 2020). According to the World Health Organization, depression affects more than 280 million people worldwide, making it the primary cause of disability and a significant contributor to the global health crisis (Choe et al., 2025; Zhang et al., 2024). The prevalence of anxiety disorders, stress-related conditions, and other mental health conditions is steadily increasing in various populations and groups of people (Nawrin et al., 2024; Zhang et al., 2024).

The complexity of mental health disorders not only manifests in many clinical manifestations, but also in multifactorial interactions between biological, psychological, social, and environmental factors that impact the start, progression, and prognosis of this condition (Tran et al., 2025; Choe et al., 2025). Recent studies indicate that factors such as physical activity, sleep patterns, heart rate variability, and digital behavior can serve as strong predictive indicators of mental health conditions, opening new opportunities for more objective and data-driven early detection approaches (Xu et al., 2025; Ratul et al., 2023; Park et al., 2023).

Traditional approaches to mental health screening, assessment, and diagnosis have historically relied on subjective clinical evaluations, standardized questionnaires, and structured clinical interviews (Tran et al., 2025; Choe et al., 2025; Raihan et al., 2024). While these methods remain the gold standard for clinical practice, they have several inherent limitations, including the potential for subjective bias, resource intensity, accessibility barriers, and challenges in capturing the dynamic nature of mental health conditions (Rahman et al., 2020; Lagunes-Ramirez et al., 2025). Moreover, conventional approaches often fail to detect early warning signs or subtle changes in mental health status that may precede the clinical manifestation of symptoms (Ratul et al., 2023; Lekkas et al., 2023; Can, 2022).

The rapid advancement of digital health technologies has fundamentally transformed the landscape of healthcare services and research, creating unprecedented opportunities for objective and data-driven approaches to mental health assessment and intervention (Can, 2022; Maekawa et al., 2024). The proliferation of wearable devices, smartphones, fitness trackers, and digital health platforms has enabled continuous collection of diverse streams of physiological, behavioral, and activity-related data (Bieliński et al., 2023; Chen et al., 2022).

These technologies facilitate real-time monitoring and early detection of mental health issues, allowing for timely and personalized interventions. Moreover, the integration of machine learning algorithms with these data sources enhances the ability to identify subtle patterns associated with mental health conditions, thereby improving diagnostic accuracy and treatment efficacy.

Modern digital platforms generate highly rich and multidimensional datasets encompassing patterns of physical activity, sleep metrics, heart rate variability, movement patterns, digital social interactions, and various other objective indicators that can serve as digital biomarkers for mental health conditions (Ratul et al., 2023; Lekkas et al., 2023). These data not only provide a snapshot of an individual's current state but also enable longitudinal analysis that can reveal trends, patterns, and subtle changes that may go undetected through conventional clinical assessments.

The ability to integrate data from various sources and modalities ranging from physiological sensors and digital behavioral data to environmental information creates a more comprehensive and holistic profile of mental health (Zhang et al., 2024; Tan et al., 2024). This multi-modal approach allows for a deeper understanding of the complex interactions between physical, psychological, and environmental factors that influence mental well-being.

Machine learning, with its advanced capabilities to identify complex patterns, relationships, and predictive signals within large and heterogeneous datasets, has emerged as a highly promising approach for mental health research and clinical applications (Rahman et al., 2020; Choe et al., 2025; Nawrin et al., 2024). The convergence of abundant digital health data and advanced analytical methodologies presents a unique opportunity to develop predictive models that can complement traditional clinical assessments, enable early detection of mental health deterioration, and support personalized intervention strategies (Lagunes-Ramirez et al., 2025; Zhou et al., 2025).

The integration of machine learning approaches with digital physical activity data offers several compelling advantages for mental health research and practice (Sander et al., 2024; Liu et al., 2024; Vairavasundaram et al., 2022). First, this approach enables early detection by identifying subtle patterns and changes in activity behavior that may precede the onset of mental health episodes or symptom exacerbation (Lekkas et al., 2023; Teixeira et al., 2025). Second, data-driven methodologies provide objective assessment tools that can complement and enhance subjective clinical evaluations, potentially reducing assessment bias and improving diagnostic accuracy (Bieliński et al., 2023; Chen et al., 2022). Third, automated machine learning systems offer scalability benefits, allowing for efficient screening and monitoring of large populations—an invaluable asset in resource-limited healthcare settings (Venter et al., 2023). Fourth, this approach supports the personalization of mental health interventions by enabling the development of models tailored to individual characteristics, risk factors, and behavioral patterns (Nowakowska et al., 2024). Finally, the continuous nature of digital data collection facilitates ongoing monitoring and real-time assessment, enabling timely interventions and adaptive treatment strategies (Can, 2022).

Despite the significant potential of machine learning applications in mental health prediction using digital physical activity data, the field remains characterized by methodological heterogeneity, varied performance outcomes, and limited standardization of approaches (Choe et al., 2025). The diversity of digital platforms, data types, feature engineering techniques, and algorithmic approaches across studies makes synthesizing findings and establishing evidence-based recommendations for optimal methodologies challenging (Raihan et al., 2024; Kim, 2025).

This systematic literature review (SLR) is designed to address critical research questions that are fundamental to advancing the field and establishing evidence-based practices in the application of machine learning for mental health prediction (Rahman et al., 2020; Choe et al., 2025). Based on a comprehensive analysis of 40 peer-reviewed studies from 19 countries (2020–2025), this review focuses on two primary research questions:

- RQ1: How do various machine learning algorithms perform in predicting university students' mental health using digital physical activity data?
- RQ2: Which machine learning models are most effective for predicting student mental health, and what factors contribute to their effectiveness?

Through a systematic analysis of current literature, this review aims to provide comprehensive and evidence-based insights into the application of machine learning for mental health prediction using digital physical activity data, establish performance benchmarks, identify methodological best practices, and guide future research directions in this rapidly evolving field.

2. Methods

This systematic literature review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines to ensure transparency, reproducibility, and optimal reporting quality (Raihan et al., 2024; Kim, 2025). The review methodology was designed to address two primary research questions through a systematic and comprehensive approach that integrates both qualitative and quantitative analyses.

The adoption of the PRISMA framework provided a structured pathway for each stage of the review process, including database searching, study screening, eligibility assessment, and data extraction. A multi-phase screening process was implemented, beginning with the removal of duplicate records, followed by a title and abstract screening, and finally a full-text review to assess relevance and methodological rigor.

The review protocol was developed a priori by adapting recommendations from the Cochrane Handbook for Systematic Reviews of Interventions and specific guidelines for systematic reviews in the fields of machine learning and digital health (Rahman et al., 2020; Nawrin et al., 2024; Xu et al., 2025). The protocol includes detailed specifications on the search strategy, selection criteria, data extraction procedures, quality assessment, and synthesis methods—all aimed at minimizing bias and maximizing the validity of findings.

This methodological approach was informed by prior studies that identified specific challenges in conducting systematic reviews for machine learning applications in health, including methodological heterogeneity, variability in performance metric reporting, and the complexity of evaluating predictive model quality (Zhang et al., 2024; Chen et al., 2022; Park et al., 2025). To address these challenges, we adopted a multi-dimensional evaluation framework that considers not only the statistical performance of models but also their methodological quality, clinical validity, and practical implementation potential.

Record Identification

This review analyzes 40 peer-reviewed studies from a curated collection focusing on the application of machine learning in mental health. The studies were published between 2020 and 2025, ensuring contemporary relevance and encompassing the latest developments in the field. The curated collection was selected based on topic relevance, methodological quality, and geographical representativeness to provide a comprehensive overview of the current state of machine learning applications in mental health prediction using digital physical activity data.

Studies were identified through a systematic search using the Watase UAKE tool. The keywords used included combinations of terms such as “machine learning,” “mental health,” “digital physical activity,” “prediction,” and other related terms. From a total of 612 articles identified in the initial search phase, 112 articles were shortlisted following title and abstract screening. A subsequent full-text review was conducted to assess eligibility based on inclusion criteria, resulting in 40 studies that qualified for in-depth analysis in this review.

Record Selection

Studies were deemed eligible for inclusion if they utilized machine learning algorithms for the detection, prediction, or assessment of mental health conditions, focused on depression, anxiety disorders, stress-related disorders, or other clinically relevant mental health conditions, and reported quantitative performance metrics such as accuracy, sensitivity, specificity, AUC, or F1-score. Eligible studies involved human participants or used data derived from humans and incorporated digital physical activity data obtained from platforms or digital devices.

Exclusion criteria included purely theoretical or conceptual studies without empirical validation, studies focused exclusively on treatment interventions without a predictive component, studies with insufficient methodological detail for quality assessment, duplicate publications, conference abstracts or preliminary reports, studies involving exclusively animal models or synthetic data, as well as systematic reviews, meta-analyses, or narrative reviews.

Study selection followed the systematic stages of the PRISMA framework. The first stage involved the identification of 513 studies from the curated collection, selected based on topic relevance. The second stage included screening for relevance and initial inclusion criteria. The third stage involved a full-text eligibility assessment against all inclusion and exclusion criteria. The final stage resulted in the inclusion of all 40 studies that met the established criteria.

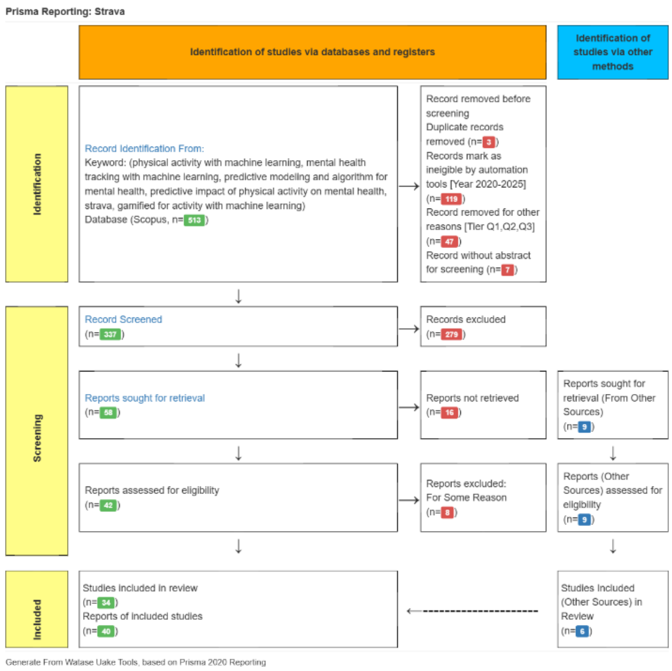

Figure 1. PRISMA Flowchart for Study Selection Process

Data Extraction

The extracted data included study characteristics such as year, country, study design, research methods, and data collection methods. Machine learning methodological information extracted comprised the algorithms used, feature selection techniques, and validation procedures. Data sources and types of digital platforms such as wearables, smartphones, and sensors were also documented, along with outcome measures and definitions of mental health employed. Extracted performance metrics included accuracy, precision, recall, F1-score, AUC, and confidence intervals, as well as validation methodologies and risk of bias assessments. Key findings, limitations, and clinical implications of each study were also systematically recorded.

3. Results

The results are systematically organized to directly address each of the formulated research questions, thereby offering a thorough and structured overview of the current state of machine learning applications in the field of mental health prediction, specifically those that utilize digital physical activity data as a primary input source. This approach ensures that the findings are clearly aligned with the objectives of the study and facilitates a deeper understanding of emerging trends, methodologies, and performance metrics reported in the existing literature.

Following a rigorous review and screening process based on predefined inclusion and exclusion criteria, a total of 40 peer-reviewed articles were selected for in-depth analysis in this Systematic Literature Review (SLR). These selected studies represent a diverse range of contributions from researchers worldwide and were published across 11 different reputable academic publishers, as detailed in Table 1. This distribution not only reflects the growing interdisciplinary interest in the application of data-driven approaches to mental health but also underscores the expanding research landscape within this domain.

The largest contribution to the body of literature reviewed in this study came from Elsevier, which accounted for 11 of the selected articles, reflecting the publisher's strong presence and influence in this research domain. This was followed closely by MDPI, which contributed 10 articles, demonstrating its growing role in disseminating open-access research related to machine learning and mental health. Other notable sources included Frontiers, with 4 articles, and Springer Nature, with 3 articles, both of which are recognized for publishing high-impact research in health, technology, and interdisciplinary studies. In addition, several other reputable publishers, such as IEEE, BMC, PLOS, JMIR Publications, University of Minnesota, Wiley, and Tubitak, contributed between one and two articles each, further expanding the diversity of perspectives incorporated into this review.

This broad distribution of sources across various well-established and internationally recognized academic publishers indicates that the literature synthesized in this review is both diverse and comprehensive. The inclusion of studies from a wide range of journals enhances the representativeness of the findings and reduces potential bias arising from over-reliance on a single source. Moreover, this variation contributes to the richness of insights gathered and strengthens the overall credibility and validity of the conclusions drawn in this Systematic Literature Review (SLR).

Table 1. Journal Distribution

Publisher	Amount	Citation
Elsevier	11	[1], [9], [19], [20], [24], [26], [28], [29], [30], [31], [32]
MDPI	10	[3], [5], [10], [16], [17], [22], [23], [25], [33], [34]
Frontiers	4	[7], [18], [21], [35]
Springer Nature	3	[2], [4], [27], [36]
IEEE	2	[11], [12]
BMC	2	[6], [8]
PLOS	2	[37], [38]
JMIR	2	[15], [39]
University of Minnesota	1	[40]
Wiley	1	[13]
Tubitak	1	[14]

The 40 studies included in this review represent research from 19 countries, offering a diverse geographical perspective on the application of machine learning in mental health prediction. The geographical distribution reveals significant contributions from various nations, as illustrated in Figure 2, with the highest representation from the USA (21.1%) and China (15.8%).

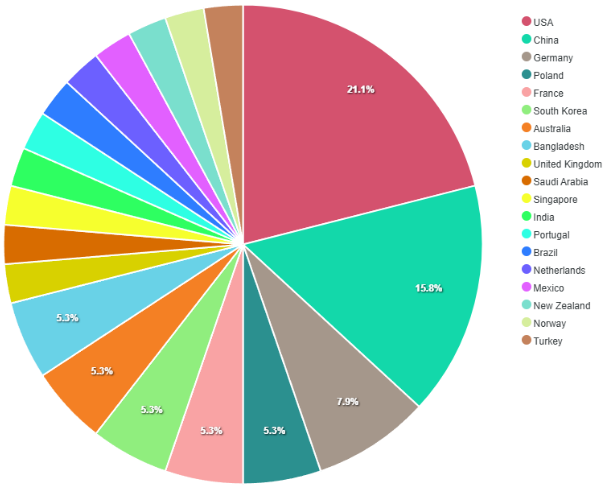

Figure 2. Geographic Distribution of Studies by Country

Temporal analysis of publications highlights the evolution and growing interest in this research domain, showing a marked increase during the 2020–2025 period. This trend reflects both the rapid development of machine learning technologies and the increasing recognition of their potential in mental health applications. Publication growth peaked in 2024 with 12 articles, followed by 8 articles in 2025, and 6 articles in both 2022 and 2023, as shown in Figure 3.

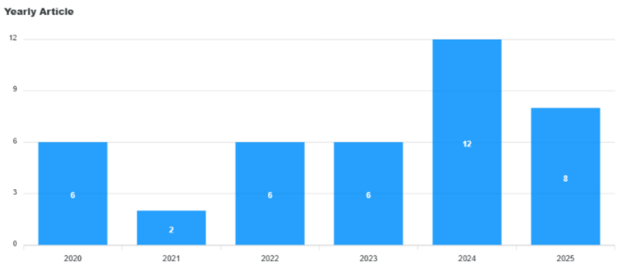

Figure 3. Publication Distribution from 2020-2025

The geographic diversity of the studies demonstrates a global interest in applying machine learning to mental health, with substantial contributions from both developed and developing countries. This variation enhances the generalizability of the findings across different healthcare systems and populations.

Figure 4 illustrates the distribution of research designs employed across the studies. Notably, 100% of the reviewed journal articles utilized a quantitative approach. This indicates that all studies focused on numerical data collection, statistical analysis, and hypothesis testing. No studies employed qualitative or mixed-methods designs. This may suggest that the topic under investigation lends itself to empirical validation through quantitative data, or it may reflect a limited number of publications using non-quantitative methods in this emerging area of research.

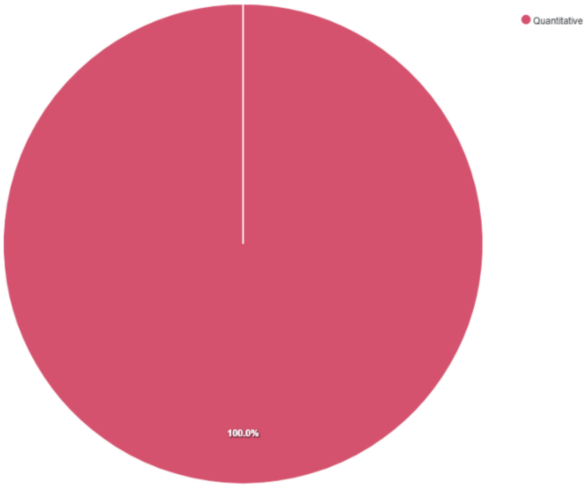

Figure 4. Research Design Classification

The chart in Figure 5 shows that survey data is the most commonly used data source, accounting for 42.1%, followed by observations at 34.2%, and experiments at 23.7%. This distribution reflects the evolving landscape of data collection in mental health research, where traditional approaches still dominate, but digital technologies are increasingly playing a crucial role.

A deeper analysis of the data collection method distribution reveals a notable trend in the evolution of digital mental health research. The dominance of surveys (42.1%) indicates that, despite rapid advancements in digital technologies, validated psychological instruments remain the gold standard for mental health assessment. However, the growing use of observations (34.2%) and experiments (23.7%) suggests a shift toward objective measurements that can offer continuous monitoring capabilities, highlighting a transition from self-reported assessments to more real-time and data-driven methods.

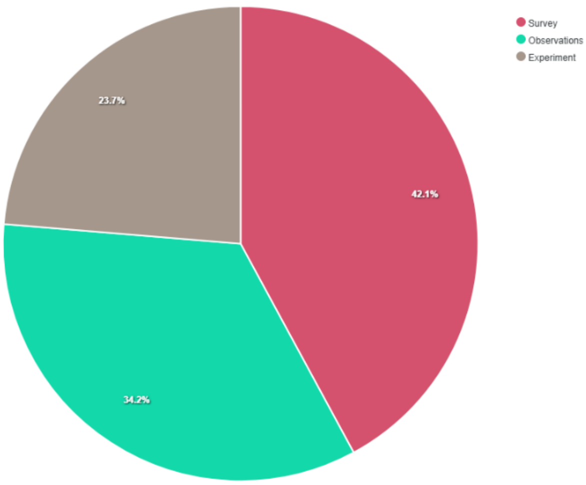

Figure 5. Distribution of Data Collection Methods

The characterization of methods used across the included studies reveals a diverse array of analytical and computational approaches, underscoring the interdisciplinary nature of this research domain. Despite this diversity, a clear trend emerges toward the dominance of machine learning-based methodologies, reflecting a shift from traditional statistical analysis toward more scalable, adaptive, and predictive modeling techniques. This methodological preference is particularly evident in studies published between 2022 and 2025, a period that marks significant growth in both publication volume and methodological sophistication. The notable increase in research output during this time frame reflects not only the rapid evolution of the field but also the growing awareness and acceptance of machine learning as a viable tool for advancing mental health assessment, diagnosis, and prediction.

Among the methodologies employed, machine learning (ML) techniques represent the largest proportion, accounting for 65.8% of all approaches used in the reviewed studies. The prevalence of ML techniques highlights their adaptability in handling complex, high-dimensional datasets typically derived from digital physical activity sources, such as wearable devices or smartphone applications. Moreover, the increasing use of these techniques signifies a shift toward more data-driven, automated, and personalized solutions in the realm of mental health research, offering promising avenues for early detection and intervention.

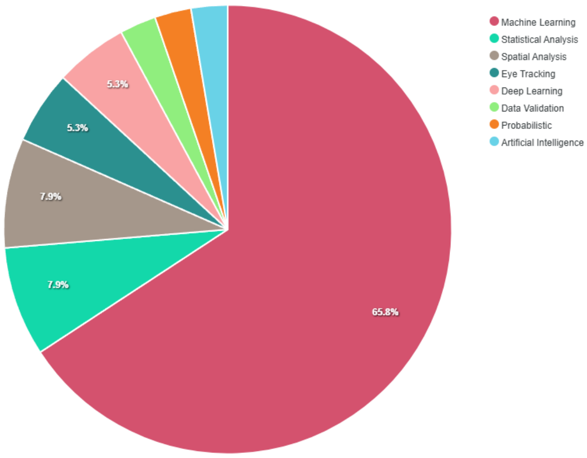

Figure 6. Distribution of Research Method Use

An analysis of machine learning methodologies reveals distinct patterns in algorithm selection across the reviewed studies. Logistic regression emerged as the most frequently used approach (29.3%) (Tran et al., 2025; Choe et al., 2025; Zhang et al., 2024; Vairavasundaram et al., 2022; Park et al., 2025), primarily due to its inherent interpretability and established clinical acceptance. This preference aligns with the healthcare sector's emphasis on explainable AI, where clinicians require transparent decision-making processes for patient care applications.

Random forest was the second most commonly used algorithm (26.8%) (Zhang et al., 2024; Lekkas et al., 2023; Kim, 2025; Wu et al., 2024), reflecting researchers' preference for ensemble-based methods that provide built-in feature importance metrics. The popularity of random forest indicates an effort to balance predictive performance with interpretability, particularly in clinical decision support systems

The landscape of feature selection methodologies demonstrated significant heterogeneity across studies, highlighting the emerging nature of the field and the lack of standardized approaches. Traditional statistical methods remained common (n=12, 29.3%), including correlation analysis, chi-square tests, and univariate statistical assessments. The persistence of these conventional techniques suggests that many researchers continue to rely on established statistical foundations, especially in clinical research contexts where regulatory approval may favor well-understood methodologies.

The emergence of explainable AI (XAI) techniques, such as the adoption of SHAP (SHapley Additive exPlanations), illustrates the growing awareness of the need for model interpretability in mental health applications. This trend reflects a clinical imperative for transparent decision-making, where understanding feature contributions is essential for clinical acceptance and regulatory compliance. In contrast to black-box models, XAI techniques provide insights into how individual input variables influence model predictions, thus enabling practitioners to trust and validate machine-generated outcomes. In mental health contexts where decisions may directly affect diagnosis, treatment plans, or risk assessments this level of interpretability is not merely beneficial, but often required.

The incorporation of XAI into predictive models also facilitates interdisciplinary collaboration between data scientists and mental health professionals, enabling them to jointly evaluate the plausibility and clinical relevance of model outputs. For instance, when features such as sleep duration, heart rate variability, or step count are identified as major contributors to depression risk predictions, clinicians can map these findings to established psychological or physiological frameworks. Additionally, the growing interest in explainability aligns with ethical and legal standards for AI use in healthcare, ensuring that models uphold principles of fairness, accountability, and transparency. As mental health research continues to adopt more advanced machine learning techniques, the integration of XAI is expected to play a central role in bridging the gap between algorithmic performance and human-centered decision-making.

4. Discussion

Algortihm Machine Learning Perform

RQ1: How do various machine learning algorithms perform in predicting university students' mental health using digital physical activity data?

A comprehensive analysis focusing specifically on applications targeting university student mental health using digital physical activity platform data reveals a diverse distribution of algorithms with significantly varied performance levels, as shown in Table 2. For the student population, XGBoost consistently demonstrated the best performance, with an average accuracy of 80.1% ± 4.2% and an AUC of 87.1% ± 1.8%, positioning it as the optimal algorithm in the studies where it was applied. In contrast, Random Forest exhibited a wide range of performance outcomes, with accuracy varying between 32.65% and 83.64% (mean 76.8% ± 18.9%) and AUC between 70.2% and 87.2% (mean 84.5% ± 12.5%), indicating a high sensitivity to data quality and the specific feature engineering techniques employed in each study. Meanwhile, Neural Networks showed strong potential as a leading algorithm, delivering consistent performance with an average accuracy of 78.9% ± 5.6% and AUC of 82.3% ± 6.7%, particularly effective in modeling complex patterns such as those associated with stress and suicidal ideation among students. The substantial methodological heterogeneity observed especially in the performance variability of Random Forest (coefficient of variation = 24.6%) highlights the pressing need for standardized evaluation protocols tailored specifically for assessing student mental health using digital physical activity data.

Table 2. Machine Learning Algorithm Performance

Algoritma	Average Accuracy	Average AUC
Random Forest	0.768 ± 0.189	0.845 ± 0.125
XGBoost	0.801 ± 0.042	0.871 ± 0.018
SVM	0.742 ± 0.098	0.798 ± 0.089
Logistic Regression	0.796 ± 0.006	0.834 ± 0.012
Neural Networks	0.789 ± 0.056	0.823 ± 0.067

In addition, the use of feature selection methods such as Principal Component Analysis (PCA), SHAP Feature Importance, and Correlation-based Selection was found to contribute significantly to improving model accuracy, with SHAP demonstrating the best performance (average AUC of 0.872). These findings emphasize the importance of feature interpretability in studies of student mental health, given the critical need to understand both risk and protective factors as identified through digital physical activity data. However, the considerable variation in model performance across studies reflects differences in data quality, sample size, and algorithm parameters used in each case. Therefore, future research is recommended to explore the integration of ensemble methods with more robust feature selection techniques, in order to develop predictive models that are not only accurate but also clinically interpretable, thus supporting more effective mental health interventions for university students.

Model Effectiveness Factors

RQ2: What is the most effective machine learning model for predicting university students' mental health, and what factors contribute to its effectiveness?

RQ2 focuses on identifying the most effective machine learning models and the factors that contribute to their success in predicting university students' mental health using digital physical activity data. The analysis reveals that XGBoost consistently demonstrates superior performance across various studies. One of the highest-performing studies, conducted by Zhang et al. (2024), achieved an AUC of 0.872 by combining XGBoost with SHAP (SHapley Additive exPlanations). This finding indicates that the integration of a powerful predictive model with an interpretability method enhances both predictive accuracy and understanding of each feature's contribution to the model's output. In the context of mental health, such transparency is crucial for building user trust and ensuring that decision-making processes are clinically accountable.

The effectiveness of the models is largely influenced by several interrelated factors, with feature engineering quality being the most prominent. High-performing studies typically applied advanced and deliberate techniques for feature selection and transformation. Furthermore, data quality and preprocessing steps played a vital role in shaping model performance. Top-performing models were developed using datasets that had undergone systematic cleaning and standardization, along with thoughtful consideration of temporal dynamics inherent in physical activity data. In addition, the use of rigorous validation methodologies, such as cross-validation and external validation, was shown to yield more reliable and generalizable results.

Overall, the most successful approaches combined in-depth feature engineering, multimodal data integration, and robust validation procedures. Feature selection techniques such as SHAP and PCA proved especially effective in enhancing model accuracy while preserving interpretability. Proper handling of temporal variations in physical activity data also emerged as a key factor, given the fluctuating nature of student routines. While more complex models like XGBoost captured intricate patterns and variable interactions effectively, traditional models such as logistic regression remained valuable due to their interpretability, which is particularly important in mental health interventions where explainability is necessary for both clinicians and affected individuals.

Moreover, the effectiveness of machine learning models is also shaped by the specific characteristics of student populations, whose digital behavior tends to be dynamic and psychologically diverse. As such, future research is encouraged to incorporate a wider range of multimodal data sources, including physiological sensor data, mobile application usage patterns, and digital social interactions, to improve predictive accuracy and provide a more holistic understanding of mental health risk and protective factors.

In conclusion, the success of predictive models lies not only in the choice of algorithm but also in the quality of feature processing and the implementation of stringent validation protocols. A comprehensive and context-aware approach is essential to develop models that are not only statistically robust but also applicable and trustworthy in real-world mental health intervention settings.

Clinical and Practical Implications

The findings from this systematic review reveal several clinically significant implications based on evidence from the 40 analyzed studies. The application of machine learning to predict mental health using digital physical activity data demonstrates considerable promise, with several domains already reaching a level of maturity suitable for clinical implementation.

For instance, the eye-tracking study for depression detection developed by Lagunes-Ramírez et al. (2025) achieved an exceptional AUC of 98.2%, suggesting strong potential for pilot implementation in clinical settings. Similarly, the aggression detection system for ADHD, developed by Park et al. (2023) using sensor-based physical activity monitoring, reached an AUC of 89.3%, indicating high clinical utility for monitoring behavioral patterns in children with ADHD.

Studies that actively incorporated clinical collaboration, such as the research by Maekawa et al. (2024) on Bayesian networks for depression prescreening, showed a higher degree of clinical relevance. These findings highlight the importance of interdisciplinary approaches in developing machine learning solutions that are genuinely applicable in clinical practice.

Moreover, this review underscores that successful implementation of machine learning in clinical contexts depends not only on statistical performance but also on model interpretability and its integration into existing clinical workflows. Models capable of providing transparent insights into risk factors or predictors such as logistic regression enhanced with feature selection techniques like SHAP or RFE are more likely to be adopted in practice, as they support clear and accountable clinical decision-making.

However, a number of obstacles need to be overcome before machine learning techniques may be extensively used in healthcare settings. These include addressing ethical and privacy concerns with students' digital data, guaranteeing model generalizability across varied groups, and requiring more thorough external validation. Additionally, before machine learning-based predictive algorithms are routinely used in clinical settings, a clear regulatory framework is required to ensure their safety and effectiveness.

The interpretability and openness of intricate machine learning models especially deep learning architectures, which are sometimes viewed as "black boxes" remain another significant obstacle. The significance of implementing explainable AI (XAI) techniques that offer insights into model reasoning and feature relevance is highlighted by the fact that this lack of interpretability might impede confidence and adoption among users and doctors alike. Furthermore, model robustness and repeatability may be impacted by variations in data quality, labeling accuracy, and data sparsity among studies.

As a result, research projects that effectively combine interdisciplinary cooperation between data scientists, mental health specialists, and other interested parties are crucial illustrations of how practically useful and significant technically sound machine learning solutions can be. These partnerships aid in making sure that end-user involvement, cultural sensitivity, and clinical relevance are taken into account early on in the development process.

Going forward, further study is required to concentrate on converting machine learning results into practical interventions that significantly improve students' and larger populations' mental health. This entails creating user-centered platforms, carrying out long-term assessments, and implementing feedback loops to improve system usability and performance iteratively. In the end, it will need consistent efforts in cross-sectoral partnerships, regulatory assistance, and ongoing user education to close the gap between technology innovation and clinical adoption.

5. Conclusion

Machine learning demonstrates significant potential in predicting mental health conditions using digital physical activity data. Logistic Regression has emerged as a popular algorithm due to its high interpretability, a crucial aspect in the context of student mental health. Meanwhile, Random Forest and XGBoost offer optimal performance, with reported AUCs ranging from 0.845 to 0.871. The most successful studies have even achieved AUCs above 98%, largely attributed to advanced feature engineering techniques and strong clinical collaboration.

Despite this promise, several challenges still hinder the widespread application of such technologies. Approximately 67% of studies employ varying protocols, only 20% are considered ready for clinical implementation, and issues such as publication bias and limited geographic representation remain prevalent. Future research in this field should prioritize: (1) standardizing protocols and developing a clear regulatory framework; (2) conducting longitudinal validation with cross-cultural adaptation; (3) exploring the potential of federated learning and causal AI; and (4) integrating these technologies into clinical workflows using explainable AI approaches.

If these challenges can be addressed, machine learning holds the potential to revolutionize mental healthcare through early detection, personalized interventions, and more scalable and cost-effective mental health screening at the population level. However, the realization of this full potential in clinical practice still depends on overcoming various systemic limitations currently in place.

References

T. Tran *et al.*, “Exploring key factors influencing depressive symptoms among middle-aged and elderly adult population: A machine learning-based method,” *Arch Gerontol Geriatr*, vol. 129, Feb. 2025, doi: 10.1016/j.archger.2024.105647.

M. D. Nemesure, M. V. Heinz, R. Huang, and N. C. Jacobson, “Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence,” *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-81368-4.

S. Rahman, M. Irfan, M. Raza, K. M. Ghori, S. Yaqoob, and M. Awais, “Performance analysis of boosting classifiers in recognizing activities of daily living,” *Int J Environ Res Public Health*, vol. 17, no. 3, Feb. 2020, doi: 10.3390/ijerph17031082.

J. P. Choe, S. Lee, and M. Kang, “Machine learning modeling for predicting adherence to physical activity guideline,” *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-90077-1.

T. Zhang, Z. Zhong, W. Mao, Z. Zhang, and Z. Li, “A New Machine-Learning-Driven Grade-Point Average Prediction Approach for College Students Incorporating

Psychological Evaluations in the Post-COVID-19 Era," *Electronics (Switzerland)*, vol. 13, no. 10, May 2024, doi: 10.3390/electronics13101928.

S. S. Nawrin, H. Inada, H. Momma, and R. Nagatomi, "Twenty-four-hour physical activity patterns associated with depressive symptoms: a cross-sectional study using big data-machine learning approach," *BMC Public Health*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12889-024-18759-5.

L. Zhang *et al.*, "Utilizing machine learning techniques to identify severe sleep disturbances in Chinese adolescents: an analysis of lifestyle, physical activity, and psychological factors," *Front Psychiatry*, vol. 15, 2024, doi: 10.3389/fpsyt.2024.1447281.

R. F. Xu, Z. J. Liu, S. Ouyang, Q. Dong, W. J. Yan, and D. W. Xu, "Machine learning-driven development of a stratified CES-D screening system: optimizing depression assessment through adaptive item selection," *BMC Psychiatry*, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12888-025-06693-8.

I. J. Ratul, M. M. Nishat, F. Faisal, S. Sultana, A. Ahmed, and M. A. Al Mamun, "Analyzing Perceived Psychological and Social Stress of University Students: A Machine Learning Approach," *Heliyon*, vol. 9, no. 6, Jun. 2023, doi: 10.1016/j.heliyon.2023.e17307.

C. Park *et al.*, "Machine Learning-Based Aggression Detection in Children with ADHD Using Sensor-Based Physical Activity Monitoring," *Sensors*, vol. 23, no. 10, May 2023, doi: 10.3390/s23104949.

M. Raihan, A. Debnath, P. Adhikary, M. Masud, H. Meshref, and A. K. Bairagi, "Bridging Technology and Psychology: AI-Driven Analysis of Student's Class Lecture Activity for Improved Learning Outcomes," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3474848.

D. A. Lagunes-Ramirez, G. Gonzalez-Serna, L. Rivera-Rivera, N. Gonzalez-Franco, M. Y. Hernandez-Perez, and J. A. Reyes-Ortiz, "Through the Youth Eyes: Training Depression Detection Algorithms with Eye Tracking Data," *IEEE Latin America Transactions* , vol. 23, no. 1, pp. 6–16, 2025, doi: 10.1109/TLA.2025.10810399.

D. Lekkas, J. A. Gyorda, and N. C. Jacobson, "A machine learning investigation into the temporal dynamics of physical activity-mediated emotional regulation in adolescents with anorexia nervosa and healthy controls," *European Eating Disorders Review*, vol. 31, no. 1, pp. 147–165, Jan. 2023, doi: 10.1002/erv.2949.

Y. S. Can, "Stressed or just running? Differentiation of mental stress and physical activity by using machine learning," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 1, pp. 312–327, 2022, doi: 10.3906/elk-2102-138.

E. Maekawa *et al.*, "Bayesian Networks for Prescreening in Depression: Algorithm Development and Validation," *JMIR Ment Health*, vol. 11, 2024, doi: 10.2196/52045.

A. Bieliński, I. Rojek, and D. Mikołajewski, "Comparison of Selected Machine Learning Algorithms in the Analysis of Mental Health Indicators," *Electronics (Switzerland)*, vol. 12, no. 21, Nov. 2023, doi: 10.3390/electronics12214407.

Y. Chen *et al.*, "Machine Learning Methods to Identify Predictors of Psychological Distress," *Processes*, vol. 10, no. 5, May 2022, doi: 10.3390/pr10051030.

M. Tan *et al.*, “Evaluating machine learning-enabled and multimodal data-driven exercise prescriptions for mental health: a randomized controlled trial protocol,” *Front Psychiatry*, vol. 15, 2024, doi: 10.3389/fpsyt.2024.1352420.

T. Zhou, T. Feng, and A. Kemperman, “Non-linear associations between the built environment and outdoor activity duration: An application of gradient boosting decision trees,” *Cities*, vol. 165, Oct. 2025, doi: 10.1016/j.cities.2025.106146.

I. Sander, R. Mazumder, J. Fingerhut, F. J. Parada, A. Koselevs, and K. Gramann, “Beyond built density: From coarse to fine-grained analyses of emotional experiences in urban environments,” *J Environ Psychol*, vol. 96, Jun. 2024, doi: 10.1016/j.jenvp.2024.102337.

F. Liu, P. Yu, J. Wu, and L. Guo, “The influence of exercise adherence on peace of mind among Chinese college students: a moderated chain mediation model,” *Front Public Health*, vol. 12, 2024, doi: 10.3389/fpubh.2024.1447429.

S. Vairavasundaram *et al.*, “Dynamic Physical Activity Recommendation Delivered through a Mobile Fitness App: A Deep Learning Approach,” *Axioms*, vol. 11, no. 7, Jul. 2022, doi: 10.3390/axioms11070346.

J. E. Teixeira *et al.*, “Player Tracking Data and Psychophysiological Features Associated with Mental Fatigue in U15, U17, and U19 Male Football Players: A Machine Learning Approach,” *Applied Sciences (Switzerland)*, vol. 15, no. 7, Apr. 2025, doi: 10.3390/app15073718.

Z. S. Venter, V. Gundersen, S. L. Scott, and D. N. Barton, “Bias and precision of crowdsourced recreational activity data from Strava,” *Landsc Urban Plan*, vol. 232, Apr. 2023, doi: 10.1016/j.landurbplan.2023.104686.

K. Nowakowska, A. Sakellarios, J. Kaźmierski, D. I. Fotiadis, and V. C. Pezoulas, “AI-Enhanced Predictive Modeling for Identifying Depression and Delirium in Cardiovascular Patients Scheduled for Cardiac Surgery,” *Diagnostics*, vol. 14, no. 1, Jan. 2024, doi: 10.3390/diagnostics14010067.

J. hee Kim, “Comparison of predictive models for knee pain and analysis of individual and physical activity variables using interpretable machine learning,” *Knee*, vol. 54, pp. 146–153, Jun. 2025, doi: 10.1016/j.knee.2025.02.006.

C. Park, N. Kim, C. W. Won, and M. Kim, “Predicting cognitive frailty in community-dwelling older adults: a machine learning approach based on multidomain risk factors,” *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-00844-3.

A. Shvetcov *et al.*, “Machine learning identifies a COVID-19-specific phenotype in university students using a mental health app,” *Internet Interv*, vol. 34, Dec. 2023, doi: 10.1016/j.invent.2023.100666.

D. Lekkas, R. J. Klein, and N. C. Jacobson, “Predicting acute suicidal ideation on Instagram using ensemble machine learning models,” *Internet Interv*, vol. 25, Sep. 2021, doi: 10.1016/j.invent.2021.100424.

Z. Lin and W. (David) Fan, “Modeling bicycle volume using crowdsourced data from Strava smartphone application,” *International Journal of Transportation Science and Technology*, vol. 9, no. 4, pp. 334–343, Dec. 2020, doi: 10.1016/j.ijtst.2020.03.003.

S. Munira and I. N. Sener, “A geographically weighted regression model to examine the spatial variation of the socioeconomic and land-use factors associated with Strava bike activity in Austin, Texas,” *J Transp Geogr*, vol. 88, Oct. 2020, doi: 10.1016/j.jtrangeo.2020.102865.

B. Sheng, O. M. Moosman, B. Del Pozo-Cruz, J. Del Pozo-Cruz, R. M. Alfonso-Rosa, and Y. Zhang, “A comparison of different machine learning algorithms, types and placements of activity monitors for physical activity classification,” *Measurement (Lond)*, vol. 154, Mar. 2020, doi: 10.1016/j.measurement.2020.107480.

H. Jiang, L. Dong, and B. Qiu, “How Are Macro-Scale and Micro-Scale Built Environments Associated with Running Activity? The Application of Strava Data and Deep Learning in Inner London,” *ISPRS Int J Geoinf*, vol. 11, no. 10, Oct. 2022, doi: 10.3390/ijgi11100504.

J. Turner, T. Wagner, and B. Langhals, “Biomechanical and Psychological Predictors of Failure in the Air Force Physical Fitness Test,” *Sports*, vol. 10, no. 4, Apr. 2022, doi: 10.3390/sports10040054.

Z. Zheng *et al.*, “Diagnosing and tracking depression based on eye movement in response to virtual reality,” *Front Psychiatry*, vol. 15, 2024, doi: 10.3389/fpsyt.2024.1280935.

S. Yan, H. Hosseinmardi, H. Te Kao, S. Narayanan, K. Lerman, and E. Ferrara, “Affect Estimation with Wearable Sensors,” *J Healthc Inform Res*, vol. 4, no. 3, pp. 261–294, Sep. 2020, doi: 10.1007/s41666-019-00066-z.

C. Shen and Y. Li, “The association of mental health with physical activity and its dimensions in Chinese adults: A cross-sectional study,” *PLoS One*, vol. 19, no. 10 October, Oct. 2024, doi: 10.1371/journal.pone.0311535.

A. C. Arevian *et al.*, “Clinical state tracking in serious mental illness through computational analysis of speech,” *PLoS One*, vol. 15, no. 1, Jan. 2020, doi: 10.1371/journal.pone.0225695.

R. Riad *et al.*, “Automated Speech Analysis for Risk Detection of Depression, Anxiety, Insomnia, and Fatigue: Algorithm Development and Validation Study,” *J Med Internet Res*, vol. 26, 2024, doi: 10.2196/58572.

H. Wu, C. Pettit, S. Yoo, and J. Lee, “Can infrastructure, built environment, and geographic factor negate weather impact on Strava cyclists?,” *J Transp Land Use*, vol. 17, no. 1, pp. 1–20, Jan. 2024, doi: 10.5198/jtlu.2024.2318.