

Review

Passive Sensing for Mental Health Monitoring Using Machine Learning With Wearables and Smartphones: Scoping Review

ShiYing Shen^{1*}, BSc; Wenhao Qi^{1*}, BSc; Jianwen Zeng¹, BSc; Sixie Li¹, BSc; Xin Liu¹, BSc; Xiaohong Zhu¹, BSc; Chaoqun Dong¹, BSc; Bin Wang¹, BSc; Yankai Shi¹, BSc; Jiani Yao¹, BSc; Bingsheng Wang¹, BSc; Xiajing Lou¹, BSc; Simin Gu², BSc; Pan Li³, MSc; Jinghua Wang⁴, PhD; Guowei Jiang⁵, MSc; Shihua Cao^{1,6}, PhD

¹School of Nursing, Hangzhou Normal University, Hangzhou, China

²School of Nursing, Zhejiang Chinese Medical University, Hangzhou, China

³Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

⁴Department of Neurology, The Affiliated Hospital of Hangzhou Normal University, Hangzhou, China

⁵Department of Psychiatry and Neuropsychology and Alzheimer Center Limburg, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Maastricht, The Netherlands

⁶Key Engineering Research Center of Mobile Health Management System, Hangzhou, China

*these authors contributed equally

Corresponding Author:

Shihua Cao, PhD

School of Nursing

Hangzhou Normal University

Yuhangtang Road 2318

Hangzhou, 311121

China

Phone: 86 13777861361

Email: cs@hznu.edu.cn

Abstract

Background: Mental health issues have become a significant global public health challenge. Traditional assessments rely on subjective methods with limited ecological validity. Passive sensing via wearable devices and smartphones, combined with machine learning (ML), enables objective, continuous, and noninvasive mental health monitoring.

Objective: This study aimed to provide a comprehensive review of the current state of passive sensing-based and ML technologies for mental health monitoring. We summarized the technical approaches, revealed the association patterns between behavioral features and mental disorders, and explored potential directions for future advancements.

Methods: This scoping review adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines and was prospectively registered on the Open Science Framework. We systematically searched 7 databases (Web of Science, PubMed, IEEE Xplore, Embase, PsycINFO, Scopus, and ACM Digital Library) for studies published between January 2015 and February 2025. We included 42 peer-reviewed studies that used passive sensing from wearables or smartphones with ML to monitor clinically diagnosed mental disorders, such as depression and anxiety. Data were synthesized across technical dimensions (data collection, preprocessing, feature engineering, and ML models) and clinical associations, with behavioral features categorized into 8 domains.

Results: The 42 included studies were predominantly cohort designs (23/42, 55%), with a median sample size of 60.5 (IQR 54–99). Most studies focused on depression (23/42, 55%) and anxiety (9/42, 21%) using primarily wrist-worn devices (32/42, 76%) collecting heart rate (28/42, 67%), movement index (25/42, 60%), and step count (17/42, 40%) as key biomarkers. Deep learning models (eg, convolutional neural networks and long short-term memory) showed high accuracy, while traditional ML (eg, random forest) remained prevalent due to better interpretability. We identified critical limitations, including small samples (32/42, 76% with N<100), short monitoring periods (19/42, 45% <7 days), scarce external validation (1/42, 2%), and limited reporting on data anonymization (6/42, 14%).

Conclusions: While passive sensing and ML demonstrate promising accuracy (eg, convolutional neural network–long short-term memory achieving 92.16% in anxiety detection), the evidence remains constrained by three key limitations: (1) methodological heterogeneity (32/42, 76% single-device studies; 19/42, 45% with <7-day monitoring), (2) high risk of bias from small samples

(median 60.5, IQR 54-99 participants) and scarce external validation (1/42, 2%), and (3) ethical gaps (only 6/42, 14% addressing anonymization). These findings underscore the technology's potential to transform mental health care through objective, continuous monitoring—particularly for depression (heart rate and step count biomarkers) and anxiety (sleep and social interaction patterns). However, clinical translation requires standardized protocols, larger longitudinal studies (≥ 3 months), and ethical frameworks for data privacy. Future work should prioritize multimodal sensor fusion and explainable artificial intelligence to bridge the gap between technical performance and clinical deployability.

(*J Med Internet Res* 2025;27:e77066) doi: [10.2196/77066](https://doi.org/10.2196/77066)

KEYWORDS

passive sensing; mental health monitoring; mental disorders; machine learning; digital biomarkers; wearable devices; scoping review; artificial intelligence; AI

Introduction

Background

Mental health issues have become a significant global public health challenge, affecting the quality of life and social functioning of hundreds of millions of people [1]. According to the World Health Organization, there are 322 million people with depression and 264 million people with anxiety worldwide, yet a large number of individuals still fail to receive timely diagnosis and intervention [2,3]. Traditional mental health assessments mainly rely on clinical interviews [4] and self-report scales [5], which, although widely used, have limitations such as high subjectivity [6], long assessment periods [7], and insufficient ecological validity [8], making it difficult to capture dynamic psychological changes in daily life [9]. In addition, many patients seek help only when symptoms become severe, hindering early intervention [10]. Therefore, there is an urgent need for an objective, continuous, and noninvasive mental health monitoring method to address the limitations of traditional approaches.

With the proliferation of smartphones, wearable devices, and other digital health technologies, passive sensing has emerged as a promising tool for mental health monitoring [11,12]. This technology uses embedded sensors such as accelerometers, GPS, and microphones to continuously and unobtrusively collect data on users' behavior, physiology, and environment without requiring active participation [13]. These data reflect an individual's daily activity patterns, social behaviors, voice features, and sleep quality, indirectly revealing their psychological state [14]. For example, GPS trajectories can be used to assess the level of social activity [14], accelerometer data can be used to analyze physical activity patterns [15], and voice analysis may detect emotional changes [16]. Compared to traditional methods, passive sensing offers significant advantages in terms of objectivity, continuity, and low intervention [12], showing great potential for monitoring and early warning of various mental disorders, including depression [17], anxiety [18], and bipolar disorder [19].

However, the data generated through passive sensing are typically high-dimensional, noisy, and unstructured, making it challenging to process using traditional data analysis methods [20]. The introduction of machine learning (ML) techniques provides a solution to this challenge. ML can automatically extract features from vast and complex sensor data and establish a mapping relationship between data and mental health status

[11]. For instance, supervised learning algorithms (eg, support vector machine [SVM] and random forest) can be trained on labeled data to classify depression or anxiety levels [21,22], unsupervised learning (eg, clustering analysis) can uncover latent behavior patterns [23], and deep learning (eg, convolutional neural networks [CNNs] and recurrent neural networks) can extract deep features related to mental health directly from time-series data such as voice or motion sensors, overcoming the limitations of manual feature engineering and significantly improving recognition performance [24]. The continuous data collection by smart devices, combined with advanced algorithms, enables real-time monitoring and early warning of mental health status, providing strong data support for personalized interventions and precision medicine [25].

Objectives

Although several reviews have addressed the application of passive sensing technologies in mental health monitoring, significant methodological gaps still exist. First, most reviews focus on the sensing technology itself rather than the integration of sensing with ML [26-28]. Second, there is limited discussion of key aspects of ML (such as data preprocessing and feature engineering), restricting the reproducibility and optimization potential of the technology [29-31]. Moreover, existing studies often focus on a single mental disorder (eg, depression [32,33] or anxiety [34]). Furthermore, they fail to include relevant databases such as PsycINFO [34-36], IEEE Xplore [37], and the ACM Digital Library [34,38]. On the basis of these gaps, this study aimed to conduct a scoping review to comprehensively assess the current application and development trends of ML techniques in mental health monitoring based on passive sensing data from wearable devices and smartphones. Specific objectives included to (1) synthesize and summarize current mainstream technical frameworks and commonly used methods for data acquisition, preprocessing, feature engineering, and algorithm selection; (2) evaluate association patterns between diverse behavioral features (eg, sleep, physical activity, and social interaction) and specific mental health conditions (eg, depression and anxiety); (3) identify the primary limitations and challenges of existing approaches with respect to sample representativeness, data quality, and model generalizability; and (4) explore potential future improvements in standardized data collection, algorithm optimization, and ethical guidelines.

Methods

Overview

This scoping review adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [39] guidelines to ensure methodological rigor and transparency. The review protocol was preregistered on the Open Science Framework (registration 10.17605/OSF.IO/74ACP). The full PRISMA-ScR checklist is provided in Multimedia Appendix 1.

Data Sources and Search Strategy

To ensure the systematic and reproducible nature of the literature search, we used an interdisciplinary database search strategy covering 7 authoritative databases across the fields of medicine, engineering, psychology, and computer science: Web of Science Core Collection, PubMed, IEEE Xplore, Embase, PsycINFO, Scopus, and the ACM Digital Library. In addition, we manually screened the reference lists of the included studies and used citation tracking to extend the search scope.

Before the official search, we collaborated with librarians, mental health experts, and health informatics specialists to develop a detailed search strategy based on the definitions in the Biomarkers, Endpoints, and Other Tools vocabulary [40]. The core keywords included “wearable devices,” “smartphones,” “mental health,” and “machine learning,” which were combined using Boolean logic. All search queries were pretested to ensure high relevance across different databases. To mitigate any bias introduced by daily updates to the databases, all searches were completed synchronously on February 3, 2025. After removing duplicates using the EndNote X20 software (Clarivate Analytics), the retrieved studies were systematically organized and analyzed. The specific search strategy is detailed in Multimedia Appendix 2.

Inclusion and Exclusion Criteria

Study eligibility was determined using the population, intervention, comparison, outcomes, and study design framework, as shown in Textbox 1, with additional methodological considerations detailed later in this section.

Textbox 1. Eligibility criteria based on the population, intervention, comparison, outcomes, and study design framework.

<p>Inclusion criteria</p> <ul style="list-style-type: none">Population: participants with mental health–related disorders (≥70% of the sample)Intervention: passive data collection using wearable and smartphone sensors and machine learning–based prediction or detection of mental health conditionsOutcomes: clinically diagnosed mental health disorders (eg, depression and anxiety) monitored via passive sensing data collected in naturalistic settings using wearable devices or smartphonesStudy design: peer-reviewed original research in English and published between 2015 and 2025 <p>Exclusion criteria</p> <ul style="list-style-type: none">Population: studies focusing solely on transient states (eg, mood or stress) without clinical diagnoses and prototype devices (eg, electronic tattoos and textile sensors)Intervention: active data (eg, questionnaires) as model input features and simulated or laboratory-only task dataOutcomes: prediction of isolated physiological signals (eg, heart rate) without mental health linkageStudy design: nonresearch articles (reviews and protocols); unavailable full text or incomplete data; and conference papers, abstracts, or nonjournal publications
--

Two key methodological considerations guided our eligibility assessment. First, given the rapid evolution of wearable artificial intelligence (AI) technology and our focus on modern methodologies, we restricted the search to articles published after 2015. Second, while changes in mental health states (eg, emotions, mood, and stress) may serve as potential indicators of disorders such as depression or anxiety [41], examining these factors in isolation fails to reflect clinical diagnoses of mental health conditions [42,43]. Thus, we excluded studies focusing solely on these states to maintain research precision.

It should be noted that the comparison component of the population, intervention, comparison, outcomes, and study design was not applied as our scoping review focused on technological applications rather than therapeutic comparisons.

Screening Strategy

Before the formal screening process, the evaluators (SYS and JZ) randomly selected 30 articles and conducted a preliminary review of titles and abstracts followed by full-text evaluation based on the inclusion and exclusion criteria. A double-blind cross-validation approach was used. The pilot screening yielded a Cohen κ of 0.91, demonstrating a high level of agreement between the evaluators, and no adjustments to the screening process or criteria were necessary. Subsequently, SYS and JZ independently screened the remaining articles, and any discrepancies were resolved through discussion with a third-party evaluator (WQ), who made the final decision. On February 15, 2025, the screening was completed, and 42 studies meeting the criteria were included.

Data Extraction and Synthesis

To explore the application of ML methods in the field of mental health and analyze how passive, noninvasive sensor data

combined with ML techniques can be used to predict, detect, or monitor mental health disorders and outcomes, we followed the process of ML model development and validation. This process clarified the key stages of data acquisition, data processing, ML method application, and health outcome screening.

The first stage related to study characteristics. We used a standardized form to extract the key characteristics of each included study. Two independent reviewers collected data on study design, sample size, participant demographics, and the specific mental health issues addressed. For gender data, raw values were converted into percentages, with median values and IQRs subsequently calculated. In addition, we documented the duration of data collection and the sources of funding for each study. Using the R packages *ggplot2*, *dplyr*, and *ggsci* (R Foundation for Statistical Computing), we visualized the distribution of sample characteristics and funding sources.

In terms of data acquisition, we focused on the measurement environment, types of devices used, and specific configuration parameters reported in the studies. Detailed information, including device brands, sensor types, and body placement, was systematically documented. Using R packages such as *ggplot2*, *tidyr*, *dplyr*, *igraph*, *fmsb*, and *tibble*, we constructed an integrated network diagram illustrating the relationships among device types, sensor categories, and behavioral features. In addition, *reshape2* and *ggplot2* were used to analyze feature use patterns across studies and generate a heat map highlighting the prevalence of specific features under different mental health conditions.

Regarding data preparation, we systematically compiled the data preprocessing techniques used across the included studies. This encompassed a variety of signal denoising methods, including specific filtering algorithms and their parameter settings; strategies for handling missing data; and approaches

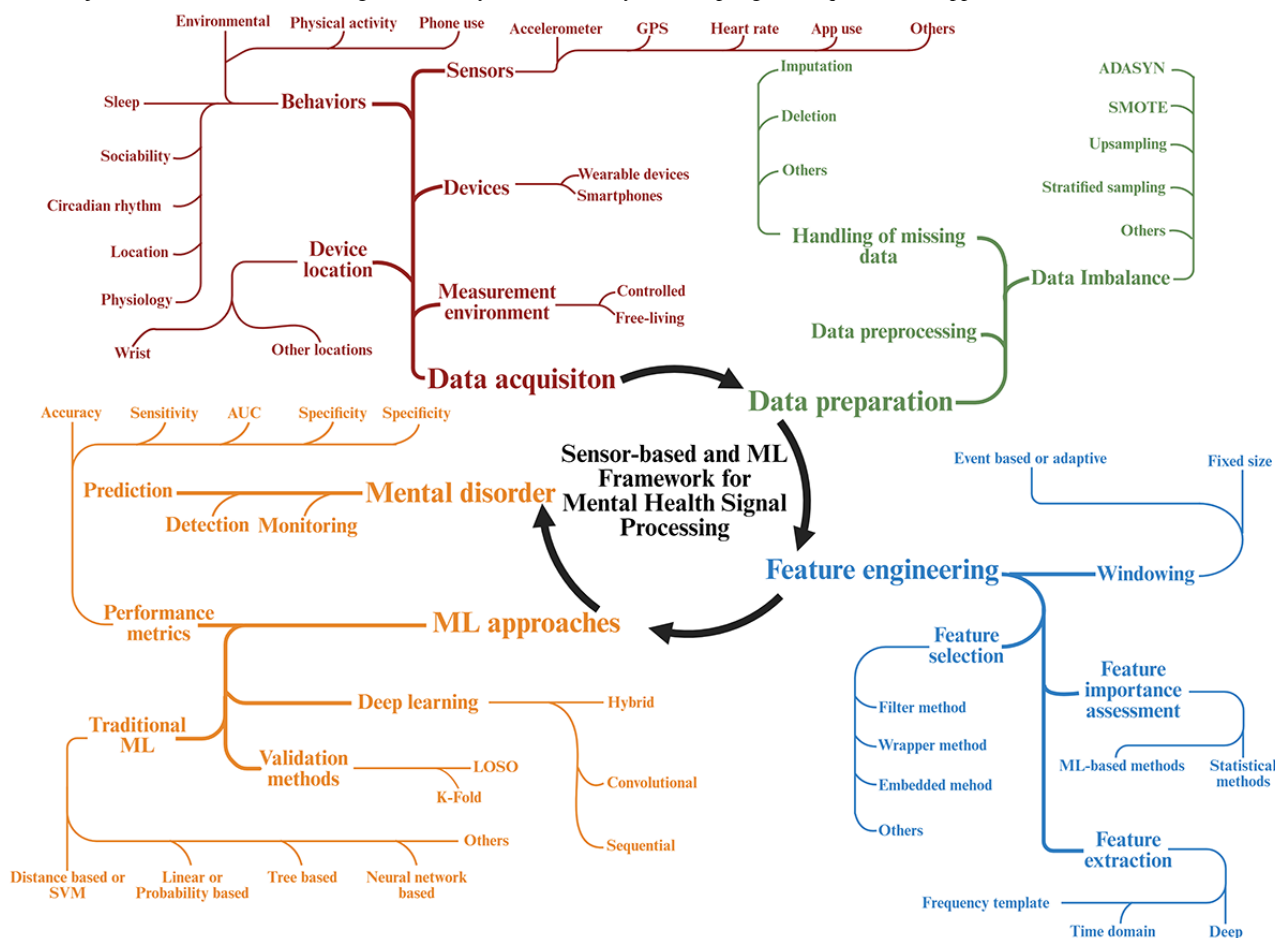
to address class imbalance. These approaches were organized into comparative tables to highlight methodological variations among research teams. Special emphasis was placed on sensor-specific preprocessing techniques tailored to different data modalities.

Regarding feature engineering, we extracted features along three dimensions: (1) time window parameters, including window length and overlap ratio; (2) feature extraction techniques, spanning time-domain and frequency-domain analyses; and (3) feature selection strategies. We also compiled methods for assessing feature importance and, across 31% (13/42) of the studies, extracted quantitative metrics such as importance rankings and effect sizes. Data were processed using the *tidyverse* package in R, and we visualized the top 3 features' associations with psychiatric disorders in each study via network graphs rendered using *ggraph* and *igraph*.

In terms of ML methods, we comprehensively documented the ML strategies used across the studies, including algorithm classes, model validation schemes, and performance metrics. Using the R packages *dplyr*, *networkD3*, *htmlwidgets*, and *jsonlite*, we generated an overview diagram summarizing the application of ML methods.

Figure 1 provides an overview of the key data items extracted from the included studies. In addition, we categorized the features into 8 main behavioral categories: sleep, physical activity, circadian rhythm, sociability, location, physiology, phone use, and environmental behavior [33]. Each behavioral category included several lower-level features that resulted from the integration of various individual features reported in each study. As a result, a single study may identify multiple associations with the same feature. Detailed definitions and classifications of these behavioral categories can be found in [Multimedia Appendix 3](#).

Figure 1. Overview of key data items extracted from the included studies. ADASYN: adaptive synthetic sampling; AUC: area under the curve; LOSO: leave one subject out; ML: machine learning; SMOTE: synthetic minority oversampling technique; SVM: support vector machine.



Results

Study Characteristics

The study selection process is detailed in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart (Figure 2). A total of 42 studies were included in this scoping review. Among them, cohort studies accounted for the largest proportion ($n=23$, 55%), followed by cross-sectional studies ($n=14$, 33%). The median overall sample size was 60.5 (IQR 54-99) participants, whereas the median proportion of female participants was 57.5% (IQR 46.25%-67.34%). The study populations were primarily focused on depression ($n=23$, 55%) and anxiety disorders ($n=9$, 21%), followed by bipolar disorder ($n=4$, 10%) and schizophrenia ($n=6$, 14%).

In terms of control group design, 48% (20/42) of the studies used healthy or population-based control groups, 17% (7/42) compared patients with different mental health conditions, and only 2% (1/42) were based entirely on a healthy population.

Notably, as many as 76% (32/42) of the studies had sample sizes of <100 participants, with 24% (10/42) of the studies having sample sizes of <50 participants. In addition, 45% (19/42) of the studies collected passive data for <7 days, and 12% (5/42) of the studies did not report the specific duration of data collection. The distribution of sample sizes and data collection durations is shown in Figure 3A. Furthermore, the distribution of participant age characteristics can be found in Figure 3B.

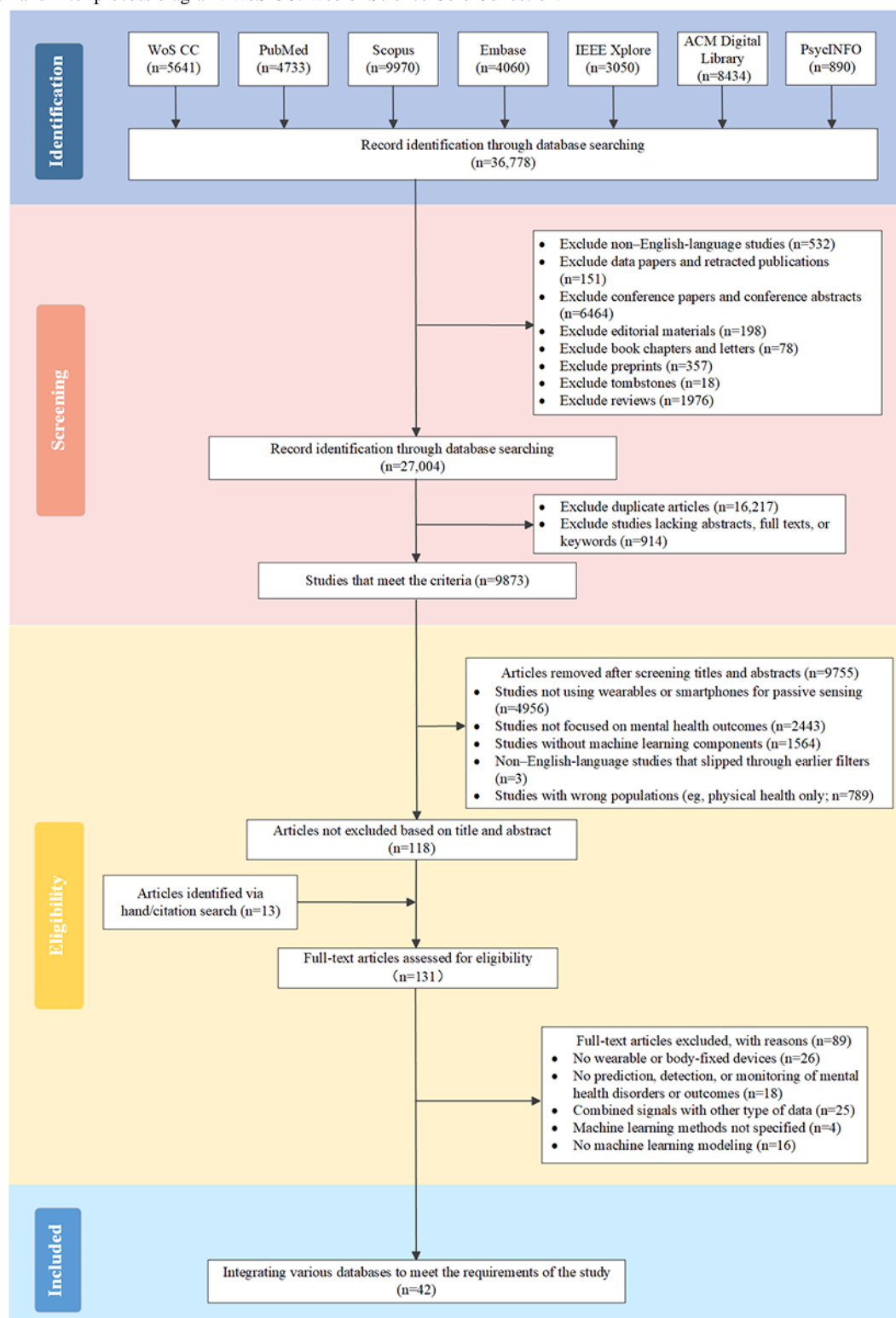
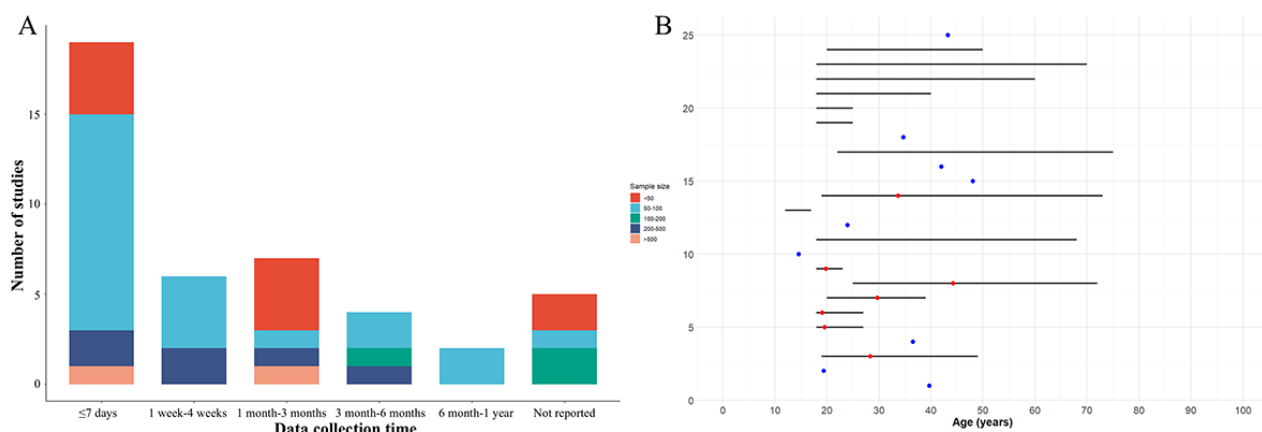
Figure 2. Search and filter process diagram. WoS CC: Web of Science Core Collection.

Figure 3. Overview of study sample characteristics and data collection duration—(A) distribution of sample size and data collection duration; (B) distribution of participant age characteristics.



Regarding the availability of datasets, 38% (16/42) of the studies did not provide explicit information, whereas the remaining studies included 31% (13/42) that provided open access datasets. A detailed list of the characteristics of the included studies is shown in [Table 1](#), and the complete list of studies can be found in [Multimedia Appendix 4](#) [14,15,17-19,24,25,44-79].

In addition, we conducted a detailed analysis of the funding sources for the included studies. The results revealed that 62% (26/42) of the included studies received funding, with these studies having a total of 69 funding instances. The breakdown of these funding instances showed that government agencies were the most prominent source, accounting for 57% (39/69) of instances. This was followed by funding from universities and research institutions (19%, 13/69) and nonprofit

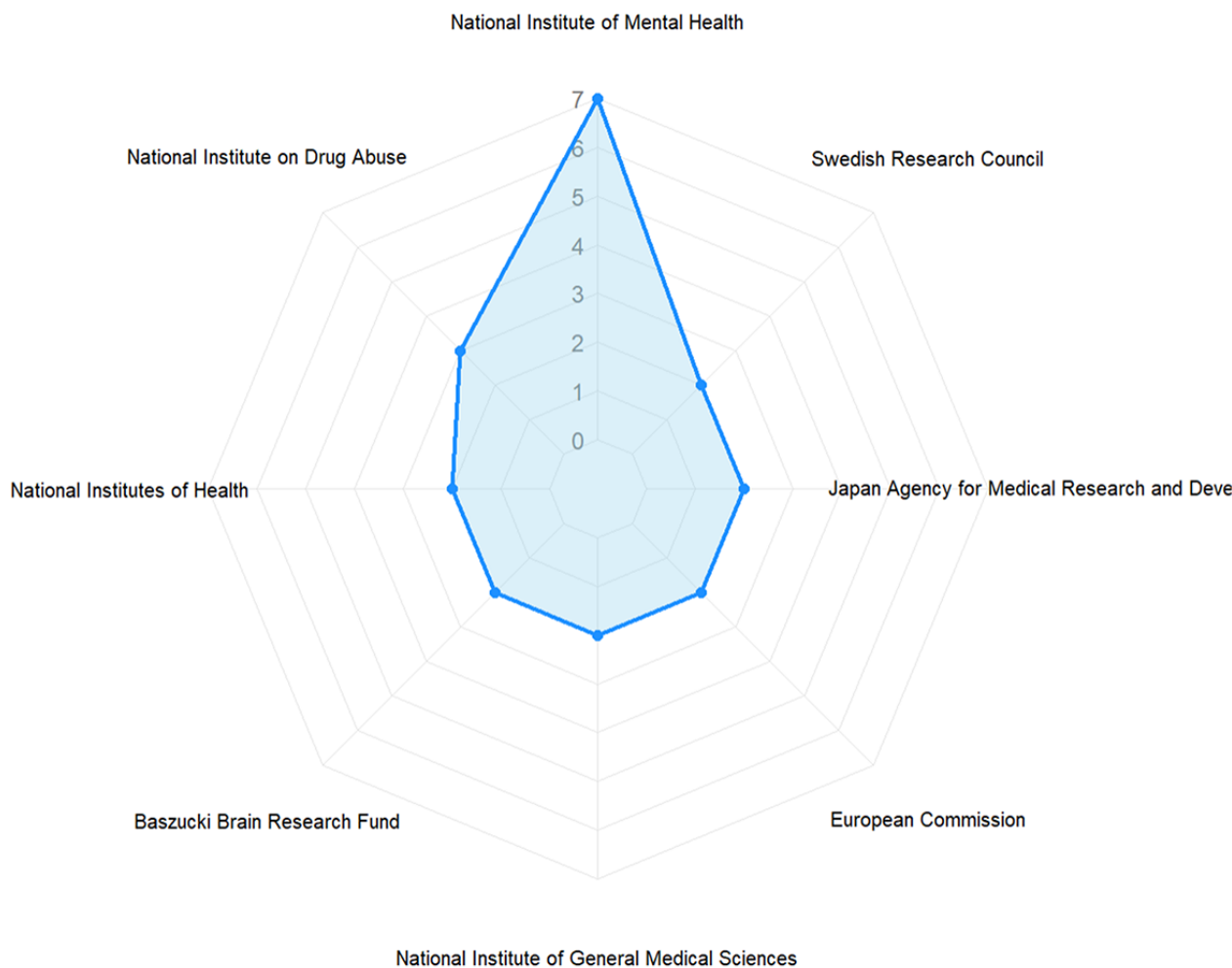
organizations or foundations (15%, 10/69). Funding from international organizations (6%, 4/69), private individuals (3%, 2/69), and corporations (1%, 1/69) was less common. This distribution indicates a high level of public sector interest and suggests that the field has relatively low appeal for industry and private capital.

Furthermore, as shown in [Figure 4](#), the major funding agencies were the National Institute of Mental Health in the United States, which provided 10% (7/69) of the funding instances, followed by the National Institute on Drug Abuse, which contributed 4% (3/69) of the funding instances. These data not only highlight the concentration of funding sources but also provide a strong basis for further exploration of research trends and policy support in this field.

Table 1. Characteristics of the included studies (N=42).

Category and subcategory	Articles, n (%)	References
Research design		
Cohort study	23 (55)	[15,17,18,44-62,80]
Cross-sectional study	14 (33)	[24,25,63-74]
RCT ^a	1 (2)	[75]
Case-control study	4 (10)	[76-79]
Disease categories		
Depression	24 (57)	[15,17,46,47,49,51-53,55-57,59-61,63,67,68,71,73,77-79]
Schizophrenia	6 (14)	[58,63,68,74-76]
Bipolar disorder	4 (10)	[19,44,48,69]
Anxiety	9 (21)	[14,18,24,45,47,50,65,66,70,72]
PTSD ^b	1 (2)	[54]
Attention-deficit/hyperactivity disorder	1 (2)	[25]
Suicidal ideation	1 (2)	[62]
Data collection duration		
6 mo-1 y	2 (5)	[19,75]
3-6 mo	4 (10)	[46,52,58,62]
1-3 mo	7 (17)	[17,48,51,53,55,61,65,70]
1-4 wk	5 (12)	[14,45,63,68,76]
≤7 d	19 (45)	[18,24,25,44,47,49,50,54,57,60,66,67,69-71,74,77-79]
Not reported	5 (12)	[14,15,56,59,72,73]
Accessibility of data		
Open access	13 (31)	[14,24,25,47,60,63,66,67,70-72,76,77]
Not open access	9 (21)	[15,44-46,51,53,54,57,62,74,78,79]
Not specified	16 (38)	[17-19,48-50,52,56,58,59,61,65,68,69,73,75]
Device types		
Mobile phone	20 (48)	[14,15,17,18,24,45,46,48,49,51,53,59,61,62,65,70,73-75]
Watch, wristband, or wrist strap	14 (33)	[19,44,46,52-55,57,60,61,65,69,74,79]
Research-grade accelerator	11 (26)	[25,47,50,56,63,66-68,71,76,77]
Others	5 (12)	[24,25,70,72,78]

^aRCT: randomized controlled trial.^bPTSD: posttraumatic stress disorder.

Figure 4. Distribution of funding frequency from major funding organizations.

Data Acquisition

Measurement Environment and Location

Regarding the data collection environment, 17% (7/42) of the studies conducted data collection in controlled environments, whereas 83% (35/42) of the studies were carried out in natural settings. Most studies (31/42, 74%) used pre-established measurement protocols, requiring participants to collect data during specified tasks or activities. This approach helps reduce variability between individuals and the environment, enhancing the standardization of data and providing more stable and reliable inputs for subsequent ML-based signal analysis [45,46,48,65]. In contrast, a few studies (10/42, 24%) did not implement strict collection protocols and aimed to capture signal variations in participants' daily real-world states [24,25,47,55,67,71-73,76,77]. Although this method may introduce noise, it more accurately reflects the real psychological dynamics, providing valuable data support for applications in real-world settings.

Regarding device placement, the use of wearable devices was mostly concentrated at the wrist, with 57% (24/42) of the studies specifically fixing the device at this location. This choice is likely driven by the practical considerations of ease of long-term wear and the efficient capture of physiological signals (such as heart rate and activity levels) [81,82]. Only 5% (2/42) of the

studies examined the use of wearable devices placed at 2 different locations on the body [25,53], whereas the remaining studies (40/42, 95%) opted for a single site for data collection, indicating a current reliance on single-point measurements in the research. For smartphones, most studies (18/20, 90%) did not provide specific details regarding device placement.

Devices and Sensors

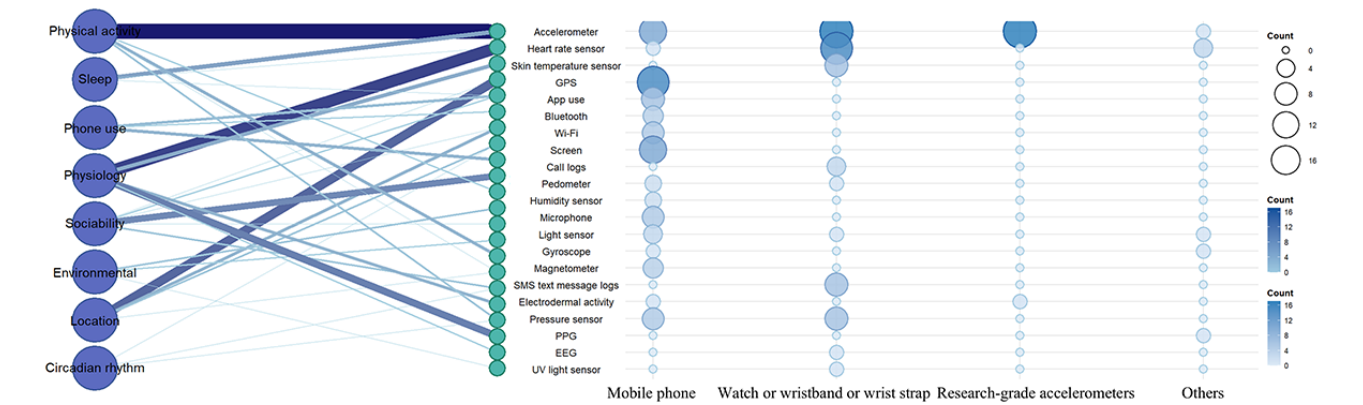
A total of 26% (11/42) of the studies used 2 types of devices to collect passive signals, whereas the remaining 76% (32/42) used a single device. Specifically, 48% (20/42) of the studies used smartphones for data collection; 33% (14/42) used watches, wristbands, or wrist straps; 26% (11/42) used research-grade accelerometers; and 12% (5/42) used other types of devices. Among wearable devices, brands such as Fitbit, Empatica, and Actiwatch were frequently used, whereas most smartphone studies (19/20, 95%) did not specify the brand, with Android being the most commonly used operating system.

Furthermore, we extracted the device types, sensor categories, and the passive feature types collected in each study and presented them in a network bubble diagram (Figure 5). The network diagram on the left of the figure lists the passive feature categories collected through smartphones and other sensors. For example, physical activity and physiological dimensions were particularly associated with accelerometers and heart rate sensors, highlighting the central role of these 2 sensors in

monitoring these dimensions. Accelerometers were used not only to assess physical activity but also to effectively monitor sleep quality. In addition, position data collected by sensors such as GPS and Bluetooth provided valuable insights into individuals' social behaviors and biological rhythms. The bubble diagram on the right of Figure 5 shows each device type (such as mobile phones, wearable devices, and research-grade accelerometers) and the types of sensors it is equipped with.

Through this analysis, we found that smartphones offered broad sensor coverage capable of collecting data from multiple mental health dimensions simultaneously. In contrast, wearable devices such as watches or wristbands focused more on collecting physiological signals, particularly accelerometers and heart rate sensors. A detailed classification of feature categories and sensor terminology can be found in Multimedia Appendix 3.

Figure 5. Application and distribution of different device types and sensors in passive feature signal collection for mental health assessment. Left: A network of passive data categories. Right: A bubble chart of device types and their sensors, where bubble size reflects the quantity and variety of sensors. EEG: electroencephalography; PPG: photoplethysmography.

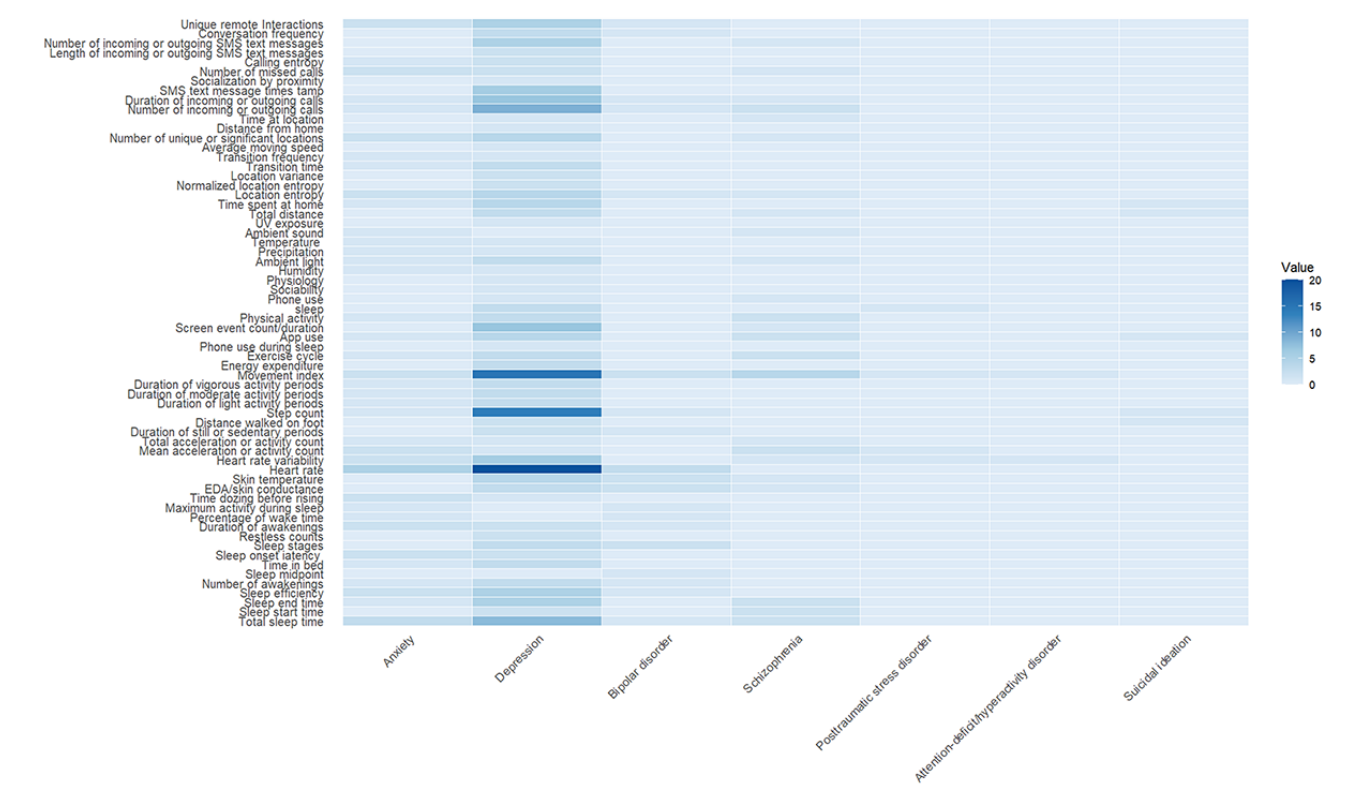


Features and Behaviors

We identified 64 passive features from wearable devices and smartphone sensors and compiled their frequency of occurrence across different mental illnesses (Figure 6). The analysis

revealed that the top 5 most frequently used features were heart rate, movement index, step count, total sleep time, and the number of incoming and outgoing calls, appearing in 67% (28/42), 60% (25/42), 40% (17/42), 33% (14/42), and 31% (13/42) of the studies, respectively.

Figure 6. Heat map of feature use frequency across different mental illnesses. EDA: electrodermal activity.



Furthermore, we categorized all features based on behavior type. The results showed that sleep-related features were the most common, with a total of 14 features. This was followed

by physical activity–related features (n=11) and location-related features (n=11). The high frequency of sleep-related features indicates that sleep status was considered a crucial indicator of

mental health in the existing studies. In addition, the frequent occurrence of physical activity and location information provided key data for understanding individuals' daily rhythms and social interactions.

From the perspective of mental disorders, the frequency of feature occurrence varied across diseases. For depression, heart rate (20/24, 83% of the studies), movement index (15/24, 62%), and step count (14/24, 58%) were the most frequently used features. Anxiety disorders were primarily associated with heart rate (5/9, 56%) and total sleep time (3/9, 33%). In studies on schizophrenia (6/42, 14%), movement index (3/6, 50%) was the most common feature. Research on bipolar disorder showed a high frequency of sleep and physiological features, highlighting the significance of these aspects in the disorder. For posttraumatic stress disorder, features related to physical activity were most frequent. Attention-deficit/hyperactivity disorder was primarily associated with physical activity and physiological features. Finally, studies on individuals with suicidal tendencies (1/42, 2%) showed the highest frequency of features related to physical activity and location information.

Data Preparation

Overview

The quality of passive sensor data directly impacts model performance. The included studies primarily focused on improving data quality through data preprocessing, missing data handling, and class balancing. Detailed information on data preparation methods can be found in [Multimedia Appendix 4](#).

Data Preprocessing

The studies included in this review extensively applied preprocessing techniques to eliminate noise, motion artifacts, and irrelevant signal components, thereby enhancing the signal-to-noise ratio and analytical stability. A detailed breakdown of these methods by data type is provided in [Table 2](#). Specifically, some studies (2/42, 5%) used rule-based filtering methods [44,57] to exclude abnormal physiological data (eg, heart rate exceeding a reasonable range) to prevent erroneous data from generating false correlations. For noise reduction in accelerometer, photoplethysmography, and electrocardiography (ECG) signals, researchers used low-pass or high-pass filtering [56,69,79] as well as external quality control algorithms [78]. Advanced techniques such as discrete wavelet transform were also used for both denoising and feature extraction, effectively reducing environmental interference and collection artifacts [60].

Table 2. Preprocessing techniques for passive sensing signals used in the included studies.

Data type	Technology	Noise or artifact type	Key parameters	Reference
Physiological	Fourth-order low-pass+second-order high-pass Butterworth filter	Baseline drift and high-frequency noise	Low-pass cutoff of 1.5 Hz, high-pass cutoff of 0.05 Hz, and extract tonic and phasic components	[69]
Physiological	Third-order band-pass Butterworth filter	Baseline drift and motion artifacts	Passband of 0.5-8 Hz, retain HR ^a -related frequencies (0.5-4 Hz corresponding to an HR of 30-240 bpm ^b)	[69]
Physiological	DRL ^c circuit noise elimination	Baseline drift and motion artifacts	Electrode type (silver or silicone rubber) and reference electrode FpZ ^d	[70]
Physiological	Adjustable Q-factor wavelet transform	Frequency aliasing and noise interference	Q-factor=1; redundancy=3; 5-level decomposition generating 6 frequency bands	[72]
Physiological	DWT ^e	High-frequency noise and trend fluctuations	Mother wavelet (sym9, db4, or rbio3.9) and decomposition levels (1, 3, or 2)	[60]
Physiological	Resampling and normalization	High-frequency noise and dimensional differences	Target sampling rate of 500 Hz, mean and SD calculated based on the training set, and no data leakage in the test set	[78]
Physiological	Z score normalization	Dimensional differences and baseline drift	Mean and SD based on the training set, standardized range of $\mu=0$, $\sigma=1$, and retain relative amplitude differences	[72]
Physiological	Signal quality index for HR variability analysis filtering	Motion artifacts and waveform distortion	Threshold=1.0; based on harmonic energy ratio	[79]
Physiological	Signal quality index for pulse rate detection filtering	Frequency-domain noise and pulse rate detection errors	Threshold=0.5; 3-min sliding window (1-min overlap)	[79]
Physiological	Band-pass filtering	Baseline drift and high-frequency or low-frequency noise	PPG ^f : 0.5-5 Hz	[79]
Physiological	Rule-based filter	Electrode detachment and circuit overload	EDA ^g (0.05-60 μ S), HR (25-250 bpm), temperature (30-40 °C), and exclude transitional segments of 5 s	[44]
Physiological	Downsampling and time alignment	Inconsistent sampling rate	Unified to 1-s time units and high-frequency signal downsampled by mean (eg, accuracy of 32 Hz \rightarrow 1 Hz)	[44]
Physiological	Global normalization	Dimensional differences and uneven distribution	Scaled to (0, 1) based on minimum and maximum values calculated from the training set	[44]
Acceleration	Resampling	High-frequency motion noise and short-term fluctuations	Downsampling interval=2 h; window=48 h	[77]
Acceleration	Physically infeasible value filtering	Device malfunction and environmental interference	Exclude acceleration peaks of >10 g, long-term invalid values, and nonhuman motion fluctuations	[71]
Acceleration	High-pass filtering+Gaussian smoothing	Low-frequency drift and high-frequency noise	High-pass threshold of 20 mg, rolling window of 30 points (1 s), and aggregated to 1-min epoch	[56]
Acceleration	Z score standardization	Dimensional differences and device sensitivity noise	Calculate mean and SD using classification; formula: $Z \text{ score} = (x - \mu)/\sigma$	[63]
Acceleration	Third-order Butterworth filter	Device vibration and gravity interference	Cutoff frequencies: 20 Hz for high-frequency noise and 0.3 Hz for gravity separation	[17]
GPS	DBSCAN ^h clustering algorithm	Short stop noise points and isolated location points	$\epsilon=40$ m; MinPts=5	[17]
GPS	Velocity threshold filtering	Invalid positioning during motion and high-frequency sampling noise	Stationary speed \leq 1 km/h	[17]
GPS	Fixed-interval downsampling	Data redundancy and position fluctuations	Target sampling rate: 15 samples per hour	[17]

Data type	Technology	Noise or artifact type	Key parameters	Reference
GPS	Spatiotemporal clustering denoising	Short-term stop and position drift	Distance=60 m (distance threshold); time=600 s (time threshold)	[45]

^aHR: heart rate.
^bbpm: beats per minute.
^cDRL: driven right leg.
^dFpZ: frontal pole zero.
^eDWT: discrete wavelet transform.
^fPPG: photoplethysmography.
^gEDA: electrodermal activity.
^hDBSCAN: density-based spatial clustering of applications with noise.

In terms of data smoothing, resampling [77,78] and framing techniques [77] were used to create stable data sequences by segmenting continuous signals into fixed windows. This approach not only smooths the data and reduces high-frequency noise but also facilitates subsequent feature extraction. In addition, 2% (1/42) of the studies reported hardware-level solutions, including noise suppression designs for electroencephalography headsets and a signal quality evaluation method that identifies and removes noise segments through statistical metrics (variance, amplitude, and kurtosis) [70].

Handling of Missing Data

Among the 48% (20/42) of the studies that explicitly addressed missing data, 2 primary strategies were used: data imputation and quality screening. Imputation methods were particularly diverse, including simple statistical replacements (mean [25] or median [15] imputation), ML approaches (k-nearest neighbor [KNN] [52,65] or random forest [19]), advanced filtering techniques (Kalman filter imputation [54] or median filtering interpolation [83]), and multiple imputation [14,18,49,84]. In addition, some studies (4/42, 10%) used probabilistic model-based strategies [62] that naturally integrate missing values to maintain the integrity of the dataset. This approach, compared to traditional imputation, better reflects the true structure and uncertainty of the data, thereby enhancing the robustness and predictive accuracy of the model. Another 17% (7/42) of the studies opted for deletion strategies, either directly removing low-quality segments [46,69] or merging adjacent data points [45]. Notably, only 14% (6/42) of the studies quantified and reported the proportion of missing data [14,15,19,46,52,75].

Data Imbalance

A total of 21% (9/42) of the studies implemented techniques to address data imbalance. The most common approach was synthetic oversampling, with 44% (4/9) of the studies using synthetic minority oversampling technique [15,60,74,76] and 22% (2/9) using adaptive synthetic sampling [47,60]. In addition, traditional upsampling [54] and downsampling [71] were applied individually. One innovative study introduced a class-weighted loss function during model training, offering a novel approach to address class imbalance that goes beyond conventional data-level techniques [55].

Feature Engineering

Window Segmentation

Window segmentation plays a crucial role in passive sensor data analysis as the choice of parameters directly impacts the effectiveness of behavior and physiological pattern recognition. Among the reviewed studies, a fixed-length sliding window was the most commonly used strategy (16/42, 38%), although its implementation varied significantly in terms of window size and overlap rate. In total, 12% (5/42) of the studies systematically compared different window durations, revealing the trade-off between time resolution and noise robustness. For instance, one study tested window lengths ranging from 1 to 2048 seconds (in powers of 2) using physiological data, finding that a 32-second window achieved the best balance between accuracy and noise suppression. Short windows (eg, <32 seconds) were prone to feature fragmentation, whereas long windows (eg, >32 seconds) resulted in noise accumulation [44].

In smartphone-based social behavior analysis, a 14-day window outperformed a 7-day window as it smoothed short-term fluctuations and captured long-term trends more effectively [73]. In contrast, windows as short as 2.5 seconds were suitable for high-frequency motion signal analysis [17].

Regarding the window overlap strategy, most studies (9/42, 21%) used overlapping segmentation (with overlap rates ranging from 33% to 98%) to maximize data use and capture temporal dependencies. For example, a 3-minute window with a 1-minute overlap provided a balance between noise resistance and dynamic capture ability in free-living environments [79]. In a 50-Hz sampling rate scenario (50 data points per second), a 12-second window (600 time points) with a 90% overlap rate enabled a CNN-long short-term memory (LSTM) model to achieve an accuracy of 92.16% [72]. Furthermore, multiscale window designs (eg, 5-min windows for analyzing exercise intensity, 30-min windows for environmental interactions, and 24-h windows for circadian rhythms) were applied for hierarchical behavior modeling. By integrating features across different temporal dimensions, these designs improved the generalization ability of the models [14].

Feature Extraction

A total of 14% (6/42) of the studies used time-domain feature extraction methods, analyzing raw data from devices such as accelerometers, gyroscopes, and heart rate sensors to extract statistical metrics such as mean, SD, skewness, and kurtosis.

These metrics were applied to assess gait [17], physical activity intensity [54,63], and sleep patterns [61], thereby providing strong support for the early diagnosis of mood disorders. Most studies (12/42, 29%) used Python as the core toolchain, implementing custom algorithms or leveraging open-source library functions (eg, NumPy and pandas) for automating the analysis of raw signals into clinical indicators. This processing workflow, standardized by design, enhanced the reproducibility of feature extraction and, through modular design, facilitated method reuse across studies.

In addition, 17% (7/42) of the studies further combined time-domain features with frequency-domain features using methods such as Fourier transform to uncover periodic fluctuations in signals, thereby facilitating deeper analysis of physiological rhythms and long-term variation patterns [14,46,56,60,66,72,79]. For example, one study used tunable Q-factor wavelet transform decomposition of ECG signals into 6 wavelet frequency bands, extracting probabilistic texture features from each band. After merging these with spatial-domain features, a highly efficient frequency domain-spatial domain feature vector was formed. These frequency-domain features effectively captured autonomic nervous system abnormalities related to anxiety, playing a critical role in achieving high model accuracy (>98.5%) [72].

Furthermore, 36% (15/42) of the studies used end-to-end learning models to automate the extraction of complex features and predict mood disorders. Autoencoder models [65,75], by constructing multilayer neural network architectures, performed nonlinear compression and reconstruction of sensor data from accelerometers and heart rate sensors, effectively extracting low-dimensional latent features to capture deep associations within behavior patterns. CNNs [24,47,77,78] excelled in modeling multichannel time-series features using 1D and 2D convolutional kernels to capture local correlations in sensor data. These models, trained end to end, directly learned the nonlinear relationships within raw data, thus avoiding the subjectivity of manual feature engineering. In addition, cross-modal fusion (eg, combining physiological signals with environmental data) further enhanced prediction accuracy, offering an efficient technical pathway for real-time monitoring using wearable devices.

Feature Selection

Among the included studies, 48% (20/42) used feature selection techniques. Specifically, 30% (6/20) of the studies applied filter-based methods to quickly identify key features through statistical tests. For example, one study reduced the feature set by >88% using the information gain method [77]. A total of 30% (6/20) of the studies used wrapper methods, where researchers used random forest [63], recursive feature elimination [59,63,73], and the neural network with weighted fuzzy membership function algorithm [51] to continuously evaluate the impact of different feature subsets on model performance, ultimately selecting the optimal feature combination. In addition, 40% (8/20) of the studies used embedded methods, performing automatic feature selection from hundreds of initial features through least absolute shrinkage and selection operator [58], L1 penalty [61], or L2 penalty [79],

as well as intrinsic feature importance evaluations based on tree models [14,55,61] to enhance model robustness. To address the issue of ultrahigh-dimensional data, 10% (2/20) of the studies used dimensionality reduction techniques, including principal component analysis [45] and deep autoencoders [50]. For instance, one study reduced 3000 features to 200 [45], whereas another reduced 800 features to 50 [50].

Feature Importance Evaluation

Among the 52% (22/42) of the studies that explicitly mentioned feature importance evaluation, the methods can be broadly categorized into statistical and ML methods. Statistical methods included the use of the Cohen *d* effect size to measure the differences between abnormal behaviors before relapse and behaviors under healthy conditions, with significant features selected by setting a threshold of >0.8 [75]. Permutation importance analysis shuffles feature data and observes changes in model performance, providing a model-agnostic evaluation approach [44]. In addition, one study used standardized regression coefficients [56] and conducted ANOVA across multiple datasets to identify variables that significantly influence classification or prediction.

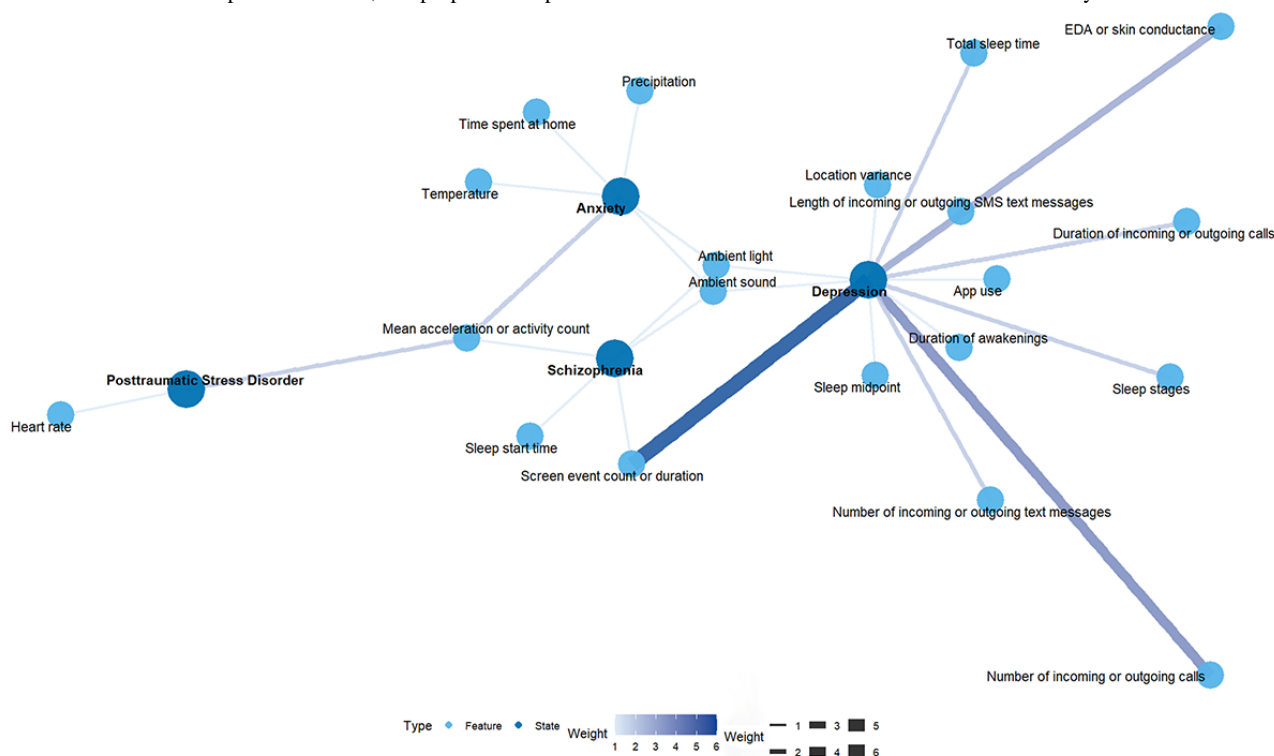
In terms of ML methods, various model interpretation and feature selection techniques were used to quantify the contribution of features. The Shapley additive explanations method based on game theory was used in 12% (5/42) of the studies due to its ability to intuitively quantify the marginal contribution of each feature to the model output. Some studies (2/42, 5%) combined deep neural networks [51] and multitask learning [58] to explore the role of features in complex tasks. Random forest [52,61], adaptive boosting [52], Extreme Gradient Boosting (XGBoost) [14,52,55,57,73], and the Boruta algorithm [53] quantified feature contributions by reducing node impurity or performing iterative validation. One study used feature ablation, which systematically evaluates changes in model performance based on feature subsets to identify critical feature combinations [46]. Additional methods included extra trees classifier [73], integrated gradients [65], and SVM recursive feature elimination strategies [59], which determine the most stable and crucial features through cross-validation and statistical feature selection frequency.

Furthermore, we reviewed 31% (13/42) of studies that explicitly ranked feature importance. Some studies (10/13, 77%) presented the ranking, whereas others (3/13, 23%) used quantitative indicators such as effect size [15,18,19,44,51-55,57,58,73,75]. To simplify the analysis and focus on the most representative features, we extracted the top 3 key features from each study and presented them in a relational diagram (Figure 7). The results indicated that, for schizophrenia, abnormalities in physical activity [75], mobile phone use behavior [58], and sleep rhythm [75] were of particular significance. For depression, changes in mobile phone use behavior [15,51-53] and social interaction [51,52,73] stood out, whereas sleep indicators [19,51,55,57] and electrodermal activity or skin conductance [44,53] also demonstrated high importance. For anxiety disorders, time at home [18] and environmental climate factors [18] were considered important features, suggesting that anxiety may be associated with reduced outdoor activity and

reliance on indoor environments. For posttraumatic stress disorder, heart rate and average acceleration or activity count were identified as important indicators [54], reflecting the

critical role of physiological and physical activity signals in this condition.

Figure 7. Relationship diagram of key features and mental health disorders. Feature importance was derived from the ranking or effect size across 12 related studies. Blue nodes represent features, and purple dots represent mental health disorders. EDA: electrodermal activity.



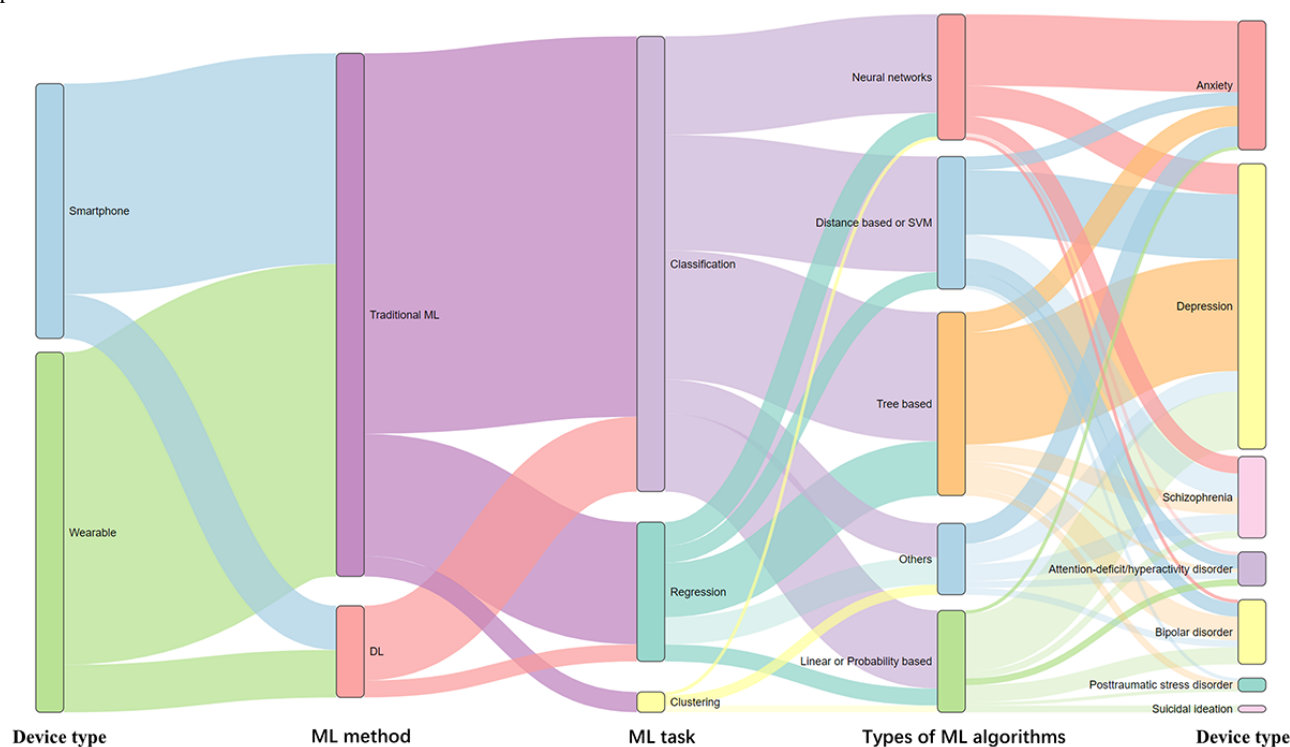
ML Methods

Overview

In the 42 studies included, various ML algorithms were extensively used to extract behavioral features, capture abnormal patterns, and predict mental health conditions from passive, noninvasive signals collected by wearable devices and

smartphones. These algorithms included both deep learning models and traditional ML methods, with many studies (12/42, 29%) achieving performance improvements through ensemble or multitask learning approaches. Figure 8 provides an overview of the application levels of these algorithms across different mental health conditions. Detailed information on the ML methods used in the included studies can be found in Multimedia Appendix 4.

Figure 8. Hierarchical overview of machine learning (ML) methods used in the studies included in this scoping review. DL: deep learning; SVM: support vector machine.



ML Algorithms

Our analysis revealed that nearly half (19/42, 45%) of the studies used deep learning models, with deep neural networks (5/19, 26%) [24,47,51,74,76] and CNNs (5/19, 26%) [24,47,76-78] being the most commonly used architectures. These models demonstrated significant effectiveness in automatically extracting features and identifying complex patterns from high-dimensional time-series data. Notably, LSTM networks [18,71] and their variants, including bidirectional LSTM [44], hybrid CNN-LSTM [24], and ConvLSTM [24], were widely used to capture temporal dependencies, further improving the accuracy of mental health predictions. Unsupervised feature representation learning was mainly achieved through autoencoders, including fully connected neural network autoencoders [75] and deep autoencoders [65]. One study innovatively combined autoencoders with ensemble techniques such as random forest and XGBoost [65]. In addition, some studies (2/42, 5%) explored CNNs integrated with attention mechanisms [76] and hybrid architectures combining neural networks with weighted fuzzy membership functions [51].

Traditional ML algorithms played a central role, with random forest (17/42, 40%), SVM (17/42, 40%), logistic regression (11/42, 26%), and KNN (10/42, 24%) frequently used in classification tasks. Boosting algorithms, including adaptive boosting (4/42, 10%), XGBoost (13/42, 31%), and CatBoost [60], demonstrated superior performance in both regression and classification tasks. Some studies (13/42, 31%) also used linear models such as least absolute shrinkage and selection operator and Elastic Net [52], as well as decision tree [48,61,65,67,69,74], naïve Bayes [61,67,74], and ensemble methods (eg, stacking ensemble [71] and decision integration strategies [46,50,62]), to further explore data relationships and

feature importance. Furthermore, spatiotemporal clustering [45] and change point detection [62] methods proved effective for extracting semantically meaningful locations from GPS data and detecting sudden behavior changes.

In terms of algorithm performance comparison, 60% (25/42) of the studies provided valuable results. Among them, 68% (17/25) focused on comparing traditional algorithms. The findings indicated that random forest demonstrated consistent performance advantages in several studies, with 12% (3/25) of the studies independently selecting random forest as the best model [52,61,67]. The average accuracy of these studies reached 0.835, and the average area under the curve (AUC) was 0.634. XGBoost also performed excellently, leading in 12% (3/25) of the studies [54,57,73], with an average AUC of 0.727 [46,48]. Methods that integrated multiple models showed unique advantages in 24% (6/25) of the studies, achieving an average accuracy of 0.816 and an average AUC of 0.701.

In comparisons between deep learning and traditional algorithms, 82% (9/11) of the studies reported that deep learning methods significantly outperformed traditional methods, with accuracy ranging from 0.85 to 0.963. Common baseline models included random forest, SVM, logistic regression, and KNN. Among these, the CNN-LSTM model demonstrated exceptional performance in one study, with an accuracy of 0.9216 in predicting anxiety disorder [72]. The 2D CNN model showed better balance across metrics, with a sensitivity of 0.75, specificity of 0.77, accuracy of 0.7672, and AUC of 0.76 [47]. Although the LSTM model exhibited a high sensitivity of 0.84, its practical application may be limited by other performance factors [71]. In total, 18% (2/11) of the studies focusing on deep learning architectures further confirmed the technical advantages

of CNN-based feature extraction models (CNN-LSTM and 2D CNN) [24,47].

Table 3 summarizes the performance of each model across 4 metrics—sensitivity, specificity, accuracy, and AUC—in the

included studies (note: for studies reporting multiple performance metrics for the same model, only the highest value is presented). Table 4 presents the performance data of the models selected as the best by the original studies, specifically including those that reported the 4 aforementioned metrics.

Table 3. Performance comparison of machine learning models across the included studies.

Model type	Sensitivity	Specificity	Accuracy	AUC ^a	Average performance
DNN ^b	0.61 [47]	0.83 [47]	0.72 [47], 0.77 [51], 0.912 [74], and 0.7029 [24]	0.73 [47]	Sensitivity: 0.61; specificity: 0.83; accuracy: 0.776; AUC: 0.73
CNN ^c	— ^d	—	0.8072 [77], 0.77 [78], and 0.7932 [24]	—	Accuracy: 0.79
AlexNet	0.45 [71]	0.82 [71]	—	0.68 [71]	Sensitivity: 0.45; specificity: 0.83; AUC: 0.68
LSTM ^e	0.84 [71]	0.26 [71]	0.8126 [77] and 0.6713 [24]	0.55 [71]	Sensitivity: 0.84; specificity: 0.26; accuracy: 0.742; AUC: 0.55
2D CNN	0.75 [47]	0.77 [47]	0.7672 [47]	0.76 [47]	Sensitivity: 0.75; specificity: 0.77; accuracy: 0.7672; AUC: 0.76
CNN-LSTM ^f	—	—	0.9216 [24]	—	Accuracy: 0.9216
Fully connected neural network autoencoder ^g	0.25 [75]	0.88 [75]	—	—	Sensitivity: 0.25; specificity: 0.88
GRU Seq2Seq ^h	0.29 [75]	0.86 [75]	—	—	Sensitivity: 0.29; specificity: 0.86
BiLSTM ⁱ	—	—	0.7 [44]	—	Accuracy: 0.7
Autoencoder ^j	—	—	0.87 [65]	—	Accuracy: 0.87
MLP ^k	—	—	0.81 [45] and 0.7692 [70]	—	Accuracy: 0.79
ConvLSTM ^l	—	—	0.6982 [72]	—	Accuracy: 0.698
Decision tree	0.9662 [67]	0.9737 [67]	0.9697 [67], 0.8056 [61], 0.798 [69], and 0.689 [74]	—	Sensitivity: 0.9662; specificity: 0.9737; accuracy: 0.816
Logistic regression	0.777 [67], 0.39 [71], 0.873 [79], and 0.682 [60]	0.8 [67], 0.74 [71], 0.84 [79], and 0.594 [60]	0.791 [46], 0.7885 [67], 0.859 [79], 0.64 [60], 0.7076 [70], and 0.7591 [61]	0.55 [71], 0.62 [54], 0.93 [79], and 0.638 [60]	Sensitivity: 0.68; specificity: 0.744; accuracy: 0.758; AUC: 0.685
Gradient boosting	—	—	0.791 [46] and 0.798 [72]	—	Accuracy: 0.795
Naïve Bayes	0.6418 [67]	0.9078 [67]	0.77516 [67], 0.8056 [61], and 0.704 [74]	—	Sensitivity: 0.6418; specificity: 0.9078; accuracy: 0.762
Random forest	0.9989 [67] and 0.673 [60]	0.95 [67] and 0.602 [60]	0.9865 [67], 0.9824 [63], 0.64 [60], 0.8769 [70], 0.8542 [61], 0.719 [74], and 0.7837 [72]	0.63 [54] and 0.638 [60]	Sensitivity: 0.836; specificity: 0.776; accuracy: 0.835; AUC: 0.634
XGBoost ^m	0.7 [66], 0.9662 [67], 0.73 [57], and 0.648 [60]	0.955 [66], 0.9931 [67], 0.81 [57], 0.82 [73], and 0.561 [60]	0.82 [15], 0.953 [67], 0.76 [57], and 0.607 [60]	0.892 [66], 0.7 [54], 0.712 [55], and 0.605 [60]	Sensitivity: 0.761; specificity: 0.828; accuracy: 0.816; AUC: 0.727
SVM ⁿ	0.9391 [67], 0.63 [60], and 0.843 [17]	0.897 [17], 0.9736 [67], 0.79 [59], 0.75 [73], and 0.571 [60]	0.75 [51], 0.872 [17], 0.8624 [67], 0.602 [60], 0.7786 [61], and 0.79 [74]	0.61 [54] and 0.601 [60]	Sensitivity: 0.804; specificity: 0.796; accuracy: 0.776; AUC: 0.606
ANN ^o	0.767 [17]	0.915 [17]	0.854 [17]	—	Sensitivity: 0.767; specificity: 0.915; accuracy: 0.854
CatBoost	0.687 [60]	0.588 [60]	0.64 [60]	0.58 [60]	Sensitivity: 0.687; specificity: 0.988; accuracy: 0.64; AUC: 0.58
KNN ^p	0.695 [17], 0.99 [67], and 0.61 [60]	0.846 [17], 0.9736 [67], and 0.55 [60]	0.825 [17], 0.9865 [67], 0.582 [60], 0.8267 [61], and 0.752 [74]	0.58 [60]	Sensitivity: 0.765; specificity: 0.79; accuracy: 0.794; AUC: 0.58

Model type	Sensitivity	Specificity	Accuracy	AUC ^a	Average performance
Ensemble model	0.846 [50], 0.712 [19], 0.95 [76], 0.74 [71], and 0.682 [60]	0.527 [50], 0.856 [19], 0.43 [71], and 0.594 [60]	0.881 [46], 0.687 [50], 0.851 [77], 0.801 [19], 0.64 [60], 0.9995 [72], 0.9159 [69], and 0.755 [74]	0.695 [50], 0.86 [19], 0.61 [71], and 0.638 [60]	Sensitivity: 0.786; specificity: 0.602; accuracy: 0.816; AUC: 0.701

^aAUC: area under the receiver operating characteristic curve.

^bDNN: deep neural network.

^cCNN: convolutional neural network.

^dNot available.

^eLSTM: long short-term memory.

^fCNN-LSTM: a hybrid model combining CNN and LSTM.

^gAn autoencoder composed of fully connected layers used for nonlinear dimensionality reduction or feature learning.

^hGRU Seq2Seq: gated recurrent unit sequence-to-sequence model.

ⁱBiLSTM: bidirectional LSTM.

^jAn unsupervised learning model.

^kMLP: multilayer perceptron.

^lConvLSTM: an LSTM integrated with convolutional operations, suitable for spatiotemporal sequence prediction.

^mXGBoost: Extreme Gradient Boosting.

ⁿSVM: support vector machine.

^oANN: artificial neural network.

^pKNN: k-nearest neighbor.

Table 4. Best reported performance of the models in the included studies.

Model type	Sensitivity	Specificity	Accuracy	Area under the ROC ^a curve
Fully connected neural network autoencoder ^b	0.25 [75]	0.88 [75]	— ^c	—
Autoencoder ^d	—	—	0.87 [65]	—
CNN-LSTM ^e	—	—	0.9216 [72]	—
Random forest	0.9989 [67]	0.95 [67]	0.9865 [67], 0.8769 [70], and 0.8542 [61]	—
XGBoost ^f	0.73 [57]	0.81 [57] and 0.82 [73]	0.76 [57]	0.7 [54]
AlexNet	0.45 [71]	0.82 [71]	—	0.68 [71]
CatBoost	0.687 [60]	0.588 [60]	0.64 [60]	0.58 [60]
SVM ^g	0.843 [17]	0.897 [17]	0.872 [17]	—
DNN ^h	—	—	0.77 [51] and 0.912 [74]	—
Ensemble model	0.846 [50] and 0.712 [19]	0.527 [50] and 0.856 [19]	0.881 [46], 0.687 [50], 0.851 [77], 0.801 [19], 0.9995 [72], and 0.9159 [69]	0.696 [50] and 0.860 [19]

^aROC: receiver operating characteristic curve.

^bAn autoencoder composed of fully connected layers used for nonlinear dimensionality reduction or feature learning.

^cNot available.

^dAn unsupervised learning model.

^eCNN-LSTM: a hybrid model combining convolutional neural network and long short-term memory.

^fXGBoost: Extreme Gradient Boosting.

^gSVM: support vector machine.

^hDNN: deep neural network.

Validation Methods and Performance Metrics

K-fold cross-validation was the primary validation strategy (19/42, 45% of the studies), with 5-fold (5/19, 26%) and 10-fold

(11/19, 58%) cross-validation being the most commonly used. Leave-one-out cross-validation (9/42, 21%) was also frequently mentioned. In addition, repeated stratified 10-fold

cross-validation [55] and sliding window cross-validation [18,49] were applied to time-series data in studies focused on dynamic prediction. A total of 12% (5/42) of the studies used multiple validation methods [15,45,52,58,67,78], and 5% (2/42) of the studies used 4 different cross-validation schemes [52,58]. Notably, only 2% (1/42) of the studies conducted external validation [56].

Performance evaluation used task-specific metrics. Classification studies primarily reported accuracy, sensitivity, specificity, precision, recall, F_1 -score, and AUC. One study introduced the true positive rate-to-false positive rate ratio to optimize the sensitivity-specificity trade-off [75]. In regression analysis, metrics such as mean squared error, root-mean-squared error, mean absolute error, mean absolute percentage error, R^2 , and the Pearson correlation coefficient (eg, $r=0.587$) were used. These were supplemented with the Brier score [66], Matthews correlation coefficient [63,79], κ statistic [70], and Bland-Altman plots [56] to assess the robustness of the results.

Discussion

Methodological Challenges and Data Limitations

For a long time, mental health monitoring has relied on subjective assessment methods such as clinical interviews and self-report scales, which are often limited by ecological validity and recall bias [4,5]. With the widespread use of smartphones and wearable devices, passive sensing technologies now enable continuous, noninvasive collection of multidimensional data on user activity, heart rate, sleep, social interactions, and more in natural environments. This provides more objective, granular, and quantitative information on mental health status [14]. However, the multisource, time-series data generated by these sensors often exhibit characteristics such as high dimensionality, redundancy, and noise, making it difficult for traditional statistical methods to effectively capture the associations between underlying behavioral patterns and psychological indicators [20].

ML, particularly deep learning models, offers advantages in automatic feature extraction and nonlinear pattern learning. These models can uncover subtle emotional and cognitive features from complex sensor data, significantly improving the accuracy of mental state detection and prediction [24]. While several reviews have focused on passive sensing technologies themselves or provided single-technology summaries for specific mental disorders, there is a lack of a comprehensive review that systematically evaluates the combination of passive sensing and ML for mental health monitoring, including the entire technical framework and practical outcomes.

This study was conducted in this context and, for the first time, comprehensively reviewed and integrated research on the combination of passive sensing and ML for mental health monitoring from 2015 to 2025. It covered key aspects such as data acquisition, preprocessing, feature engineering, algorithm selection, and validation. The aim was to reveal the relationships between digital biomarkers and various mental disorders; assess the strengths and weaknesses of different ML models; and summarize existing challenges regarding sample size, data

quality, model generalization, and ethical compliance. The following discussion will focus on methodological limitations, technical optimization pathways, clinical validation requirements, and ethical considerations.

Analysis revealed that cohort studies dominate the field (23/42, 55%), reflecting a consensus on the importance of longitudinal data collection. However, the sample sizes were generally small, with a median of only 60.5 (IQR 54-99) participants. Furthermore, 76% (32/42) of the studies included <100 participants, with 24% (10/42) of the studies having <50. This small sample size increases the risk of model overfitting, which undermines the robustness of the conclusions, especially in the validation of complex models such as bidirectional LSTM [44]. In addition, nearly half (19/42, 45%) of the studies had data collection periods of <7 days, whereas the pathological features of mental disorders typically require extended observation periods (eg, manic episodes in bipolar disorder often last several weeks [85]). Short-term monitoring may struggle to effectively distinguish transient mood fluctuations from true disease states.

Regarding data sharing, only 31% (13/42) of the studies provided open access data, and 38% (16/42) of the studies did not clearly specify data availability, which, coupled with a lack of standardized assessment protocols in the field, hinders result reproducibility [28]. It is noteworthy that only 2% (1/42) of the studies conducted external validation. The widespread absence of external validation, which is crucial for assessing model generalizability and avoiding overestimation of performance due to sample bias or overfitting, further highlights the limitations in model dissemination and validation in the current research [86].

The bottlenecks related to insufficient dataset size and lack of diversity are primarily constrained by the sensitivity of mental health data, the complexity of cross-study collaborations, and the challenges of long-term monitoring. The sensitivity of mental health data necessitates strict privacy protection requirements, significantly limiting the availability of data [87]. Moreover, conducting cross-center, cross-cultural studies requires coordinating multiple institutions and facing issues such as inconsistent technical standards and high data heterogeneity, which further increase collaboration costs [88]. More critically, the pathological features of mental disorders often require long-term monitoring for accurate capture, yet patient compliance, device battery life, and the continuous investment of research resources present practical barriers [89]. In response to these challenges, recent research has explored solutions from both system architecture and algorithm optimization perspectives.

At the system level, scalable digital data collection platforms (eg, Intelligent Sensing to Inform and Learn [90] and AwarNS [91]) have reduced the implementation barriers for multicenter studies through modular design and privacy protection mechanisms. Intelligent Sensing to Inform and Learn provides standardized sensor toolkits for iOS and Android devices, whereas AwarNS integrates a *sense-analyze-act* framework, supporting the closed-loop process from data collection to intervention and providing infrastructure for large-scale heterogeneous data collection.

At the algorithm level, researchers have reduced reliance on labeled data through unsupervised learning [92] and used data augmentation and generative adversarial networks to synthesize virtual samples, thereby expanding training datasets [93]. In addition, federated learning has enabled cross-institution data collaboration through distributed training, improving model generalization capabilities while maintaining privacy protection [94]. Although these methods have not fully resolved the data scarcity issue, they offer feasible paths to balance data scale and quality.

Technological Implementation and Multimodal Integration

In terms of technological implementation, 76% (32/42) of the studies used a single device. While this reduces implementation complexity, it may overlook other potentially important signal dimensions. In addition, most studies (24/42, 57%) opted for wrist-worn devices, which, while convenient, have limitations in fully monitoring physiological and behavioral signals. For example, head-worn devices can capture electroencephalography activity [70], whereas chest sensors are better suited for monitoring cardiopulmonary function [25]. Future research should consider the integration of multilocation devices (eg, head, chest, and waist) to enhance the multidimensionality and consistency of the data, as well as explore multimodal fusion architectures (eg, CNN-LSTM combining physiological and GPS data [24]).

The quality of passive sensing data faces dual challenges in device compliance and technological reliability. Studies have shown that individual differences in device use compliance—such as irregular wear, forgotten charging, or voluntary discontinuation—often result in significant gaps and noise in the data. For instance, Mullick et al [52] reported a 67.74% data loss rate in their adolescent sample due to compliance issues such as forgetting to wear the device, charging requirements, and skin irritation. This issue may be further exacerbated by variations in user motivation. The intrusiveness of long-term monitoring may lead to “device fatigue” [95,96], whereas socioeconomic factors such as the availability of charging infrastructure, as well as privacy concerns across different cultural backgrounds, can also affect users’ willingness to continue participation.

From a technological perspective, consumer-grade wearable devices (eg, Fitbit) often use closed algorithms that may obscure the true quality of the data [55], and device malfunctions can exacerbate data sparsity [18,44,47,65]. To address these challenges, future research could focus on 3 main areas: the development of adaptive reminder systems based on individual use patterns, the establishment of effective gamification incentives, and the optimization of lightweight device designs (eg, smart rings) [97–99]. When combined with intermittent active feedback methods such as ecological momentary assessment [100], these solutions may help establish a more robust multimodal data validation system.

Regarding data preprocessing, noise reduction techniques such as outlier detection based on statistical features and time-domain filtering are effective at reducing interference. However, due to the differing signal characteristics of various sensors (eg,

accelerometers, photoplethysmography, and ECG), there remains a lack of standardized practices in the field [101]. More importantly, few studies have systematically evaluated the actual impact of these preprocessing methods on the final model performance, resulting in a lack of empirical evidence to guide method selection.

The handling of missing data also presents challenges. Traditional methods such as mean imputation or KNN imputation can partially restore data completeness, but they may distort the true characteristics of physiological signals. On the other hand, directly deleting low-quality data may lead to the loss of critical clinical information [102]. Recent studies on motor activity data in schizophrenia have proposed a day-night difference-sensitive heuristic imputation method that preserves the key biological marker of nocturnal activity suppression while achieving a diagnostic accuracy of 94% [76]. This provides a reference framework for preprocessing similar data types.

In addition, techniques such as synthetic oversampling offer potential solutions for addressing class imbalance. However, the synthetic samples generated often fail to adequately reflect the complex temporal patterns in real data, leading to potential model overfitting or poor performance in practical applications [103,104]. A deeper challenge lies in the fact that many studies [15,19,45,46,48,51,57,73] overlook the details of time granularity, simplifying sensor data into daily summaries. This may miss micropatterns that signal symptom changes. For example, sleep fragmentation features in patients with bipolar disorder may be entirely masked in daily average sleep duration metrics [19].

The selection of window length, overlap rate, and feature selection strategies significantly impacts the performance and generalization ability of the model. The studies included in this review indicate that the window size in time-series analysis ranges from a few seconds to several days, and its choice directly affects signal integrity and model discriminability. For example, in high-frequency motion data analysis, a 2.5-second window effectively captures fine-grained behavior fluctuations such as walking and resting (with step count accuracy reaching 99.2%). However, it is also prone to interference from motion artifacts and other noise, requiring the application of a third-order Butterworth filter to enhance signal quality [17]. In contrast, a 48-hour window in circadian rhythm analysis can effectively capture periodic movement patterns. However, a long window may obscure short-term behavioral abnormalities in patients with depression—such as sudden decreases in activity or disruptions in movement patterns, which could be smoothed out by the 48-hour window, leading to the model missing transient features related to fluctuations in the patients’ condition [77]. This suggests that there is an *optimal threshold* for window length—shorter windows may capture fine-grained behavioral fluctuations but may amplify noise, whereas longer windows may mask transient psychological state changes, requiring dynamic adjustments based on the target behavior’s timescale.

The range of overlap rates also plays a critical role in model performance, balancing sample independence and data use efficiency. Khan et al [24] systematically analyzed the impact

of overlap rates (0%-90%) on time-series model performance in a study on anxiety-related behaviors. The results indicate that a high overlap rate increases sample density through resampling, improving the model's ability to capture brief behaviors, but it also introduces data redundancy, significantly increasing the risk of overfitting. A low overlap rate reduces sample correlation but may miss key state transitions. This research provides empirical evidence for the dynamic optimization of overlap rates in time-series analysis, highlighting the need to consider the target behavior's duration and the model's generalization requirements when adjusting parameters.

Differences in feature selection methods directly influence model complexity and interpretability. Variations in feature space search strategies, previous assumptions, and computational costs across different techniques can lead to significant divergence in the key variables captured by the model and their clinical relevance. For instance, brute force search, which exhaustively tests feature combinations (generating 14,892 models), can cover more potential patterns but may introduce redundant features, leading to overfitting and reduced interpretability [56]. In contrast, stepwise ML algorithms based on biological priors (such as circadian rhythm theory) progressively select features. While this sacrifices some model diversity, it allows the focus to remain on core variables with clinical mechanisms, and the resulting feature interpretability is stronger [15]. In cross-modal scenarios, embedded methods (eg, Depress-Deep Convolutional Neuro-Fuzzy combining CNN with statistical features) enhance accuracy to 85.1% by removing 88% of redundant features, whereas traditional SVM recursive feature elimination methods, reliant on manual feature engineering, perform more limitedly when multimodal data are missing [77]. Therefore, the choice of method should balance data-driven pattern discovery with theory-driven mechanism interpretation, optimizing based on the research goals (exploratory or confirmatory) and data characteristics (such as modality and completeness).

Associations Between Passive Features and Mental Disorders

The included studies suggest underlying patterns linking passive digital features to mental disorders; however, these findings are predominantly exploratory and require further validation. Notably, the significant methodological heterogeneity across studies—in terms of study quality, data acquisition methods, and reporting standards—limits the interpretability of pooled associations between digital features and affective states. Nevertheless, we would be remiss in ignoring the growing consensus on the relationship between digital features and emotional states. Consequently, we provide a synthesis of these preliminary findings, emphasize their preliminary nature, and urge readers to consider these limitations when interpreting the results.

In depression-related research, heart rate (20/42, 48%), movement index (15/42, 36%), and step count (14/42, 33%) were the most frequently reported features; these metrics correlate strongly with core depressive symptoms such as psychomotor retardation associated with anhedonia [105]. However, although smartphone use behaviors ranked highly in

feature importance analyses, only 17% (7/42) of the studies actually collected these data, likely reflecting data acquisition challenges—phone call records raise privacy concerns, whereas step counts can be easily extracted from commercial wearables [19]. In anxiety disorder research, physiological and sleep characteristics such as heart rate (5/42, 12%) and sleep duration (3/42, 7%) were commonly observed. However, the importance of environmental and behavioral factors such as time spent at home suggests that future studies should focus on enhancing the integration of environmental sensors. In schizophrenia research, the movement index (3/42, 7%) was predominant, which may be associated with motor impairments caused by antipsychotic medications [106].

Feature Engineering and Model Selection

Window segmentation lays the foundation for constructing stable data sequences by structuring inputs for feature extraction and, via overlapping windows, increasing the number of training samples to enhance model generalizability. However, there were variations in the selection of window sizes across the studies, and no unified guidelines have been established. In feature extraction, deep learning-based approaches have emerged as the dominant strategy because they automatically capture complex feature interactions and markedly improve predictive performance. CNNs in particular excel at automated feature extraction but rely on large-scale training datasets, increasing the risk of overfitting [107]. To mitigate this issue, previous studies have suggested that layerwise dropout regularization is an effective strategy, although the generalizability of such architecture optimization solutions across different datasets still requires further validation [47].

It is important to note that, although end-to-end deep models offer performance advantages, their *black box* nature limits their interpretability in clinical applications [108]. Some studies have introduced Shapley additive explanations [25,54] and integrated gradients [47] to enhance transparency, but core issues of model explainability persist. In some research, efforts to improve interpretability have led to oversimplified feature spaces (eg, retaining only 3 core features [65]). While this may facilitate clinical decision-making, it could result in the loss of multidimensional information essential for mental health monitoring. This highlights the ongoing challenge in model development to balance clinical interpretability with the need for sufficient representation of symptom heterogeneity.

In model selection, current studies retain the application of traditional ML methods (such as random forest, SVM, and logistic regression) while continuously exploring the potential of deep learning models. CNNs and their variants deliver superior performance for time-series data processing, nonlinear relationship modeling, and automated feature derivation; however, their substantial computational demands hinder deployment on edge and wearable devices, further intensifying the complexity-interpretability trade-off [24,47]. Autoencoders offer an unsupervised dimensionality reduction approach that opens new avenues for feature engineering. Notably, Baygin et al [72] developed a lightweight model based on a probabilistic binary pattern, which not only achieved >98.5% classification accuracy in ECG signal-based anxiety detection but also

significantly reduced computational load. Their model, with a limited but highly interpretable feature set, offers a new paradigm for balancing performance and interpretability.

Furthermore, there is mounting evidence suggesting that personalized models yield superior predictive accuracy. For example, individualized LSTM models explain up to 39% of within-subject variance in anxiety symptoms—far exceeding the explanatory power of traditional cohort-based models [18]. These findings highlight the critical importance of modeling individual differences in mental health prediction.

Clinical Translation and Ethical Considerations

Passive sensing data often involve highly sensitive personal information. However, only 14% (6/42) of the studies explicitly mentioned anonymization procedures [44,48,51,55,59,62] despite the fact that data anonymization and encryption are fundamental requirements for ensuring privacy and security. Moreover, as research may involve repeated monitoring and prolonged data collection, informing participants and obtaining their informed consent, as well as ensuring their right to withdraw consent for data use, are crucial issues that must be addressed when translating such technologies into clinical practice.

Insufficient sample representativeness is another significant limitation. Most studies (23/42, 55%) [53] relied on single-center or specific datasets (such as university student populations [14,18,45,49] or outpatient hospital patients [15,19,46,48,51,61,62,69]), which often exhibit high homogeneity in dimensions such as age, gender, and ethnicity, making it difficult to represent a broader population. This distribution imbalance may lead to decreased model performance in other populations and even trigger systemic biases. Future research should enhance model generalizability by expanding sample diversity and using statistical correction methods (such as multiple comparison corrections). In addition, insufficient model interpretability may undermine physicians' trust in algorithmic results, potentially affecting clinical decision-making [109].

The generalizability of current research findings is further limited by the lack of representation across cultural and socioeconomic backgrounds. Due to differences in cultural norms, stigma, and access to technology, mental health symptoms and behavioral patterns (eg, social interactions and mobility) may differ significantly across cultures [110]. For example, in some Asian cultures, the concept of *family honor* may lead individuals to reduce their use of public social applications (eg, social media engagement) due to psychological distress while increasing the use of private communication apps (eg, one-on-one messaging) [111]. In African American communities, distrust of mental health services may manifest as a preference for avoiding mental health monitoring applications that require sensitive data authorization or increased reliance on anonymous communication platforms that do not require real name registration, resulting in systematic bias in passively collected data such as location tracking and voice features [112]. In addition, socioeconomic disparities such as lower device ownership and unstable internet access in

low-income or rural communities often lead to sparse or noisy data, affecting data quality [113].

In addition, current research primarily validates the correlation between passive monitoring features and mental health, but due to the lack of causal inference, it remains unclear whether these features can serve as effective intervention targets or clinical prediction biomarkers [18,45,65]. Furthermore, most studies in this review (26/42, 62%) relied on government funding, with industry involvement at only 1% (1/69), which may delay the technology translation and practical application process. Moreover, most studies (39/42, 93%) focused on the analysis of a single disease, neglecting common comorbidities. For example, the comorbidity rate between depression and anxiety is as high as 25% [114]. Therefore, future algorithm designs should aim to identify cross-diagnostic features, thereby providing a more comprehensive assessment of an individual's mental health risks.

In terms of ethical deployment, addressing user-centered challenges is critical. Long-term adherence is often hindered by concerns about data privacy or perceived utility, especially in clinical populations. Transparent, dynamic consent interfaces and the provision of personalized feedback can effectively alleviate privacy concerns and enhance technology acceptance [115]. Furthermore, when integrating passive sensing technologies into existing health care workflows, it is essential to consider the digital literacy differences between clinicians and patients. Therefore, developing specialized training programs is crucial to ensure equitable adoption [116].

Although this study systematically reviewed the potential of passive sensing technologies in mental health monitoring, it must be acknowledged that the field currently lacks unified device approval standards, which has become a key bottleneck in the clinical translation of these technologies. Current regulatory frameworks primarily address traditional medical devices, whereas mental health monitoring technologies such as smartwatches and smartphone sensors do not yet have established uniform regulatory requirements due to their passive data collection methods, device heterogeneity, and diverse application contexts [117,118]. For example, while the US Food and Drug Administration has launched a pilot program for digital health technology precertification, specific standards for continuous mental health monitoring are still absent [119]. Similarly, the European Union's Medical Device Regulation includes AI algorithms but does not specify how to assess the reliability of passively collected data [120]. This situation leads to potential fluctuations in model performance due to device changes, making it challenging to meet the clinical diagnostic requirements for consistency and reproducibility.

Strengths and Limitations

To the best of our knowledge, this is the first scoping review assessing the current status and development trends of ML techniques in passive mental health monitoring based on wearable device and smartphone sensor data. Compared to previous similar studies, this review has the following advantages. First, it is the first to provide a comprehensive analysis of the entire technological process from data collection to model validation, addressing the gap in previous research

that only covered partial technological aspects. Second, by integrating data from 42 empirical studies, we explored the associative patterns between different mental disorders and behavioral traits, covering a broader range of disease types and feature dimensions than most similar reviews. In addition, this study highlights the limitations of existing methods in areas such as sample representativeness, data quality, and model generalizability while proposing paths for future technological improvements, clinical validation, and ethical practice. These contributions provide a more comprehensive assessment of the current state of methodologies in the field and offer technical references and practical insights for further optimizing the application of passive sensing technologies in mental health monitoring.

However, this study also has some limitations. First, in terms of literature search, while we used a cross-disciplinary database co-search strategy, the language and literature selection criteria (only including English-language journal articles and excluding conference papers and other nonjournal literature) may have led to the omission of important research from non-English-speaking countries and time-sensitive technological solutions from conference reports. Moreover, although MeSH (Medical Subject Headings) terms were considered, a keyword-based approach was prioritized to capture emerging terminology and maintain consistency across databases; this may have introduced some noise but ensured broader coverage of interdisciplinary literature. Second, regarding the algorithm performance comparisons presented in our results, it should be noted that these quantitative findings reflect model performance under specific study conditions. Due to substantial methodological variations in experimental designs, evaluation metrics, and data characteristics across the studies, these results are more suitable for descriptive analysis of current research trends rather than definitive conclusions about algorithmic superiority. Third, our findings may be partially influenced by the broader methodological limitations of the included studies, such as small sample sizes, potential overfitting of models, unclear generalizability of results, and differences in measurement metrics and outcome definitions across studies. These factors may affect the robustness and applicability of the conclusions. Moreover, although we strictly adhered to the PRISMA-ScR guidelines, applied clear inclusion and exclusion criteria, and systematically recorded risk of bias and other relevant factors, the broad scope of health outcomes and the differences in study designs somewhat limit our ability to comprehensively assess the overall quality of the studies [121].

Future Research Directions

Despite the significant potential of passive sensing technologies in mental health monitoring, substantial breakthroughs are still needed to translate exploratory research into clinically practical tools. On the basis of a systematic evaluation of current limitations and opportunities, we propose a comprehensive development road map aimed at addressing methodological, technical, clinical, and ethical challenges while promoting the translation of research outcomes into practice.

Enhancing Data Quality and Standardization

Current research is often limited by small sample sizes and short observation periods. Future studies should focus on large-scale longitudinal research with cohorts that encompass diverse population characteristics. Standardized protocols for data collection, preprocessing, and reporting should be established to ensure comparability and reproducibility across studies. In addition, extending the monitoring window is crucial for capturing the dynamic evolution of mental health states. There is also a need to develop cross-platform open databases to facilitate data sharing among researchers and provide support for algorithm validation and cross-study integration.

Facilitating Clinical Translation and Ethical Practices

Integrating passive sensing technologies into clinical practice requires rigorous external validation in real-world health care settings, along with the design of user interfaces that are compatible with clinical workflows. Establishing clear regulatory approval pathways and clinical application standards is essential for transitioning technology from research to application. Future research should collaborate with regulatory bodies (eg, the Food and Drug Administration and International Medical Device Regulators Forum), clinical experts, and technology developers to jointly establish approval standards for passive mental health monitoring. These standards should define device performance parameters, data quality control processes (eg, setting thresholds for missing data), and clinical validation criteria to expedite the safe and reliable integration of passive sensing technologies into real-world health care environments [122]. Implementing scientific approaches to understand clinicians' adoption barriers and patients' willingness to use these technologies will enhance acceptance. From an ethical perspective, privacy protection measures (such as differential privacy [123] and dynamic informed consent mechanisms [115]), bias detection, and fairness assurances must be enforced to ensure the fairness of algorithms across different genders, races, and socioeconomic groups.

Optimizing Algorithm Architecture and Interpretability

Future algorithm development should seek an optimal balance among model complexity, interpretability, and computational efficiency. By combining deep learning with explainable AI techniques, transparent and efficient hybrid models can be designed. In addition, solutions such as lightweight models should be explored to improve algorithm efficiency while maintaining predictive performance. The adoption of frameworks such as federated learning should prioritize the development of personalized modeling approaches to capture individual differences in symptom manifestations. Furthermore, introducing causal inference methods will help identify digital biomarkers with true clinical intervention value from passive data, overcoming the limitations of current correlation-based research.

Promoting Interdisciplinary Collaboration and Technological Implementation

The successful implementation of these technologies depends on deep interdisciplinary collaboration among clinical medicine, data science, engineering, ethics, and policy making [108]. It

is recommended to form interdisciplinary alliances involving psychiatrists, data scientists, ethicists, and biomedical engineers to jointly establish technical standards and ethical guidelines (such as those from international organizations such as the International Medical Device Regulators Forum [124]). Strengthening strategic collaborations among governments, industry, and academia is also critical. Only with the dual support of policy and market forces can the technology be scaled and eventually translated into practical tools that enhance public health and improve patient outcomes.

Strengthening Cross-Cultural Validation and Inclusive Research

Future research should prioritize inclusive recruitment strategies such as collaborating with community health organizations, using multilingual sensing platforms, and validating models within low-income or rural populations. Causal frameworks for cross-cultural studies (eg, standardized data generation mechanisms and multigroup stratified analysis) provide a methodological foundation for cross-cultural validation in global digital phenotyping research. By systematically integrating cultural differences, these approaches can enhance the translational relevance of research findings across diverse populations [125].

Conclusions

This study provides a comprehensive assessment of the current status, technical pathways, and future directions of passive

sensing technologies based on wearable devices and smartphones combined with ML for mental health monitoring. We found that passive features such as sleep, physical activity, physiological signals, and social behavior showed stable associations with various mental disorders, including depression, anxiety, bipolar disorder, and schizophrenia. Among these features, heart rate, movement index, and step count were the most commonly used indicators. At the algorithmic level, deep learning models (eg, CNN and LSTM) demonstrated excellent performance in mental health prediction due to their robust temporal data processing capabilities. In contrast, traditional methods such as random forest and XGBoost continue to play a significant role in clinical research due to their strong interpretability.

Despite these advancements, challenges such as small sample sizes, short monitoring durations, limited device types, data noise and missing values, lack of model interpretability, and insufficient external validation hinder the generalizability and clinical translation of these studies. To advance the field, future research should focus on large-scale, multicenter, longitudinal studies across diverse populations; establish standardized protocols for data collection and preprocessing; strengthen multimodal integration and personalized modeling; and introduce explainable AI and privacy protection mechanisms to ensure the safe and efficient application of passive sensing technologies in real-world health care environments.

Acknowledgments

This research was supported by the Medical and Health Technology Plan of Zhejiang Province (2022507615); Key Research Project for Laboratory Work in Zhejiang Province Colleges (ZD202202); Zhejiang Province Traditional Chinese Medicine Inheritance and Innovation Project (2023ZX0950); 2024 Research Project of Engineering Research Center of Mobile Health Management System, Ministry of Education; and 2022 Zhejiang Province First-Class Undergraduate Courses, Zhejiang Provincial Department of Education (1133).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SS and WQ made equal contributions to the study. SYS was responsible for literature retrieval and screening, drafting the initial manuscript, and contributing significantly to the manuscript's structure and content. WQ handled the literature analysis and integration, provided the research framework, and made key revisions to the final manuscript. JZ and SL primarily contributed to the literature analysis and discussion, assisting in organizing the theoretical background of the paper and revising relevant sections. XL and XZ assisted in literature collection and screening, contributing to the analysis of some of the literature and writing tasks. CD, JW, GJ, and BW participated in the structural design of the manuscript, helped with in-depth literature analysis, and made important contributions to the final review and revision of the manuscript. YS, PL, and JY assisted with the design and argumentation of the research framework and helped organize relevant literature. BW, SG, and XL primarily contributed to the proofreading and revision of the manuscript and provided key professional insights throughout the study. SC provided guidance on the overall design of the review, the analytical framework for the literature, and the final revision and editing of the manuscript, taking on the crucial role of manuscript review and approval.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-ScR checklist.

[\[DOCX File , 20 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Search strategy.

[\[DOCX File , 20 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Detailed definitions and classifications of behavioral categories.

[\[DOCX File , 36 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Detailed information on the included studies.

[\[XLSX File \(Microsoft Excel File\), 27 KB-Multimedia Appendix 4\]](#)

References

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. *Lancet Psychiatry*. Feb 2016;3(2):171-178. [doi: [10.1016/s2215-0366\(15\)00505-2](#)]
2. Depression and other common mental disorders. World Health Organization. Jan 03, 2017. URL: <https://www.who.int/publications/i/item/depression-global-health-estimates> [accessed 2025-07-29]
3. Mental health ATLAS 2020. World Health Organization. Oct 08, 2021. URL: <https://www.who.int/publications/i/item/9789240036703> [accessed 2025-07-29]
4. Mittal VA, Walker EF. Diagnostic and statistical manual of mental disorders. *Psychiatry Res*. Aug 30, 2011;189(1):158-159. [FREE Full text] [doi: [10.1016/j.psychres.2011.06.006](#)] [Medline: [21741095](#)]
5. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. Sep 2001;16(9):606-613. [FREE Full text] [doi: [10.1046/j.1525-1497.2001.016009606.x](#)] [Medline: [11556941](#)]
6. Garb HN. Clinical judgment and decision making. *Annu Rev Clin Psychol*. 2005;1:67-89. [FREE Full text] [doi: [10.1146/annurev.clinpsy.1.102803.143810](#)] [Medline: [17716082](#)]
7. Trull TJ, Ebner-Priemer U. Ambulatory assessment. *Annu Rev Clin Psychol*. 2013;9:151-176. [FREE Full text] [doi: [10.1146/annurev-clinpsy-050212-185510](#)] [Medline: [23157450](#)]
8. Pantelis C, Bartholomeusz CF. Social neuroscience in psychiatry: pathways to discovering neurobiological risk and resilience. *World Psychiatry*. Jun 2014;13(2):146-147. [FREE Full text] [doi: [10.1002/wps.20123](#)] [Medline: [24890063](#)]
9. Kuppens P, Verduyn P. Emotion dynamics. *Curr Opin Psychol*. Oct 2017;17:22-26. [FREE Full text] [doi: [10.1016/j.copsyc.2017.06.004](#)] [Medline: [28950968](#)]
10. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. *BMC Psychiatry*. Dec 30, 2010;10:113. [FREE Full text] [doi: [10.1186/1471-244x-10-113](#)]
11. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA*. Oct 03, 2017;318(13):1215-1216. [FREE Full text] [doi: [10.1001/jama.2017.11295](#)] [Medline: [28973224](#)]
12. Ebner-Priemer U, Santangelo P. Digital phenotyping: hype or hope? *Lancet Psychiatry*. Apr 2020;7(4):297-299. [FREE Full text] [doi: [10.1016/s2215-0366\(19\)30380-3](#)]
13. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology*. Jun 2016;41(7):1691-1696. [FREE Full text] [doi: [10.1038/npp.2016.7](#)] [Medline: [26818126](#)]
14. Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. *J Med Internet Res*. May 29, 2020;22(5):e16875. [FREE Full text] [doi: [10.2196/16875](#)] [Medline: [32348284](#)]
15. Ikäheimonen A, Luong N, Baryshnikov I, Darst R, Heikkilä R, Holmen J, et al. Predicting and monitoring symptoms in patients diagnosed with depression using smartphone data: observational study. *J Med Internet Res*. Dec 03, 2024;26:e56874. [FREE Full text] [doi: [10.2196/56874](#)] [Medline: [39626241](#)]
16. Karlgren K, Mcmillan D. Sleep planning with Awari: uncovering the materiality of body rhythms using research through design. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023. Presented at: CHI '23; April 23-28, 2023; Hamburg, Germany. [doi: [10.1145/3544548.3581502](#)]
17. Masud MT, Mamun MA, Thapa K, Lee DH, Griffiths MD, Yang SH. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. *J Biomed Inform*. Mar 2020;103:103371. [FREE Full text] [doi: [10.1016/j.jbi.2019.103371](#)] [Medline: [31935462](#)]
18. Jacobson NC, Bhattacharya S. Digital biomarkers of anxiety disorder symptom changes: personalized deep learning models using smartphone sensors accurately predict anxiety symptoms from ecological momentary assessments. *Behav Res Ther*. Feb 2022;149:104013. [FREE Full text] [doi: [10.1016/j.brat.2021.104013](#)] [Medline: [35030442](#)]

19. Lipschitz JM, Lin S, Saghaian S, Pike CK, Burdick KE. Digital phenotyping in bipolar disorder: using longitudinal Fitbit data and personalized machine learning to predict mood symptomatology. *Acta Psychiatr Scand*. Mar 2025;151(3):434-447. [FREE Full text] [doi: [10.1111/acps.13765](https://doi.org/10.1111/acps.13765)] [Medline: [39397313](https://pubmed.ncbi.nlm.nih.gov/39397313/)]
20. Lane ND, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell AT. A survey of mobile phone sensing. *IEEE Commun Mag*. Sep 2010;48(9):140-150. [FREE Full text] [doi: [10.1109/mcom.2010.5560598](https://doi.org/10.1109/mcom.2010.5560598)]
21. Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: a survey. *Pervasive Mob Comput*. Dec 2018;51:1-26. [FREE Full text] [doi: [10.1016/j.pmcj.2018.09.003](https://doi.org/10.1016/j.pmcj.2018.09.003)]
22. Opoku Asare K, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR Mhealth Uhealth*. Jul 12, 2021;9(7):e26540. [FREE Full text] [doi: [10.2196/26540](https://doi.org/10.2196/26540)] [Medline: [34255713](https://pubmed.ncbi.nlm.nih.gov/34255713/)]
23. Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med*. Feb 12, 2019;49(09):1426-1448. [FREE Full text] [doi: [10.1017/s0033291719000151](https://doi.org/10.1017/s0033291719000151)]
24. Khan NS, Ghani MS, Anjum G. ADAM-sense: anxiety-displaying activities recognition by motion sensors. *Pervasive Mob Comput*. Dec 2021;78:101485. [FREE Full text] [doi: [10.1016/j.pmcj.2021.101485](https://doi.org/10.1016/j.pmcj.2021.101485)]
25. Misgar MM, Bhatia MP. Advancing ADHD diagnosis: using machine learning for unveiling ADHD patterns through dimensionality reduction on IoMT actigraphy signals. *Int J Inf Technol*. May 07, 2024. [FREE Full text] [doi: [10.1007/s41870-024-01895-x](https://doi.org/10.1007/s41870-024-01895-x)]
26. Choi A, Ooi A, Lottridge D. Digital phenotyping for stress, anxiety, and mild depression: systematic literature review. *JMIR Mhealth Uhealth*. May 23, 2024;12:e40689. [FREE Full text] [doi: [10.2196/40689](https://doi.org/10.2196/40689)] [Medline: [38780995](https://pubmed.ncbi.nlm.nih.gov/38780995/)]
27. Hassan L, Milton A, Sawyer C, Casson AJ, Torous J, Davies A, et al. Utility of consumer-grade wearable devices for inferring physical and mental health outcomes in severe mental illness: systematic review. *JMIR Ment Health*. Jan 07, 2025;12:e65143. [FREE Full text] [doi: [10.2196/65143](https://doi.org/10.2196/65143)] [Medline: [39773905](https://pubmed.ncbi.nlm.nih.gov/39773905/)]
28. Bladon S, Eisner E, Bucci S, Oluwatayo A, Martin GP, Sperrin M, et al. A systematic review of passive data for remote monitoring in psychosis and schizophrenia. *NPJ Digit Med*. Jan 27, 2025;8(1):62. [FREE Full text] [doi: [10.1038/s41746-025-01451-2](https://doi.org/10.1038/s41746-025-01451-2)] [Medline: [39870797](https://pubmed.ncbi.nlm.nih.gov/39870797/)]
29. Khoo LS, Lim MK, Chong CY, McNaney R. Machine learning for multimodal mental health detection: a systematic review of passive sensing approaches. *Sensors (Basel)*. Jan 06, 2024;24(2):348. [FREE Full text] [doi: [10.3390/s24020348](https://doi.org/10.3390/s24020348)] [Medline: [38257440](https://pubmed.ncbi.nlm.nih.gov/38257440/)]
30. Abd-Alrazaq A, AlSaad R, Aziz S, Ahmed A, Denecke K, Househ M, et al. Wearable artificial intelligence for anxiety and depression: scoping review. *J Med Internet Res*. Jan 19, 2023;25:e42672. [FREE Full text] [doi: [10.2196/42672](https://doi.org/10.2196/42672)] [Medline: [36656625](https://pubmed.ncbi.nlm.nih.gov/36656625/)]
31. Abd-Alrazaq A, Alajlani M, Ahmad R, AlSaad R, Aziz S, Ahmed A, et al. The performance of wearable AI in detecting stress among students: systematic review and meta-analysis. *J Med Internet Res*. Jan 31, 2024;26:e52622. [FREE Full text] [doi: [10.2196/52622](https://doi.org/10.2196/52622)] [Medline: [38294846](https://pubmed.ncbi.nlm.nih.gov/38294846/)]
32. Abd-Alrazaq A, AlSaad R, Shuweihi F, Ahmed A, Aziz S, Sheikh J. Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. *NPJ Digit Med*. May 05, 2023;6(1):84. [FREE Full text] [doi: [10.1038/s41746-023-00828-5](https://doi.org/10.1038/s41746-023-00828-5)] [Medline: [37147384](https://pubmed.ncbi.nlm.nih.gov/37147384/)]
33. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. *NPJ Digit Med*. Jan 11, 2022;5(1):3. [FREE Full text] [doi: [10.1038/s41746-021-00548-8](https://doi.org/10.1038/s41746-021-00548-8)] [Medline: [35017634](https://pubmed.ncbi.nlm.nih.gov/35017634/)]
34. Park Y, Park S, Lee M. Effectiveness of artificial intelligence in detecting and managing depressive disorders: systematic review. *J Affect Disord*. Sep 15, 2024;361:445-456. [FREE Full text] [doi: [10.1016/j.jad.2024.06.035](https://doi.org/10.1016/j.jad.2024.06.035)] [Medline: [38889858](https://pubmed.ncbi.nlm.nih.gov/38889858/)]
35. Saccaro LF, Amatori G, Cappelli A, Mazzotti R, Dell'Osso L, Rutigliano G. Portable technologies for digital phenotyping of bipolar disorder: a systematic review. *J Affect Disord*. Dec 01, 2021;295:323-338. [FREE Full text] [doi: [10.1016/j.jad.2021.08.052](https://doi.org/10.1016/j.jad.2021.08.052)] [Medline: [34488086](https://pubmed.ncbi.nlm.nih.gov/34488086/)]
36. Gomes N, Pato M, Lourenço AR, Datia N. A survey on wearable sensors for mental health monitoring. *Sensors (Basel)*. Jan 25, 2023;23(3):1330. [FREE Full text] [doi: [10.3390/s23031330](https://doi.org/10.3390/s23031330)] [Medline: [36772370](https://pubmed.ncbi.nlm.nih.gov/36772370/)]
37. Maatoug R, Oudin A, Adrien V, Saudreau B, Bonnot O, Millet B, et al. Digital phenotype of mood disorders: a conceptual and critical review. *Front Psychiatry*. Jul 26, 2022;13:895860. [FREE Full text] [doi: [10.3389/fpsy.2022.895860](https://doi.org/10.3389/fpsy.2022.895860)] [Medline: [35958638](https://pubmed.ncbi.nlm.nih.gov/35958638/)]
38. Rahman RA, Omar K, Mohd Noah SA, Danuri MS, Al-Garadi MA. Application of machine learning methods in mental health detection: a systematic review. *IEEE Access*. Oct 06, 2020;8:183952-183964. [FREE Full text] [doi: [10.1109/access.2020.3029154](https://doi.org/10.1109/access.2020.3029154)]
39. Tricco A, Lillie E, Zarin W, O'Brien K, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. Oct 02, 2018;169(7):467-473. [FREE Full text] [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
40. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring and Bethesda, MD. National Institutes of Health and Food and Drug Administration; 2016.

41. Gong Y, Poellabauer C. Topic modeling based multi-modal depression detection. In: Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge. 2017. Presented at: AVEC '17; October 23, 2017; Mountain View, CA. [doi: [10.1145/3133944.3133945](https://doi.org/10.1145/3133944.3133945)]
42. van Praag HM. Can stress cause depression? *Prog Neuropsychopharmacol Biol Psychiatry*. Aug 2004;28(5):891-907. [doi: [10.1016/j.pnpbp.2004.05.031](https://doi.org/10.1016/j.pnpbp.2004.05.031)] [Medline: [15363612](https://pubmed.ncbi.nlm.nih.gov/15363612/)]
43. Power MJ, Tarsia M. Basic and complex emotions in depression and anxiety. *Clin Psychol Psychother*. Jan 29, 2007;14(1):19-31. [doi: [10.1002/cpp.515](https://doi.org/10.1002/cpp.515)]
44. Anmella G, Corponi F, Li B, Mas A, Sanabra M, Pacchiarotti I, et al. Exploring digital biomarkers of illness activity in mood episodes: hypotheses generating and model development study. *JMIR Mhealth Uhealth*. May 04, 2023;11:e45405. [FREE Full text] [doi: [10.2196/45405](https://doi.org/10.2196/45405)] [Medline: [36939345](https://pubmed.ncbi.nlm.nih.gov/36939345/)]
45. Boukhechba M, Chow P, Fua K, Teachman BA, Barnes LE. Predicting social anxiety from global positioning system traces of college students: feasibility study. *JMIR Ment Health*. Jul 04, 2018;5(3):e10101. [FREE Full text] [doi: [10.2196/10101](https://doi.org/10.2196/10101)] [Medline: [29973337](https://pubmed.ncbi.nlm.nih.gov/29973337/)]
46. Chikersal P, Doryab A, Tumminia M, Villalba DK, Dutcher JM, Liu X, et al. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing. *ACM Trans Comput Hum Interact*. Jan 20, 2021;28(1):1-41. [FREE Full text] [doi: [10.1145/3422821](https://doi.org/10.1145/3422821)]
47. Espino-Salinas CH, Galván-Tejada CE, Luna-García H, Gamboa-Rosales H, Celaya-Padilla JM, Zanella-Calzada LA, et al. Two-dimensional convolutional neural network for depression episodes detection in real time using motor activity time series of Depresjon dataset. *Bioengineering (Basel)*. Sep 09, 2022;9(9):458. [FREE Full text] [doi: [10.3390/bioengineering9090458](https://doi.org/10.3390/bioengineering9090458)] [Medline: [36135004](https://pubmed.ncbi.nlm.nih.gov/36135004/)]
48. Grunerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Troster G, et al. Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE J Biomed Health Inform*. Jan 2015;19(1):140-148. [FREE Full text] [doi: [10.1109/jbhi.2014.2343154](https://doi.org/10.1109/jbhi.2014.2343154)]
49. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors (Basel)*. Jun 24, 2020;20(12):3572. [FREE Full text] [doi: [10.3390/s20123572](https://doi.org/10.3390/s20123572)] [Medline: [32599801](https://pubmed.ncbi.nlm.nih.gov/32599801/)]
50. Jacobson NC, Lekkas D, Huang R, Thomas N. Deep learning paired with wearable passive sensing data predicts deterioration in anxiety disorder symptoms across 17-18 years. *J Affect Disord*. Mar 01, 2021;282:104-111. [FREE Full text] [doi: [10.1016/j.jad.2020.12.086](https://doi.org/10.1016/j.jad.2020.12.086)] [Medline: [33401123](https://pubmed.ncbi.nlm.nih.gov/33401123/)]
51. Kim JS, Wang B, Kim M, Lee J, Kim H, Roh D, et al. Prediction of diagnosis and treatment response in adolescents with depression by using a smartphone app and deep learning approaches: usability study. *JMIR Form Res*. May 24, 2023;7:e45991. [FREE Full text] [doi: [10.2196/45991](https://doi.org/10.2196/45991)] [Medline: [37223978](https://pubmed.ncbi.nlm.nih.gov/37223978/)]
52. Mullick T, Radovic A, Shaaban S, Doryab A. Predicting depression in adolescents using mobile and wearable sensors: multimodal machine learning-based exploratory study. *JMIR Form Res*. Jun 24, 2022;6(6):e35807. [FREE Full text] [doi: [10.2196/35807](https://doi.org/10.2196/35807)] [Medline: [35749157](https://pubmed.ncbi.nlm.nih.gov/35749157/)]
53. Pedrelli P, Fedor S, Ghandeharioun A, Howe E, Ionescu DF, Bhatena D, et al. Monitoring changes in depression severity using wearable and mobile sensors. *Front Psychiatry*. Dec 18, 2020;11:584711. [FREE Full text] [doi: [10.3389/fpsy.2020.584711](https://doi.org/10.3389/fpsy.2020.584711)] [Medline: [33391050](https://pubmed.ncbi.nlm.nih.gov/33391050/)]
54. Sadeghi M, McDonald AD, Sasangohar F. Posttraumatic stress disorder hyperarousal event detection using smartwatch physiological and activity data. *PLoS One*. 2022;17(5):e0267749. [FREE Full text] [doi: [10.1371/journal.pone.0267749](https://doi.org/10.1371/journal.pone.0267749)] [Medline: [35584096](https://pubmed.ncbi.nlm.nih.gov/35584096/)]
55. Saito T, Suzuki H, Kishi A. Predictive modeling of mental illness onset using wearable devices and medical examination data: machine learning approach. *Front Digit Health*. 2022;4:861808. [FREE Full text] [doi: [10.3389/fdgh.2022.861808](https://doi.org/10.3389/fdgh.2022.861808)] [Medline: [35493532](https://pubmed.ncbi.nlm.nih.gov/35493532/)]
56. Spulber S, Elberling F, Svensson J, Tiger M, Ceccatelli S, Lundberg J. Patterns of activity correlate with symptom severity in major depressive disorder patients. *Transl Psychiatry*. Jun 02, 2022;12(1):226. [FREE Full text] [doi: [10.1038/s41398-022-01989-9](https://doi.org/10.1038/s41398-022-01989-9)] [Medline: [35654778](https://pubmed.ncbi.nlm.nih.gov/35654778/)]
57. Tazawa Y, Liang KC, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, et al. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. *Heliyon*. Mar 2020;6(2):e03274. [FREE Full text] [doi: [10.1016/j.heliyon.2020.e03274](https://doi.org/10.1016/j.heliyon.2020.e03274)] [Medline: [32055728](https://pubmed.ncbi.nlm.nih.gov/32055728/)]
58. Tseng VW, Sano A, Ben-Zeev D, Brian R, Campbell AT, Hauser M, et al. Using behavioral rhythms and multi-task learning to predict fine-grained symptoms of schizophrenia. *Sci Rep*. Sep 15, 2020;10(1):15100. [FREE Full text] [doi: [10.1038/s41598-020-71689-1](https://doi.org/10.1038/s41598-020-71689-1)] [Medline: [32934246](https://pubmed.ncbi.nlm.nih.gov/32934246/)]
59. Ware S, Yue C, Morillo R, Lu J, Shang C, Bi J, et al. Predicting depressive symptoms using smartphone data. *Smart Health*. Mar 2020;15:100093. [FREE Full text] [doi: [10.1016/j.smhl.2019.100093](https://doi.org/10.1016/j.smhl.2019.100093)]
60. Ahmed A, Ramesh J, Ganguly S, Aburukba R, Sagahyroon A, Aloul F. Investigating the feasibility of assessing depression severity and valence-arousal with wearable sensors using discrete wavelet transforms and machine learning. *Information*. Aug 27, 2022;13(9):406. [FREE Full text] [doi: [10.3390/info13090406](https://doi.org/10.3390/info13090406)]

61. Bai R, Xiao L, Guo Y, Zhu X, Li N, Wang Y, et al. Tracking and monitoring mood stability of patients with major depressive disorder by machine learning models using passive digital data: prospective naturalistic multicenter study. *JMIR Mhealth Uhealth*. Mar 08, 2021;9(3):e24365. [FREE Full text] [doi: [10.2196/24365](https://doi.org/10.2196/24365)] [Medline: [33683207](https://pubmed.ncbi.nlm.nih.gov/33683207/)]
62. Barrigon ML, Romero-Medrano L, Moreno-Muñoz P, Porras-Segovia A, Lopez-Castroman J, Courtet P, et al. One-week suicide risk prediction using real-time smartphone monitoring: prospective cohort study. *J Med Internet Res*. Sep 01, 2023;25:e43719. [FREE Full text] [doi: [10.2196/43719](https://doi.org/10.2196/43719)] [Medline: [37656498](https://pubmed.ncbi.nlm.nih.gov/37656498/)]
63. Rodríguez-Ruiz JG, Galván-Tejada CE, Luna-García H, Gamboa-Rosales H, Celaya-Padilla JM, Arceo-Olague JG, et al. Classification of depressive and schizophrenic episodes using night-time motor activity signal. *Healthcare (Basel)*. Jul 05, 2022;10(7):1256. [FREE Full text] [doi: [10.3390/healthcare10071256](https://doi.org/10.3390/healthcare10071256)] [Medline: [35885784](https://pubmed.ncbi.nlm.nih.gov/35885784/)]
64. Jacobson N, Summers B, Wilhelm S. Digital Biomarkers of Social Anxiety Severity: Digital Phenotyping Using Passive Smartphone Sensors. *J Med Internet Res*. May 29, 2020;22(5):e16875. [FREE Full text] [doi: [10.2196/16875](https://doi.org/10.2196/16875)] [Medline: [32348284](https://pubmed.ncbi.nlm.nih.gov/32348284/)]
65. Choi H, Cho Y, Min C, Kim K, Kim E, Lee S, et al. Multiclassification of the symptom severity of social anxiety disorder using digital phenotypes and feature representation learning. *Digit Health*. 2024;10:20552076241256730. [FREE Full text] [doi: [10.1177/20552076241256730](https://doi.org/10.1177/20552076241256730)] [Medline: [39114113](https://pubmed.ncbi.nlm.nih.gov/39114113/)]
66. Jacobson NC, Feng B. Digital phenotyping of generalized anxiety disorder: using artificial intelligence to accurately predict symptom severity using wearable sensors in daily life. *Transl Psychiatry*. Aug 17, 2022;12(1):336. [FREE Full text] [doi: [10.1038/s41398-022-02038-1](https://doi.org/10.1038/s41398-022-02038-1)] [Medline: [35977932](https://pubmed.ncbi.nlm.nih.gov/35977932/)]
67. Misgar MM, Bhatia MP. Hopping-mean: an augmentation method for motor activity data towards real-time depression diagnosis using machine learning. *Multimed Tools Appl*. Jul 11, 2024;84(18):18781-18799. [FREE Full text] [doi: [10.1007/s11042-024-19631-9](https://doi.org/10.1007/s11042-024-19631-9)]
68. Price GD, Heinz MV, Zhao D, Nemesure M, Ruan F, Jacobson NC. An unsupervised machine learning approach using passive movement data to understand depression and schizophrenia. *J Affect Disord*. Nov 01, 2022;316:132-139. [FREE Full text] [doi: [10.1016/j.jad.2022.08.013](https://doi.org/10.1016/j.jad.2022.08.013)] [Medline: [35964770](https://pubmed.ncbi.nlm.nih.gov/35964770/)]
69. Cote-Allard U, Jakobsen P, Stautland A, Nordgreen T, Fasmer OB, Oedegaard KJ, et al. Long-short ensemble network for bipolar manic-euthymic state recognition based on wrist-worn sensors. *IEEE Pervasive Comput*. Apr 1, 2022;21(2):20-31. [FREE Full text] [doi: [10.1109/mpmv.2022.3155728](https://doi.org/10.1109/mpmv.2022.3155728)]
70. Arsalan A, Majid M. A study on multi-class anxiety detection using wearable EEG headband. *J Ambient Intell Human Comput*. Apr 11, 2021;13(12):5739-5749. [FREE Full text] [doi: [10.1007/s12652-021-03249-y](https://doi.org/10.1007/s12652-021-03249-y)]
71. Price GD, Heinz MV, Collins AC, Jacobson NC. Detecting major depressive disorder presence using passively-collected wearable movement data in a nationally-representative sample. *Psychiatry Res*. Mar 2024;332:115693. [FREE Full text] [doi: [10.1016/j.psychres.2023.115693](https://doi.org/10.1016/j.psychres.2023.115693)] [Medline: [38194801](https://pubmed.ncbi.nlm.nih.gov/38194801/)]
72. Baygin M, Barua PD, Dogan S, Tuncer T, Hong TJ, March S, et al. Automated anxiety detection using probabilistic binary pattern with ECG signals. *Comput Methods Programs Biomed*. Apr 2024;247:108076. [FREE Full text] [doi: [10.1016/j.cmpb.2024.108076](https://doi.org/10.1016/j.cmpb.2024.108076)] [Medline: [38422891](https://pubmed.ncbi.nlm.nih.gov/38422891/)]
73. Ware S, Yue C, Morillo R, Shang C, Bi J, Kamath J, et al. Automatic depression screening using social interaction data on smartphones. *Smart Health*. Dec 2022;26:100356. [FREE Full text] [doi: [10.1016/j.smhl.2022.100356](https://doi.org/10.1016/j.smhl.2022.100356)]
74. Hassantabar S, Zhang J, Yin H, Jha NK. MHDeep: mental health disorder detection system based on wearable sensors and artificial neural networks. *ACM Trans Embed Comput Syst*. Dec 12, 2022;21(6):1-22. [FREE Full text] [doi: [10.1145/3527170](https://doi.org/10.1145/3527170)]
75. Adler DA, Ben-Zeev D, Tseng VW, Kane J, Brian R, Campbell AT, et al. Predicting early warning signs of psychotic relapse from passive sensing data: an approach using encoder-decoder neural networks. *JMIR Mhealth Uhealth*. Aug 31, 2020;8(8):e19962. [FREE Full text] [doi: [10.2196/19962](https://doi.org/10.2196/19962)] [Medline: [32865506](https://pubmed.ncbi.nlm.nih.gov/32865506/)]
76. Misgar MM, Bhatia MP. Utilizing deep convolutional neural architecture with attention mechanism for objective diagnosis of schizophrenia using wearable IoMT devices. *Multimed Tools Appl*. Oct 03, 2023;83(13):39601-39620. [FREE Full text] [doi: [10.1007/s11042-023-17119-6](https://doi.org/10.1007/s11042-023-17119-6)]
77. Kumar A, Sangwan SR, Arora A, Menon VG. Depress-DCNF: a deep convolutional neuro-fuzzy model for detection of depression episodes using IoMT. *Appl Soft Comput*. Jun 2022;122:108863. [FREE Full text] [doi: [10.1016/j.asoc.2022.108863](https://doi.org/10.1016/j.asoc.2022.108863)]
78. Habib A, Vaniya SN, Khandoker A, Karmakar C. MDDBranchNet: a deep learning model for detecting major depressive disorder using ECG signal. *IEEE J Biomed Health Inform*. Jul 2024;28(7):3798-3809. [FREE Full text] [doi: [10.1109/jbhi.2024.3390847](https://doi.org/10.1109/jbhi.2024.3390847)]
79. Sato S, Hiratsuka T, Hasegawa K, Watanabe K, Obara Y, Kariya N, et al. Screening for major depressive disorder using a wearable ultra-short-term HRV monitor and signal quality indices. *Sensors (Basel)*. Apr 10, 2023;23(8):3867. [FREE Full text] [doi: [10.3390/s23083867](https://doi.org/10.3390/s23083867)] [Medline: [37112208](https://pubmed.ncbi.nlm.nih.gov/37112208/)]
80. Lipschitz J, Lin S, Saghaian S, Pike C, Burdick KE. Digital phenotyping in bipolar disorder: Using longitudinal Fitbit data and personalized machine learning to predict mood symptomatology. *Acta Psychiatr Scand* 2024. 2025. URL: <http://doi.org/10.1111/acps.13765> [accessed 2025-07-29]

81. Guk K, Han G, Lim J, Jeong K, Kang T, Lim EK, et al. Evolution of wearable devices with real-time disease monitoring for personalized healthcare. *Nanomaterials* (Basel). May 29, 2019;9(6):813. [FREE Full text] [doi: [10.3390/nano9060813](https://doi.org/10.3390/nano9060813)] [Medline: [31146479](https://pubmed.ncbi.nlm.nih.gov/31146479/)]
82. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. *Front Physiol*. 2018;9:743. [FREE Full text] [doi: [10.3389/fphys.2018.00743](https://doi.org/10.3389/fphys.2018.00743)] [Medline: [30002629](https://pubmed.ncbi.nlm.nih.gov/30002629/)]
83. Liu Y, Hankey J, Cao B, Chokka P. Screening for major depressive disorder in a tertiary mental health centre using EarlyDetect: a machine learning-based pilot study. *J Affect Disord Rep*. Jan 2021;3:e100062. [FREE Full text] [doi: [10.1016/j.jadr.2020.100062](https://doi.org/10.1016/j.jadr.2020.100062)]
84. Saylam B, İncel Ö. Multitask learning for mental health: depression, anxiety, stress (DAS) using wearables. *Diagnostics* (Basel). Mar 26, 2024;14(5):501. [FREE Full text] [doi: [10.3390/diagnostics14050501](https://doi.org/10.3390/diagnostics14050501)] [Medline: [38472973](https://pubmed.ncbi.nlm.nih.gov/38472973/)]
85. Bipolar disorder: assessment and management. National Institute for Health and Care Excellence. Sep 24, 2014. URL: <https://www.nice.org.uk/guidance/cg185> [accessed 2025-07-30]
86. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. Jan 2010;21(1):128-138. [FREE Full text] [doi: [10.1097/EDE.0b013e3181c30fb2](https://doi.org/10.1097/EDE.0b013e3181c30fb2)] [Medline: [20010215](https://pubmed.ncbi.nlm.nih.gov/20010215/)]
87. Shen N, Sequeira L, Silver MP, Carter-Langford A, Strauss J, Wiljer D. Patient privacy perspectives on health information exchange in a mental health context: qualitative study. *JMIR Ment Health*. Nov 13, 2019;6(11):e13306. [FREE Full text] [doi: [10.2196/13306](https://doi.org/10.2196/13306)] [Medline: [31719029](https://pubmed.ncbi.nlm.nih.gov/31719029/)]
88. Spake L, Hassan A, Schaffnit SB, Alam N, Amoah AS, Badjie J, et al. A practical guide to cross-cultural and multi-sited data collection in the biological and behavioural sciences. *Proc Biol Sci*. Apr 30, 2024;291(2021):20231422. [FREE Full text] [doi: [10.1098/rspb.2023.1422](https://doi.org/10.1098/rspb.2023.1422)] [Medline: [38654647](https://pubmed.ncbi.nlm.nih.gov/38654647/)]
89. Sheikh M, Qassem M, Kyriacou PA. Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. *Front Digit Health*. 2021;3:662811. [FREE Full text] [doi: [10.3389/fdgth.2021.662811](https://doi.org/10.3389/fdgth.2021.662811)] [Medline: [34713137](https://pubmed.ncbi.nlm.nih.gov/34713137/)]
90. Barnett S, Huckvale K, Christensen H, Venkatesh S, Mouzakis K, Vasa R. Intelligent sensing to inform and learn (InSTIL): a scalable and governance-aware platform for universal, smartphone-based digital phenotyping for research and clinical applications. *J Med Internet Res*. Nov 06, 2019;21(11):e16399. [FREE Full text] [doi: [10.2196/16399](https://doi.org/10.2196/16399)] [Medline: [31692450](https://pubmed.ncbi.nlm.nih.gov/31692450/)]
91. González-Pérez A, Matey-Sanz M, Granell C, Díaz-Sanahuja L, Bretón-López J, Casteleyn S. AwarNS: a framework for developing context-aware reactive mobile applications for health and mental health. *J Biomed Inform*. May 2023;141:104359. [FREE Full text] [doi: [10.1016/j.jbi.2023.104359](https://doi.org/10.1016/j.jbi.2023.104359)] [Medline: [37044134](https://pubmed.ncbi.nlm.nih.gov/37044134/)]
92. Gui J, Chen T, Zhang J, Cao Q, Sun Z, Luo H, et al. A survey on self-supervised learning: algorithms, applications, and future trends. *IEEE Trans Pattern Anal Mach Intell*. Dec 2024;46(12):9052-9071. [FREE Full text] [doi: [10.1109/tpami.2024.3415112](https://doi.org/10.1109/tpami.2024.3415112)]
93. Yoon J, Jarrett D, van der Schaar M. Time-series generative adversarial networks. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. 2019. Presented at: NIPS'19; December 8-14, 2019; Vancouver, BC.
94. Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: challenges, methods, and future directions. *IEEE Signal Process Mag*. May 2020;37(3):50-60. [FREE Full text] [doi: [10.1109/msp.2020.2975749](https://doi.org/10.1109/msp.2020.2975749)]
95. Moshawrab M, Adda M, Bouzouane A, Ibrahim H, Raad A. Smart wearables for the detection of occupational physical fatigue: a literature review. *Sensors* (Basel). Oct 02, 2022;22(19):7472. [FREE Full text] [doi: [10.3390/s22197472](https://doi.org/10.3390/s22197472)] [Medline: [36236570](https://pubmed.ncbi.nlm.nih.gov/36236570/)]
96. Wilkowska W, Offermann J, Colonna L, Florez-Revuelta F, Climent-Pérez P, Mihailidis A, et al. Interdisciplinary perspectives on privacy awareness in lifelogging technology development. *J Ambient Intell Humaniz Comput*. 2023;14(3):2291-2312. [FREE Full text] [doi: [10.1007/s12652-022-04486-5](https://doi.org/10.1007/s12652-022-04486-5)] [Medline: [36530469](https://pubmed.ncbi.nlm.nih.gov/36530469/)]
97. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building new computational models to support health behavior change and maintenance: new opportunities in behavioral research. *Transl Behav Med*. Sep 2015;5(3):335-346. [FREE Full text] [doi: [10.1007/s13142-015-0324-1](https://doi.org/10.1007/s13142-015-0324-1)] [Medline: [26327939](https://pubmed.ncbi.nlm.nih.gov/26327939/)]
98. Hamari J, Koivisto J, Sarsa H. Does gamification work? -- A literature review of empirical studies on gamification. In: *Proceedings of the 47th Hawaii International Conference on System Sciences*. 2014. Presented at: HICSSC 2014; January 6-9, 2014; Waikoloa, HI. [doi: [10.1109/hicss.2014.377](https://doi.org/10.1109/hicss.2014.377)]
99. Asgari Mehrabadi M, Azimi I, Sarhaddi F, Axelin A, Niela-Vilén H, Myllyntausta S, et al. Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. *JMIR Mhealth Uhealth*. Nov 02, 2020;8(10):e20465. [FREE Full text] [doi: [10.2196/20465](https://doi.org/10.2196/20465)] [Medline: [33038869](https://pubmed.ncbi.nlm.nih.gov/33038869/)]
100. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British J Health Psychol*. Jan 07, 2011;15(1):1-39. [FREE Full text] [doi: [10.1348/135910709x466063](https://doi.org/10.1348/135910709x466063)]
101. Heikenfeld J, Jajack A, Rogers J, Gutruf P, Tian L, Pan T, et al. Wearable sensors: modalities, challenges, and prospects. *Lab Chip*. Jan 16, 2018;18(2):217-248. [FREE Full text] [doi: [10.1039/c7lc00914c](https://doi.org/10.1039/c7lc00914c)] [Medline: [29182185](https://pubmed.ncbi.nlm.nih.gov/29182185/)]
102. Le LP, Nguyen T, Riegler MA, Halvorsen P, Nguyen BT. Multimodal missing data in healthcare: a comprehensive review and future directions. *Comput Sci Rev*. May 2025;56:100720. [FREE Full text] [doi: [10.1016/j.cosrev.2024.100720](https://doi.org/10.1016/j.cosrev.2024.100720)]

103. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: review of methods and applications. *Expert Syst Appl*. May 2017;73:220-239. [FREE Full text] [doi: [10.1016/j.eswa.2016.12.035](https://doi.org/10.1016/j.eswa.2016.12.035)]
104. Kovács G. An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets. *Appl Soft Comput*. Oct 2019;83:105662. [FREE Full text] [doi: [10.1016/j.asoc.2019.105662](https://doi.org/10.1016/j.asoc.2019.105662)]
105. Der-Avakian A, Markou A. The neurobiology of anhedonia and other reward-related deficits. *Trends Neurosci*. Jan 2012;35(1):68-77. [FREE Full text] [doi: [10.1016/j.tins.2011.11.005](https://doi.org/10.1016/j.tins.2011.11.005)] [Medline: [22177980](https://pubmed.ncbi.nlm.nih.gov/22177980/)]
106. Pieters LE, Deenik J, Tenback DE, van Oort J, van Harten PN. Exploring the relationship between movement disorders and physical activity in patients with schizophrenia: an actigraphy study. *Schizophr Bull*. Jul 08, 2021;47(4):906-914. [FREE Full text] [doi: [10.1093/schbul/sbab028](https://doi.org/10.1093/schbul/sbab028)] [Medline: [33764476](https://pubmed.ncbi.nlm.nih.gov/33764476/)]
107. Gavrilov AD, Jordache A, Vasdani M, Deng J. Preventing model overfitting and underfitting in convolutional neural networks. *Int J Softw Sci Comput Intell*. 2018;10(4):19-28. [FREE Full text] [doi: [10.4018/IJSSCI.2018100102](https://doi.org/10.4018/IJSSCI.2018100102)]
108. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. Jan 2019;25(1):44-56. [FREE Full text] [doi: [10.1038/s41591-018-0300-7](https://doi.org/10.1038/s41591-018-0300-7)] [Medline: [30617339](https://pubmed.ncbi.nlm.nih.gov/30617339/)]
109. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2019;9(4):e1312. [FREE Full text] [doi: [10.1002/widm.1312](https://doi.org/10.1002/widm.1312)] [Medline: [32089788](https://pubmed.ncbi.nlm.nih.gov/32089788/)]
110. Naderbagi A, Loblay V, Zahed IU, Ekambareshwar M, Poulsen A, Song YJ, et al. Cultural and contextual adaptation of digital health interventions: narrative review. *J Med Internet Res*. Jul 09, 2024;26:e55130. [FREE Full text] [doi: [10.2196/55130](https://doi.org/10.2196/55130)] [Medline: [38980719](https://pubmed.ncbi.nlm.nih.gov/38980719/)]
111. Lauber C, Rössler W. Stigma towards people with mental illness in developing countries in Asia. *Int Rev Psychiatry*. Apr 2007;19(2):157-178. [FREE Full text] [doi: [10.1080/09540260701278903](https://doi.org/10.1080/09540260701278903)] [Medline: [17464793](https://pubmed.ncbi.nlm.nih.gov/17464793/)]
112. Mishra SI, Lucksted A, Gioia D, Barnet B, Baquet CR. Needs and preferences for receiving mental health information in an African American focus group sample. *Community Ment Health J*. Apr 17, 2009;45(2):117-126. [FREE Full text] [doi: [10.1007/s10597-008-9157-4](https://doi.org/10.1007/s10597-008-9157-4)] [Medline: [18633704](https://pubmed.ncbi.nlm.nih.gov/18633704/)]
113. Gonzales A. The contemporary US digital divide: from initial access to technology maintenance. *Inf Commun Soc*. Jun 02, 2015;19(2):234-248. [FREE Full text] [doi: [10.1080/1369118x.2015.1050438](https://doi.org/10.1080/1369118x.2015.1050438)]
114. Tiller JW. Depression and anxiety. *Med J Aust*. Sep 16, 2013;199(S6):S28-S31. [doi: [10.5694/mja12.10628](https://doi.org/10.5694/mja12.10628)] [Medline: [25370281](https://pubmed.ncbi.nlm.nih.gov/25370281/)]
115. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. *Eur J Hum Genet*. Mar 2015;23(2):141-146. [FREE Full text] [doi: [10.1038/ejhg.2014.71](https://doi.org/10.1038/ejhg.2014.71)] [Medline: [24801761](https://pubmed.ncbi.nlm.nih.gov/24801761/)]
116. Rogan J, Bucci S, Firth J. Health care professionals' views on the use of passive sensing, AI, and machine learning in mental health care: systematic review with meta-synthesis. *JMIR Ment Health*. Jan 23, 2024;11:e49577. [FREE Full text] [doi: [10.2196/49577](https://doi.org/10.2196/49577)] [Medline: [38261403](https://pubmed.ncbi.nlm.nih.gov/38261403/)]
117. Paek AY, Brantley JA, Evans BJ, Contreras-Vidal JL. Concerns in the blurred divisions between medical and consumer neurotechnology. *IEEE Syst J*. Jun 2021;15(2):3069-3080. [FREE Full text] [doi: [10.1109/jsyst.2020.3032609](https://doi.org/10.1109/jsyst.2020.3032609)] [Medline: [35126800](https://pubmed.ncbi.nlm.nih.gov/35126800/)]
118. Lui GY, Loughnane D, Polley C, Jayarathna T, Breen PP. The apple watch for monitoring mental health-related physiological symptoms: literature review. *JMIR Ment Health*. Sep 07, 2022;9(9):e37354. [FREE Full text] [doi: [10.2196/37354](https://doi.org/10.2196/37354)] [Medline: [36069848](https://pubmed.ncbi.nlm.nih.gov/36069848/)]
119. Digital health software precertification (pre-cert) pilot program. U.S. Food & Drug Administration. Sep 26, 2022. URL: <https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-pilot-program> [accessed 2025-07-30]
120. Guidance on qualification and classification of software in regulation (EU) 2017/745 – MDR and regulation (EU) 2017/746 – IVDR. European Commission. 2019. URL: <https://ec.europa.eu/docsroom/documents/37581> [accessed 2025-08-06]
121. Bown MJ, Sutton AJ. Quality control in systematic reviews and meta-analyses. *Eur J Vasc Endovasc Surg*. Nov 2010;40(5):669-677. [FREE Full text] [doi: [10.1016/j.ejvs.2010.07.011](https://doi.org/10.1016/j.ejvs.2010.07.011)] [Medline: [20732826](https://pubmed.ncbi.nlm.nih.gov/20732826/)]
122. Hopkin G, Branson R, Campbell P, Coole H, Cooper S, Edelmann F, et al. Considerations for regulation and evaluation of digital mental health technologies. *Digit Health*. 2024;10:20552076241293313. [FREE Full text] [doi: [10.1177/20552076241293313](https://doi.org/10.1177/20552076241293313)] [Medline: [39502489](https://pubmed.ncbi.nlm.nih.gov/39502489/)]
123. Dwork C, Roth A. The algorithmic foundations of differential privacy. *Found Trends Theor Comput Sci*. 2014;9(3-4):211-407. [doi: [10.1561/04000000042](https://doi.org/10.1561/04000000042)]
124. International Medical Device Regulators Forum homepage. International Medical Device Regulators Forum. URL: <https://www.imdrf.org/> [accessed 2025-07-30]
125. Deffner D, Rohrer JM, McElreath R. A causal framework for cross-cultural generalizability. *Adv Methods Pract Psychol Sci*. Sep 21, 2022;5(3). [FREE Full text] [doi: [10.1177/25152459221106366](https://doi.org/10.1177/25152459221106366)]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
CNN: convolutional neural network
ECG: electrocardiography
KNN: k-nearest neighbor
LSTM: long short-term memory
MeSH: Medical Subject Headings
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SVM: support vector machine
XGBoost: Extreme Gradient Boosting

Edited by A Coristine; submitted 07.05.25; peer-reviewed by Z Chen, MMR Sanjeewa, M Liu; comments to author 04.06.25; revised version received 17.06.25; accepted 10.07.25; published 14.08.25

Please cite as:

Shen S, Qi W, Zeng J, Li S, Liu X, Zhu X, Dong C, Wang B, Shi Y, Yao J, Wang B, Lou X, Gu S, Li P, Wang J, Jiang G, Cao S
Passive Sensing for Mental Health Monitoring Using Machine Learning With Wearables and Smartphones: Scoping Review
J Med Internet Res 2025;27:e77066

URL: <https://www.jmir.org/2025/1/e77066>

doi: [10.2196/77066](https://doi.org/10.2196/77066)

PMID:

©ShiYing Shen, Wenhao Qi, Jianwen Zeng, Sixie Li, Xin Liu, Xiaohong Zhu, Chaoqun Dong, Bin Wang, Yankai Shi, Jiani Yao, Bingsheng Wang, Xiajing Lou, Simin Gu, Pan Li, Jinghua Wang, Guowei Jiang, Shihua Cao. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 14.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.