



Original Article

Digital Phenotyping of High-Risk Gaming Behavior Using Wearable Devices

Yijie Song^{1,†}, Yawen Shi^{2,†}, Yiqun Tu¹, Haidi Shan¹, Chuaning Huang¹,
Yue Wang¹, Shuo Li¹, Zhekai Li¹, Hang Su¹, Na Zhong¹, Tianzhen Chen¹,
Min Zhao^{1,3,4,*}, Jiang Du^{1,*}

¹Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, 200030 Shanghai, China

²Shanghai Changning Mental Health Center, Affiliated Mental Health Center of East China Normal University, 200335 Shanghai, China

³Shanghai Key Laboratory of Psychotic Disorders, 200030 Shanghai, China

⁴CAS Center for Excellence in Brain Science and Intelligence Technology (CEBSIT), Chinese Academy of Sciences, 200031 Shanghai, China

*Correspondence: drminzhao@smhc.org.cn (Min Zhao); dujiangdou@163.com (Jiang Du)

†These authors contributed equally.

Academic Editor: Francesco Bartoli

Submitted: 14 November 2025 Revised: 19 January 2026 Accepted: 30 January 2026 Published: 30 June 2026

Abstract

Background: Gaming addiction is a serious mental health issue. Early identification of high-risk gaming behaviors (HRGB) is crucial for prevention, yet research using wearable data for detection remains limited. This study analyzed the temporal dynamics of gaming behavior and health indicators using 28-day summary statistics derived from longitudinal data in young adults with HRGB, thereby providing a basis for gaming disorder prevention and intervention. **Methods:** This 28-day longitudinal study employed a self-designed system implemented on the Huawei Research Platform for participant recruitment and data collection. Participants completed daily mobile app questionnaires between 18:00 and 23:00 assessing gaming time, mood, and sleep, while a smart band continuously collected physiological and activity data. Emotional states were assessed weekly using the Patient Health Questionnaire-9 (PHQ-9) and the Generalized Anxiety Disorder-7. Machine learning models were applied to compare the predictive performance of active (self-report), passive (wearable), and combined data in identifying HRGB, and to determine key predictive factors relevant for intervention. **Results:** The final sample comprised 37 participants in the HRGB group and 23 in the non-HRGB group. Significant between-group differences were observed across multiple indicators. Combined data yielded the highest area under the receiver operating characteristic curve (AUROC) in five of six models (logistic regression: 0.86 vs 0.81). Wearable-derived data enhanced early classification of HRGB, with depression severity (PHQ-9), heart rate, and sleep metrics emerging as potential markers. A one-standard-deviation increase in PHQ-9 scores was associated with a 3.51-point increase in Gaming Disorder Screening Scale (GDSS) scores (89% credible interval [CrI], 1.85–5.20), while a higher heart rate index was associated with a 1.36-point increase in GDSS scores (89% CrI, 0.13–2.60). The population attributable fraction was 15.51% for PHQ-9. **Conclusions:** Wearable data enhances early HRGB classification. PHQ-9, heart rate, and sleep metrics are key predictors, showing clinical utility.

Keywords: internet gaming disorder; wearable electronic devices; machine learning; digital health

Main Points

1. Multimodal fusion of wearable-derived physiological and self-reported psychobehavioral data boosts the binary classification performance of the machine learning model, validating digital phenotyping for early detection of high-risk gaming behaviors (HRGB).

2. Patient Health Questionnaire-9 (PHQ-9), heart rate index, and deep sleep proportion were the strongest predictors among all collected variables, with each standard deviation increase in PHQ-9 score associated with a 3.51-point rise in the Gaming Disorder Screening Scale score.

3. Alleviating depressive symptoms in high-risk individuals could prevent approximately 15.5% of HRGB cases, suggesting that depression intervention is a potential key prevention strategy.

4. Elevated heart rate and reduced deep sleep proportion in HRGB may suggest heightened sympathetic activity and disrupted sleep architecture, which could implicate autonomic nervous dysregulation in gaming behavior mechanisms.

1. Introduction

Gaming addiction has increasingly been recognized as a major public health challenge, particularly among adolescents and young adults [1,2]. Problematic gaming behaviors are rising due to the widespread availability of digital devices and immersive games, resulting in psychological impairment, social disruption, and poor academic or occupational performance [3,4]. The World Health Organization's inclusion of "gaming disorder" in the International Classification of Diseases underscores the seriousness of



this condition [5,6]. Early identification of high-risk gaming behaviors (HRGB) is clinically critical because it provides a window for intervention before patterns become entrenched. Detecting subclinical vulnerability enables targeted prevention, such as psychoeducation and cognitive-behavioral techniques, to mitigate the progression to a full-blown disorder [7]. Despite these compelling reasons, current screening methods often rely on self-reported retrospective data, which are susceptible to bias and underreporting [8,9]. Developing objective, continuous, and ecologically valid assessment tools is critical for capturing dynamic real-world data, thereby revolutionizing early detection and supporting scalable public health interventions.

Recently, there has been growing interest in leveraging digital technologies to enhance mental health monitoring and diagnostics. Wearable devices provide a critical window into autonomic and behavioral patterns that are typically altered in mental health conditions by passively collecting continuous, multidimensional data [10]. Changes in sleep continuity and reduced physical activity have been consistently linked to depressive disorders [11,12], while elevated resting heart rate variability has been associated with anxiety [13]. However, most research to date has focused on traditional mental health disorders, such as depression and anxiety [14], with less attention paid to behavioral addictions, particularly HRGB [15]. This gap is noteworthy given the global rise in gaming engagement and its associated harms. This study aimed to bridge this gap by examining whether wearable data can improve the identification of university students with HRGB. We aimed to benchmark multimodal (self-report + wearable) predictions against traditional methods (only self-report) using machine learning to identify key wearable-derived risk factors and explore their preventive utility. An overview of this study is presented in Fig. 1. This approach integrates digital phenotyping with predictive analytics, contributing to digital psychiatry and paving the way for early and tailored interventions.

2. Methods and Materials

2.1 Participants and Recruitment

Participants were recruited simultaneously between September 2023 and February 2024 through advertisements placed at a university in Shanghai and a local university in Zhuzhou, Hunan Province. The participants were divided into two groups: HRGB and non-HRGB. The inclusion criteria for both groups were as follows: (1) current enrollment as a university student; (2) no restriction on gender; (3) voluntary participation in HRGB screening, with a Gaming Disorder Screening Scale (GDSS) total score ≥ 30 for the HRGB group or < 25 for the non-HRGB group [16]; (4) ownership of an Android smartphone, agreement to wear a Huawei smartwatch (Huawei Watch GT3 or Wristband6, Huawei Technologies Co., Ltd., Shenzhen, Guangdong, China) with physiological monitoring functions, and consent to download the research app and upload

data; (5) agreement to participate in the study with written informed consent provided. GDSS is an 18-item instrument with a total score range of 18–72 that was developed and psychometrically validated specifically for screening gaming disorder in Chinese adolescents and young adults. It has demonstrated robust reliability and validity in clinical assessments. Based on previous findings from our research team, individuals with GDSS scores in the top 20% (≥ 30) were classified as the HRGB group, while those with scores below the mean (< 25) were classified as the non-HRGB group [17]. Participants who scored between 25 and 29 were excluded from the study. The exclusion criteria for both groups were as follows: (1) history of head trauma with loss of consciousness, epilepsy, or other severe or unstable medical conditions; (2) past or current diagnosis of severe psychiatric disorders such as schizophrenia, mood disorders, intellectual disability, and others; (3) previous use of psychoactive substances, such as cocaine and methamphetamine, excluding tobacco and alcohol, or the use of other dopaminergic medications. Baseline data were collected in three domains: (1) demographic characteristics (gender and age); (2) gaming behaviors (gaming days per week, daily gaming duration and gaming expenditure); (3) clinical scales, including GDSS, Patient Health Questionnaire-9 (PHQ-9) to assess depressive symptoms, Generalized Anxiety Disorder-7 (GAD-7) to measure anxiety symptoms, and Pittsburgh Sleep Quality Index (PSQI) to evaluate sleep quality.

2.2 Time-Based Ecological Momentary Assessment (t-EMA)

Participants were instructed to log in daily to complete the t-EMA within a designated window (18:00–23:00), which took approximately 2 min. PHQ-9 and GAD-7 scales were administered on days 7, 14, 21, and 28. PHQ-9 is a self-report instrument used to screen for major depressive disorder and assess the severity of depressive symptoms. GAD-7 is a brief self-administered questionnaire designed to screen for generalized anxiety disorder and measure the severity of anxiety symptoms. These two scales were used to assess the general emotional state of the participants during the study period and represent important dimensions to evaluate gaming addiction behaviors. To ensure compliance, the completion status was monitored, and reminders were sent for missed submissions. A tiered incentive scheme was used: weekly bonuses (\$0.70–1.40) were awarded for full compliance, and a final bonus (\$11–16) was awarded for completing all 28 days, with the exact amounts based on overall compliance and data quality. Technical details regarding the development of the t-EMA application are described in **Supplementary Methods**.

2.3 Monitoring Metrics

In this study, physiological and psychological data were collected using validated active and passive mea-

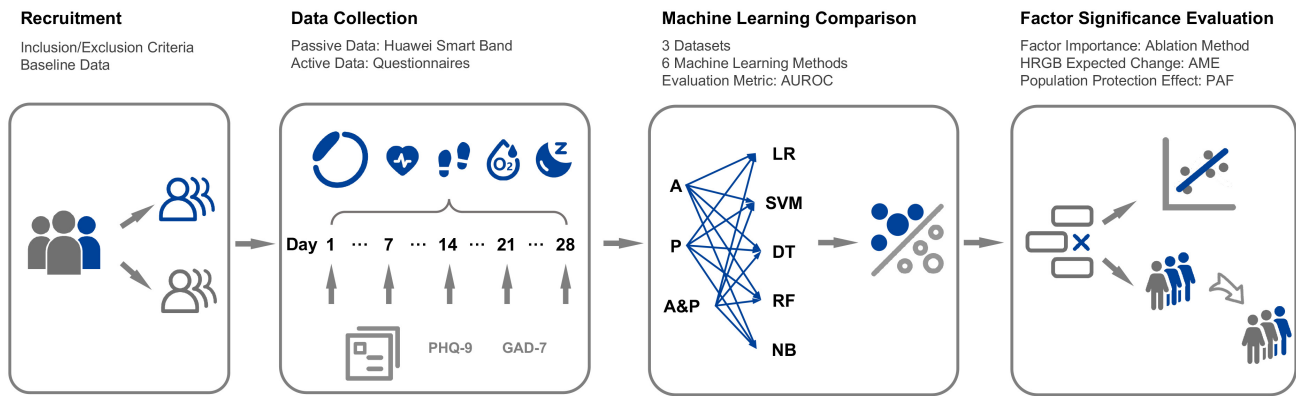


Fig. 1. Study design and data analysis flow. Recruitment: HRGB and non-HRGB controls are recruited according to predefined inclusion/exclusion criteria, and baseline data are collected at the same time. Data Collection: Game behavior-related physiological indicators such as heart rate, daily steps, blood oxygen saturation, and sleep metrics are passively collected via Huawei Smart Band. Emotion and stress scores are actively collected through weekly questionnaires. Machine Learning Comparison: three datasets (active, passive, and combined) are used to evaluate the classification performance of six machine learning methods in distinguishing HRGB and non-HRGB behaviors, with AUROC as the evaluation metric. Factor Significance Evaluation: The importance of active and passive features is assessed using an ablation method. For significant factors, AME is applied to estimate their effect on HRGB severity, and PAF is used to evaluate the proportion of HRGB cases that could be prevented by intervening on these factors. Abbreviations: A, active data; AME, average marginal effects; AUROC, area under the receiver operating characteristic curve; A&P, active and passive data; DT, Decision Tree; HRGB, high-risk gaming behaviors; LR, Logistic Regression; NB, Naïve Bayes; P, passive data; PAF, population attributable fraction; RF, Random Forest; SVM, Support Vector Machine.

surement tools, resulting in 24 continuous variables. To enhance the performance of machine learning classification, we draw on methods from previous studies [18,19,20]. All variables were aggregated over time into mean and variance features. Active assessments included PHQ-9 for depression and GAD-7 for anxiety symptoms. Passive data were collected using a smartwatch or wristband. The study utilized daily average values obtained from Huawei’s health platform, encompassing the following metrics: blood oxygen (peripheral capillary oxygen saturation (SpO₂): max/min/mean), sleep architecture (N1/N2, N3, and rapid eye movement (REM) percentages), physical activity (daily calories and steps), and a heart rate index (max/min ratio) [21,22]. Crucially, all data represent daily averages, and no modeling of rhythmic, trend-based, or change-point temporal structures was performed. The complete list of variables is presented in **Supplementary Table 1**.

2.4 Data Preprocessing

Following a rigorous recruitment and data filtering process, 60 participants were enrolled in the study, yielding 15,738 valid temporal data points. Details on participant attrition and data filtering are described in the **Supplementary Methods** section. The MissForest algorithm was used to handle missing values while preserving as many of the original data characteristics as possible. This non-parametric, machine-learning-based imputation method is suitable for mixed-type data (continuous and categorical)

[23]. After imputation, descriptive statistics and distributional comparisons were conducted to assess the consistency between the original and imputed datasets.

2.5 Statistical Analysis

2.5.1 Difference Comparison

Group differences in the 24 metrics between non-HRGB and HRGB groups were assessed using Independent Samples *t*-tests (for normally distributed data) or Mann-Whitney U tests (for non-parametric data). To account for multiple comparisons, *p*-values were false discovery rate (FDR)-corrected using the Benjamini–Hochberg procedure, with adjusted *p*-values < 0.05 deemed significant. All analyses were conducted in R (version 4.2.0; R Foundation for Statistical Computing, Vienna, Austria).

2.5.2 Machine Learning Classification Task

Binary classification prediction was performed using Python (version 3.12.3; Python Software Foundation, Wilmington, DE, USA) with the Scikit-learn library (version 1.5.2; scikit-learn developers, Inria, Saclay, France) to distinguish between the participant groups. Accordingly, three distinct classification trials were conducted for each binary classification task using only actively collected data as input features, only passively collected data, and the full combined dataset. Multiple machine learning models were tested, including Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), and XGBoost. The data were

randomly split into independent training and test sets with stratification (1:1 ratio), and standardization was performed during training to avoid data leakage. The model training and hyperparameter tuning were performed using nested cross-validation. The inner layer employed repeated stratified five-fold cross-validation (with three repeats) for GridSearchCV from scikit-learn to select the optimal parameters using the area under the receiver operating characteristic curve (AUROC) as the metric, and the outer layer conducted the final evaluation on the independent test set. To further quantify performance stability, 1000 bootstrap resamplings were applied to the test set predictions, and the 95% confidence interval (CI) for AUROC was calculated. All analyses were performed using Python along with libraries, including Scikit-learn and XGBoost (version 2.1.0; XGBoost Contributors, University of Washington, Seattle, WA, USA). A fixed random seed was set throughout the study to ensure the reproducibility of the results.

2.5.3 Importance Prediction

In this study, an ablation method was used to assess feature importance by randomly permuting each feature's values across subjects and breaking its association with the target, such as a psychiatric diagnosis. The resulting decrease in model performance was measured using AUROC, with a larger decrease indicating greater importance. The importance score for each variable was defined as the average percentage decrease in the AUROC across the six machine learning models after randomization.

2.5.4 Average Marginal Effects (AME)

AME was calculated to assess the impact of various indicators on the risk of gaming behavior. AME is the sample mean of the partial derivatives of the regression equation with respect to a given explanatory variable. For continuous outcomes, such as the GDSS score, AME corresponds to the change per one-standard-deviation increase in the predictor. Predicted outcomes for the target sleep feature were displayed in adjusted prediction plots, with other covariates held at their median values and varied by ± 2 standard deviations. All models were estimated using a Bayesian framework with Stan (version 2.33.0; Stan Development Team, Columbia University, New York, NY, USA), accessed via the R package brms (version 2.22.0; Paul-Christian Bürkner, University of Stuttgart, Stuttgart, Germany) [24]. All models were estimated using 40,000 sampling iterations, with thinning applied by retaining every 10th draw. Weakly informative priors centered at zero were employed to constrain the estimates within plausible ranges and to provide statistical regularization by shrinking coefficients toward zero [25]. The posterior samples of AME were summarized using an 89% credible interval (CrI). All available outcome data were retained under the assumption of missing data at random. To evaluate the potential impact of gender on the results, a sensitivity analysis

was performed using data from males only, while controlling for gender as a covariate.

2.5.5 Population Attributable Fraction (PAF)

To assess the potential population-level preventive impact of modifying the indicators, the adjusted PAF was calculated as the proportion of incident HRGB cases that could be prevented if a specific exposure were eliminated. Each continuous feature was dichotomized by classifying the upper quartile as exposed (1) and the lower three quartiles as unexposed (0), yielding binary variables, such as “high” versus “low depressive symptoms”. A confounder-adjusted binary LR model was fitted using the glm function in R, and the adjusted PAF was computed using AFglm function from the R package AF (version 0.1.5; Karolinska Institutet, Stockholm, Sweden). Additionally, a sensitivity analysis was conducted using data from male participants only, with gender included as a covariate, to assess its potential influence on the outcomes.

3. Results

Initially, 83 undergraduate students were enrolled in this study. After rigorous exclusion procedures (Fig. 2 and **Supplementary Methods**), a final sample of 60 participants with complete and valid data was retained, yielding 15,738 valid dynamic data points.

Overall, this study included 23 participants in the non-HRGB group and 37 in the HRGB group. Significant differences were observed between the two groups in terms of demographic characteristics (gender), gaming behaviors (daily gaming time), and clinical scale scores (GDSS, PHQ-9, GAD-7, and PSQI). The details are presented in Table 1.

3.1 Differences Between HRGB and Non-HRGB Groups on Active and Passive Indicators

Multiple active (self-reported) and passive (wearable-sensor-derived) indicators were compared between HRGB and non-HRGB groups to characterize group-specific profiles. In terms of active indicators assessed via psychological questionnaires, the HRGB group exhibited significantly higher mean scores on the PHQ-9 (5.00 versus 1.79; $p_{FDR} = 0.01$) and GAD-7 (3.37 versus 1.12; $p_{FDR} = 0.02$) than the non-HRGB group (Fig. 3a), indicating a greater burden of psychological distress in the HRGB group. Additionally, the mean heart rate index (HRidx_mean) was significantly elevated in the HRGB group (2.03 versus 1.68; $p_{FDR} = 0.01$; Fig. 3a), further supporting heightened physiological or psychological stress. Although differences in blood oxygen (SpO₂: max/min/mean), sleep architecture (N1/N2, N3, and REM percentages), and physical activity (daily calories and steps) were observed, they were not statistically significant after FDR correction. Details of the differences between non-HRGB and HRGB groups are provided in **Supplementary Table 2**.

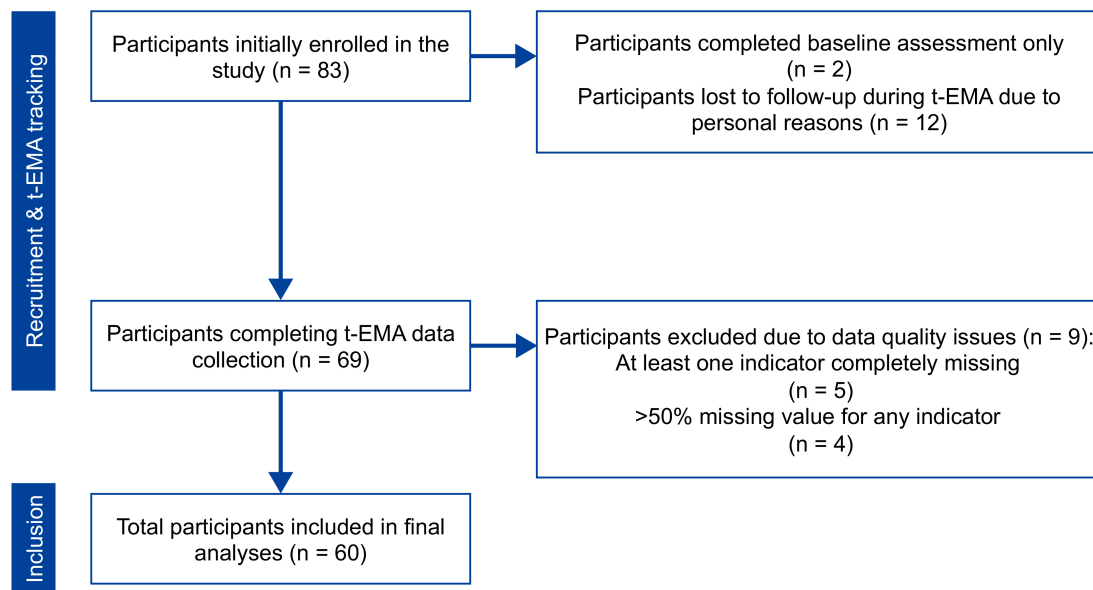


Fig. 2. Flowchart of participant recruitment, t-EMA tracking, and inclusion in the study. A total of 83 participants were recruited and assigned to either the HRGB (n = 52) or non-HRGB group (n = 31). For the HRGB group, 2 participants completed baseline assessments only, while 50 continued to the t-EMA phase. Of these, 43 completed t-EMA data collections, and 37 provided complete data for all indicators, resulting in 37 participants included in the final analysis. For the non-HRGB group, all 31 participants continued to the t-EMA phase, with 26 completing t-EMA data collection and 23 providing complete data for all indicators, resulting in 23 participants included in the final analysis. Abbreviations: t-EMA, time-based Ecological Momentary Assessment.

Table 1. Demographic and baseline characteristics of the study participants.

		Control group (n = 23)	High risk group (n = 37)	χ^2 or T value	p value
Demographic characteristics	Gender			$\chi^2 (1) = 12.50$	<0.00
	Male	12 (52.17%)	34 (91.89%)		
	Female	11 (47.83%)	3 (8.11%)		
	Age	21.09 ± 3.03	18.43 ± 1.34	4.66	<0.00
Gaming behaviors	Game days per week	3.74 ± 2.83	6.16 ± 1.48	4.35	<0.00
	Game time per day	1.28 ± 1.22	4.26 ± 2.32	5.67	<0.00
	Game spending	29.78 ± 52.75	65.11 ± 85.08	1.79	0.08
Clinical scales	GDSS	20.78 ± 2.50	36.92 ± 4.69	15.17	<0.00
	PHQ-9	2.57 ± 3.11	5.16 ± 3.91	2.70	0.01
	GAD-7	1.44 ± 2.29	3.65 ± 3.28	2.83	0.01
	PSQI	11.00 ± 3.78	15.32 ± 6.19	3.02	<0.01

Abbreviations: GDSS, Gaming Disorder Screening Scale; PHQ-9, Patient Health Questionnaire-9; GAD-7, Generalized Anxiety Disorder-7; PSQI, Pittsburgh Sleep Quality Index.

3.2 Passive Indicators Enhance Machine Learning Performance in Classifying HRGB Group

We used multiple machine learning approaches to investigate the utility of passive indicators for identifying individuals with HRGB. Specifically, we framed the task as a binary classification problem to distinguish between the HRGB and non-HRGB groups. We used six algorithms to build the predictive models: LR, SVM, DT, RF, NB, and XGBoost. Three sets of predictor variables were evaluated: (1) four active indicators only, (2) 20 passive indicators only, and (3) all 24 combined indicators (**Supplementary Materials**).

The results indicated that combining active and passive indicators yielded the best predictive performance in five of the six models (LR, SVM, DT, RF, and XGBoost), outperforming the models using either data type alone (Fig. 3b,c; **Supplementary Tables 3–10**). The AUROC for LR increased to 0.86 (combined) from 0.81 (active) and 0.67 (passive), a trend that was mirrored by the SVM, DT, and RF.

For the XGBoost model, the combined data approach yielded only a marginally higher performance (by 0.02) than using active data alone. However, for the NB model, the combined performance was slightly lower than the

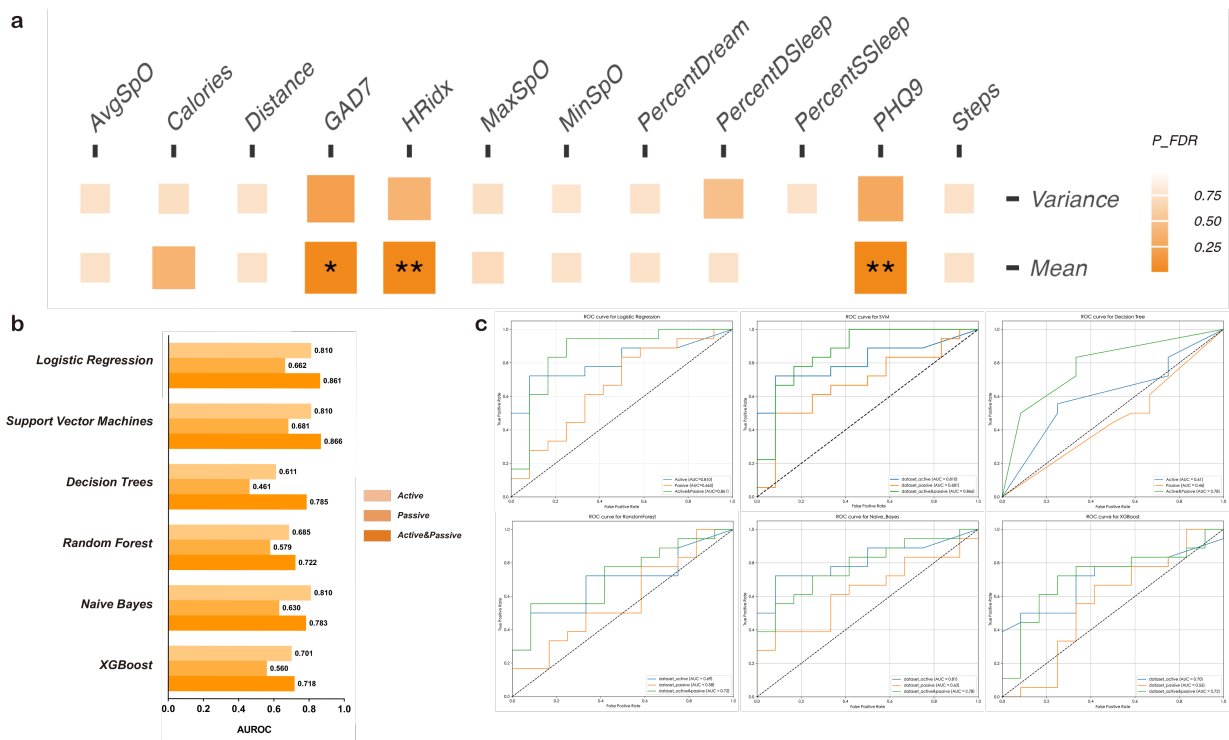


Fig. 3. Group comparisons of active and passive indicators and machine learning classification of HRGB. (a) Comparative analysis of 24 active and passive indicators between the HRGB and non-HRGB groups. Square size and color depth correspond to p_{FDR} values (smaller values \rightarrow larger squares and darker colors). (b) Performance comparison of machine learning models for classifying HRGB versus non-HRGB groups under different predictor conditions (active indicators only, passive indicators only, and combined). (c) ROC curves for six machine learning models in predicting HRGB: Logistic Regression, SVM, Decision Trees, Random Forest, Naive Bayes, and XGBoost. Each panel displays model performance using active indicators only (blue), passive indicators only (orange), and combined indicators (green). In panel (a), * indicates $p_{FDR} < 0.05$, and ** indicates $p_{FDR} < 0.01$. Abbreviations: AvgSpO, average SpO₂; SpO₂, peripheral capillary oxygen saturation; HRidx, heart rate index; MaxSpO, maximum SpO₂; MinSpO, minimum SpO₂; PercentDream, percentage of REM sleep in total sleep time; PercentDSleep, percentage of deep sleep (N3 stage of Non-Rapid Eye Movement Sleep, NREM) in total sleep time; PercentSSleep, percentage of shallow sleep (N1/N2 stages of NREM) in total sleep time; ROC, receiver operating characteristic.

active-only model. These findings underscore that integrating passive sensing data with self-reports generally enhances the accuracy of identifying HRGB.

3.3 Heart Rate, Sleep, and Depressive Symptoms as Key Predictors

We performed an ablation analysis to quantify the relative importance of each feature, identify the most influential predictors and enhance model interpretability. The results revealed that the top five most important predictors were the mean heart rate index (HRidx_mean; mean AUROC decrease, 0.12; 95% CI, 0.00 to 0.25), mean PHQ-9 score (PHQ9_mean, 0.06; 95% CI, -0.05 to 0.18), variance of minimum SpO₂ (MinSpO_var, 0.06; 95% CI, -0.06 to 0.17), mean percentage of deep sleep (PercentDSleep_mean, 0.05; 95% CI, -0.07 to 0.16), and variance of percentage of light sleep (PercentSSleep_var, 0.04; 95% CI, -0.11 to 0.19; Fig. 4; **Supplementary Table 11**). These findings corroborate prior evidence highlighting the rele-

vance of psychological and physiological markers in predicting HRGB and provide empirical support for refining predictive models to improve their accuracy and reliability.

3.4 Protective Effects of Improved Mood, Sleep, and Heart Rate Regulation on Gaming Behavior

We calculated the AME of each variable on GDSS scores to quantify the magnitude of the association between key predictors and gaming severity. Higher mean PHQ-9 scores were positively associated with GDSS scores: a one-standard-deviation increase in the PHQ-9 score was associated with an average increase of 3.508 in the GDSS score (89% CrI [1.85, 5.20]). Similarly, a higher mean heart rate index was associated with increased GDSS scores (AME = 1.36; 89% CrI [0.13, 2.60]; **Supplementary Table 12**). The findings were not significantly influenced by gender (**Supplementary Tables 13,14** for gender-related analyses).

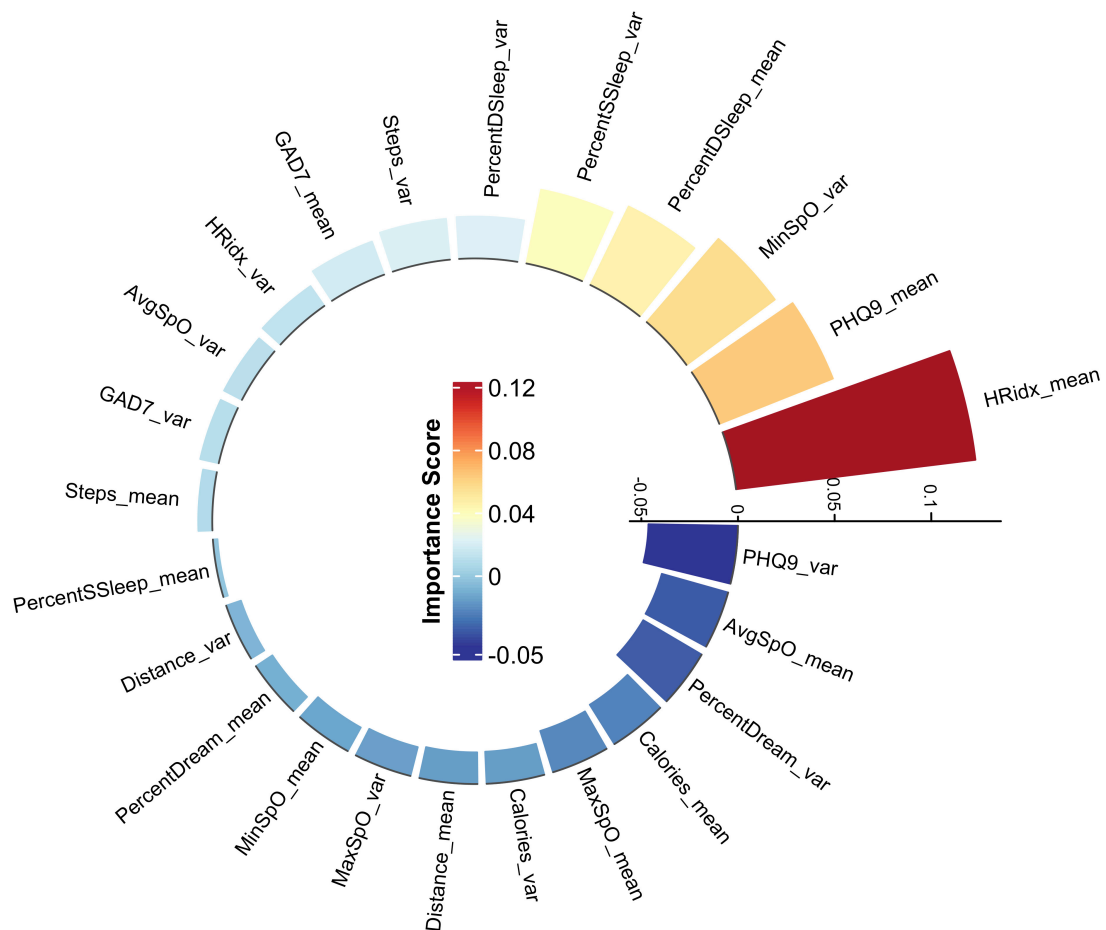


Fig. 4. Feature importance ranking derived from ablation analysis across machine learning models. The importance of each variable is quantified by the mean decrease in AUROC following permutation. Full abbreviations are provided in **Supplementary Material**.

The adjusted prediction plots (Fig. 5) illustrate the modeled relationships between the selected active and passive features and GDSS scores, with all other covariates held at their median values. Above-average PHQ-9 scores and heart rate index were associated with higher GDSS scores, whereas higher proportions of deep sleep and greater variance in minimum SpO₂ were associated with lower GDSS scores. The effect of the variance in the light sleep percentage on the GDSS scores was relatively small.

Furthermore, we estimated the potential public health impact of interventions targeting these modifiable factors (**Supplementary Table 15**). For the PHQ-9 mean, PAF was 15.51% (95% CI, 2.04%–28.99%), indicating that reducing depressive symptoms could prevent a significant proportion of HRGB at the population level. In contrast, the 95% CIs for PAFs of the other four predictors included zero, suggesting insufficient evidence for a significant preventive effect through intervention on these parameters. The findings were not significantly influenced by gender (**Supplementary Tables 16,17** for gender related analysis).

4. Discussion

This study innovatively used passively collected data from wearable wristbands to track and predict HRGB patterns among young adults. Six machine learning models were employed to compare three data modalities (active, passive, and combined) for classifying individuals as HRGB versus non-HRGB. After hyperparameter tuning, models using only active or passive data exhibited moderate performance (AUROC). In contrast, integrating both data types improved the predictive accuracy: the combined approach outperformed the active-only models in five of the six classifiers. The superior performance of multimodal digital phenotyping stems from the advantages of wearable-based data acquisition, which provides continuous, objective measurements less affected by recall and social desirability biases. Data collection using wearables is convenient, less biased, objective, highly comparable, and easily scalable and replicable, facilitating cross-study validation and comparison [26]. Moreover, physiologically grounded metrics from wearables may help patients understand the biological basis of their diagnosis and the connection be-

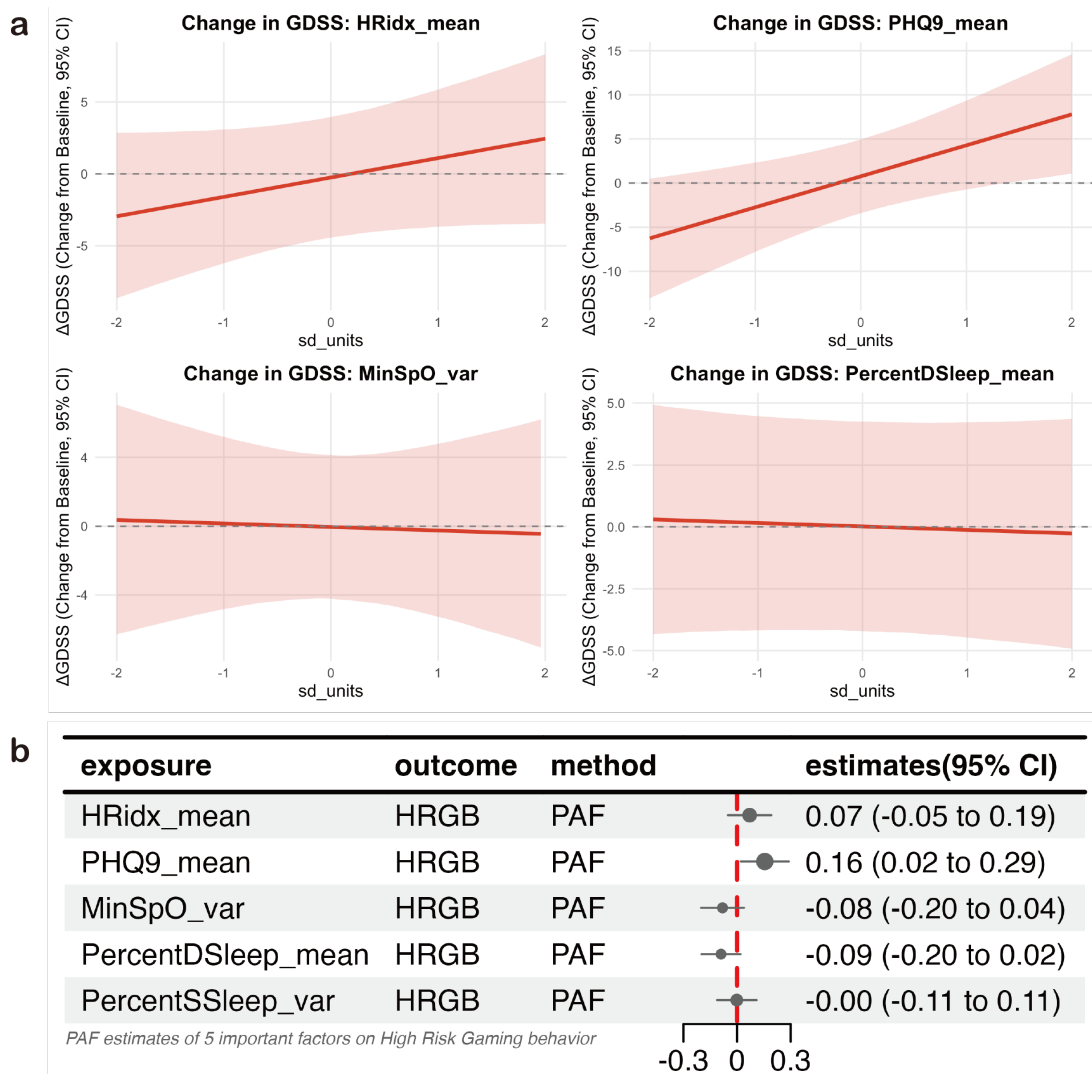


Fig. 5. Adjusted associations with GDSS scores and PAFs for modifiable risk factors of HRGB. (a) Adjusted prediction plots for GDSS scores across a two-standard-deviation range of key predictors, with all other covariates held at their median values. Effect of mean heart rate index (HRidx_mean), mean PHQ-9 score (PHQ9_mean), variance of minimum oxygen saturation (MinSpO_var), and mean percentage of deep sleep (PercentDSleep_mean). (b) Forest plot of PAF and corresponding 95% confidence intervals for modifiable risk factors associated with HRGB. Abbreviations: CrI, credible interval; CI, confidence interval; PercentSSleep_var: Variance of Percentage of Light Sleep.

tween behavior and symptoms. In this study, the classification of HRGB and non-HRGB groups was based on an extreme-group approach using within-sample GDSS scores to enhance phenotypic contrast for machine-learning classification. Previous studies have also demonstrated that HRGB individuals identified by this method significantly differ from controls in areas such as time perception and perceived social support. However, the intermediate score range of 25–29 may still contain feature information relevant to HRGB individuals. Future studies should explore the relationship between continuous GDSS scores and individual digital phenotypes in larger samples to facilitate precise prevention strategies.

These findings identified the level of depressive symptoms as a significant predictor in machine learning models, differentiating the HRGB group from the non-HRGB group. The AME analysis revealed that higher depressive symptoms, indicated by elevated PHQ-9 scores, were significantly associated with increased GDSS scores, consistent with the existing literature. A longitudinal follow-up study identified depressive symptoms as independent predictors of gaming disorder incidence [27]. Furthermore, at the individual level, childhood depressive symptoms have been demonstrated to positively predict subsequent gaming disorder symptoms [28]. Consequently, early identification, assessment, and intervention targeting depressive symptoms may be critical strategies for preventing the onset

of gaming disorders. This study further quantified the potential preventive effect of reducing depression levels using PAF analysis. The results estimated that effectively lowering the depressive symptoms in the HRGB group to the level of the non-HRGB group, via evidence-based interventions such as cognitive behavioral therapy [7], prevented approximately 15.5% (95% CI, 2.04%–28.99%) of individuals from developing HRGB. These findings substantiate the therapeutic potential of targeting depressive symptoms as a complementary strategy for mitigating HRGB, highlighting the centrality of emotional regulation in gaming disorder etiology and supporting integrated treatment approaches.

The heart rate index was also identified as important for the machine-learning classification task. Within two standard deviations, an increase in the heart rate index was associated with an increase in the estimated GDSS score, indicating a positive correlation between the heart rate index and HRGB. Based on the available device data, we defined the heart rate index as the ratio of the maximum to minimum heart rate within a day. Previous research has linked sympathetic nervous system activity to this index, which reflects a broader range of autonomic nervous fluctuations. Inhibiting sympathetic activity (using beta-blockers) in elderly hypertensive patients reduced the 24-h heart rate amplitude and the max/min ratio, whereas enhancing sympathetic tone (using calcium channel blockers) significantly increased the heart rate amplitude and raised this ratio [29]. Studies in animal models and simulations have similarly demonstrated that suppressing sympathetic activity reduces the daily heart rate amplitude and max/min ratio in mice [30,31]. Multiple studies have found increased sympathetic activity and vagal inhibition in individuals with gaming or other psychiatric disorders [22,32]. Accordingly, the higher max/min heart rate ratio observed in the HRGB group in this study may be associated with elevated sympathetic excitability. These findings suggest a potential link between the HRGB and autonomic nervous system dysregulation. However, they fail to clarify whether the widespread autonomic changes are a direct manifestation of HRGB or a compensatory response. This distinction is critical, as the concepts of dynamic biological interactions and compensatory pathways are highly relevant across neurological and behavioral disorders. Given evidence of such compensatory mechanisms in other contexts [33], future studies should specifically investigate whether the autonomic changes observed in HRGB represent a direct effect or a compensatory response. Elucidating this distinction is crucial, as it could facilitate the identification of more specific digital biomarkers for HRGB.

Sleep has been consistently associated with mood, stress, and Internet gaming disorders. Adolescents and young adults with gaming disorders tend to have higher PSQI scores (indicating poorer sleep quality) [4]. Neuroimaging evidence revealed that sleep influences gaming

behavior via disrupted hippocampal connectivity, particularly with the nucleus accumbens and prefrontal-cortical regions, implicating memory, reward, and cognitive control pathways [34]. Poor sleep quality is a recognized risk factor for gaming disorders [35]. This study further revealed that sleep disturbances manifest concurrently with the emergence of HRGB in young adults. At baseline, affected individuals demonstrated significantly higher PSQI scores, while subsequent machine-learning analyses identified a reduced proportion of deep sleep (N3 stage of Non-Rapid Eye Movement Sleep) as a significant predictor of HRGB. These findings position sleep architecture as a viable biomarker for early risk identification. Wearable devices offer a practical means of detecting these physiological signatures during the initial stage. Future investigations should examine the dynamic evolution of sleep parameters as the progression from HRGB to gaming disorder and evaluate the preventive efficacy of sleep-focused interventions in mitigating the development of pathological gaming.

Limitations

This study has several methodological and ethical limitations that warrant careful consideration. The limited sample size and significant gender imbalance, such as male predominance in the HRGB group, may compromise both the representativeness of the identified predictive features and model generalizability, particularly in female populations. Future studies should address the limitations of sample size and gender imbalance, such as male predominance in the HRGB group, using gender-balanced sampling and conducting gender-stratified analyses, thereby enhancing the robustness of predictive models and validating the external generalizability of these findings across more diverse populations. From a data quality perspective, irregularities in device wear and occasional sensor detachment introduce non-random missingness and measurement noise. Although rigorous data-cleaning protocols were applied, residual biases may persist, potentially compromising feature robustness and model performance. Furthermore, the continuous passive monitoring of physiological and behavioral parameters raises substantive ethical questions regarding data privacy, secure storage infrastructure, the scope of informed consent, and the boundaries of commercial data application. In the absence of established regulatory frameworks, reconciling data utility with the protection of participants' rights remains a critical prerequisite for ethically sustainable large-scale implementation.

5. Conclusions

This study systematically evaluated the utility of digital phenotyping in identifying HRGB by integrating passively collected physiological data from wearable devices with actively reported psychobehavioral indicators. The results have shown that machine learning models incorporating multimodal data performed best at distinguishing be-

tween HRGB and non-HRGB individuals, with depressive symptoms, heart rate index, and deep sleep proportion presenting significant predictive values. Early intervention targeting depressive symptoms may have the potential to prevent HRGB. These findings not only confirm the advantages of wearable devices in enabling continuous and objective monitoring of behavior and physiological states and provide empirical support for early screening and preventive interventions based on digital phenotyping. The feasibility of this approach is enhanced by the non-invasive nature and widespread adoption of consumer-grade wearables (e.g., smart bands) among adolescents, which facilitates large-scale monitoring in naturalistic settings and strengthens ecological validity. Future research should further validate the robustness and generalizability of these biobehavioral markers by increasing sample sizes, extending observation periods, and balancing gender distribution. Additionally, exploring their potential applications in clinical and public health contexts, such as integrating them into routine school health screenings or outpatient follow-up monitoring, could advance precision prevention and personalized interventions for gaming disorders.

Availability of Data and Materials

The datasets used and analyzed during the current study are available from the corresponding authors on reasonable request.

Author Contributions

YJS, YQT, TZC, and YWS contributed to the design of the data analysis plan. YWS, YQT, YW, HDS, CNH, SL, ZKL, NZ, MZ, JD, and HS collected the data. YJS performed the data analysis. TZC, NZ, MZ, and JD supervised the study. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

This study was reviewed and approved by the Ethics Committee of Shanghai Mental Health Center (Approval No. 2023-63). All procedures performed in this study were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all participants.

Acknowledgment

We thank the Innovative Research Team of High-level Local Universities in Shanghai for their support and appreciate Huawei's research platform for providing computational resources and technical support for this study. We also gratefully acknowledge Catherine Chen, a high school

intern from The Nueva School (USA), for her dedicated assistance with data analysis during her supervised internship and for her thoughtful insights from an adolescent perspective, which enriched the interpretation of findings and informed practical considerations for supporting youth at risk of gaming-related challenges. The authors also thank all the participants who took part in the sample collection and clinical study.

Funding

This work was supported by STI2030-major projects (2022ZD0211100, NZ), the National Key R&D Program of China (2023YFC3304204, JD), National Natural Science Foundation of China (82571707, JD; 82571704, TZC), Shanghai Jiao Tong University 'Star of Jiao Tong' Program Medical-Engineering Interdisciplinary Research Fund (YG20237D25, JD), Shanghai "Rising Stars of Medical Talents" Youth Development Program (SHWSRS(2025)_071, TZC), Shanghai Clinical Research Center for Mental Health (19MC1911100, MZ) and Xuhui District Technology Cooperation Project (23XHYD-09, MZ). None of the funding bodies played any role in the study's conceptualization, design, data collection, analysis, decision to publish or manuscript preparation.

Conflicts of Interest

The authors declare no conflicts of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/AP48256>.

References

- [1] Limone P, Ragni B, Toto GA. The epidemiology and effects of video game addiction: A systematic review and meta-analysis. *Acta Psychologica*. 2023; 241: 104047. <https://doi.org/10.1016/j.actpsy.2023.104047>
- [2] Macur M, Pontes HM. Internet Gaming Disorder in adolescence: investigating profiles and associated risk factors. *BMC Public Health*. 2021; 21: 1547. <https://doi.org/10.1186/s12889-021-11394-4>
- [3] Ko CH, Lin HC, Lin PC, Yen JY. Validity, functional impairment and complications related to Internet gaming disorder in the DSM-5 and gaming disorder in the ICD-11. *The Australian and New Zealand Journal of Psychiatry*. 2020; 54: 707–718. <https://doi.org/10.1177/0004867419881499>
- [4] Wong HY, Mo HY, Potenza MN, Chan MNM, Lau WM, Chui TK, et al. Relationships between Severity of Internet Gaming Disorder, Severity of Problematic Social Media Use, Sleep Quality and Psychological Distress. *International Journal of Environmental Research and Public Health*. 2020; 17: 1879. <https://doi.org/10.3390/ijerph17061879>
- [5] Billieux J, Stein DJ, Castro-Calvo J, Higushi S, King DL. Rationale for and usefulness of the inclusion of gaming disorder in the ICD-11. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*. 2021; 20: 198–199. <https://doi.org/10.1002/wps.20848>.
- [6] Long J, Bhadr R, Potenza MN, Orsolini L, Phan V, Kanabar M, et

- al. Public health approaches and policy changes after the inclusion of gaming disorder in ICD-11: global needs. *BJPsych International*. 2022; 19: 63–66. <https://doi.org/10.1192/bji.2021.57>
- [7] Lindenberg K, Kindt S, Szász-Janoch C. Effectiveness of Cognitive Behavioral Therapy-Based Intervention in Preventing Gaming Disorder and Unspecified Internet Use Disorder in Adolescents: A Cluster Randomized Clinical Trial. *JAMA Network Open*. 2022; 5: e2148995. <https://doi.org/10.1001/jamanetworkopen.2021.48995>
- [8] Corneille O, Gawronski B. Self-reports are better measurement instruments than implicit measures. *Nature Reviews Psychology*. 2024; 3: 835–846. <https://doi.org/10.1038/s44159-024-00376-z>
- [9] Wolvers PJD, Bruin SC, Mairuhu WM, de Leeuw-Terwijn M, Hutten BA, Brandjes DPM, et al. Self-Reported Smoking Compared to Serum Cotinine in Bariatric Surgery Patients: Smoking Is Underreported Before the Operation. *Obesity Surgery*. 2020; 30: 23–37. <https://doi.org/10.1007/s11695-019-04128-4>
- [10] Gullett N, Zajkowska Z, Walsh A, Harper R, Mondelli V. Heart rate variability (HRV) as a way to understand associations between the autonomic nervous system (ANS) and affective states: A critical review of the literature. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*. 2023; 192: 35–42. <https://doi.org/10.1016/j.ijpsycho.2023.08.001>
- [11] Riemann D, Krone LB, Wulff K, Nissen C. Sleep, insomnia, and depression. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*. 2020; 45: 74–89. <https://doi.org/10.1038/s41386-019-0411-y>
- [12] Pearce M, Garcia L, Abbas A, Strain T, Schuch FB, Golubic R, et al. Association Between Physical Activity and Risk of Depression: A Systematic Review and Meta-analysis. *JAMA Psychiatry*. 2022; 79: 550–559. <https://doi.org/10.1001/jamapsychiatry.2022.0609>
- [13] Cheng YC, Su MI, Liu CW, Huang YC, Huang WL. Heart rate variability in patients with anxiety disorders: A systematic review and meta-analysis. *Psychiatry and Clinical Neurosciences*. 2022; 76: 292–302. <https://doi.org/10.1111/pcn.13356>
- [14] Choi A, Ooi A, Lottridge D. Digital Phenotyping for Stress, Anxiety, and Mild Depression: Systematic Literature Review. *JMIR MHealth and UHealth*. 2024; 12: e40689. <https://doi.org/10.2196/40689>
- [15] Cho K, Kim M, Cho Y, Hur JW, Kim DH, Park S, et al. Digital Phenotypes for Early Detection of Internet Gaming Disorder in Adolescent Students: Explorative Data-Driven Study. *JMIR Mental Health*. 2024; 11: e50259. <https://doi.org/10.2196/50259>
- [16] Lyu X, Chen T, Wang Z, Lu J, Ma C, Tan H, et al. The reliability and validity of a screening scale for online gaming disorder among Chinese adolescents and young adults. *BMC Psychiatry*. 2022; 22: 28. <https://doi.org/10.1186/s12888-021-03678-1>
- [17] Shan H, Su H, Guo L, Li X, Huang C, Jiang H, et al. Network analysis of time perspective and its interaction with internet gaming disorder in college students. *Computers in Human Behavior*. 2023; 147: 107844. <https://doi.org/10.1016/j.chb.2023.107844>
- [18] Langholm C, Breiting S, Gray L, Goes F, Walker A, Xiong A, et al. Classifying and clustering mood disorder patients using smartphone data from a feasibility study. *NPJ Digital Medicine*. 2023; 6: 238. <https://doi.org/10.1038/s41746-023-00977-7>
- [19] Liu Y, Peng S, Wu X, Liu Z, Lian Z, Fan H, et al. Neural, cognitive and psychopathological signatures of a prosocial or delinquent peer environment during early adolescence. *Developmental Cognitive Neuroscience*. 2025; 73: 101566. <https://doi.org/10.1016/j.dcn.2025.101566>
- [20] Yu G, Wu X, Liu Z, Shi M, Fan H, Liu Y, et al. Genetic influence and neural pathways underlying the dose-response relationships between wearable-measured physical activity and mental health in adolescence. *Psychiatry Research*. 2025; 349: 116503. <https://doi.org/10.1016/j.psychres.2025.116503>
- [21] Choi J, Gutierrez-Osuna R. Using heart rate monitors to detect mental stress. 2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks. 2009; 219–223. <https://doi.org/10.1109/BSN.2009.13>
- [22] Park SM, Lee JY, Choi AR, Kim BM, Chung SJ, Park M, et al. Maladaptive neurovisceral interactions in patients with Internet gaming disorder: A study of heart rate variability and functional neural connectivity using the graph theory approach. *Addiction Biology*. 2020; 25: e12805. <https://doi.org/10.1111/adb.12805>
- [23] Tang F, Ishwaran H. Random Forest Missing Data Algorithms. *Statistical Analysis and Data Mining*. 2017; 10: 363–377. <https://doi.org/10.1002/sam.11348>
- [24] Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*. 2017; 80: 1–28. <https://doi.org/10.18637/jss.v080.i01>
- [25] Lemoine NP. Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*. 2019; 128: 912–928. <https://doi.org/10.1111/oik.05985>
- [26] Iqbal SMA, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. *NPJ Flexible Electronics*. 2021; 5: 9. <https://doi.org/10.1038/s41528-021-00107-x>
- [27] Liu Y, Gong R, Yu Y, Xu C, Yu X, Chang R, et al. Longitudinal predictors for incidence of internet gaming disorder among adolescents: The roles of time spent on gaming and depressive symptoms. *Journal of Adolescence*. 2021; 92: 1–9. <https://doi.org/10.1016/j.adolescence.2021.06.008>
- [28] Zhang M, Nie Q, Ye W, Wang Y, Yang Z, Teng Z. Longitudinal Dynamic Relationships Between Videogame Use and Symptoms of Gaming Disorder and Depression Among Chinese Children and Adolescents. *Journal of Youth and Adolescence*. 2025; 54: 426–438. <https://doi.org/10.1007/s10964-024-02068-6>
- [29] Munakata M, Imai Y, Hashimoto J, Sakuma H, Sekino H, Abe K, et al. The influence of antihypertensive agents on circadian rhythms of blood pressure and heart rate in patients with essential hypertension. *The Tohoku Journal of Experimental Medicine*. 1992; 166: 217–227. <https://doi.org/10.1620/tjem.166.217>
- [30] Anderson C, Forte G, Hu W, Zhang H, Boyett MR, D’Souza A. Non-canonical role of the sympathetic nervous system in the day-night rhythm in heart rate. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 2023; 378: 20220179. <https://doi.org/10.1098/rstb.2022.0179>
- [31] Li P, Kim JK. Circadian regulation of sinoatrial nodal cell pacemaking function: Dissecting the roles of autonomic control, body temperature, and local circadian rhythmicity. *PLoS Computational Biology*. 2024; 20: e1011907. <https://doi.org/10.1371/journal.pcbi.1011907>
- [32] Ramesh A, Nayak T, Beestrum M, Quer G, Pandit JA. Heart Rate Variability in Psychiatric Disorders: A Systematic Review. *Neuropsychiatric Disease and Treatment*. 2023; 19: 2217–2239. <https://doi.org/10.2147/NDT.S429592>
- [33] Fan H, Wang H, Lian Z, Yu Q, Wu X, Kuang N, et al. Dynamic Interactions Between Hemispheres Reveal a Compensatory Pathway for Motor Recovery in Moderate-to-Severe Subcortical Stroke. *Journal of Stroke*. 2026; 28: 97–114. <https://doi.org/10.5853/jos.2025.01725>
- [34] Zheng H, Wang M, Zheng Y, Dong GH. How sleep disturbances affect internet gaming disorder: The mediating effect of hippocampal functional connectivity. *Journal of Affective Disorders*. 2022; 300: 84–90. <https://doi.org/10.1016/j.jad.2021.12.085>
- [35] Kristensen JH, Pallesen S, King DL, Hysing M, Erevik EK. Problematic Gaming and Sleep: A Systematic Review and Meta-Analysis. *Frontiers in Psychiatry*. 2021; 12: 675237. <https://doi.org/10.3389/fpsy.2021.675237>