Scientific Research Publishing

Health, 2026, 18(1), 94-111
https://www.scirp.org/journal/health
ISSN Online: 1949-5005
ISSN Print: 1949-4998

Development of a 4-Horizon Prognostic System for Mental States via Multimodal Circadian Rhythm Imaging

Eri Matsuyama

Faculty of Informatics, The University of Fukuchiyama, Kyoto, Japan
Email: matsuyama-eri@fukuchiyama.ac.jp

How to cite this paper: Matsuyama, E. (2026) Development of a 4-Horizon Prognostic System for Mental States via Multimodal Circadian Rhythm Imaging. *Health*, **18**, 94-111.
https://doi.org/10.4236/health.2026.181007

Received: December 25, 2025
Accepted: January 19, 2026
Published: January 22, 2026

Abstract

Background: The diagnosis and follow-up of mental disorders still rely heavily on subjective clinical assessments, highlighting the need for objective and quantitative monitoring methods. Although wearable devices enable continuous collection of digital behavioral data, the predictive characteristics associated with different psychiatric conditions remain insufficiently understood. **Objective:** This study proposes a novel four-horizon regression model that predicts future mental-state risk at 1, 3, 7, and 10 days using multimodal inputs integrating wearable activity data with comorbidity information (migraine). **Methods:** Fifteen circadian-rhythm features and one static comorbidity feature were extracted from time-series activity data and converted into 7-day heatmap images. These images served as inputs for a fine-tuned ResNet50 model, which predicted a composite risk score across multiple future time horizons. **Results:** The model demonstrated stable predictive performance (mean RMSE = 0.081). Notably, condition-specific predictive characteristics emerged. In the bipolar disorder group, the 7-day-ahead prediction achieved the highest accuracy (R^2 = 0.65), reflecting the disorder's intrinsic cyclic fluctuations. In contrast, in the unipolar depression group, the 10-day-ahead prediction substantially outperformed short-term forecasts (R^2 = 0.48), suggesting that changes in activity levels may precede subsequent mood alterations. **Conclusions:** These findings indicate that the proposed approach enables quantitative mental-state monitoring and captures condition-specific temporal dynamics, such as periodicity and early predictive signals, embedded within daily activity patterns.

Keywords

Deep Learning CNN, Multi-Horizon Prognostic, Mental Disorder, Circadian Rhythms

1. Introduction

Mental disorders represent a serious public health challenge in modern society. In particular, psychiatric illnesses such as unipolar depression and bipolar disorder exhibit high global prevalence and are major contributors to social and economic burden [1]. Early detection and continuous monitoring of disease trajectories are crucial; however, diagnosis still relies heavily on subjective assessments based on clinical rating scales and patient-reported symptoms, leading to issues related to symptom fluctuation and inter-rater variability [2]. Consequently, there is a strong need for novel indices that enable objective and quantitative evaluation of an individual's mental state.

Recent advances in smartphones and wearable devices have enabled non-invasive, long-term acquisition of physiological data such as activity levels, heart rate, and sleep patterns. These data objectively reflect daily behavioral rhythms and lifestyle characteristics, and they have attracted growing attention for their potential to support monitoring and severity estimation in psychiatric disorders [3] [4]. Numerous studies have investigated psychiatric characteristics using actigraphy-derived activity data. For example, Krane-Gartiser *et al.* reported pronounced disruptions in circadian motor activity rhythms among acutely admitted patients with bipolar disorder [4]. In contrast, actigraphy-based studies of depression have linked major depressive disorder/unipolar depression to reduced diurnal stability and increased activity-rhythm fragmentation [5]. Furthermore, Van Someren *et al.* proposed non-parametric circadian rhythm analysis (NPCRA), which evaluates the regularity of behavioral rhythms using indices such as interdaily stability (IS), intradaily variability (IV), and relative amplitude (RA) [6]-[9]. These indices are particularly relevant to the onset of bipolar episodes, with characteristic changes often emerging several days to weeks in advance [10].

Despite this progress, most existing studies have focused on group-level statistical comparisons or cross-sectional classification. Prior work using digital phenotyping data has frequently documented early anomalies occurring days to weeks before clinical episodes [11], and several studies have reported increased anomaly frequency in the month preceding onset or relapse [12]. Nevertheless, only a limited number of studies have established methodologies capable of capturing these complex temporal patterns and predicting future prognosis across multiple time horizons using deep learning techniques. Collectively, prior findings underscore that early changes occur on a time scale of days to weeks—an important "lead time" during which clinical intervention and anomaly detection become feasible.

Against this background, the present study proposes a new four-horizon forecasting framework that predicts future psychiatric risk by converting multimodal data—primarily circadian activity metrics supplemented with minimal comorbidity information (specifically, migraine history)—into time-series heatmap images. This work serves as a foundation for a planned future project aimed at developing a multi-task model for psychiatric disorder classification and risk prediction.

Ultimately, our goal is to construct a data-driven analytical platform that contributes to objective assessment and early detection of psychiatric disorders, with potential applications in clinical decision support and personalized mental healthcare.

2. Materials and Methods

2.1. Dataset

We used the OBF-Psychiatric dataset [13], a publicly available collection of actigraphy data obtained using a wrist-worn accelerometer (Actiwatch AW4, Cambridge Neurotechnology Ltd., UK). In addition to activity data, we extracted clinical metadata regarding comorbid conditions, specifically the presence or absence of migraine, provided within the dataset. Because the dataset was fully anonymized prior to release, no additional ethical approval was required for this study.

The Actiwatch is a piezoelectric accelerometer that records the intensity, magnitude, and duration of movement in units of gravitational acceleration (g). Raw signals were sampled at 32 Hz, and activity counts were incremented whenever the acceleration exceeded 0.05 g. The device automatically summarizes activity into 1-min epochs, yielding 1440 measurements per day. Continuous recordings were obtained over multiple consecutive days for each participant.

Participants in the OBF-Psychiatric dataset were clinically diagnosed by senior psychiatrists according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) and the International Classification of Diseases, Tenth Revision (ICD-10). We included three groups in our analysis: patients with bipolar disorder (bipolar), those with major depressive disorder (unipolar), and healthy controls. An example of the raw activity profiles for each group is shown in Figure 1.

For the present study, we defined a 7-day consecutive recording period as one analysis window. Within each window, raw activity counts were processed to calculate circadian rhythm indices. Thus, data for each 7-day window were organized into a 24-hour × 7-day grid structure, which served as the basis for the input heatmap images (Figure 2). The number of available windows for each group is summarized in Table 1.

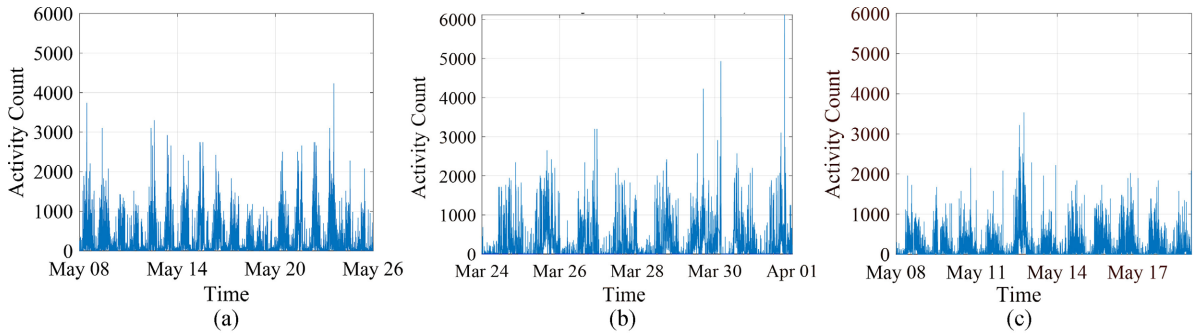

Figure 1. Example actigraphy data from the control group. (a) 19 days for patients with bipolar disorder; (b) 9 days for healthy individuals, and (c) 11 days for patients with unipolar disorder.

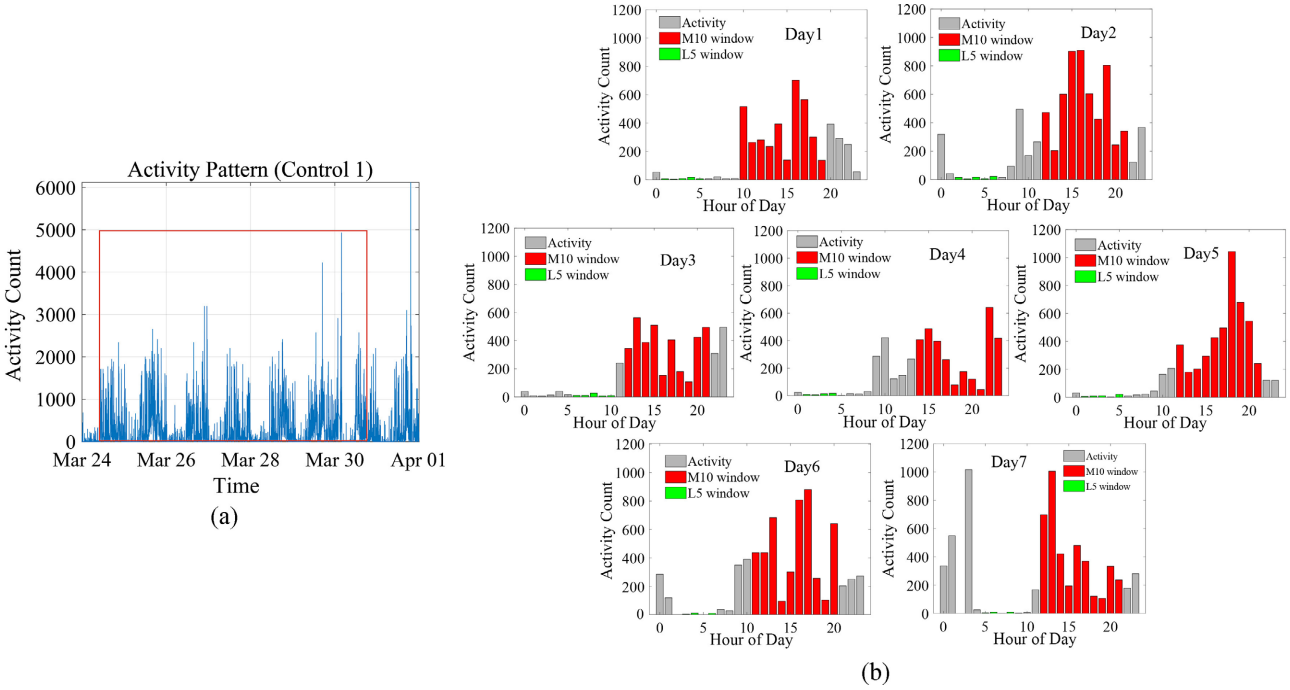

Figure 2. Data division used in the experiment. (a) Measurements for one patient over seven consecutive days are defined as one analysis window; (b) Seven days of the hourly activity profile within one window. The vertical axis represents activity level, and the horizontal axis represents time (24 hours). The window consists of 24 slots. The red slots represent the 10 hours of highest activity (M10), the gray slots represent the 5 hours of lowest activity (L5), and the green slots represent sleep epochs.

Table 1. Data breakdown.

	Number of windows
Bipolar Disorder (BD)	61
Major Depressive Disorder (MD)	105
Healthy Control (HC)	219

To maximize sample size from the continuous actigraphy recordings (mean duration: approx. 13 days), 7-day (168-hour) analysis windows were generated using a sliding window approach with a 1-day step size. This method preserves the temporal structure of the data while capturing the chronological transitions of physiological risk states within each participant.

2.2. Calculation of Multimodal Indices

To capture the multifaceted characteristics of behavioral rhythms, we calculated 15 circadian rhythm indices from each 7-day activity window. These indices were selected to evaluate rhythm stability, activity magnitude, and sleep quality. In addition, the presence of migraine was included as a static binary feature to account for comorbidity-related risk. The definitions of these 16 indicators used in this study are summarized in Table 2.

2.2.1. Rhythm Stability and Fragmentation (NPCRA)

We employed non-parametric circadian rhythm analysis (NPCRA) to quantify

the regularity of rest-activity rhythms [6]. The key indices include interdaily stability (IS), which measures the consistency of 24-hour patterns across days, and intradaily variability (IV), which reflects rhythm fragmentation within a day. We also computed relative amplitude (RA) to assess the contrast between high-activity (daytime) and low-activity (nighttime) periods. These NPCRA metrics are particularly important for detecting rhythm disturbances that are characteristic of bipolar disorder [7].

Table 2. The 16 indicators used in this study.

No.	Indicator Name	Description
1	Bed time	Time at which the individual went to bed (not sleep onset).
2	Sleep latency (SL)	Time required to transition from going to bed to the onset of the lowest-activity period (approximate sleep onset).
3	L5_start	Start time of the least active consecutive 5-hour period.
4	L5_ave	Average activity level during the least active 5-hour period.
5	Fragmentation index	Frequency of awakenings relative to the total sleep time.
6	Sleep efficiency (SE)	Percentage of time in bed spent in sleep (activity below threshold).
7	Wake time	Time at which the individual gets out of bed (activity sharply increases).
8	M10_start	Start time of the most active consecutive 10-hour period.
9	M10_ave	Average activity level during the most active 10-hour period.
10	L5/M10 ratio	Ratio between L5_ave and M10_ave, indicating the balance between rest and activity.
11	Relative amplitude (RA)	Relative difference between rest (L5_ave) and activity (M10_ave) levels.
12	Intra-daily variability (IV)	Degree of fragmentation in daily rest–activity transitions.
13	Inter-daily stability (IS)	Regularity of rest–activity patterns across days.
14	Day total	Average daytime (10:00-18:00) activity level across 7 days.
15	DayNight_ratio7	Ratio of daytime (10:00-18:00) to nighttime (22:00-04:00) activity over 7 days.
16	Migraine	Comorbid risk factor.

2.2.2. Activity Magnitude and Sleep Quality (with Threshold Optimization)

To evaluate the intensity and quality of rest-activity patterns, we calculated indices such as L5, M10, sleep efficiency (SE), and sleep latency (SL). To ensure accurate sleep-wake classification specifically tailored to this dataset, we optimized the activity threshold using the healthy control group. We performed a receiver operating characteristic (ROC) analysis to assess the discriminative ability of activity counts in distinguishing sleep (nighttime) from wakefulness (daytime). As shown in **Figure 3(a)**, the area under the curve (AUC) indicated strong classification per-

formance. We then computed the Youden Index (J = sensitivity + specificity − 1) and identified the cut-off value that maximized J. Based on this analysis, the optimal activity threshold was determined to be 102 counts/min. Figure 3(b) shows the distribution of daytime and nighttime activity, demonstrating the effective separation achieved by this threshold. Using this optimized threshold, each epoch was binary-labeled as sleep or wake, enabling accurate computation of SE, SL, and related sleep metrics.

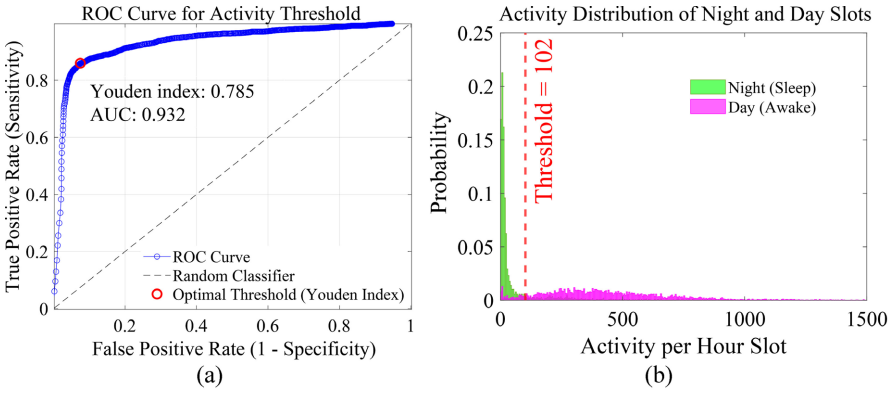

Figure 3. Optimization of the sleep-wake detection threshold using the Youden Index. (a) ROC curve analysis for sleep-wake classification based on activity counts in the Healthy Control group. The optimal cut-off point was determined by maximizing the Youden Index; (b) Histograms of activity counts during daytime (wake) and nighttime (sleep). The vertical line indicates the determined optimal threshold of 102 counts/min, which effectively separates sleep and wake states.

The detailed explanations and calculations of the metrics employed in this study are described below.

- **Intra-daily Variability (IV)**

Intra-daily variability (IV) quantifies the degree and frequency of transitions between rest and activity within a day, computed at an hourly resolution [14]. As shown in Equation (1), IV is derived from the first-order differences of the hourly activity series x_i, which represent abrupt changes in activity levels. These differences are normalized by the unbiased variance of the entire series, allowing IV to capture the fragmentation of the rest-activity rhythm. The theoretical range of IV is 0 - 2, with higher values indicating more frequent or abrupt transitions (*e.g.*, nighttime awakenings or daytime napping). Healthy individuals typically show IV values below 1.

$$IV = \frac{n\sum_{i=2}^{24}\left(x_i - x_{i-1}\right)^2}{(n-1)\sum_{i=1}^{24}\left(x_i - \bar{x}\right)^2}, \tag{1}$$

where x_i is the hourly activity count, $\bar{x}$ is the mean activity across the day, and $n = 24$.

- **Inter-daily Stability (IS)**

Inter-daily stability (IS) quantifies how strongly an individual's rest–activity rhythm is synchronized with the external 24-hour light-dark (zeitgeber) cycle [14]. As defined in Equation (2), IS represents the ratio of the variance of the mean 24-hour activity profile to the total variance of the activity series. Higher IS values indicate greater day-to-day regularity in activity timing.

$$IS = \frac{n\sum_{h=1}^{p}\left(\overline{x_h} - \overline{x}\right)^2}{p\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2}, \tag{2}$$

where $P = 24$ is the number of hourly slots per day, N is the total number of data points, x_i is the activity at time i, and $\overline{x_h}$ is the mean activity at hour h across all days.

- **Least Active Five-Hour Period (L5): L5_ave and L5_start**

L5_ave is the mean activity during the least active continuous 5-hour period within a 24-hour cycle, typically corresponding to nocturnal rest. Lower values reflect consolidated and less fragmented sleep. L5_start denotes the onset time of this least active period, representing the core timing of nocturnal rest.

- **Most Active Ten-Hour Period (M10): M10_ave and M10_start**

M10_ave is the mean activity during the most active continuous 10-hour period, generally capturing daytime behavioral intensity. Reduced M10_ave may indicate motor slowing or diminished daytime engagement. M10_start is the onset time of this high-activity period.

- **Relative Amplitude (RA)**

Relative amplitude (RA) reflects the contrast between daytime and nighttime activity by combining M10_ave and L_5ave (Equation (3)) [15] [16]. Higher RA values indicate robust diurnal rhythms characterized by high daytime activity and low nighttime activity.

$$RA = \frac{M10\_ave - L5\_ave}{M10\_ave + L5\_ave}, \tag{3}$$

- **Ratio of L5_ave to M10_ave**

The L5/M10 ratio (Equation (4)) is a complementary indicator of rhythm robustness, capturing the relative difference between nighttime rest and daytime activity. Values closer to 0 indicate a clear day-night contrast, whereas values approaching 1 suggest diminished diurnal rhythmicity.

$$L5/M10\ ratio = \frac{L5_{ave}}{M10_{ave}}, \tag{4}$$

- **Bed Time**

Bed time is defined as the onset of the resting period preceding the L5_start. For each day, we traced backward up to six hours from L5_start and identified the earliest time point where activity continuously fell below the sleep-wake threshold. This estimate reflects the time a subject went to bed, rather than when sleep onset occurred.

- **Wake Time**

Wake time varies across individuals and days. In this study, wake time was defined as the point of maximal activity increase (largest positive derivative) within the time window from 04:00 to 12:00. This algorithm captures the transition from rest to sustained wakefulness.

- **Sleep Latency (SL)**

Sleep Latency (SL) was computed as the time difference between bedtime and L5_start. This value approximates the latency from going to bed to reaching the deepest and most stable portion of sleep.

- **Sleep Efficiency (SE)**

Sleep efficiency (SE) is a measure of the quality and consolidation of the sleep period. It quantifies the proportion of time a person spends actually sleeping relative to the total time spent in bed. SE is calculated by dividing the total time recorded as sleep (sleep epochs) by the total duration of the nocturnal rest period (time in bed), as follows (Equation (5)):

$$SE = \frac{Number\ of\ sleep\ epochs\left(activity < threshold\right)}{wake\ time - L5_{start}\left(or\ bed\ time\right)}, \tag{5}$$

- **Fragmentation Index**

The fragmentation index reflects the degree to which sleep is continuous or interrupted. As shown in Equation (6), it is defined as the ratio of total sleep time to the number of nighttime awakenings. Lower values indicate more fragmented and unstable sleep.

$$Fragmentation\ index = \frac{sleep\ duration}{number\ of\ awakenings}, \tag{6}$$

- **Day Total and DayNight_ratio7**

Day Total represents the 7-day mean activity during daytime hours (07:00-18:59). DayNightr_atio7 is the 7-day mean ratio of daytime activity (07:00-18:59) to nighttime activity (19:00-06:59). These measures quantify overall behavioral activation and the strength of diurnal preference.

2.3. Construction of Input Data

In this study, we developed a multi-horizon prediction system based on ResNet50 [17] [18], a convolutional neural network with well-established performance in image recognition tasks. The input data consisted of 16 multimodal features: 15 computed circadian rhythm indices and the binary status of migraine comorbidity. Migraine diagnoses were based on the International Headache Society criteria [19] [20]. The binary "migraine" feature (1 for presence, 0 for absence) is assigned to a specific cell coordinate within the 2D grid.

To integrate these heterogeneous features, all scalar values were mapped onto a predefined two-dimensional grid layout to form a single-channel heatmap-like image. This representation enabled the model to process multimodal temporal-behavioral information using convolutional operations. Example inputs for each

diagnostic group: (a) bipolar, (b) control, and (c) unipolar are shown in Figure 4. The gray regions represent zero-padding cells inserted to maintain a fixed spatial resolution and aspect ratio across all samples.

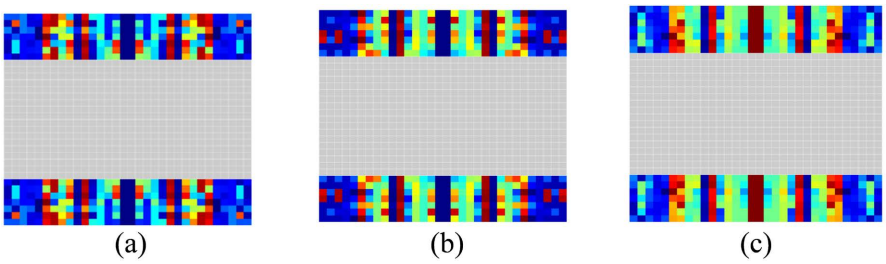

Figure 4. Examples of input heatmap images for the three diagnostic groups: (a) Bipolar, (b) Control, and (c) Unipolar. The gray regions represent zero-padding cells used to standardize the spatial layout across samples.

2.4. Definition of Risk Labels

Because no standardized risk score exists for evaluating mental state severity comprehensively across different psychiatric conditions, we constructed risk labels directly from the circadian rhythm indices. First, to identify indices that strongly reflect pathological states, we calculated Cohen's d for each index as shown in Equation (7).

$$Cohen's\ d = \frac{\mu_{disorder} - \mu_{control}}{SD_{pooled}} \tag{7}$$

where $\mu_{disorder}$ and $\mu_{control}$ represent the mean values for the psychiatric disorder groups (bipolar and unipolar groups) and the healthy control group, respectively, and SD_{pooled} is the pooled standard deviation. Seven indices with the largest effect sizes (IS, RA, IV, SL, SE, SF, and L5/M10) were selected as the core contributors to the risk label.

The selected indices were min-max normalized within the dataset. A weighted sum was then computed using Equation (8) to generate a continuous total risk score ranging from 0 to 1. The weights were heuristically determined based on the magnitude of Cohen's d and clinical relevance.

$$Total_{Risk} = 0.1(1-IS) + 0.3(1-RA) + 0.05IV + 0.3SL + 0.1(1-SE) + 0.1SF + 0.05\left(\frac{L5}{M10}\right), \tag{8}$$

The resulting total risk distribution aligned with clinically expected severity levels (bipolar > unipolar > control), supporting the validity of the computed label. The distributions for each group are shown in Figure 5. For multi-horizon prediction, each input window was assigned four regression targets corresponding to the predicted risk at 1, 3, 7, and 10 days following the observation window.

3. Design of the Multi-Horizon Regression Model

3.1. Network Architecture

We employed ResNet50, a deep residual learning convolutional neural network

(CNN), as the backbone for our multi-horizon regression model. The network was initialized with ImageNet-pretrained weights to leverage feature representations learned from large-scale image datasets (transfer learning).

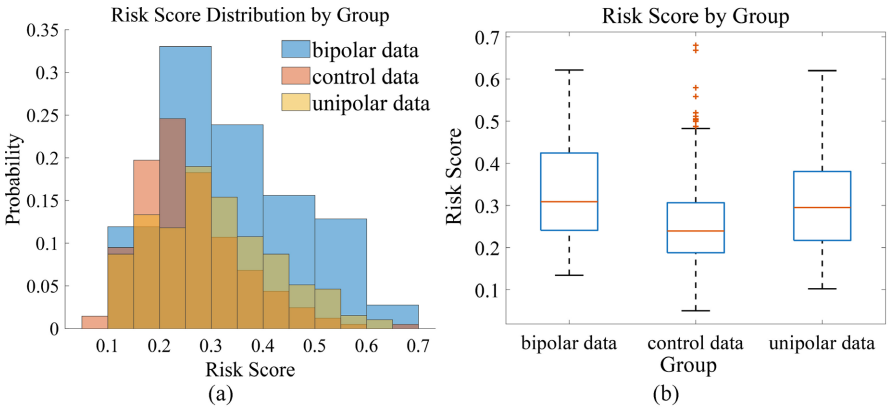

Figure 5. Distribution of total risk values. (a) The horizontal axis is total risk, and the vertical axis is frequency. (b) The horizontal axis is, from left, bipolar disorder, healthy, and unipolar depression, and the vertical axis is risk score.

The architecture was adapted to process the multimodal heatmap images described in Section 2.3. Specifically, the original fully connected classification layer (1000 Softmax outputs) was removed and replaced with a task-specific fully connected layer consisting of four linear output units, corresponding to the predicted risk scores for the four forecasting horizons (1, 3, 7, and 10 days). The global average pooling layer preceding the regression head aggregates high-level spatial features extracted by the convolutional blocks, enabling stable regression performance even under small spatial variations in the input heatmaps. An overview of the complete architecture is shown in Figure 6.

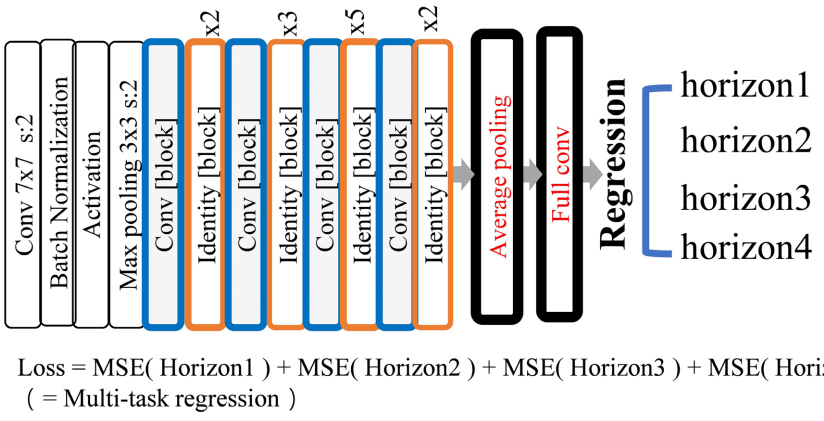

Figure 6. Architecture of the multi-horizon forecasting model. The model utilizes ResNet50 as its backbone architecture, adapted for multi-horizon forecasting via transfer learning. The original fully connected classification layer (softmax output) was removed. A new task-specific regression head, consisting of a fully connected layer with four linear output units, was implemented to predict the risk scores corresponding to the four forecasting horizons (1, 3, 7, and 10 days).

3.2. Loss Function and Training Procedure

The model was trained to minimize the discrepancy between the predicted and ground-truth risk scores. We used the mean squared error (MSE) loss function, defined as:

$$L = \frac{1}{N}\sum_{i=1}^{N}\sum_{h\in\{1,3,7,10\}}\left(y_{i,}, h - \hat{y}_i, h\right)^2, \tag{9}$$

where N is the batch size, y_i, h is the ground-truth risk value for the i-th input at horizon h and $\hat{y}_i, h$ is the corresponding model prediction. To evaluate generalization performance and ensure robustness against inter-individual variability, we conducted 10-fold cross-validation. In each fold, nine subsets were used for training (full fine-tuning of all layers), and the remaining subset was used for validation.

The network parameters were optimized using the Adam algorithm with an initial learning rate of 8×10^{-4} and a mini-batch size of 16. Training was conducted for a maximum of 30 epochs. An early-stopping criterion based on validation loss was included to prevent overfitting; however, in our experiments, training consistently progressed to the full epoch limit, indicating stable convergence without premature stagnation.

To assess generalization performance and ensure robustness to inter-individual variability, we performed 10-fold cross-validation. The dataset was divided into 10 subsets; in each iteration, nine subsets were used for training with all layers fine-tuned, and the remaining subset served as the validation set. All analyses, including preprocessing, model construction, and training, were performed using MATLAB R2024b (The MathWorks, Inc., Natick, MA, USA).

4. Results

4.1. Overall Prediction Performance

The proposed multi-horizon regression model achieved a mean RMSE of 0.081 across all folds and prediction horizons, indicating stable predictive performance on the multimodal circadian heatmaps.

Figure 7 presents the grouped scatter plots of predicted versus true risk scores, overlaid with regression lines for each diagnostic group. As indicated by the regression lines, the predictions generally follow the ideal trajectory ($y = x$), demonstrating a strong correlation between the circadian features and the risk labels. Specifically, the bipolar group (blue) exhibits a high coefficient of determination (R^2), confirming that the model effectively captures the variance in risk associated with bipolar activity patterns.

4.2. Analysis of Prediction Bias

To further investigate the prediction characteristics, Figure 8 illustrates the distribution of residuals (true risk—predicted risk). While the overall distribution is centered around zero, characteristic deviations were observed across groups. In

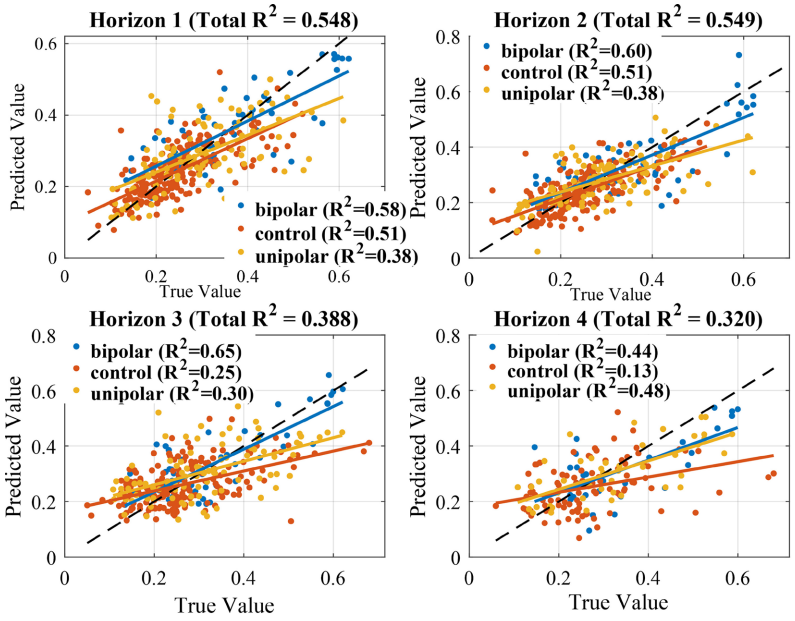

Figure 7. Scatter plots of predicted versus true risk scores across the four prediction horizons. The panels correspond to (a) Horizon 1 (1 day), (b) Horizon 2 (3 days), (c) Horizon 3 (7 days), and (d) Horizon 4 (10 days). Colored points and solid lines represent the individual samples and linear regression fits for each group: bipolar (blue), unipolar (yellow), and healthy control (orange). The black dashed line indicates the ideal prediction ($y = x$). The coefficient of determination (R^2) is provided for each group, highlighting the distinct predictive peaks for Bipolar (Horizon 3) and Unipolar (Horizon 4) groups.

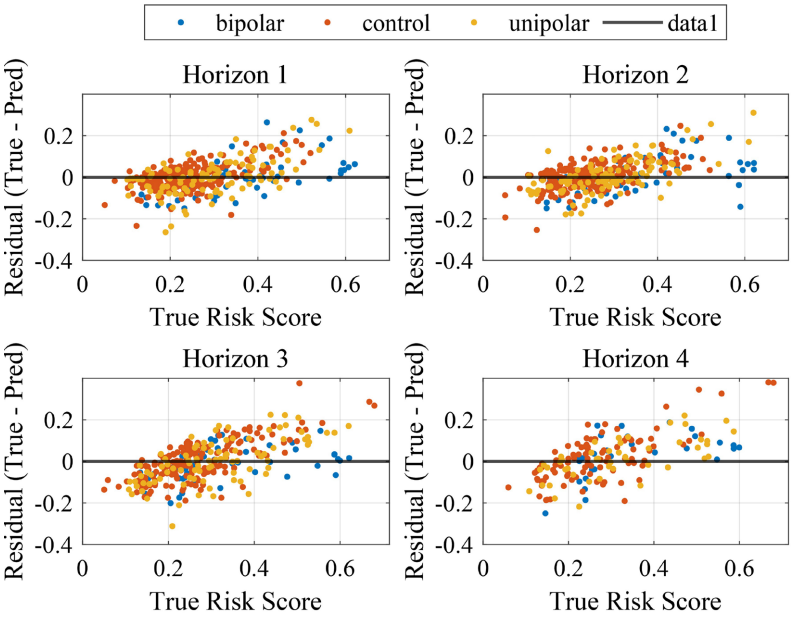

Figure 8. Distribution of prediction residuals (True Risk—Predicted Risk) by horizon. The layout follows the same order as in Figure 7: (a) Horizon 1, (b) Horizon 2, (c) Horizon 3, and (d) Horizon 4. The horizontal black line at zero represents perfect prediction. Positive residuals indicate underestimation (True > Predicted), while negative residuals indicate overestimation (Predicted > True). Note the tendency for overestimation in the bipolar group (blue) in early horizons, and the presence of significant underestimation outliers in the control group (orange) at Horizon 4.

the bipolar group, the residuals tend to shift slightly below zero in the short-term horizons, indicating a systematic tendency toward overestimation (predicting higher risk than actual). Conversely, the healthy control group shows sporadic large positive residuals at Horizon 4, reflecting an underestimation of sudden high-risk episodes. These group-specific bias patterns suggest differing levels of predictability across population subtypes.

4.3. Horizon-Wise Performance by Disorder Group

The predictive accuracy varied significantly by forecast horizon and disorder type, as quantified by the R^2 values shown in Figure 7.

- Bipolar Disorder: Performance peaked at Horizon 3 (7 days), suggesting the capture of weekly periodicity.
- Unipolar Depression: Accuracy improved over time, peaking at Horizon 4 (10 days), suggesting activity as a leading indicator.
- Healthy Controls: Performance dropped sharply at Horizon 4, indicating difficulty in predicting non-periodic fluctuations.

4.4. Representative Individual Cases

To illustrate these trends at the individual level, Figure 9 presents representative prediction trajectories for individual subjects selected from the test set. Figure 9(a) (bipolar disorder) demonstrates a case where the model successfully tracked the risk fluctuation across the horizons. Consistent with the group-level analysis, the prediction at Horizon 3 (7 days) shows particularly close agreement with the ground truth. In contrast, Figure 9(b) (healthy control) illustrates a case where the model prediction remained stable despite a sudden increase in the true risk score at Horizon 4. This deviation corresponds to the underestimation tendency observed in Figure 8.

5. Discussion

5.1. Interpretation of Disease-Specific Predictive Horizons

The most significant finding of this study is that the optimal prediction horizon varies distinctly depending on the psychiatric disorder, as illustrated by the R^2 trajectories in Figure 7. In the bipolar disorder group, the prediction accuracy peaked at Horizon 3 (7 days), as shown in Figure 7. This aligns with clinical observations that bipolar disorder is characterized by cyclical fluctuations in energy and sleep patterns. The high performance at the 7-day horizon suggests that the model successfully extracted the inherent weekly periodicity from the 7-day input window, enabling it to project the rhythmic phase accurately one cycle ahead. The representative case in Figure 9(a) further supports this, showing the model's ability to track periodic risk elevations closely.

Conversely, in the unipolar depression group, accuracy improved over longer horizons, peaking at Horizon 4 (10 days) (Figure 7). This inversion pattern supports the hypothesis that physiological activity alterations serve as a leading indicator

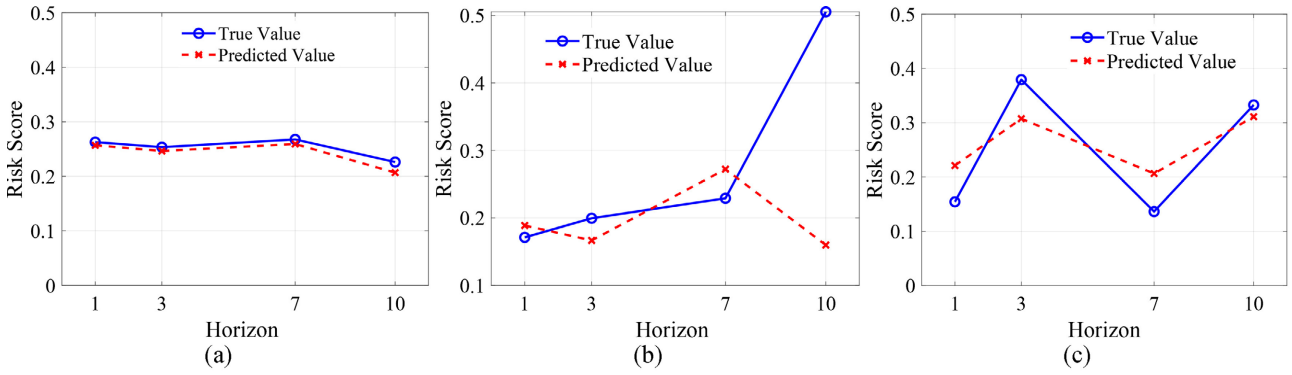

Figure 9. Representative prediction trajectories for individual subjects from each diagnostic group. Panels show test cases from (a) bipolar disorder, (b) healthy controls, and (c) unipolar depression. The dashed (red) line denotes the model-predicted risk score, and the solid (blue) line denotes the ground-truth risk (True Risk) derived from circadian indices. (a) In the Bipolar case, the model tracks periodic fluctuations in risk across prediction horizons. (b) In the Control case, the model predicts consistently low and stable risk values; sudden fluctuations in true risk, when present, may not be fully captured, consistent with the group-level residual analysis. (c) In the Unipolar case, the model follows the overall trend of the true risk trajectory.

for mood changes. According to the theory of behavioral activation, changes in physical activity often precede subjective improvements or deteriorations in depressive mood. Our results imply that the model detected these early prodromal signs in the activity heatmaps, which manifested as quantifiable risk changes only after a time lag of approximately 10 days.

5.2. Prediction Challenges in Healthy Controls

While the model showed high stability for the healthy control group in short-term horizons, performance dropped significantly at Horizon 4 (Figure 7). This indicates a limitation of using activity data alone for long-term prediction in healthy individuals. Unlike the pathological fluctuations seen in psychiatric disorders, the risk variations in healthy controls are often driven by acute stressors that do not leave a distinct prodromal footprint. This limitation is clearly reflected in the residual analysis (Figure 8), where sporadic high-risk episodes in the Control group resulted in significant underestimation outliers.

5.3. Efficacy of Multimodal Circadian Imaging

Our approach of converting time-series actigraphy data into multimodal heatmaps proved effective for prognosis forecasting. By organizing 15 circadian indices into a 2D grid structure (Figure 4), the CNN could capture not only the temporal evolution of individual metrics but also the inter-metric relationships. Furthermore, the integration of static comorbidity data (migraine) demonstrated a feasible framework for multimodal fusion, enhancing the model's context-awareness.

5.4. Clinical Validity of the Risk Score

The risk label was designed as a composite digital biomarker reflecting established physiological deterioration rather than a mere data reconstruction. By aggregating multiple circadian and activity dimensions weighted by their clinical effect sizes,

the score serves as a proxy for objective severity. The high performance of the model demonstrates its capability to capture these complex clinical nuances directly from raw, unstructured data, providing a foundation for automated, real-time prognostic monitoring via wearable sensors.

5.5. Theoretical Superiority Over Traditional Tabular Models

The decision to convert tabular actigraphy data into 2D images, rather than employing traditional machine learning models (e.g., Random Forest or SVM), is based on the theoretical advantage of capturing spatiotemporal dependencies. Unlike conventional models that treat features as discrete, independent inputs in a flattened vector, the ResNet50 architecture utilizes convolutional kernels to recognize localized "morphological" patterns within the activity data. This allows for the effective extraction of complex circadian rhythms and subtle phase shifts that are inherently lost during the flattening process required by traditional tabular formats. Furthermore, this deep learning framework minimizes the need for manual feature engineering and establishes a scalable foundation for the future integration of high-dimensional, multi-modal clinical data.

5.6. Theoretical Rationale and Ablation Study

We employed an image-based ResNet50 framework to preserve the spatiotemporal 'morphology' of activity patterns, which is often lost in traditional tabular models. To quantify the impact of this approach, we performed an ablation study (Table 3). The proposed multimodal model (RMSE = 0.813) outperformed the actigraphy-only baseline (RMSE = 0.862), demonstrating a 5.7% improvement. These results justify the image-conversion complexity and confirm the model's ability to successfully fuse dynamic sensor data with static clinical attributes like migraine status.

Table 3. Ablation study results.

Input Modality	Feature Size	RMSE
Actigraphy Only	Image (Input)	0.862
Proposed (Fusion)	Image + Migraine (Binary)	0.813

5.7. Limitations

Several limitations should be acknowledged. First, the "risk label" used as the ground truth was heuristically defined based on circadian-rhythm indices and effect sizes (Cohen's d), as no continuous gold-standard clinical score was available for day-to-day tracking. Although the resulting label distribution was generally consistent with diagnostic severity, future studies should validate this approach against daily ecological momentary assessments (EMA) of mood. Second, although migraine history was included as comorbidity information, other potential confounding factors—such as medication effects, sleep medications, or changes in treatment regimens—were not explicitly modeled due to dataset constraints.

Third, it is important to clarify that the "risk label" used as the ground truth in this study is a synthetic composite score derived from actigraphic features, rather than a direct evaluation by clinicians using standard diagnostic scales (e.g., HAM-D or YMRS). Consequently, the high prediction accuracy (R2) reflects the model's fidelity in estimating this physiological proxy from raw data, rather than a direct prediction of clinical diagnosis. Future studies are required to validate the correlation between this digital biomarker and standard clinical ratings in an independent cohort to establish its clinical utility. Fourth, this study is the lack of detailed medication data (type and dosage), which may influence motor activity levels. However, since all participants were under clinical care, the model is considered to have learned rhythm patterns that manifest under standard pharmacological treatment. Future research should integrate comprehensive clinical metadata to further isolate the effects of specific medications and other comorbidities.

Finally, we performed cross-validation on a window-wise basis to maximize data utilization for learning state-dependent patterns. We acknowledge that window-level splitting may yield higher performance metrics compared to subject-wise splitting, potentially limiting generalizability to unseen individuals. To address this, future studies should prioritize validation on independent external cohorts. Furthermore, we intend to expand the multimodal framework to incorporate broader clinical metadata and evaluate the system's utility within prospective clinical settings.

6. Conclusions

In this study, we proposed a novel four-horizon prognostic framework for forecasting mental states using Multimodal Circadian Rhythm Imaging—a method that converts heterogeneous physiological data into unified visual representations. By transforming time-series activity features and static comorbidity information into unified heatmap images, we successfully applied a fine-tuned deep convolutional neural network (ResNet50) to predict psychiatric risk across multiple future time horizons.

Our findings indicate that the optimal prediction horizon is intrinsically linked to the underlying pathophysiology of each disorder. Specifically, the model captured the weekly periodicity in bipolar disorder (peaking at 7 days) and identified activity alterations as a leading indicator in unipolar depression (peaking at 10 days). These results provide quantitative evidence that wearable activity data encodes disorder-specific prodromal signatures—patterns that consistently precede clinical symptoms by days to weeks.

Although the current risk-labeling strategy relies on heuristically defined metrics, the proposed framework establishes a foundation for objective and continuous mental-state monitoring. Future work will refine ground-truth labels through ecological momentary assessments and expand the multimodal inputs to incorporate broader clinical metadata. Ultimately, this technology lays the groundwork for personalized, non-invasive mental healthcare, enabling early intervention strategies aligned with individual biological rhythms and behavioral patterns.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] World Health Organization (2017) Depression and Other Common Mental Disorders: Global Health Estimates. https://www.who.int/publications/i/item/depression-global-health-estimates

[2] American Psychiatric Association (2022) Diagnostic and Statistical Manual of Mental Disorders, 5th Edition, Text Revision (DSM-5-TR). https://www.psychiatry.org/psychiatrists/practice/dsm

[3] Cornet, V.P. and Holden, R.J. (2018) Systematic Review of Smartphone-Based Passive Sensing for Health and Wellbeing. *Journal of Biomedical Informatics*, **77**, 120-132. https://doi.org/10.1016/j.jbi.2017.12.008

[4] Krane-Gartiser, K., Henriksen, T.E.G., Morken, G., Vaaler, A. and Fasmer, O.B. (2014) Actigraphic Assessment of Motor Activity in Acutely Admitted Inpatients with Bipolar Disorder. *PLOS ONE*, **9**, e89574. https://doi.org/10.1371/journal.pone.0089574

[5] Luik, A.I., Zuurbier, L.A., Hofman, A., Van Someren, E.J.W. and Tiemeier, H. (2013) Stability and Fragmentation of the Activity Rhythm across the Sleep-Wake Cycle: The Importance of Age, Lifestyle, and Mental Health. *Chronobiology International*, **30**, 1223-1230. https://doi.org/10.3109/07420528.2013.813528

[6] Van Someren, E.J.W., Kessler, A., Mirmiran, M. and Swaab, D.F. (1997) Indirect Bright Light Improves Circadian Rest-Activity Rhythm Disturbances in Demented Patients. *Biological Psychiatry*, **41**, 955-963. https://doi.org/10.1016/s0006-3223(97)89928-3

[7] Van Someren, E.J.W., Swaab, D.F., Colenda, C.C., Cohen, W., McCall, W.V. and Rosenquist, P.B. (1999) Bright Light Therapy: Improved Sensitivity to Its Effects on Rest-Activity Rhythms in Alzheimer Patients by Application of Nonparametric Methods. *Chronobiology International*, **16**, 505-518. https://doi.org/10.3109/07420529908998724

[8] Gonçalves, B.S.B., Cavalcanti, P.R.A., Tavares, G.R., Campos, T.F. and Araujo, J.F. (2014) Nonparametric Methods in Actigraphy: An Update. *Sleep Science*, **7**, 158-164. https://doi.org/10.1016/j.slsci.2014.09.013

[9] Gonçalves, B., Adamowicz, T., Louzada, F.M., Moreno, C.R. and Araujo, J.F. (2015) A Fresh Look at the Use of Nonparametric Analysis in Actimetry. *Sleep Medicine Reviews*, **20**, 84-91. https://doi.org/10.1016/j.smrv.2014.06.002

[10] Witting, W., Kwa, I.H., Eikelenboom, P., Mirmiran, M. and Swaab, D.F. (1990) Alterations in the Circadian Rest-Activity Rhythm in Aging and Alzheimer's Disease. *Biological Psychiatry*, **27**, 563-572. https://doi.org/10.1016/0006-3223(90)90523-5

[11] Bufano, P., Laurino, M., Said, S., Tognetti, A. and Menicucci, D. (2023) Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *Journal of Medical Internet Research*, **25**, e46778. https://doi.org/10.2196/46778

[12] Cohen, A., Naslund, J.A., Chang, S., Nagendra, S., Bhan, A., Rozatkar, A., *et al.* (2023) Relapse Prediction in Schizophrenia with Smartphone Digital Phenotyping during COVID-19: A Prospective, Three-Site, Two-Country, Longitudinal Study. *Schizophrenia*, **9**, Article No. 6. https://doi.org/10.1038/s41537-023-00332-5

[13] Garcia-Ceja, E., Stautland, A., Riegler, M.A., Halvorsen, P., Hinojosa, S., Ochoa-Ruiz, G., *et al.* (2025) OBF-Psychiatric, a Motor Activity Dataset of Patients Diagnosed with

Major Depression, Schizophrenia, and ADHD. *Scientific Data*, **12**, Article No. 32. https://doi.org/10.1038/s41597-025-04384-3

[14] van Someren, E.J.W., Hagebeuk, E.E.O., Lijzenga, C., Scheltens, P., de Rooij, S.E.J.A., Jonker, C., *et al.* (1996) Circadian Rest—Activity Rhythm Disturbances in Alzheimer's Disease. *Biological Psychiatry*, **40**, 259-270. https://doi.org/10.1016/0006-3223(95)00370-3

[15] Liang, H., Wu, C., Lin, C., Chang, H., Lin, Y., Chen, S., *et al.* (2024) Rest-Activity Rhythm Differences in Acute Rehabilitation between Poststroke Patients and Non-Brain Disease Controls: Comparative Study. *Journal of Medical Internet Research*, **26**, e49530. https://doi.org/10.2196/49530

[16] Ali, F.Z., Parsey, R.V., Lin, S., Schwartz, J. and DeLorenzo, C. (2023) Circadian Rhythm Biomarker from Wearable Device Data Is Related to Concurrent Antidepressant Treatment Response. *npj Digital Medicine*, **6**, Article No. 81. https://doi.org/10.1038/s41746-023-00827-6

[17] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[18] Koonce, B. (2021) ResNet 50. In: Koonce, B., Ed., *Convolutional Neural Networks with Swift for Tensorflow*. Apress, 63-72.

[19] Raieli, V., Raimondo, D., Gangitano, M., D'Amelio, M., Cammalleri, R. and Camarda, R. (1996) The IHS Classification Criteria for Migraine Headaches in Adolescents Need Minor Modifications. *Headache*: *The Journal of Head and Face Pain*, **36**, 362-366. https://doi.org/10.1046/j.1526-4610.1996.3606362.x

[20] Ornello, R., Caponnetto, V., Ahmed, F., Al-Khazali, H.M., Ambrosini, A., Ashina, S., *et al.* (2025) Evidence-Based Guidelines for the Pharmacological Treatment of Migraine. *Cephalalgia*, **45**, No. 4. https://doi.org/10.1177/03331024241305381