



Temporal Connectomics for actigraphy-based psychiatric stratification: a comparative study of circadian, entropy, graph, and sequence representations

Chung J

Submitted: April 10, 2026, Revised: version 1, May 18, 2026, version 2, May 26, 2026

Accepted: May 27, 2026

Abstract

Wrist actigraphy is increasingly used for digital phenotyping in psychiatry, yet most analytical approaches compress week-long recordings into static summary statistics that overlook higher-order temporal organization. We introduce Temporal Connectomics, an interpretable representation that characterizes not only how much individuals move, but how their activity is organized across circadian time. The framework combines multiscale entropy with shrinkage-regularized clock-hour coordination graphs to capture complementary aspects of behavioral complexity and temporal coordination. Using a standardized cohort of 117 participants from the public OBF-Psychiatric dataset (ADHD, depression, schizophrenia, and healthy controls), we compared Temporal Connectomics against conventional activity statistics, classical circadian features, graph-only representations, handcrafted feature fusion, and convolutional neural networks under subject-level cross-validation. Temporal Connectomics significantly outperformed conventional static activity summaries while achieving performance comparable to graph-based and deep-learning baselines, demonstrating that temporal organization provides diagnostically relevant information beyond activity magnitude alone. Although graph-only features achieved the highest internal classification performance, the proposed framework offers a unified, interpretable representation that integrates temporal complexity and circadian coordination within a reproducible analytical pipeline. These findings suggest that behavioral organization across time is an informative dimension of digital phenotyping while emphasizing that external validation is required before clinical application.

Keywords

Digital phenotyping, Actigraphy, Wearable sensors, Circadian rhythms, Behavioral biomarkers, Machine learning, Time series analysis, Graph analytics, Multiscale entropy, Explainable artificial intelligence

Jiwon Chung, Northwood High School, 4515 Portola Pkwy, Irvine, CA 92620, USA. *Current affiliation:* Orange County School of the Arts, 1010 N Main St, Santa Ana, CA 9270, USA. jiwonjc@gmail.com

1. Introduction

1.1 Actigraphy as a longitudinal behavioral phenotype

Motor activity is clinically relevant across psychiatric disorders because psychomotor slowing, agitation, irregular sleep-wake organization, and fragmented daily routines are observable components of illness expression. Wrist actigraphy provides a practical method for measuring these behaviors over multiple days in naturalistic settings (1,2). Compared with brief clinical ratings, actigraphy captures time-varying behavioral organization and can reveal whether movement is regular, fragmented, delayed, compressed, or irregular across temporal scales.

The central methodological challenge is representational. A 7-day minute-level recording contains 10,080 observations per subject, but many published analyses reduce that record to daily means or other low-dimensional summaries. Such reductions are useful for interpretability, yet they can discard clinically meaningful information about circadian phase, day-to-day regularity, transitions between rest and activity, and the coordination among different hours of the day (3).

Temporal Connectomics was developed to address this limitation. Rather than asking only how much a subject moved, the framework asked how activity had been organized in time. The present study retained that conceptual aim while specifying the empirical pipeline, feature definitions, model architectures, and validation design in reproducible detail.

1.2 Circadian, entropy, and graph-based representations

Three methodological traditions motivated the feature design. Classical circadian analysis summarizes rest-activity rhythms using measures such as interdaily stability, intradaily variability, relative amplitude, and cosinor parameters (4,5). These features are interpretable and clinically familiar, but they compress the time series into a limited set of rhythm descriptors.

Complexity analysis offers a complementary view. Sample entropy and multiscale entropy quantify how predictable a behavioral signal remains after progressive coarse-graining (6,7). This can be relevant in psychiatry because pathological motor behavior may be excessively rigid at some time scales and irregular at others (8). However, entropy does not directly indicate whether different clock hours are coordinated across days.

Graph-based representations address that remaining question by treating clock hours as nodes and their regularized cross-day relationships as edges. In this framing, a subject's 24-hour activity pattern becomes a small temporal network. This graph-based logic is conceptually related to network approaches used in neuroscience, although the present graph is a behavioral coordination graph rather than a neural connectome (9). The term Temporal Connectomics is therefore used as an analogy for time-of-day coordination, not as a claim of direct brain connectivity measurement.

1.3 Objectives and scope

The study had four objectives. First, the public actigraphy archive was converted into a reproducible subject-level 7-day cohort with explicit data handling rules. Second, static, circadian, entropy, graph, fused handcrafted, and raw-sequence representations were compared under the same validation protocol. Third, the feature definitions and model architecture were specified with enough detail for independent reproduction. Fourth, additional circadian derivative, spectral, transition, and fusion extensions were tested to determine whether more elaborate features improved performance.

The study was not designed to establish a deployable diagnostic biomarker, prove physiological complementarity, or demonstrate external generalization. Claims were therefore restricted to internal validation on the available public dataset.

2. Materials and Methods

2.1 Data source, device, and wear protocol

The analysis used the public OBF-Psychiatric dataset, an aggregated actigraphy archive containing recordings from individuals diagnosed with major depression, schizophrenia, ADHD, healthy controls, and a separate mixed clinical group (10). The dataset had been derived from previously collected psychiatric actigraphy studies (11-15) and was released as anonymized comma-separated value files with one file per subject. The present analysis used the ADHD, depression, schizophrenia, and healthy-control groups; the mixed clinical group was excluded because the

target task was a four-class diagnostic comparison.

According to the dataset descriptor, motor activity had been recorded using an Actiwatch worn on the right wrist (Actiwatch, Cambridge Neurotechnology Ltd., England, model AW4) (10). The device used a piezoelectric accelerometer, sampled at 32 Hz, registered magnitudes above 0.05 g, and stored activity values in one-minute intervals. Thus, the analysis did not use raw tri-axial acceleration but the device-derived one-minute activity-count series provided by the archive.

The public files did not contain individual removal diaries, battery-charging logs, bathing or sports removal indicators, or formal non-wear flags. Wear behavior therefore could not be reconstructed at the event level. This limitation was treated as a data constraint rather than silently corrected. Missing timestamps were explicitly reconstructed, and zero activity was treated cautiously because it could reflect sleep, quiet rest, device removal, or device-specific count behavior.

2.2 Cohort construction and fixed-window extraction

The primary analysis cohort included 117 subjects with at least 7 recorded days. Group counts were ADHD (n=40), control (n=32), depression (n=23), and schizophrenia (n=22). The mean available recording duration before truncation was 15.85 ± 7.83 days.

The unit of analysis was the subject, not the day. This choice was made to avoid pseudo-replication from repeated daily observations

nested within individuals. All features were computed within subject, and all cross-validation splits assigned whole subjects to either training or testing folds.

For each included subject, timestamps were sorted and duplicate timestamps were collapsed by averaging. A complete minute-level grid was reconstructed from the first recorded timestamp. The primary analysis window consisted of the first 7 consecutive days, yielding 10,080 expected minute-level observations per subject.

Residual missing values within the 7-day window were imputed using the subject-specific median activity value at the same minute of day. The mean missing ratio within analyzed windows was 0.0102 ± 0.0312 . This imputation strategy assumed that short gaps were conditionally exchangeable within subject and time of day; the assumption was explicitly considered in the limitations because missing-data handling can materially affect 24-hour activity-pattern analyses (16).

2.3 Detailed feature definitions and implementation

Static features consisted of mean minute-level activity, standard deviation of minute-level activity, and zero ratio. Zero ratio was defined as the proportion of one-minute activity values equal to exactly zero; no epsilon threshold was applied.

Classical circadian features were computed from hourly activity means. Interdaily stability

(IS) was calculated as,
$$\frac{N \sum_h (\bar{x}_h - \bar{x}_{all})^2}{24 \sum_i (x_i - \bar{x}_{all})^2}$$
, where

h indexes the 24 clock hours ($h = 0, \dots, 23$), i indexes all hourly observations within the 7-day window ($i = 1, \dots, N$), N is the number of hourly observations, x_i is the activity value at hour i , $\bar{x}_h$ is the mean activity for clock hour h across days, and $\bar{x}_{all}$ is the overall mean activity across the full 7-day hourly series.

Intradaily variability (IV) was calculated as,
$$\frac{N \sum_i (x_i - x_{i-1})^2}{(N-1) \sum_i (x_i - \bar{x}_{all})^2}$$
. Relative amplitude (RA)

was defined as $(M10 - L5) / (M10 + L5)$, where M10 was the maximum circular 10-hour mean and L5 was the minimum circular 5-hour mean of the 24-hour profile.

Cosinor features were estimated using a single fixed 24-hour harmonic fitted by ordinary least squares: activity was modeled as,
$$\beta_0 + \beta_1 \cos\left(\frac{2\pi t}{24}\right) + \beta_2 \sin\left(\frac{2\pi t}{24}\right)$$
. The extracted features were mesor, amplitude, acrophase in hours, and R^2 . No detrending was applied before cosinor fitting because the goal was to preserve within-window rest-activity structure rather than remove slow behavioral differences.

Multiscale entropy features were computed from log1p-transformed minute-level activity after subject-level z-scoring. This normalization step was included because entropy tolerance is scale dependent. For each scale from 1 to 20 minutes, the series was coarse-grained by non-overlapping window

means, and sample entropy was computed with embedding dimension $m=2$, tolerance $r=0.2$ times the standard deviation of the coarse-grained series, and Chebyshev distance [6,7]. The resulting entropy curve was summarized by trapezoidal area under the curve, mean entropy, and linear slope across valid scales.

Temporal graph features were computed from a 7×24 day-by-clock-hour matrix whose columns were aligned to clock hours (0=00:00). Columns were standardized, and a Ledoit-Wolf shrinkage covariance estimator was used to regularize the 24×24 hourly covariance matrix before conversion to correlation (17). Mean absolute off-diagonal correlation and mean positive off-diagonal correlation were retained as dense summaries. For thresholded graph summaries, only positive correlations at or above 0.20 were retained as weighted edges; negative edges were not included in the thresholded graph. Density, weighted average clustering, and weighted modularity were then computed. Modularity used NetworkX greedy modularity communities.

The primary Temporal Connectomics representation consisted of the static features plus multiscale entropy summaries and shrinkage graph summaries. FullFusion added classical circadian features to this representation. The design therefore tested whether entropy-plus-graph descriptors added value beyond static activity statistics without implying an information-theoretic proof of non-redundancy.

Exploratory rich features additionally included periodogram-derived 24-hour power, 12-hour power, spectral entropy, 24-hour power ratio, median-split rest-activity transition rates, daily profile correlation and RMSE, lag-1 and lag-24 autocorrelation, M10 and L5 start hours, day-night activity ratio, daytime and nighttime means, and clock-profile derivative summaries. The slopes between 00:00 to 05:00, 10:00 to 15:00 and 15:00 to 23:00 were calculated as $slope_{a,b} = \frac{x_b - x_a}{b - a}$, maximum hourly derivative, minimum hourly derivative, and mean absolute hourly derivative.

2.4 Machine Learning models and architecture

Tabular feature families were evaluated with the same random-forest pipeline. Missing feature values were imputed using the training-fold median. The classifier used 400 trees, class-balanced sample weighting, maximum depth of 5, minimum leaf size of 3, random seed 42, and parallel fitting across available cores.

The HourlyCNN baseline used a 168-element loglp hourly sequence from the first 7 consecutive recording days. Within each outer split, sequence values were standardized using training-fold means and standard deviations only. The architecture was Conv1D(1 to 16 channels, kernel size 5, padding 2), ReLU, max pooling by 2; Conv1D(16 to 32, kernel size 5, padding 2), ReLU, max pooling by 2; Conv1D(32 to 64, kernel size 3, padding 1), ReLU, adaptive average pooling to length 8; flattening; dense 64-unit ReLU layer; dropout 0.20; and a final linear softmax classifier.

The CNN was optimized with Adam, learning rate 0.001, weight decay 0.0001, and class-weighted cross-entropy. Training used a stratified 20% validation split within each training fold, a maximum of 80 epochs, and early stopping with patience of 10 epochs based on validation loss. Because the number of subjects per inner fold was modest, full-batch optimization was used rather than mini-batching. Randomness was controlled by seed 42 for NumPy, Python, and PyTorch.

Exploratory sequence modeling used a CNNBiGRU architecture consisting of two 32-channel Conv1D layers with kernel size 5, ReLU activations, max pooling by 2, a bidirectional GRU with hidden size 32, mean pooling over time, a 64-unit ReLU layer, dropout 0.25, and a final linear classifier. StackedFusion combined out-of-fold rich-feature probabilities and CNNBiGRU probabilities with a class-balanced logistic-regression meta-classifier.

2.5 Validation, statistical analysis, and leakage controls

All evaluations used repeated stratified 5-fold cross-validation with three repeats. Splits were performed at the subject level, so records from the same subject could not appear in both training and testing sets. No feature selection, imputation parameter, scaling parameter, or

neural-network validation split used test-fold information.

Accuracy, macro-F1, and macro-recall were summarized as means with bootstrap 95% confidence intervals across repeated splits. Macro-F1 was the primary metric because the groups were imbalanced and because balanced diagnostic performance was more relevant than raw accuracy. Exploratory pairwise comparisons used one-sided Wilcoxon signed-rank tests on matched split-level macro-F1 values, with Temporal Connectomics as the anchor. These tests were interpreted as internal descriptive comparisons, not as confirmatory clinical inference.

Graph robustness was assessed by day-level bootstrap resampling within subject. This diagnostic estimated how much mean absolute correlation varied when the seven available days were resampled. The purpose was to quantify the fragility of correlation-based graphs estimated from a short observation window.

2.6 Analytical assumptions

Several assumptions (Table 1) were necessary because the public archive did not contain all acquisition and compliance metadata that would be available in a prospective protocol.

Table 1. Analytical assumptions, justification, and expected impact.

Assumption	Justification	Potential impact
One-minute activity counts were comparable within the OBF archive.	The dataset descriptor reported identical Actiwatch AW4 devices across original datasets and homogenized CSV output.	Between-device generalization to modern tri-axial devices remains uncertain.
Short gaps could be imputed using subject-specific time-of-day medians.	To preserve circadian structure while avoiding length differences across subjects.	If missingness reflected systematic device removal, imputation could bias low-activity periods.
Exactly zero activity was not automatically equivalent to non-wear.	The released files did not include non-wear flags or removal diaries.	Zero-ratio features were interpreted cautiously and not treated as direct sleep or compliance measures.
Internal repeated cross-validation estimated within-dataset performance only.	No external cohort with identical labels and protocol was available in the project directory.	Generalizability across populations, devices, and recording protocols was not established.

3. Results

3.1 Cohort characteristics and data completeness

Figure 1 summarizes the four-group cohort after enforcing the minimum 7-day recording

criterion. Although total available recording duration differed across groups before truncation, all downstream models used the same 7-day window length per subject.

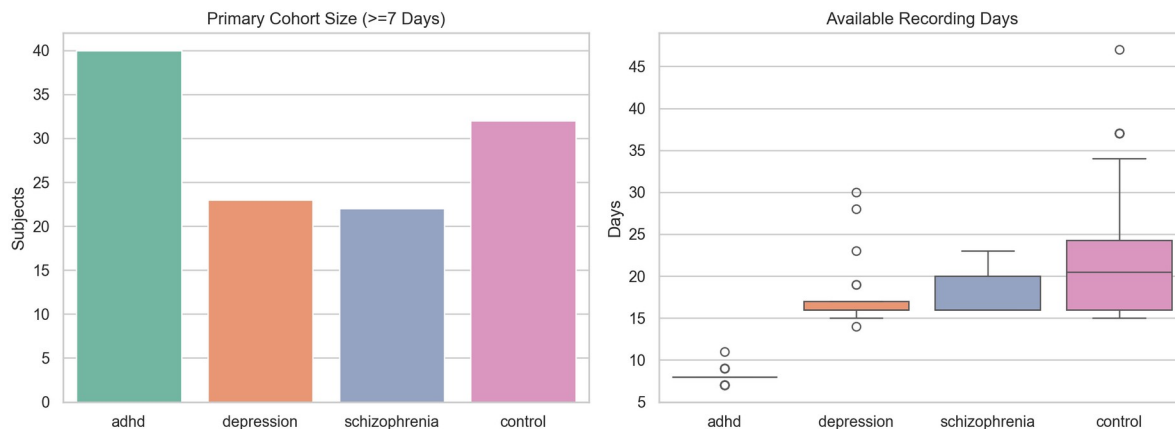


Figure 1. Cohort overview for the primary 7-day analysis set. Left: included subjects per diagnostic group. Right: total available recording days before truncation.

Table 2. Cohort characteristics for the primary 7-day analysis set (mean $\pm$ SD).

Group	n	Available recording days	Missing ratio in 7-day window
ADHD	40	8.03 $\pm$ 0.73	0.0299 $\pm$ 0.0479
Control	32	23.09 $\pm$ 8.67	0.0000 $\pm$ 0.0000
Depression	23	17.61 $\pm$ 4.02	0.0000 $\pm$ 0.0000
Schizophrenia	22	17.73 $\pm$ 2.39	0.0000 $\pm$ 0.0000
Overall	117	15.85 $\pm$ 7.83	0.0102 $\pm$ 0.0312

3.2 Descriptive clock-hour profiles and temporal graphs

Figure 2 displays group-level clock-hour activity profiles and shrinkage-regularized temporal correlation matrices. Hour 0 corresponds to 00:00. The shaded bands around activity profiles represent ± 1 SD across subjects within each group. The plots are descriptive and were not used as standalone inferential evidence.

The profiles suggested visible group differences in morning rise, daytime plateau, evening decline, and between-subject variability. The derivative features added in the exploratory benchmark were motivated by these descriptive patterns. However, because visual group averages can obscure individual heterogeneity, classification models were evaluated at the subject level rather than by heuristic inspection of group curves.

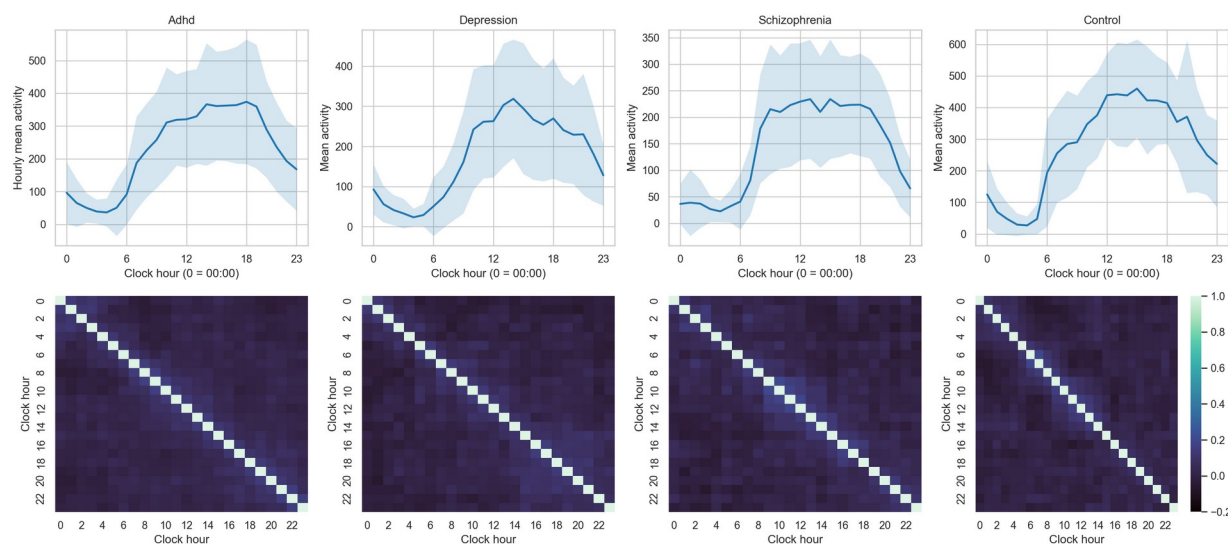


Figure 2. Group-level activity profiles and temporal graph structure. Top row: mean clock-hour activity across the first 7 days, with shaded ± 1 SD bands across subjects. Bottom row: average Ledoit-Wolf shrinkage temporal correlation matrices across clock hours.

3.3 Primary internal-validation benchmark

Across the primary benchmark, GraphOnly showed the highest internal mean macro-F1 (0.430, 95% CI 0.270-0.608). HourlyCNN followed closely (0.418, 95% CI 0.343-0.527). Temporal Connectomics achieved macro-F1 0.416 (95% CI 0.283-0.583) and exceeded StaticStats (mean delta=0.090, p=0.0128).

Temporal Connectomics did not outperform GraphOnly (mean delta=-0.014, p=0.7003) or HourlyCNN (mean delta=-0.002, p=0.5980). Therefore, the appropriate interpretation is that entropy-plus-graph fusion was competitive and better than static activity summaries, but not superior to every strong baseline.

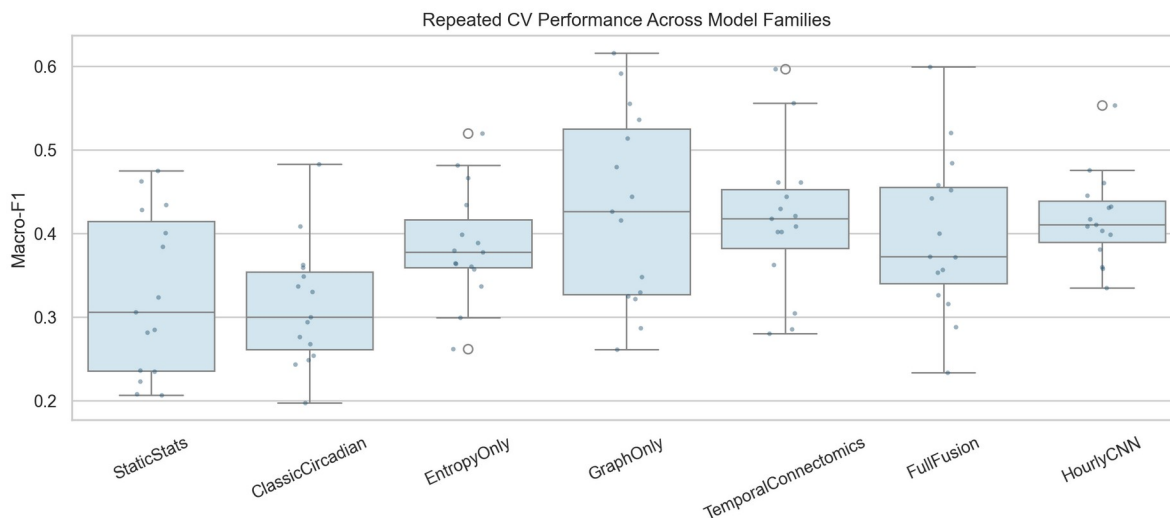


Figure 3. Repeated cross-validation macro-F1 across primary model families. Each point represents one subject-level split from repeated stratified 5-fold cross-validation with three repeats.

Table 3. Internal-validation performance across primary model families.

Model family	Accuracy	Macro-F1	Macro-recall
GraphOnly	0.450 (0.286-0.604)	0.430 (0.270-0.608)	0.455 (0.279-0.627)
HourlyCNN	0.450 (0.381-0.542)	0.418 (0.343-0.527)	0.456 (0.371-0.535)
Temporal Connectomics	0.441 (0.304-0.610)	0.416 (0.283-0.583)	0.452 (0.315-0.624)
FullFusion	0.424 (0.269-0.585)	0.399 (0.253-0.572)	0.438 (0.291-0.587)
EntropyOnly	0.407 (0.291-0.535)	0.386 (0.275-0.507)	0.425 (0.301-0.546)
StaticStats	0.328 (0.212-0.463)	0.326 (0.207-0.471)	0.350 (0.222-0.509)
ClassicCircadian	0.327 (0.229-0.471)	0.314 (0.214-0.457)	0.344 (0.231-0.501)

Values are mean (bootstrap 95% CI) across repeated subject-level splits.

3.4 Exploratory rich-feature and fusion experiments

Exploratory extensions tested circular circadian

, spectral, transition, derivative, and fusion representations. The best extended model was StackedFusion with macro-F1 0.386 (95% CI

0.270-0.537). These extensions did not exceed the best primary internal benchmark and therefore did not support the claim that additional handcrafted complexity alone improved classification in this cohort.

The derivative heuristic suggested by visual inspection was implemented as explicit slope and first-difference summaries rather than as an informal visual rule. Its inclusion in the rich-feature set made the exploratory comparison more responsive to visible profile differences while preserving the same leakage-safe subject-level validation.

Derivative-based descriptors were included in the exploratory rich-feature models rather than in the primary FullFusion model shown in Figure 4. In a separate importance analysis of the exploratory rich-feature model, mean absolute hourly derivative ranked 2nd among rich features (importance=0.039), maximum hourly derivative ranked 10th

(importance=0.032), minimum hourly derivative ranked 11th (importance=0.030), and slope_01_15 ranked 13th (importance=0.027). The remaining derivative slope features ranked lower: slope_00_05 ranked 30th (importance=0.018), and slope_15_23 ranked 33rd (importance=0.017). Ablation analysis showed that ChronoRich_ET with derivative features achieved macro-F1=0.384 (95% CI 0.271-0.478), whereas ChronoRich_ET without derivative features achieved macro-F1=0.382 (95% CI 0.282-0.505), corresponding to a mean paired delta of +0.002. When derivative features were added to the primary FullFusion random-forest feature set, macro-F1 changed from 0.399 to 0.393 (delta=-0.005). These results indicate that derivative descriptors contained discriminative signal, but provided little additional marginal improvement after activity-level, circadian, entropy, transition, and graph descriptors were already included.

Table 4. Exploratory benchmark with rich handcrafted, sequence, and stacked-fusion models.

Exploratory model	Accuracy	Macro-F1	Macro-recall
StackedFusion	0.416 (0.291-0.550)	0.386 (0.270-0.537)	0.442 (0.320-0.578)
Temporal Connectomics_RF	0.424 (0.243-0.623)	0.386 (0.203-0.590)	0.425 (0.245-0.628)
ChronoRich_ET	0.388 (0.254-0.471)	0.384 (0.271-0.478)	0.404 (0.282-0.505)
CNNBiGRU	0.416 (0.296-0.535)	0.381 (0.262-0.542)	0.436 (0.304-0.545)

3.5 Interpretability and graph-stability diagnostics

Supplementary feature-importance and graph-stability analyses were used as interpretability diagnostics (Figure 5). Feature importance was descriptive because random-forest impurity

importance can be biased by correlated predictors. Graph-stability analysis indicated that correlation summaries varied under day-level resampling, which supported cautious interpretation of graph-only performance.

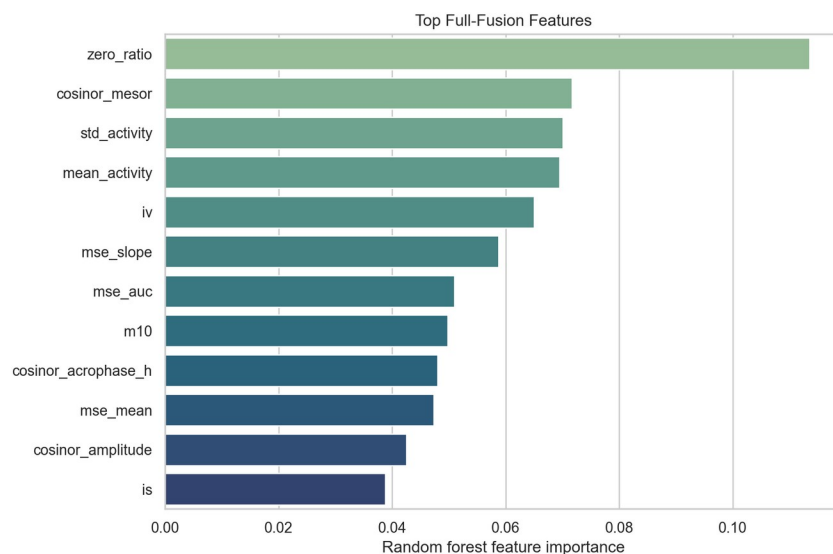


Figure 4. Descriptive feature-importance profile from the full-fusion random-forest model.

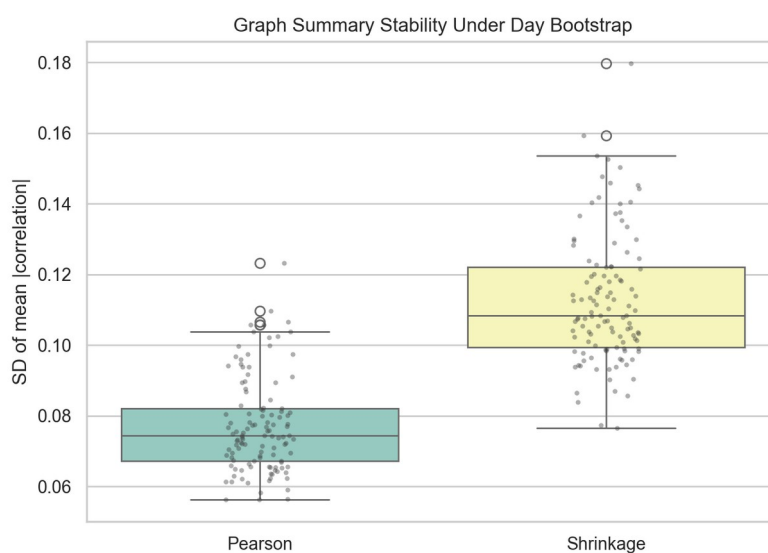


Figure 5. Day-level bootstrap diagnostic for temporal graph summaries.

4. Discussion

4.1 Principal findings

The analysis supports a precise and conservative conclusion. Week-long actigraphy

contained measurable information for four-group psychiatric stratification under subject-level internal validation. Temporal Connectomics improved over static activity summaries and remained competitive with raw-

sequence and graph-only baselines. However, the evidence did not establish that entropy-plus-graph fusion was uniquely superior.

The graph-only result was notable because it suggested that clock-hour coordination may carry useful diagnostic information. At the same time, graph features were estimated from only seven day-level observations per subject. Even with Ledoit-Wolf shrinkage, such correlations can be high-variance estimates. The appropriate interpretation is therefore methodological promise, not clinical readiness.

4.2 Methodological implications

The methodological contribution was not merely representational. The analytic pipeline was specified so that each quantitative claim was linked to a concrete dataset, feature definition, model architecture, and validation protocol. This was necessary because actigraphy features are highly sensitive to device type, body placement, epoch length, non-wear handling, normalization, entropy parameters, graph thresholding, and cross-validation structure.

The study also avoided treating repeated days as independent samples. Days were nested within subjects throughout feature construction, and cross-validation was performed only at the subject level. This design reduced the risk that performance estimates were inflated by leakage across repeated measures.

The study did not demonstrate information-theoretic synergy. No mutual-information decomposition, conditional performance analysis, or formal redundancy-

complementarity proof was performed. The term complementarity is therefore used operationally: different feature families represented different aspects of the same actigraphy signal, and their predictive value was tested empirically under a common validation protocol.

4.3 Limitations

Several limitations should be considered when interpreting these findings. First, the study evaluated a relatively small cohort of 117 participants drawn from a single public dataset. Although repeated subject-level cross-validation reduced optimistic performance estimates, the modest sample size limits statistical precision and does not establish that the observed performance would generalize to broader clinical populations.

Second, the analysis relied on internally validated models only. Consequently, the reported classification performance reflects reproducibility within the OBF-Psychiatric dataset rather than robustness across independent cohorts, wearable devices, or acquisition protocols. External validation remains necessary before any claims of clinical applicability can be made.

Third, the temporal coordination graphs were estimated from only seven daily observations per subject. Although Ledoit-Wolf shrinkage regularization improved covariance estimation, network topology derived from such short observation windows may remain sensitive to short-term behavioral fluctuations. The reported graph measures should therefore be interpreted as approximate representations of

behavioral coordination rather than stable individual network properties.

Fourth, incomplete metadata regarding device removal, charging periods, non-wear intervals, and participant compliance limited the ability to distinguish true behavioral inactivity from missing observations. Consequently, missing-data handling and zero-activity periods may have introduced uncertainty into features describing rest-activity organization.

Fifth, the public dataset was collected using Actiwatch AW4 activity-count recordings rather than raw tri-axial accelerometry. As a result, feature distributions and model performance may differ when the same analytical pipeline is applied to contemporary wearable devices or smartphone-derived movement data.

Sixth, the diagnostic groups represented broad clinical categories with substantial heterogeneity in medication exposure, illness severity, comorbidity, and clinical setting. Some of the observed differences in behavioral organization may therefore reflect these underlying factors rather than diagnosis alone.

Finally, although Temporal Connectomics consistently outperformed conventional static activity summaries, it did not consistently outperform graph-only or raw-sequence representations. Accordingly, the present study supports the proposed framework as a reproducible and interpretable representation of temporal behavioral organization, but it does not establish its superiority over alternative modeling approaches.

4.4 Future directions

Several important research directions emerge from the present work. First, external validation across independent cohorts, wearable devices, and acquisition protocols is essential to determine whether Temporal Connectomics generalizes beyond the OBF-Psychiatric dataset. Second, future studies should extend the framework beyond cross-sectional diagnosis to longitudinal monitoring. Although Temporal Connectomics did not consistently outperform competing models for cross-sectional psychiatric classification, its emphasis on temporal organization may prove more sensitive to within-subject changes in behavioral structure over time. Evaluating temporal connectomic representations during symptom progression, treatment response, or relapse could determine whether changes in behavioral coordination provide earlier or more robust indicators of clinical change than conventional activity-based representations. Third, extending the framework to multimodal digital phenotyping by integrating actigraphy with sleep metrics, physiological signals, smartphone-derived behavioral data, or environmental context may provide a more comprehensive representation of psychiatric state. Methodologically, future work should evaluate dynamic graph representations computed over sliding time windows, allowing behavioral coordination networks to evolve over time rather than remaining static over a fixed seven-day period. Finally, formal analyses of redundancy, complementarity, calibration, and feature attribution, together with correlations between behavioral temporal graphs and clinical outcomes or neurobiological measures, would help clarify

the biological and methodological significance of Temporal Connectomics.

5. Conclusion

This study demonstrates that week-long wrist actigraphy contains sufficient temporal information to support internally validated psychiatric stratification when analyzed using rigorous subject-level validation and explicitly defined feature extraction procedures. Temporal Connectomics introduces an interpretable representation that combines behavioral complexity and circadian coordination within a unified analytical framework. Although this representation improved substantially over conventional static activity summaries, it did not consistently outperform graph-only or raw-sequence approaches, indicating that multiple representations of behavioral dynamics remain competitive on the current dataset. Rather than establishing a clinical diagnostic biomarker, the present work provides a reproducible methodological framework for representing

temporal behavioral organization. Future studies should evaluate the framework across independent cohorts, modern wearable devices, and multimodal digital phenotyping datasets to determine whether temporal connectomic representations provide robust and generalizable behavioral biomarkers for psychiatric disorders.

Acknowledgments

The authors acknowledge the maintainers of the OBF-Psychiatric dataset and the participants whose actigraphy recordings made the secondary analysis possible. The authors also disclose the use of online tools for grammatical correction during manuscript preparation. The authors further thank Jihwan Bae, formerly affiliated with the Agency for Defense Development in the field of low-level computer vision research, for his valuable advice and insightful discussions regarding the theoretical framework and experimental validation of this study.

Data and code availability

The actigraphy data analyzed in this manuscript are available from the public OBF-Psychiatric archive described by Garcia-Ceja et al. (10). The exact local scripts used to generate the reported tables and figures are retained in the project directory under `scripts/rebuild_empirical_study.py`, `scripts/expanded_benchmark.py`, and `scripts/generate_revised_manuscript_r1r2.py`. Versioned code, environment information, and figure-generation scripts can be found in the GitHub repository at https://github.com/jiwooleah/temporal_connectomics

Ethics statement

The present study used a fully anonymized public dataset. No new participant recruitment, intervention, or identifiable human-subject data collection was performed for this secondary analysis. The original dataset descriptor reported ethics approval, participant consent for the original studies, and anonymization before public release.

6. References

1. Price, G. D., Heinz, M. V., Collins, A. C., & Jacobson, N. C. (2023). Detecting major depressive disorder presence using passively collected wearable movement data in a nationally representative sample. *Psychiatry Research*, 328, 115693.
<https://doi.org/10.1016/j.psychres.2023.115693>
2. Stamatis, C. A., Meyerhoff, J., Liu, T., Sherman, G., Wang, Y., & Mohr, D. C. (2024). Differential temporal utility of passively sensed smartphone features for depression and anxiety symptom prediction: A longitudinal cohort study. *npj Mental Health Research*, 3, Article 1.
<https://doi.org/10.1038/s44184-023-00041-y>
3. Lyall, L. M., Wyse, C. A., Graham, N., Ferguson, A., Lyall, D. M., Cullen, B., Celis-Morales, C. A., Biello, S. M., Mackay, D. F., & Pell, J. P. (2023). Subjective and objective sleep and circadian parameters as predictors of depression-related outcomes: A machine learning approach in UK Biobank. *Journal of Affective Disorders*, 335, 83–94.
<https://doi.org/10.1016/j.jad.2023.04.138>
4. van Someren, E. J. W., Swaab, D. F., Colenda, C. C., Cohen, W., McCall, W. V., & 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(4), 505–518. <https://doi.org/10.3109/07420529908998724>
5. Murray, G., Gottlieb, J., Hidalgo, M. P., Etain, B., Ritter, P., Skene, D. J., Garbazza, C., Bullock, B., Merikangas, K., Zipunnikov, V., Shou, H., Gonzalez, R., Scott, J., Geoffroy, P. A., & Frey, B. N. (2020). Measuring circadian function in bipolar disorders: Empirical and conceptual review of physiological, actigraphic, and self-report approaches. *Bipolar Disorders*, 22(7), 693–710. <https://doi.org/10.1111/bdi.12963>
6. Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6), H2039–H2049.
<https://doi.org/10.1152/ajpheart.2000.278.6.H2039>
7. Costa, M., Goldberger, A. L., & Peng, C.-K. (2002). Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, 89(6), 068102.
<https://doi.org/10.1103/PhysRevLett.89.068102>

8. Goldberger, A. L., Peng, C.-K., & Lipsitz, L. A. (2002). What is physiologic complexity and how does it change with aging and disease? *Neurobiology of Aging*, 23(1), 23–26.
[https://doi.org/10.1016/S0197-4580\(01\)00266-4](https://doi.org/10.1016/S0197-4580(01)00266-4)
9. Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, 20(3), 353–364. <https://doi.org/10.1038/nn.4502>
10. Garcia-Ceja, E., Stautland, A., Riegler, M. A., Hinojosa, S., Ochoa-Ruiz, G., Berle, J. Ø., Tørresen, J., Fasmer, O. B., & Jakobsen, P. (2025). OBF-Psychiatric, a motor activity dataset of patients diagnosed with major depression, schizophrenia, and ADHD. *Scientific Data*, 12, 32.
<https://doi.org/10.1038/s41597-025-04384-3>
11. Berle, J. Ø., Hauge, E. R., Oedegaard, K. J., Holsten, F., & Fasmer, O. B. (2010). Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. *BMC Research Notes*, 3, 149.
<https://doi.org/10.1186/1756-0500-3-149>
12. Hauge, E. R., Berle, J. Ø., Oedegaard, K. J., Holsten, F., & Fasmer, O. B. (2011). Nonlinear analysis of motor activity shows differences between schizophrenia and depression: A study using Fourier analysis and sample entropy. *PLoS ONE*, 6(1), e16291.
<https://doi.org/10.1371/journal.pone.0016291>
13. Fasmer, O. B., Halmøy, A., Eagan, T. M., Oedegaard, K. J., & Haavik, J. (2015). Motor activity in adult patients with attention-deficit/hyperactivity disorder. *Psychiatry Investigation*, 12(4), 474–482. <https://doi.org/10.4306/pi.2015.12.4.474>
14. Fasmer, O. B., Mjeldheim, K., Førland, W., Jakobsen, P., Oedegaard, K. J., & Halmøy, A. (2020). Diurnal variation of motor activity in adult ADHD patients analyzed with methods from graph theory. *PLoS ONE*, 15(11), e0241991. <https://doi.org/10.1371/journal.pone.0241991>
15. Jakobsen, P., Garcia-Ceja, E., Riegler, M. A., Stabell, L. A., Nordgreen, T., Tørresen, J., Fasmer, O. B., & Oedegaard, K. J. (2020). Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls. *PLoS ONE*, 15(8), e0231995. <https://doi.org/10.1371/journal.pone.0231995>
16. Weed, L., Lok, R., Chawla, D., & Zeitzer, J. M. (2022). The impact of missing data and imputation methods on the analysis of 24-hour activity patterns. *Clocks & Sleep*, 4(4), 497–507.
<https://doi.org/10.3390/clockssleep4040039>

17. Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2), 365–411.
[https://doi.org/10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4)