

Peer-Review

Chung, Jiwon. 2026. "Temporal Connectomics for Actigraphy-Based Psychiatric Stratification: A Comparative Study of Circadian, Entropy, Graph, and Sequence Representations." *Journal of High School Science* 10 (3): 296–312. <https://doi.org/10.64336/001c.165510>.

Your paper is methodologically strong at the ML evaluation level, but under-specified at the data acquisition and feature-construction level. Some observations and concerns follow;

1. From the methods section, the paper does not specify:

Actigraphy manufacturer/model

Sampling frequency / epoch length (raw vs aggregated)

Body placement (wrist, hip, ankle)

Wear protocol (24h? daytime only?)

Removal conditions (bathing, sports, charging)

Non-wear detection / filtering

This needs to be specified because different devices (e.g., ActiGraph vs GENEActiv vs Fitbit-like devices) use different filters, dynamic ranges, and proprietary activity counts and thus produce non-comparable signals. Without specifying device, external replication becomes ambiguous and feature distributions may not generalize.

2. Body placement changes the signal distribution. Wrist captures higher baseline movement, more noise, hip captures signals related to locomotion whereas ankle captures gait-dominant signal. This affects mean activity, zero ratio, circadian amplitude and entropy. Feature interpretations (especially entropy and graph structure) are placement-dependent.

3. Wear compliance and removal bias: If subjects remove device during bathing or sleep, or intermittently forget to wear it, then missingness is not random, hence your imputation (time-of-day median) may introduce bias.

4. You do not define how feature families are computed. For example, You state: "multiscale entropy summaries". But MSE is highly parameter-sensitive. You must specify entropy type (sample entropy? approximate entropy?), Parameters: (embedding dimension (m), tolerance (r, often fraction of SD), scales used (1-20?), coarse-graining method, Aggregation across scales: mean? slope? AUC? Different parameter choices can result in qualitatively different results especially in short time spans. Please specify compute parameters else your entropy results cannot be reproduced.

5. You mention interdaily stability (IS), intradaily variability (IV), relative amplitude (RA) and cosinor features. But you do not define exact formulas (there are variants) related to

Preprocessing:

Detrending?

Normalization?

Cosinor model:

Single harmonic?

Fixed 24h period?

Fitting method?

Small implementation differences can change feature values significantly.

6. For graphs, you describe 7×24 hourly matrix, Shrinkage covariance $\rightarrow$ correlation, and graph summaries (density, clustering, modularity)

You are missing the following descriptors:

Which shrinkage estimator? (Ledoit–Wolf?)

Thresholding rule for edges

Signed vs unsigned correlations

Modularity algorithm (e.g., Louvain?)

7. Static features: Even simple ones need definition:

"Zero ratio" $\rightarrow$ threshold = exactly zero? or below ϵ ?

"Activity variability" $\rightarrow$ SD? variance? log-scale?

I recommend you add a subsection called "Detailed Feature Definitions and Implementation" or equivalent. Include exact formulas, parameter values, references (standard definitions) and code

availability (deposited into GitHub). In addition, add device + protocol details and fully define entropy, circadian, and graph computations.

You need to specify minimum ML architecture so that your work can be reproduced or critically analyzed. The validity of your main conclusion depends on the CNN being reasonably well-specified and fairly trained

1. You need to specify

Number of layers

Filter sizes / kernel widths

Channel dimensions

Activation functions

Pooling strategy

Sequence length handling

Regularization (dropout, weight decay)

Optimizer, learning rate, batch size

Training epochs / early stopping

Initialization / randomness control

for your ML models.

1. In figure 2, it may help if you specify a time along with the numbers on the x-axis. For example, Zero hours corresponds to what time? 7 AM? Noon? 9 PM? In addition, since these are averages, maybe a shaded portion along the line with 1 SD may present more relevant information to the reader as regards variability for the different disorders.

2. In section 3.4, do you specify what exactly these ‘richer’ circadian and spectral features were? If not, please do so.

3. The entropy parameter is not scale invariant. Did you normalize for scale across the disorders studied? I did not see this procedure presented in the manuscript.

4. Just by visually examining the graphs in figure 2,

(i) Schizophrenia and control upward lines between 15 and 25 hours are steeper than those for ADHD and depression.

(ii) The slope of schizophrenia from 1-15 is steeper than that of ADHD, depression and control at approximately the same time.

(iii) Depression and control have upward positive slopes from 0 to 5. In contrast, ADHD and schizophrenia have negative and zero slopes respectively at these times.

I am assuming the pattern repeats after 24 hours? If this is the case, can you not set up some heuristic (some algebraic derivation of these slopes) that maps the differences from (i), (ii) and (iii) along with (say a first derivative of the graphs) to more accurately categorize these diseases from the graphs alone? You can include in the limitations section that this may not generalize across populations. However, if you do find that your generated heuristic categorizes these diseases significantly better, you can then design an informed CNN or ML model to explicitly take these features into consideration, as well as investigate further into ‘first principles’ as to why these features may matter from a physiological or biological standpoint.

Dear Editor and Reviewers,

We sincerely thank the reviewers for their thoughtful and constructive comments. The manuscript has been substantially revised to improve reproducibility, methodological transparency, and the caution of the interpretation. In particular, we expanded the data acquisition and preprocessing description, added detailed feature definitions and model specifications, clarified assumptions and limitations, revised Figure 2, added exploratory derivative and richer circadian/spectral features, and revised the Discussion to avoid overreaching claims.

Reviewer 1

Comment 1: The manuscript did not specify actigraphy device, sampling frequency, placement, wear protocol, removal conditions, or non-wear handling.

Response: We agree. This information was intentionally omitted, as the analysis was based on an open-access dataset, but we added a new subsection, “2.1 Data Source, Device, and Wear Protocol,” specifying that the dataset used Actiwatch AW4 devices worn on the right wrist, sampled at 32 Hz and stored as one-minute activity-count intervals. We also clarified that the public archive did not contain individual removal diaries, charging logs, bathing/sports removal indicators, or formal non-wear flags. This limitation is now explicitly discussed as a data constraint rather than assumed away.

Comment 2: Wear compliance and missingness may not be random, and time-of-day median imputation may introduce bias.

Response: We added explicit discussion of missingness and imputation in Sections 2.1, 2.2, Table 1, and 4.3. The manuscript now states that missing timestamps were reconstructed, residual gaps were imputed using subject-specific median time-of-day values, and this procedure assumes short gaps are conditionally exchangeable within subject and time of day. We also added a limitation noting that systematic device removal could bias imputed low-activity periods.

Comment 3: Feature families were under-specified, especially multiscale entropy, IS/IV/RA, cosinor, graph features, and static features.

Response: We added a detailed subsection, “2.3 Detailed Feature Definitions and Implementation.” This section now defines static features, zero ratio, IS, IV, RA, cosinor fitting, multiscale sample entropy parameters, graph construction, graph thresholding, signed/unsigned edge handling, and graph metrics. We specified MSE preprocessing, log1p transformation, subject-level z-scoring, scales 1-20, sample entropy embedding dimension $m=2$, tolerance $r=0.2$ SD, Chebyshev distance, and entropy-curve aggregation.

Comment 4: The graph construction was fragile and under-specified.

Response: We clarified that graph features were computed from a 7 x 24 day-by-clock-hour matrix, with clock hour 0 corresponding to 00:00. We specified Ledoit-Wolf shrinkage covariance estimation, conversion to correlation, positive-edge thresholding at $r \geq 0.20$, exclusion of negative edges from thresholded graph summaries, weighted clustering, and NetworkX greedy modularity communities. We also added graph-stability analysis using day-level bootstrap resampling and discussed the residual fragility of seven-day correlation estimates.

Comment 5: The manuscript did not specify the ML architecture sufficiently.

Response: We added “2.4 Machine-Learning Models and Architecture.” This section now specifies the random forest parameters, CNN input length, convolutional layers, channels, kernel widths, pooling strategy, dense layer, dropout, optimizer, learning rate, weight decay, class-weighted loss, validation split, maximum epochs, early stopping, full-batch training, and random seed control. The exploratory CNNBIGRU and stacked-fusion models are also specified.

Comment 6: Figure 2 should clarify what hour zero means and should show variability.

Response: Figure 2 and its caption were revised. The x-axis now uses clock hour with 0=00:00, and shaded bands represent ± 1 SD across subjects within each diagnostic group. Section 3.2 now explicitly states that these plots are descriptive and not standalone inferential evidence.

Comment 7: The manuscript should specify richer circadian and spectral features.

Response: We added the richer features in Section 2.3 and reported the corresponding exploratory benchmark in Section 3.4. These features include 24-hour power, 12-hour power, spectral entropy, 24-hour power ratio, transition rates, daily profile correlation/RMSE, lag-1 and lag-24 autocorrelation, M10/L5 start hours, day-night ratio, daytime/nighttime means, and derivative summaries.

Comment 8: Entropy is not scale invariant; normalization should be described.

Response: We now state that multiscale entropy was computed after $\log_{1p}$ transformation and subject-level z-scoring. This normalization step was added specifically because entropy tolerance is scale dependent.

Comment 9: The visible slopes in Figure 2 might be useful as heuristic features.

Response: We implemented this suggestion as explicit derivative features rather than as an informal visual rule. Section 2.3 defines slope_00_05, slope_01_15, slope_15_23, maximum hourly derivative, minimum hourly derivative, and mean absolute hourly derivative. Section 3.4 reports that these were included in the exploratory rich-feature benchmark, but did not improve beyond the best primary internal benchmark.

Comment 10: The study risked pseudo-replication and leakage because days were nested within subject.

Response: We clarified throughout Sections 2.2 and 2.5 that the unit of analysis was the subject, not the day. All features were computed within subject, and all cross-validation splits assigned whole subjects to training or testing folds. Section 4.2 now explicitly states that repeated days were not treated as independent samples.

Comment 11: The claim of synergy or complementarity was not rigorously demonstrated.

Response: We revised the manuscript to avoid overclaiming. The Discussion now states that the study did not perform mutual-information decomposition, conditional performance analysis, or a formal redundancy-complementarity proof. The term complementarity is used operationally, and conclusions are limited to internal empirical comparison.

Reviewer 2

Comment 1: Reference links should be verified and references should be credible and relevant.

Response: We revised the reference list to include foundational and recent peer-reviewed sources directly supporting actigraphy, circadian rhythm analysis, entropy, graph/network methods, covariance shrinkage, and psychiatric digital phenotyping. DOI links were checked and formatted in APA style.

Comment 2: The manuscript should disclose assistance received in manuscript preparation.

Response: We added an Acknowledgments section disclosing the use of online tools for grammatical correction during manuscript preparation, the use of the OBF-Psychiatric dataset for secondary analysis, and the contribution of Mr. Jihwan Bae, formerly affiliated with the Agency for Defense Development, South Korea in low-level computer vision research, who provided valuable advice regarding the theoretical framework and experimental validation of the study.

Comment 3: Assumptions should be explicitly stated and justified.

Response: We added Section 2.6, “Explicit Analytical Assumptions,” including a table that lists each major assumption, its justification, and its potential impact. These assumptions include comparability of one-minute activity counts, time-of-day median imputation, interpretation of zero activity, and the internal-validation scope of the study.

Comment 4: Conclusions should not overreach beyond the data.

Response: The Abstract, Results, Discussion, Limitations, and Conclusion were revised to state that Temporal Connectomics is a candidate interpretable representation rather than a validated standalone biomarker. We explicitly note that external validation is required before clinical or deployment claims can be made.

Comment 5: The manuscript should use journal-style language, third-person phrasing, and appropriate tense.

Response: We revised the manuscript throughout to use a more formal journal style, removed informal section titles, reduced first-person phrasing, and used cautious past-tense or present-perfect phrasing where appropriate.

Thank you for addressing my comments.

I have one remaining comment:

Although you say that derivative-based features were added in response to my feedback, the manuscript does not report their individual importance, discriminative contribution, or ablation impact in figure S1. Were these not added to the FullFusion model? Or did they contribute too weakly and hence are not among the top features?

Dear Editor and Reviewer,

We sincerely thank the reviewer for their thoughtful and constructive comments. We have revised the manuscript according to the reviewer’s comment.

Reviewer 1

Comment 1: Although you say that derivative-based features were added in response to my feedback, the manuscript does not report their individual importance, discriminative contribution, or ablation impact in figure S1. Were these not added to the FullFusion model? Or did they contribute too weakly and hence are not among the top features?

Response:

Thank you for pointing this out. You are correct that the previous version did not make the derivative-feature contribution sufficiently clear. The derivative-based descriptors were included in the exploratory rich-feature models, not in the primary FullFusion model used for Figure S1. Therefore, their absence from Figure S1 did not mean that they were omitted from the exploratory analysis; rather, Figure S1 reflected the primary FullFusion feature set, whereas derivative

descriptors were evaluated in the exploratory ChronoRich/stacked-fusion feature space. We have revised Section 3.4 to clarify this distinction.

To directly address the reviewer's question, we added the derivative-feature contribution values to the response and clarified the corresponding result in the manuscript. In the exploratory rich-feature model, derivative descriptors were not uniformly weak. 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). However, the ablation analysis indicated that their incremental contribution beyond the full rich-feature set was small. 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. Similarly, 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). Thus, the derivative descriptors appeared to contain discriminative signal, but they provided little additional marginal improvement once activity-level, circadian, entropy, transition, and graph descriptors were already included.

We have therefore revised the manuscript to state that the derivative-based features were included in the exploratory rich-feature models, that several derivative descriptors ranked highly in feature-importance analysis, and that their ablation impact was minimal, suggesting partial redundancy rather than complete absence of signal.

Thank you for addressing my comments. Accepted.