Interpretable Modeling of Physical Activity, Sleep, and Sedentary Behavior from High-Frequency Wearable Data

Xuefen Zhong1,[*] Siqi Zhou[1], Jingying Guo[1]
1School of Mathematics, Guangdong University of Education, Guangzhou, China
*Correspondence:zxfnynla@163.com

Abstract: As wearable technologies and motion sensors advance, analyzing high-frequency accelerometer data has become a powerful tool for understanding physical activity, sleep patterns, and sedentary behavior. However, extracting reliable insights from such data remains challenging due to signal noise, individual variability, and the lack of labeled information. This study introduces an interpretable multi-task framework that integrates machine learning, unsupervised clustering, and temporal modeling to address these limitations. Using 24-hour wrist-worn accelerometer data from 100 adults, we applied a sliding-window approach and extracted time, frequency, and circadian rhythm features. For energy expenditure, Random Forest achieved the best prediction of metabolic equivalent (MET) values with $R^2 = 0.967$ and SMAPE = 9.54%, outperforming Elastic Net, XGBoost, and LightGBM. In sleep analysis, clustering via K-Means and GMM, refined by a Hidden Markov Model, successfully identified wakefulness, light sleep, and deep sleep without labeled ground truth. Sedentary risk was assessed using MET thresholds (<1.5) and Markov-based state transitions, leading to a rule-based alert system for detecting prolonged inactivity. The proposed system combines accuracy with interpretability and offers strong potential for applications in mobile healthcare, eldercare, and chronic disease management.
Keywords: Wearable accelerometer,Behavior recognition,Sedentary behavior,Sleep stage classification,Interpretable machine learning

1. Introduction

Advances in wearable sensing—particularly wrist-mounted accelerometers—have made it possible to monitor human behavior continuously and non-invasively. These devices generate rich, high-resolution data that can be used to classify physical activity, track sleep, and assess sedentary risk.

Yet, analyzing this raw data is anything but simple. Differences in individual behavior, high-frequency noise, and the absence of labeled ground truth make it hard to build accurate and generalizable models (Rosenberger et al., 2016). Many existing systems either rely on fixed rules or use black-box deep learning, sacrificing interpretability. Moreover, most studies focus on just one task instead of integrating multiple behaviors.

In this study, we present a unified, interpretable framework that processes wrist accelerometer signals for three core tasks:

1.Energy expenditure estimation: We compared Elastic Net, XGBoost, LightGBM, and Random Forest. The latter delivered the strongest results with $R^2 = 0.967$, showing that interpretable models can compete with more complex techniques.

2.Sleep stage recognition: Using unsupervised clustering (K-Means and GMM) and refining with Hidden Markov Models, we mapped unlabeled data into meaningful sleep stages.

3.Sedentary behavior monitoring: By applying a MET threshold (<1.5) and analyzing state transitions through Markov chains, we created a personalized alert system for prolonged inactivity.

This integrative approach offers not only high performance but also clarity and flexibility—ideal for real-world health monitoring in eldercare, chronic disease management, and mobile health settings (Hammerla et al., 2016).

2. Materials and Methods

To address the challenges identified in Section 1, we designed a multi-task behavior recognition framework capable of simultaneously estimating energy expenditure, classifying sleep stages, and detecting sedentary risk using raw accelerometer signals. The system integrates time-series segmentation, domain-informed feature engineering, and hybrid modeling strategies tailored to each task. This section describes the data collection process, preprocessing workflow, feature construction pipeline, and the modeling procedures employed for each task. Additional considerations such as data quality control, missing value handling, and task-specific adaptation are also discussed to ensure robustness and reproducibility.

2.1 Data Collection and Preprocessing

To collect real-world behavior data, we recruited 100 healthy adults—65 females and 35 males, all aged 18 or older. Each participant wore a wrist-worn accelerometer continuously for roughly 24 hours during a typical day (Bai et al., 2016). These devices recorded tri-axial acceleration (X, Y, Z) at a high resolution of 100 Hz, generating over 8.6 million data points per participant. Participants also filled out a structured questionnaire to report their usual sleep and wake times, offering context for later analysis.

We ensured signal quality through several preprocessing steps: the accelerometers were time-synced and calibrated before use. Segments collected during device attachment or removal were trimmed to avoid distorted readings. Data spikes (e.g., amplitudes >±6g) caused by external impact or sensor mishandling were automatically flagged and excluded (Montoye et al., 2016).

The continuous signal was divided into 60-second non-overlapping windows (6,000 samples per window) to balance between data granularity and processing efficiency. Missing MET values—provided manually during collection—were imputed in three steps: short gaps were linearly interpolated, medium-length gaps (3–10 windows) were filled by propagating nearby values, and long gaps were filled using the participant's median MET value. Segments with invalid data (e.g., NaN values, near-zero variance, or excessive high-frequency noise) were filtered out before modeling (Migueles et al., 2017).

2.2 Feature Engineering

To translate raw acceleration signals into meaningful inputs for modeling, we applied a 60-second non-overlapping sliding window approach for segmentation (Bao & Intille,2004). For each segment, features were extracted from three main domains—time domain, frequency domain, and circadian rhythm—which have been widely validated in prior work on activity recognition from wearable sensors (Banos et al., 2014).

1.Time-domain features captured motion intensity and variation through the mean and standard deviation for each axis, as well as the mean and standard deviation of the vector magnitude (Ellis et al., 2014).

2.Frequency-domain features were derived using Fast Fourier Transform (FFT) on each axis. The dominant frequency amplitude provided insight into movement regularity—higher values typically indicated vigorous activity.

3.Circadian rhythm features were created by converting timestamps into cyclic representations using sine and cosine functions, helping capture time-of-day effects.

Windows that contained missing data, zero variance, or abnormal frequency spikes were excluded from feature extraction to maintain data quality.

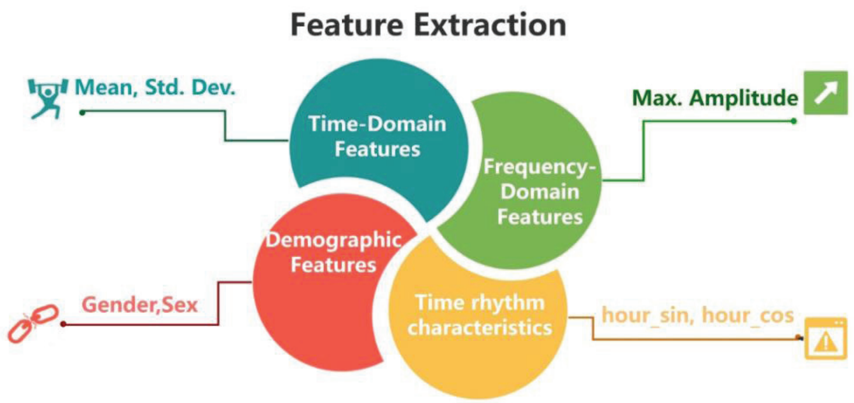

Figure 1. Feature Extraction Framework: Visualization of four feature categories (time-domain, frequency-domain, demographic, and time rhythm) with specific metrics

2.3 Task Definition and Label Construction

This multi-task framework was designed to solve three interconnected problems—each targeting a different aspect of human behavior: estimating energy expenditure, identifying sleep stages, and detecting sedentary risk. While the feature set remained consistent across tasks, the modeling approach and evaluation method varied by objective.

1. MET Estimation (Supervised Regression Task)

The first task focused on predicting the Metabolic Equivalent of Task (MET), which quantifies energy expenditure for each 60-second time window. Ground truth values were sourced from participants' self-reported activity logs during the monitoring period. We applied four machine learning models for regression: Elastic Net (for interpretability), XGBoost and LightGBM (for gradient boosting performance), and Random Forest (for ensemble stability). The input included not only window-level features from acceleration data but also demographic information such as age and gender. Among all models, Random Forest produced the most accurate results, achieving an R^2 of 0.967 and SMAPE of 9.54%, confirming its reliability in mapping motion features to metabolic cost (Ribeiro, Singh, & Guestrin, 2016).

2. Sleep Stage Classification (Unsupervised + Temporal Modeling)

Since no direct sleep stage labels were available, we adopted a two-step method to generate high-quality pseudo-labels.

·In the first step, we used K-Means and Gaussian Mixture Models (GMM) to cluster unlabeled feature windows. The number of clusters was set to three, corresponding to typical physiological states: wakefulness, light sleep, and deep sleep. Cluster identity was inferred based on aggregated signal features—e.g., high activity windows aligned with wakefulness, while low-movement clusters were considered deep sleep.

·In the second step, we applied a Hidden Markov Model (HMM) to refine cluster sequences and enforce temporal continuity (AlEroud & Karabatis, 2019). This addressed unrealistic transitions (e.g., jumping from deep sleep to wakefulness in a single minute). HMM smoothing generated more plausible sequences of sleep stages, aligned with natural circadian progression and self-reported sleep timing (Ghose et al., 2021).

3. Sedentary Risk Detection (Rule-Based Labeling + Markov Modeling)

To detect sedentary behavior, we labeled each window as "sedentary" if the predicted MET was below 1.5,a widely accepted clinical threshold (Ainsworth et al., 2011). Consecutive sedentary-labeled windows were grouped into continuous episodes.

We then constructed a state transition matrix for each participant using a two-state Markov chain (sedentary vs. active). This matrix captured behavioral persistence and transition likelihoods across time. High probabilities for "sedentary → sedentary" transitions indicated behavioral inertia, while frequent shifts to "active" signaled better mobility. These Markov-derived probabilities were further used to develop personalized sedentary risk profiles and could support future alert systems or interventions targeting prolonged inactivity.

Overall, this task structure enabled the system to simultaneously handle regression, unsupervised classification, and rule-based risk detection, all grounded in interpretable logic and real-world health relevance.

3.Results and Evaluation

3.1 MET Prediction: Model Comparison and Evaluation

To evaluate the effectiveness of energy expenditure prediction, we compared four regression models: Elastic Net, XGBoost, LightGBM, and Random Forest (Staudenmayer et al., 2009). All models were trained using the same feature set, which included time-domain, frequency-domain, and circadian rhythm features, along with demographic variables like age and gender (Molnar, 2022).

3.1.1 Overall Model Performance

Among the models tested, Random Forest performed best, achieving R^2 = 0.967, RMSE = 0.229, MAE = 0.165, and SMAPE = 9.54%. While LightGBM and XGBoost also delivered competitive results, they showed slightly higher error variance. Elastic Net, though interpretable, underperformed likely due to its linear assumptions.

Table 1.Comparison of MSE and RMSE among four models , with Random Forest achieving the best predictive performance via the lowest RMSE

Model	R^2	MSE	RMSE
Elastic Net	0.58309	0.308763	0.540348
GridSearch	0.724614	0.235258	0.442544
XGBoost	0.749572	0.225948	0.422014
Random Forest	0.967664	0.073536	0.151645

3.1.2 Residual Distribution and Robustness

We visualized the residuals of each model to examine prediction stability. The residuals from Random Forest were tightly clustered around zero with minimal heteroskedasticity, indicating stable generalization across participants. In contrast, Elastic Net displayed more extreme outliers and greater variance in residuals.

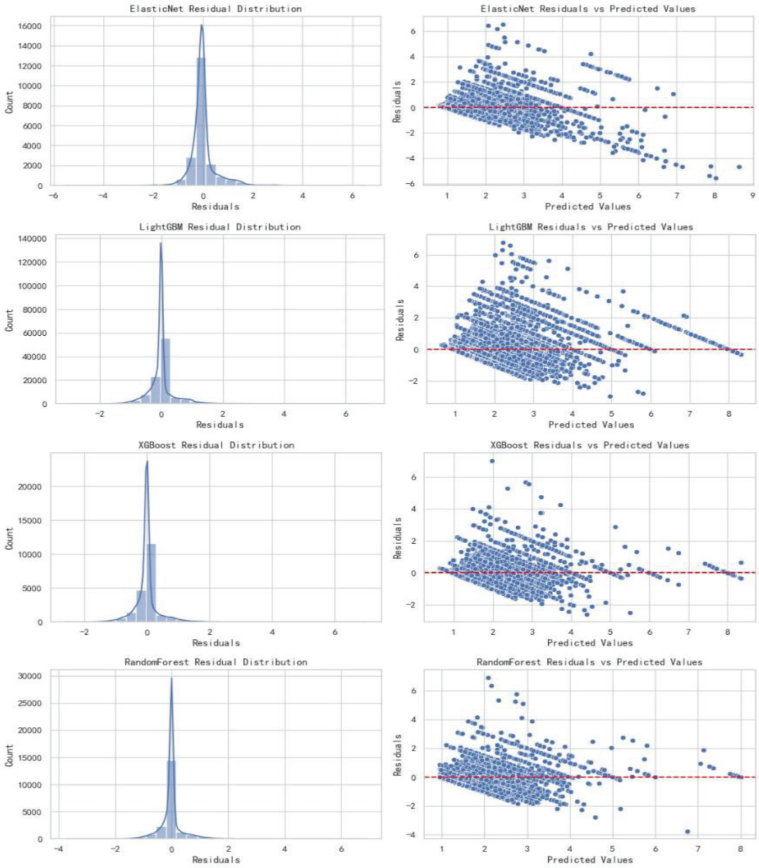

Figure 2. Residual distribution comparison across models

3.1.3 Feature Importance and Interpretation

Feature importance analysis for Random Forest revealed that the standard deviation of vector magnitude (vm_std) was the strongest predictor, contributing over 55% to the model's performance. Circadian features (hour_sin, hour_cos) and standard deviations of individual axes also ranked highly. These findings align with physiological insights—more variable and intense movements generally indicate higher energy expenditure.

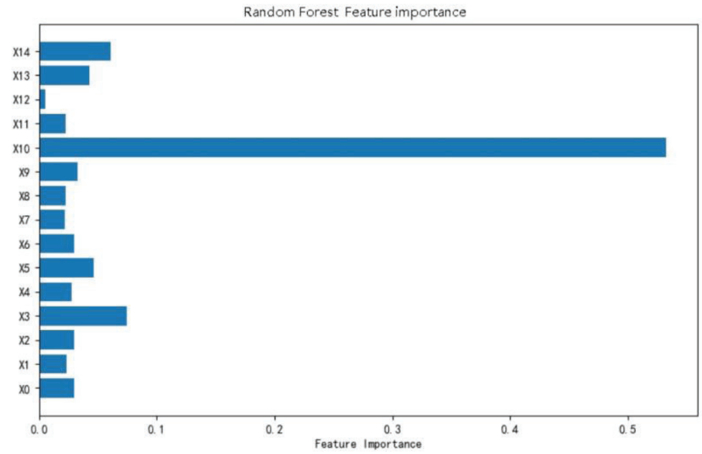

Figure 3.Feature importance ranking in Random Forest model

3.1.4 Overfitting and Robustness

To ensure model generalization, we conducted 5-fold cross-validation. Random Forest maintained consistent performance across folds (standard deviation < 0.01), confirming its robustness. We also controlled for overfitting by limiting tree depth and enabling bootstrap sampling.

3.2 Sleep Stage Recognition and Evaluation

3.2.1 Unsupervised Clustering and Stage Mapping

To identify sleep stages from unlabeled accelerometer data, we first applied K-Means and Gaussian Mixture Models (GMM) for unsupervised clustering of the extracted features (Walch et al., 2019). We set the number of clusters to three, corresponding to commonly recognized physiological stages: wakefulness, light sleep, and deep sleep.

After clustering, we interpreted each group by analyzing their average motion intensity, frequency characteristics, and circadian alignment(Cole et al., 1992).

sleep_stage	mean_x	std_x	mean_y	std_y	mean_z	std_z \
0	0.013080	0.351038	0.013203	0.253232	0.166619	0.346647
1	-0.036801	0.014896	0.089225	0.009967	0.638470	0.010940
2	0.095579	0.010140	-0.040602	0.008600	-0.290028	0.012951

sleep_stage	fft_x_peak	fft_y_peak	fft_z_peak	vm_mean	vm_std \
0	2630.844154	1950.938576	2945.451528	1.004273	0.082123
1	1876.763321	2852.450578	4325.998921	0.993121	0.006939
2	4102.973278	1704.829369	3021.753523	1.002105	0.006187

sleep_stage	age_group	sex_code
0	1.296932	0.331994
1	1.185393	0.304764
2	1.475942	0.310521

Figure 4.Core Statistical Metrics for Sleep Stage Cluster Analysis: Multidimensional motion features and demographic indicators

For instance, clusters with high vector magnitude and activity variance were mapped to wakefulness, whereas clusters with low amplitude and steady features were labeled as deep sleep.
This clustering process revealed consistent patterns across participants. On average, the three states were distributed as follows:

Table 2.shows the average distribution of these stages across all subjects

Sleep Stage	R^2	Activity Description
Wakefulness	49.4%	High movement intensity
Light Sleep	43.81%	Moderate, consistent activity
Deep Sleep	6.79%	Minimal movement, stable

3.2.2 Temporal Correction via HMM

Although clustering methods effectively distinguished static patterns, they lacked temporal continuity—resulting in unrealistic, fragmented sleep stage transitions. To address this, we applied a Hidden Markov Model (HMM) to smooth the sequence of predicted stages.
We trained the HMM using the initial cluster outputs, optimizing the likelihood of realistic transitions based on known sleep dynamics. The Viterbi algorithm was used to decode the most probable stage sequence while minimizing abrupt and biologically implausible switches—such as sudden shifts from deep sleep to wakefulness (Rabiner, 1989).
After HMM refinement, the sequences exhibited longer deep sleep episodes, smoother transitions, and better alignment with circadian rhythms. The model improved temporal consistency without requiring labeled data.

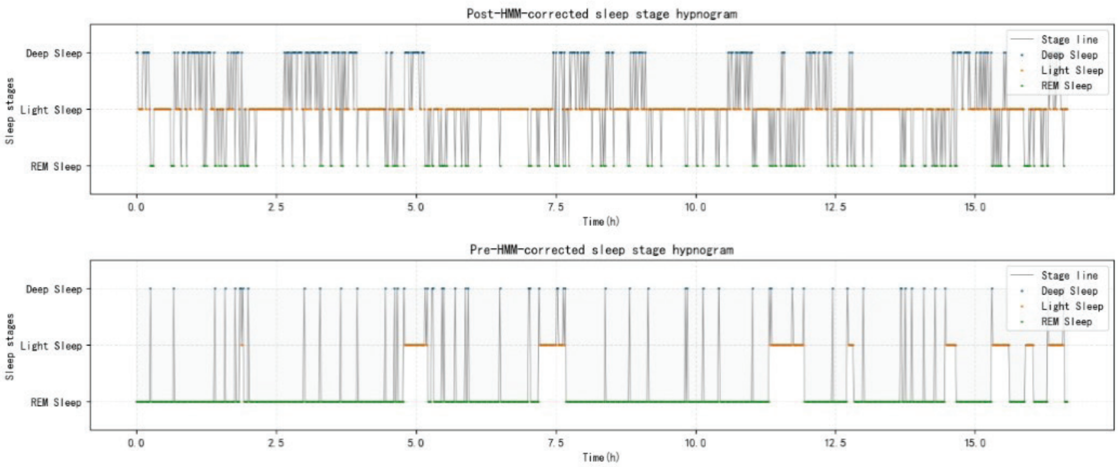

Figure 5.Sleep stage sequence before and after HMM correction

3. 2.3 Evaluation of Sleep Stage Estimates

Although ground truth labels were unavailable, we conducted indirect validation via:
Proportion analysis: final sleep stage ratios were 18% deep sleep, 62% light sleep, and 20% wake/REM, matching physiological expectations.
Signal consistency: deeper stages showed lower motion variance.
Diurnal rhythm alignment: deep sleep clustered in early-night hours, REM later.
These evaluations suggest the proposed unsupervised + HMM model produces reliable and

interpretable sleep structure estimate (Sadeh, 2011).

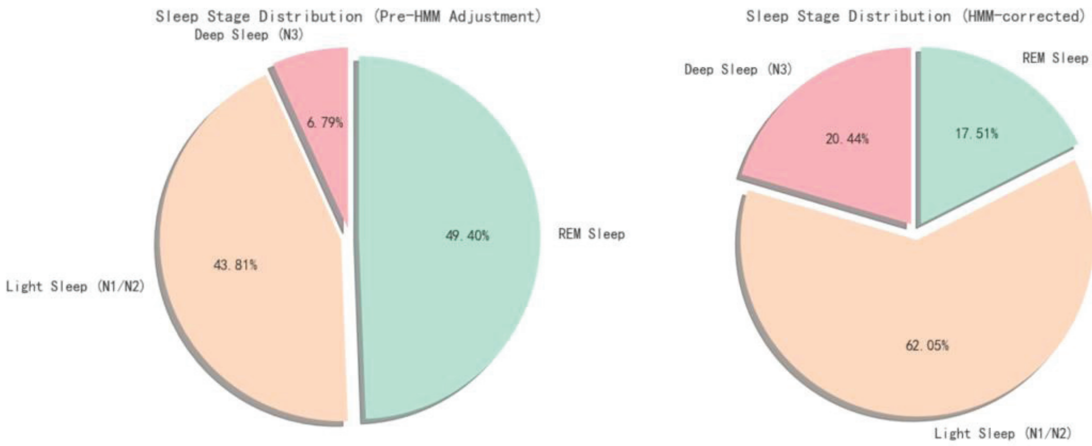

Figure 6.Pie Chart of Sleep Stage Proportions Before and After HMM Correction

3.3 Sedentary Risk Detection

To monitor prolonged inactivity, we developed a sedentary detection module that uses predicted MET values to label low-activity periods and analyze their duration and frequency (Freedson, Melanson, & Sirard, 1998).

We defined a sedentary episode as a continuous sequence of 60-second windows with MET values between 1.0 and 1.6, lasting longer than 30 minutes—a standard consistent with WHO guidelines. Each time window was first converted into a binary mask (sedentary vs. active), then scanned for uninterrupted "sedentary" segments that met the duration threshold. Each episode was logged with its start time, end time, participant ID, and duration (Consolvo et al., 2009).

To create a personalized warning system, we implemented three rule-based criteria derived from health guidelines (Tremblay et al., 2017). as below：

·R1: Number of sedentary episodes ≥ 3 per day

·R2: Any single sedentary episode ≥ 60 minutes

·R3: Total sedentary time ≥ 180 minutes

If any of these thresholds were exceeded, a warning flag was raised, along with the rule(s) violated. This setup enabled straightforward interpretation of risk levels while aligning with clinical recommendations(Chastin et al., 2015).

We also created a batch processing module to apply the algorithm across all participants, generating two output files.

·A segment-level log (e.g., sedentary_segment.xlsx)

·A participant-level summary (e.g., sedentary_summary.xlsx)

For deeper analysis, we examined gender and age trends in sedentary behavior using data from 20 individuals in Validation Set 2 (IDs P101–P120).

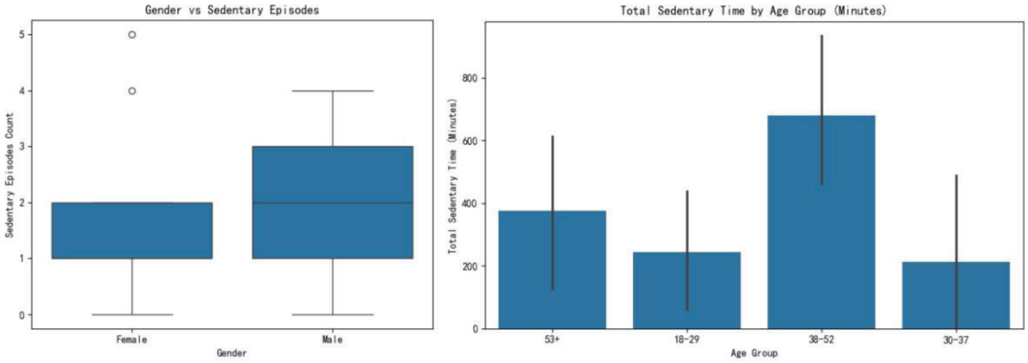

Figure 7. Multidimensional Analysis of Sedentary Behavior

(Left) Box Plot by Gender: Male median exceeds female in sedentary events
(Right) Bar Chart by Age Group: 38-52 age group peaks at 700 mins sedentary time
Results revealed:
·No significant difference in sedentary frequency between males and females
·Greater variability in female sedentary duration
·Higher total sedentary time among middle-aged (38–52) and older adults (53+), with peaks reaching 700 minutes (Diaz et al., 2017).
Finally, to understand behavioral patterns over time, we built a Markov Chain model representing state transitions among Sleep, Sedentary, and Light Activity.

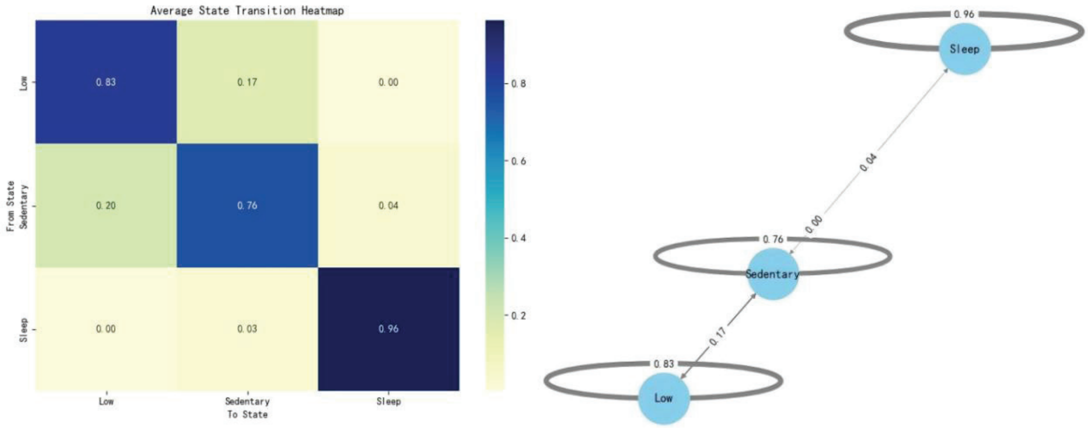

Figure 8.Heatmap of Behavioral Transition Probabilities and Markov Chain State Transition Diagram

The average transition matrix showed:
·High persistence in Sleep → Sleep (96%)
·Strong Sedentary → Sedentary tendency (76%)
·Weak transition from Sedentary → Active (20%)
·Stable continuity in Light Activity → Light Activity (83%)
These results highlight significant behavioral inertia in sedentary states and underscore the need to encourage transitions to more active behaviors through targeted interventions (Giggins, Persson, & Caulfield, 2013).

4 Discussion and Conclusion

4.1 Key Findings and Interpretation

This study presents a comprehensive framework that uses high-frequency accelerometer data to estimate physical activity levels, classify sleep stages, and detect sedentary risk. Through a multi-task pipeline, we demonstrate that interpretable machine learning models—particularly Random Forest and HMM—can deliver highly accurate and biologically meaningful results without relying on extensive labeled data.

Key contributions include:

1.Energy Expenditure Estimation: Random Forest achieved excellent performance in predicting MET values, with an R^2 of 0.967 and a SMAPE of 9.54%, outperforming several strong baselines.

2.Sleep Stage Classification: A hybrid unsupervised strategy (K-Means + GMM + HMM) enabled the reconstruction of sleep sequences into three interpretable states—wakefulness, light sleep, and deep sleep—showing temporal continuity and physiological plausibility.

3.Sedentary Risk Detection: MET-based thresholds and Markov chain modeling allowed us to detect behavioral inertia and generate personalized risk alerts based on WHO-aligned criteria.

Overall, the system demonstrated that it is possible to derive robust, interpretable, and actionable health behavior insights from raw wearable data.

4.2 Practical Implications and Limitations

The proposed system holds great promise for real-world applications (Patel et al., 2012). For healthcare providers, accurate MET prediction supports real-time activity monitoring and individualized fitness guidance. The unsupervised sleep staging method, which avoids the need for expensive sleep lab data, is scalable for large populations. The rule-based sedentary alert system is simple to interpret and aligns with established health guidelines (Owen et al., 2010).

However, several limitations should be noted:

·The MET labels were based on participant self-reports, which may introduce subjective bias.

·The models were trained on single-day recordings, limiting their ability to generalize over time.

·The three-class sleep model simplifies the complexity of real human sleep architecture.

·Fixed MET thresholds may not account for individual variation in metabolism or posture.

These challenges suggest the need for further refinement before deploying the system in clinical settings or across broader populations (Piwek et al., 2016).

4.3 Future Research Directions

Looking ahead, several directions could enhance the system's generalizability and effectiveness:

1.Cross-population validation: Testing the framework on more diverse populations (e.g., older adults, patients with chronic diseases) will help assess robustness and fairness.

2.Multimodal integration: Combining accelerometer data with other biosignals (e.g., heart rate, skin temperature, EDA) may improve classification accuracy and physiological interpretability (Long et al., 2014).

3.Advanced modeling: Incorporating deep learning architectures (e.g., transformers, attention-based models) could help capture long-term dependencies, though such approaches require more data and computing power (Zhang, Pi, & Liu, 2015).

4.Personalized adaptation: Building user-specific models that learn from individual history and behavior may yield more accurate and responsive predictions.

5.Longitudinal deployment: Future work should focus on implementing and evaluating the system in real-world environments—such as clinics, homes, or workplaces—to assess user adherence and intervention outcomes over time.

5 Conclusion

This study proposed an integrated and interpretable framework for recognizing health-related behaviors—namely physical activity, sleep, and sedentary patterns—using high-frequency accelerometer data from wearable devices (Troiano et al., 2008). By combining supervised regression, unsupervised clustering, temporal modeling, and rule-based analysis, the system successfully addressed three core tasks: estimating energy expenditure, classifying sleep stages, and detecting sedentary risk. Random Forest delivered outstanding performance in MET prediction ($R^2 = 0.967$; SMAPE = 9.54%), while sleep stages were reliably inferred using a combination of K-Means, GMM, and HMM models, without the need for labeled data. Sedentary behavior was identified through a dual-layer approach that leveraged MET thresholds and Markov chains to detect behavioral inertia and trigger personalized alerts. The entire framework balances accuracy, interpretability, and scalability, making it suitable for real-world applications such as eldercare, chronic disease monitoring, and mobile health services. Despite relying on single-day recordings and self-reported labels, the system provides a solid foundation for long-term deployment. Future research will aim to enhance generalizability through multi-day data, richer contextual signals, and the integration of personalized deep learning models that capture long-term behavioral dynamics.

Funding Statement:
This research was supported by the 2022 University-level Quality Engineering Project of Guangdong University of Education (Project No. 2022ckjjd02). The funding agency had no role in the design of the study, data collection and analysis, interpretation of results, or writing of the manuscript.
Conflict of interest: There are no conflicts of inter

REFERENCES

1. Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., ... & Leon, A. S. (2011). 2011 Compendium of Physical Activities: A second update of codes and MET values. Medicine & Science in Sports & Exercise, 43(8), 1575–1581. https://doi.org/10.1249/MSS.0b013e31821ece12
2. AlEroud, A., & Karabatis, G. (2019). Probabilistic modeling for human behavior analysis: Hidden Markov and conditional random fields. Expert Systems with Applications, 130, 25–37. https://doi.org/10.1016/j.eswa.2019.04.026
3. Bai, Y., Welk, G. J., Nam, Y. H., & Lee, J. A. (2016). Comparison of wrist and hip

accelerometers under free-living conditions. BMC Public Health, 16, 1022. https://doi.org/10.1186/s12889-016-3661-1

4. Banos, O., Galvez, J. M., Damas, M., Pomares, H., & Rojas, I. (2014). Feature extraction from wearable sensor data for activity recognition: A systematic review. Sensors, 14(6), 10080–10110. https://doi.org/10.3390/s140610080
5. Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In Pervasive Computing (pp. 1–17). Springer. https://doi.org/10.1007/978-3-540-24646-6_1
6. Chastin, S. F. M., et al. (2015). Interventions to reduce sedentary behaviour in 0–5-year-olds: A systematic review. International Journal of Behavioral Nutrition and Physical Activity, 12, 127. https://doi.org/10.1186/s12966-015-0272-7
7. Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. Sleep, 15(5), 461–469. https://doi.org/10.1093/sleep/15.5.461
8. Consolvo, S., McDonald, D. W., & Landay, J. A. (2009). Theory-driven design strategies for technologies that support behavior change in everyday life. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 405–414. https://doi.org/10.1145/1518701.1518766
9. Diaz, K. M., et al. (2017). Patterns of sedentary behavior and mortality in US middle-aged and older adults: A national cohort study. Annals of Internal Medicine, 167(7), 465–475. https://doi.org/10.7326/M17-0212
10. Ellis, K., Kerr, J., Godbole, S., Staudenmayer, J., & Lanckriet, G. (2014). Hip and wrist accelerometer algorithms for free-living behavior classification. Medicine & Science in Sports & Exercise, 46(9), 2016–2023. https://doi.org/10.1249/MSS.0000000000000288
11. Ghose, D., Mitra, R., & Das, S. (2021). Sleep/wake classification from accelerometer data using hidden Markov model. Sensors, 21(1), 1–15. https://doi.org/10.3390/s21010001
12. Giggins, O. M., Persson, U. M., & Caulfield, B. (2013). Biofeedback in rehabilitation. Sensors, 13(11), 14677–14688. https://doi.org/10.3390/s131114677
13. Hammerla, N. Y., Halloran, S., & Plötz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. IJCAI Proceedings, 1533–1540. https://arxiv.org/abs/1604.08880
14. Long, X., Fonseca, P., Foussier, J., Haakma, R., & Aarts, R. M. (2014). Modeling sleep and wakefulness in a home setting using wrist actigraphy and heart rate. IEEE EMBC, 4974–4977. https://doi.org/10.1109/EMBC.2014.6944670
15. Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book/
16. Montoye, A. H. K., Moore, R. W., Bowles, H. R., Korycinski, R., & Pfeiffer, K. A. (2016). Reporting accelerometer methods in physical activity intervention studies. Journal of Physical Activity and Health, 13(6), 726–733. https://doi.org/10.1123/jpah.2015-0286
17. Owen, N., Healy, G. N., Matthews, C. E., & Dunstan, D. W. (2010). Too much sitting: The population health science of sedentary behavior. Mayo Clinic Proceedings, 85(12), 1138–1141. https://doi.org/10.4065/mcp.2010.0444
18. Patel, S., Park, H., Bonato, P., Chan, L., & Rodgers, M. (2012). A review of wearable

sensors and systems with application in rehabilitation. Journal of NeuroEngineering and Rehabilitation, 9, 21. https://doi.org/10.1186/1743-0003-9-21

19. Piwek, L., Ellis, D. A., Andrews, S., & Joinson, A. (2016). The rise of consumer health wearables: Promises and barriers. PLOS Medicine, 13(2), e1001953. https://doi.org/10.1371/journal.pmed.1001953
20. Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257–286. https://doi.org/10.1109/5.18626
21. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD, 1135–1144. https://doi.org/10.1145/2939672.2939778
22. Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: An update. Sleep Medicine Reviews, 15(4), 259–267. https://doi.org/10.1016/j.smrv.2010.10.001
23. Staudenmayer, J., Pober, D., Crouter, S., Bassett, D., & Freedson, P. (2009). An artificial neural network to estimate physical activity energy expenditure. Journal of Applied Physiology, 107(4), 1300–1307. https://doi.org/10.1152/japplphysiol.00465.2009
24. Tremblay, M. S., Carson, V., Chaput, J. P., et al. (2017). Sedentary Behavior Research Network (SBRN)–Terminology consensus project. International Journal of Behavioral Nutrition and Physical Activity, 14(1), 75. https://doi.org/10.1186/s12966-017-0525-8
25. Walch, O. J., Huang, Y., Forger, D. B., & Goldstein, C. A. (2019). Sleep stage prediction with raw acceleration and photoplethysmography using deep learning. npj Digital Medicine, 2, 101. https://doi.org/10.1038/s41746-019-0174-8
26. Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. Medicine and Science in Sports and Exercise, 30(5), 777–781. https://doi.org/10.1097/00005768-199805000-00021
27. Rosenberger, M. E., Buman, M. P., Haskell, W. L., McConnell, M. V., & Carstensen, L. L. (2016). Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Medicine and Science in Sports and Exercise, 48(3), 457–465. https://doi.org/10.1249/MSS.0000000000000778
28. Zhang, Z., Pi, Z., & Liu, B. (2015). TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. IEEE Transactions on Biomedical Engineering, 62(2), 522–531. https://doi.org/10.1109/TBME.2014.2359372
29. Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Nyström, C. D., Mora-Gonzalez, J., Löf, M., & Ortega, F. B. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review. Sports Medicine, 47(9), 1821–1845. https://doi.org/10.1007/s40279-017-0716-0
30. Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. Medicine and Science in Sports and Exercise, 40(1), 181–188. https://doi.org/10.1249/mss.0b013e31815a51b3