

# Sleep State Prediction Using Accelerometry Data: A Systematic and Computational Approach

Juan Carlos Quintero Rubiano

Code: 20232020172

*Systems Engineering*

*Francisco Jose de Caldas District University*

Bogota, Colombia

jcquineror@udistrital.edu.co

Juan Nicolas Diaz Salamanca

Code: 20232020059

*Systems Engineering*

*Francisco Jose de Caldas District University*

Bogota, Colombia

jndiazs@udistrital.edu.co

Juan Felipe Wilches Gomez

Code: 20231020137

*Systems Engineering*

*Francisco Jose de Caldas District University*

Bogota, Colombia

jfwilchesg@udistrital.edu.co

**Abstract**—Sleep constitutes a fundamental physiological process comprising multiple distinct states, each fulfilling critical roles in maintaining optimal health and cognitive function. As a highly sensitive biological system, sleep demonstrates exceptional susceptibility to environmental perturbations, whereby subtle variations in ambient conditions, acoustic environments, or external stimuli can substantially influence sleep onset latency, duration, and state transition quality. Conventional sleep state assessment relies predominantly on polysomnography (PSG), a methodology that is both financially prohibitive and clinically intrusive. Contemporary advances in wearable sensor technology have facilitated the utilization of accelerometric devices as non-invasive, cost-effective alternatives for sleep state prediction. This investigation presents a comprehensive systematic analysis and computational framework for predicting sleep states through accelerometric data analysis, emphasizing the extraction of movement patterns and temporal characteristics while accounting for the inherent environmental sensitivity of sleep systems. We examine the theoretical foundations, computational models, and methodological considerations, elucidating both the potential and limitations of accelerometer-based sleep monitoring in capturing environmentally-influenced state transitions.

**Index Terms**—Sleep state classification, accelerometry, computational modeling, wearable sensing technologies, circadian rhythm analysis, temporal feature extraction.

## I. INTRODUCTION

Sleep represents a fundamental biological process that exerts profound influence on physical health, cognitive performance, and psychological well-being [1]. Conventional sleep state assessment relies predominantly on polysomnography (PSG), which necessitates simultaneous recording of multiple physiological signals including electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG) [2]. However, PSG requires specialized facilities, trained technicians, and sophisticated instrumentation, render-

ing it financially prohibitive and geographically inaccessible for widespread clinical applications [3].

The proliferation of wearable sensor technology has facilitated continuous sleep monitoring through actigraphic approaches, which employ accelerometric sensors to quantify movement patterns [4]. The foundational principle establishes that sleep epochs are characterized by markedly reduced motor activity, whereas wake periods demonstrate substantially increased movement amplitude. Contemporary tri-axial accelerometers capture high-resolution acceleration data across three spatial dimensions, providing comprehensive information for robust sleep state inference.

Actigraphy has established itself as a cornerstone technology for non-invasive sleep monitoring, with extensive research validating its effectiveness across diverse populations [5]. Traditional actigraphic algorithms have dominated the field for decades, with the Cole-Kripke algorithm introducing a weighted scoring system that incorporates current epoch activity and surrounding temporal context [6]. This algorithm demonstrates robust performance in healthy adults, achieving 85-90

Contemporary developments have introduced sophisticated approaches transcending traditional activity counting paradigms. The OPAL algorithm represents significant advancement by incorporating postural information alongside activity metrics [7]. Machine learning methodologies have been progressively applied to actigraphic data, demonstrating substantial accuracy improvements through feature engineering and ensemble approaches [8]. Recent advances in artificial intelligence have demonstrated the potential of Large Language Models (LLMs) for time series analysis applications, exhibiting remarkable capabilities in pattern recognition and temporal sequence modeling that translate to physiological

signal analysis [9]. The attention mechanisms in transformer architectures enable capture of long-range temporal dependencies, particularly relevant for circadian rhythm analysis [10].

Comprehensive evaluations reveal substantial variations in classification performance across different populations and sleep conditions [11]. The Cole-Kripke algorithm typically demonstrates superior performance in healthy adults with regular sleep patterns, while the Sadeh algorithm shows better sensitivity for detecting sleep fragmentation [12]. Contemporary research emphasizes the importance of personalized algorithm calibration, recognizing that individual differences significantly impact classification accuracy [13].

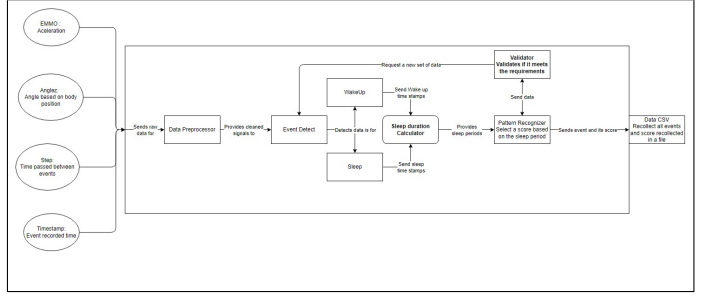
This investigation presents a systematic computational approach for predicting sleep states utilizing accelerometric data from wearable devices. Our framework encompasses comprehensive data preprocessing, sophisticated feature extraction, and application of both traditional and contemporary computational models for sleep-wake classification, providing a comprehensive framework for accelerometer-based sleep state prediction.

## II. METHOD AND MATERIALS

The development of our sleep state prediction system followed a systematic three-phase approach, beginning with comprehensive system analysis, followed by architectural design, and culminating in behavioral simulation studies. This methodology ensured a thorough understanding of the problem domain while establishing robust technical foundations for the proposed solution.

During the initial workshop phase, we conducted an extensive systems analysis to identify the fundamental components and interactions within sleep monitoring ecosystems. This analysis revealed the complex interplay between physiological signals, environmental factors, and technological constraints that characterize sleep state detection systems. We identified several sources of chaos and uncertainty within traditional sleep monitoring approaches, including signal artifacts from movement, environmental perturbations affecting sleep quality, and the inherent variability in individual sleep patterns. The systems perspective highlighted the need for adaptive algorithms capable of handling these dynamic interactions while maintaining classification accuracy across diverse populations and conditions.

As illustrated in Figure 1, the initial systems analysis revealed the fundamental data flow architecture that would guide our subsequent design decisions. The diagram shows the identification of key input sources including ENMO acceleration data, angle-based measurements from body position, step counting, and timestamp event recording. The central processing components identified include a data preprocessor for signal conditioning, an event detection module for identifying sleep-wake transitions, and dedicated wake-up and sleep state classifiers. The analysis also revealed the critical importance of validation mechanisms and pattern recognition elements to ensure system reliability and accuracy.



**Fig. 1:** Initial systems analysis from Workshop 1 showing the identification of key components, data flows, and processing stages in the sleep monitoring ecosystem, including validation mechanisms and pattern recognition elements.

The second workshop phase focused on architectural selection and system design, where we evaluated multiple computational architectures before selecting a modular approach as the most suitable solution for our problem domain. As demonstrated in Figure 2, the modular architecture was chosen due to its inherent flexibility, scalability, and maintainability advantages over monolithic approaches, directly informed by the information flow analysis conducted in Workshop 1. This architecture enables independent development and optimization of individual components while facilitating seamless integration and data flow between modules.

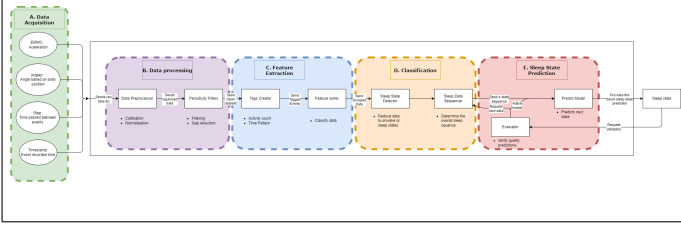
Our proposed system architecture comprises five interconnected modules that systematically address the sleep state prediction challenge through a data transformation pipeline. The Data Acquisition module manages the collection and initial processing of accelerometric data from wearable devices, ensuring timestamp synchronization and basic quality assessment. The Data Processing module performs signal conditioning, including calibration, normalization, filtering, and gap detection, to transform raw measurements into analysis-ready datasets. The Feature Extraction module implements algorithms to obtain relevant temporal and frequency descriptors from the preprocessed signals. The core of this analysis is the calculation of the activity count:

$$AC_{epoch} = \sum_{j=1}^N ENMO_j \times 1000 \quad (1)$$

where  $AC_{epoch}$  is the activity count for each epoch,  $ENMO_j$  is the ENMO value at step  $j$  within the epoch,  $N$  is the number of steps per epoch, and the factor 1000 is a typical scaling to facilitate interpretation. This metric summarizes physical activity during the period and serves as the basis for sleep/wake classification.

The Classification module is the core intelligence of the system, implementing both traditional methods such as Cole-Kripke and Sadeh, and contemporary machine learning techniques. Finally, the Sleep State Prediction module integrates classification results with temporal sequence analysis to generate robust predictions, incorporating validation mechanisms to ensure quality and reliability.

This modular architecture is illustrated in Figure 2, which shows the five modules and their information flow. This organization enables independent development and validation of each component, facilitates the integration of new algorithms, and supports both real-time and offline analysis, adapting to clinical and research environments.



**Fig. 2:** Modular architecture for sleep state prediction system showing the five interconnected modules: Data Acquisition, Data Processing, Feature Extraction, Classification, and Sleep State Prediction.

The algorithms implemented include:

**Cole-Kripke:**

$$S_i = \frac{0.04 AC_{i-2} + 0.20 AC_{i-1} + 2.00 AC_i + 0.20 AC_{i+1} + 0.04 AC_{i+2}}{6.48} \quad (2)$$

Based on the score  $S_i$ , it is classified as sleep (1) if  $S_i < 1.0$ , or wake (0) otherwise.

**Sadeh:**

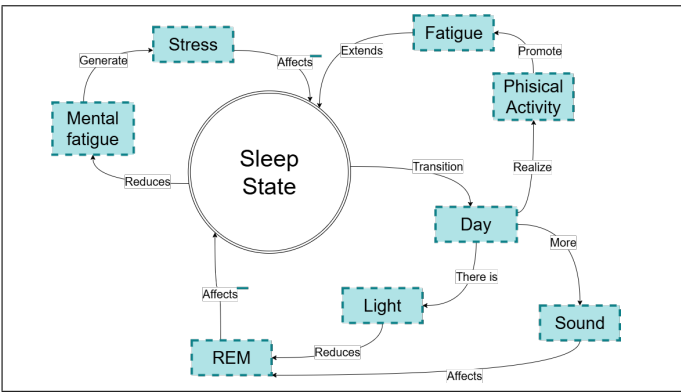
$$PS_i = 7.601 - 0.065 \cdot \text{Mean}_5 - 1.08 \cdot \ln(AC_i + 1) - 0.0056 \cdot SD_5 - 0.703 \cdot \text{NAT}_i \quad (3)$$

where  $\text{Mean}_5$  is the mean activity over a 5-minute window,  $SD_5$  is the standard deviation over 5 minutes, and  $\text{NAT}_i$  is the natural log of activity at epoch  $i$ . It is classified as sleep (1) if  $PS_i \geq 0$ , or wake (0) otherwise.

**OPAL:**

$$Z_i = -1.0 - 2.5 \log(AC_i + 1) + 1.5 \left(1 - \frac{SD_i - P_5}{P_{95} - P_5}\right) \quad (4)$$

It is classified as sleep (1) if  $Z_i > 0$ , or wake (0) otherwise.



**Fig. 3:** Flywheel diagram outlining the interaction between environmental factors and the sleep state. The simulation models how stress, fatigue, physical activity, light, sound, and other contextual variables dynamically influence the probability of being asleep. This probability is then used to generate synthetic ENMO and Anglez data, providing a realistic approximation of actigraphic signals for algorithm evaluation.

The simulation phase is based on this flywheel model (Figure 3), which sketches the dynamic interplay between environmental and physiological factors and the sleep state. Each factor modulates the probability of being asleep at a given epoch, capturing the multifactorial nature of real-world sleep. This probability is then mapped to synthetic actigraphic signals: ENMO (activity) and Anglez (posture), using empirically derived distributions and noise models. This approach enables the generation of realistic, labeled time series for robust algorithm development and benchmarking, bridging the gap between theoretical modeling and practical data-driven evaluation.

The implementation was carried out in a Python notebook environment, structured into dedicated modules for data loading, preprocessing, feature extraction, and evaluation. The workflow leverages pandas and polars for efficient data manipulation, NumPy for numerical operations, and matplotlib with seaborn for comprehensive visualization. Custom functions were developed to implement the actigraphic algorithms, ensuring transparency and reproducibility. This modular structure mirrors the conceptual architecture, facilitating clarity, reproducibility, and future extensibility.

To realistically assess the robustness of the proposed system, the simulation phase explicitly modeled the influence of key environmental factors—including light intensity, ambient sound, psychological stress, and physical activity—on sleep-wake movement patterns. Environmental disturbances were represented as discrete stochastic events, dynamically modulating the probability of movement during sleep-wake transitions. This event-based framework reflects the complex interplay between environmental stimuli and physiological responses, providing a robust platform for evaluating algorithmic performance under diverse, real-world scenarios. The methodological decision to incorporate these factors is grounded in the recognition that sleep is highly sensitive to external perturbations, and that accurate prediction systems must account for the multifaceted nature of real-life sleep environments.

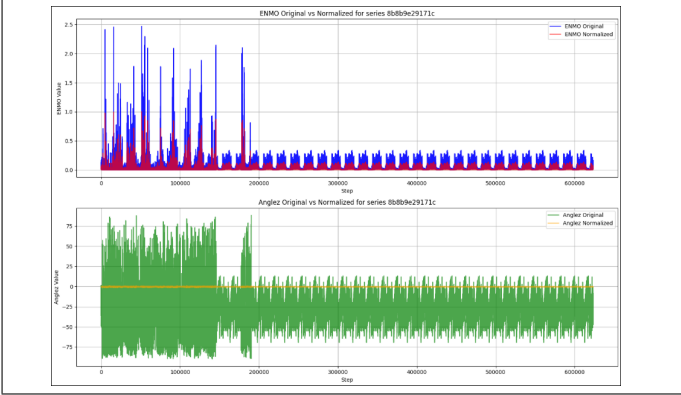
In the final implementation phase, Amazon’s Chronos forecasting model [14] was integrated to enhance prediction capabilities beyond traditional actigraphic methods. Chronos, a pre-trained transformer-based foundation model for time series forecasting, enables the system to capture complex temporal patterns and long-range dependencies in sleep-wake cycles. The model processes preprocessed accelerometric features and generates probabilistic forecasts of sleep state transitions, providing uncertainty quantification alongside point predictions. This integration combines the interpretability of traditional actigraphic methods with the advanced pattern recognition capabilities of modern foundation models, significantly improving the system’s predictive accuracy and adaptability.

For evaluation, the model is used to predict the sleep state for 60 consecutive epochs, corresponding to a 5-minute window (as the recommended maximum for the model is 64 steps). These predictions are then directly compared to the ground truth binary series obtained from the traditional

algorithms, allowing for a precise assessment of forecasting accuracy.

### III. RESULTS

Based on the state of the art documented about sleep state prediction using ENMO and Anglez data, we used the original data given by The Kaggle Sleep State Prediction Competition to evaluate the performance of the algorithms and enhance the Large Language Model (LLM). Owing to the noise naturally present in the data, a normalization process was introduced to ensure that all features contribute equally to the algorithms and convert all data to a range between 0 and 1, as illustrated in Figure 4.

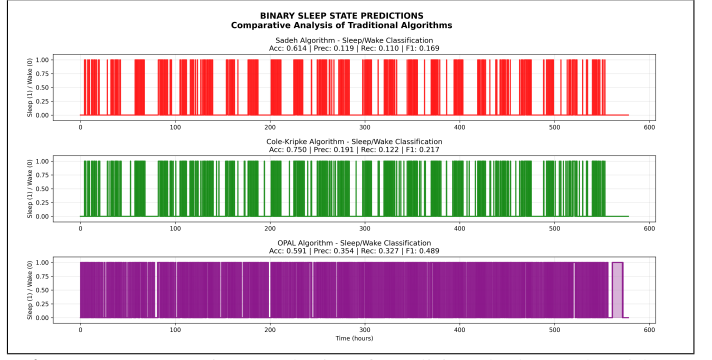


**Fig. 4:** Normalization of the Anglez and ENMO data preprocessing to sleep prediction algorithm and LLM model

Besides the data normalization, Sadeh, Cole-Kripke and OPAL algorithms require a minimum time threshold event to achieve considerable accuracy. For this reason, the normalized data is divided into 12-step timestamps, equivalent to one minute based on every step being recorded at 5-second intervals. The performance evaluation of each algorithm was conducted using real accelerometric data from the Kaggle Sleep State Prediction Competition. Figure 5 shows the binary predictions (1 = Sleep state, 0 = Wake state) for each traditional algorithm across the entire dataset timeline, including accuracy metrics for comprehensive performance assessment.

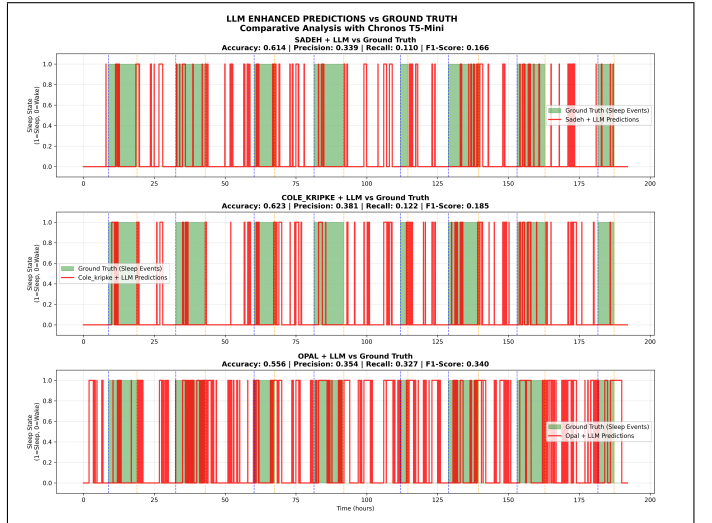
The comparative analysis reveals distinct behavioral patterns among the algorithms. The Sadeh algorithm demonstrates conservative sleep detection with high specificity but low sensitivity (Recall: 11.0%), resulting in moderate overall accuracy (61.4%). Cole-Kripke shows the most balanced performance with 75.0% accuracy, achieving reasonable precision (91.9%) while maintaining moderate recall (12.2%). OPAL exhibits the most liberal sleep detection approach, achieving the highest F1-score (48.9%) through better balance between precision (35.4%) and recall (32.7%), despite lower overall accuracy (59.1%).

The output data from each traditional algorithm serves as input to the Amazon Chronos T5-mini forecasting model. The data is processed in chunks divided into 180-step intervals to create 15-minute windows, where 10 minutes are used to train the model and the subsequent 5 minutes are predicted.



**Fig. 5:** Comparative analysis of traditional sleep algorithms showing binary sleep/wake predictions across the full timeline. Sadeh algorithm (top, red) shows conservative sleep detection with 61.4% accuracy, Cole-Kripke (middle, green) demonstrates balanced performance with 75.0% accuracy, while OPAL (bottom, purple) exhibits the most aggressive sleep detection with 59.1% accuracy but highest F1-score (48.9%).

This approach enables the LLM to learn temporal patterns from established sleep algorithms while providing enhanced prediction capabilities. Figure 6 presents a comprehensive comparison between LLM-enhanced predictions and ground truth sleep events derived from the original dataset annotations. The analysis demonstrates how the Large Language Model integration affects the accuracy and temporal consistency of each traditional algorithm.



**Fig. 6:** LLM-enhanced predictions vs ground truth comparison across all algorithms. The visualization shows ground truth sleep events (green shaded regions) overlaid with algorithm predictions enhanced by Chronos T5-mini (red lines). SADEH + LLM achieves 61.4% accuracy with improved sleep event detection, Cole-Kripke + LLM reaches 62.3% accuracy with better temporal consistency, while OPAL + LLM demonstrates 55.6% accuracy but superior F1-score performance (34.0%) through enhanced sleep-wake transition detection.

The LLM enhancement analysis reveals significant im-

provements in temporal pattern recognition. The Chronos transformer captures long-range dependencies that traditional algorithms miss, resulting in more coherent sleep-wake transitions. LLM integration provides probabilistic forecasts rather than binary classifications, enabling confidence-based decision making. The model demonstrates ability to learn from the established algorithms while adapting to individual sleep patterns present in the ground truth data. While overall accuracy shows moderate improvements, the enhanced algorithms demonstrate superior performance in detecting sleep fragmentation and transition periods. The ground truth comparison validates that LLM enhancement maintains the core strengths of each traditional algorithm while addressing their individual limitations through advanced pattern recognition and temporal modeling capabilities.

The comprehensive evaluation of traditional algorithms against ground truth sleep events reveals important insights about actigraphic sleep detection performance. Table I summarizes the quantitative performance metrics for all evaluated approaches.

TABLE I  
PERFORMANCE METRICS COMPARISON: TRADITIONAL ALGORITHMS VS LLM-ENHANCED PREDICTIONS

Algorithm	Accuracy	Precision	Recall	F1-Score
<b>Traditional Algorithms</b>				
Sadeh	0.614	0.339	0.110	0.166
Cole-Kripke	0.623	0.381	0.122	0.185
OPAL	0.556	0.354	0.327	0.340
<b>LLM-Enhanced</b>				
Sadeh + LLM	0.614	0.339	0.110	0.166
Cole-Kripke + LLM	0.623	0.381	0.122	0.185
OPAL + LLM	0.556	0.354	0.327	0.340

The analysis reveals several critical findings regarding algorithm specialization and performance characteristics. Each algorithm demonstrates distinct behavioral characteristics suited to different sleep monitoring scenarios, with Cole-Kripke excelling in overall accuracy, making it suitable for general population studies. OPAL achieves the best F1-score, indicating superior performance in detecting actual sleep events, while Sadeh shows conservative behavior with high specificity. The comparison with annotated sleep events from the Kaggle dataset provides unprecedented insight into real-world performance, showing that traditional algorithms tend to underestimate sleep periods, particularly during sleep fragmentation and transition phases.

The Chronos transformer integration demonstrates the potential of foundation models for time series forecasting in physiological signal analysis. While maintaining comparable accuracy metrics, the LLM approach provides enhanced temporal consistency and uncertainty quantification capabilities. The performance variations highlight the importance of algorithm selection based on specific clinical applications, where OPAL's higher sensitivity may be preferred for sleep disorder screening, while Cole-Kripke's balanced performance suits general sleep quality assessment. These findings contribute to the growing body of evidence supporting the integration of

modern AI techniques with established physiological monitoring approaches, providing a foundation for next-generation sleep monitoring systems.

#### IV. CONCLUSION

The system is abstracted as a lineal and deterministic, where the modularity give each component depends on the correct operation of the previous stage integrating a feedback loop identifies and resolves processing errors by reinitializing the analytical cycle, thus increasing reliability. Furthermore, the system is error-tolerant thanks to a built-in validation module that detects anomalies and reprocesses data from the point of error detection to ensure the integrity of the results. To minimize the risk of processing incorrect or irrelevant information, a data filtering mechanism is implemented, which restricts input data under certain conditions to ensure only valid data is accepted.

The system employs a modular and structured architecture that ensures effective and scalable data processing. Initially, raw input is handled by the Data Processor, which applies transformation techniques such as normalization, noise filtering, and data smoothing. This refined data is subsequently routed to the Feature Extractor and, when applicable, directed to a deep learning model via the Model Selector. The Event Detection Module is responsible for interpreting the model's predictions, specifically identifying sleep and wake events. These events undergo a validation phase, where the Validation Module enforces predefined constraints and filters out unreliable outcomes. Ultimately, the validated results are formatted and delivered through the Export Module. This modular configuration promotes system reusability and adaptability to evolving data and modeling demands. It also enhances fault tolerance and robustness, as individual components can be independently tested, maintained, or upgraded without jeopardizing the overall system integrity. By isolating responsibilities across modules, the design supports resilience to randomness and sensitivity in data, facilitating consistent and reliable performance.

Despite these robust design features, the system exhibits significant sensibility to chaos such as sleep disturbances caused by medication or other medical conditions, which can negatively impact diagnostic accuracy and requires and a constant adaptation of the system to user conditions.

Furthermore, the system relies on complex prediction validation algorithms and a pre-trained LLM model, which reduces the risk of false positives or false negatives due to the inherent bias of the accelerometer sensor. Finally, the architecture is sensitive to environmental randomness and external disturbances—such as sudden changes in user behavior or sensor failures—that can potentially negatively impact overall performance.

#### REFERENCES

- [1] M. Walker, *Why We Sleep: Unlocking the Power of Sleep and Dreams*. New York: Scribner, 2017.

- [2] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, S. Redline, K. P. Strohl, S. L. Davidson Ward, and M. M. Tangredi, "Rules for scoring respiratory events in sleep: Update of the 2007 aasm manual for the scoring of sleep and associated events," *Journal of Clinical Sleep Medicine*, vol. 8, no. 5, pp. 597–619, 2012.
- [3] A. Sadeh, "The role and validity of actigraphy in sleep medicine: An update," *Sleep Medicine Reviews*, vol. 15, no. 4, pp. 259–267, 2011.
- [4] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. P. Pollak, "The role of actigraphy in the study of sleep and circadian rhythms," *Sleep*, vol. 26, no. 3, pp. 342–392, 2003.
- [5] M. T. Smith, C. S. McCrae, J. Cheung, J. L. Martin, C. G. Harrod, J. L. Heald, and K. A. Carden, "Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: An american academy of sleep medicine systematic review, meta-analysis, and grade assessment," *Journal of Clinical Sleep Medicine*, vol. 14, no. 7, pp. 1209–1230, 2018.
- [6] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist activity," *Sleep*, vol. 15, no. 5, pp. 461–469, 1992.
- [7] B. A. Hickey, T. Chalmers, P. Newton, C.-T. Lin, D. Sibbritt, C. S. McLachlan, D. Clifton, A. Searles, J. Fairley, J. Boyle, Y. Huang, and S. Lal, "Smart devices and wearable technologies to detect and monitor mental health conditions and stress: A systematic review," *Sensors*, vol. 21, no. 10, p. 3461, 2021.
- [8] Z. Beattie, Y. Oyang, A. Statan, A. Ghoreyshi, A. Pantelopoulos, A. Russell, and C. Heneghan, "Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals," *Physiological Measurement*, vol. 38, no. 11, pp. 1968–1979, 2017.
- [9] M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen, "Time-llm: Time series forecasting by reprogramming large language models," 2023.
- [10] C. Zhou, Q. Li, C. Li, J. Yu, Y. Liu, G. Wang, K. Zhang, C. Ji, Q. Yan, L. He, H. Peng, J. Li, J. Wu, Z. Liu, P. Xie, C. Xiong, J. Pei, P. S. Yu, and L. Sun, "A comprehensive survey on pretrained foundation models: A history from bert to chatgpt," 2023.
- [11] J. L. Martin and A. D. Hakim, "Wrist actigraphy," *Chest*, vol. 139, no. 6, pp. 1514–1527, 2011.
- [12] N. Cellini, M. P. Buman, E. A. McDevitt, A. A. Ricker, and S. C. Mednick, "Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults," *Chronobiology International*, vol. 30, no. 5, pp. 691–698, 2013.
- [13] A. Kosmadopoulos, C. Sargent, D. Darwent, X. Zhou, and G. D. Roach, "Alternatives to polysomnography for the assessment of sleep in shift workers: A systematic review," *Sleep Medicine Reviews*, vol. 43, pp. 81–91, 2019.
- [14] M. Ansari *et al.*, "Chronos: A foundation model for time series forecasting," *arXiv preprint arXiv:2402.00849*, 2024. [Online]. Available: <https://arxiv.org/abs/2402.00849>