Machine Learning-based Sleep Stage Classification Using Wearable Technology

Brandon Ismalej and Dr. Xunfei Jiang
Department of Computer Science
California State University, Northridge

Abstract

Quality sleep is essential for cognitive function, physical health, and overall well-being, yet accurately tracking sleep outside of clinical settings remains a challenge. Traditional sleep studies, such as polysomnography, require expensive and inconvenient monitoring. This research explores the use of data science and deep learning to classify sleep stages—light, deep, and REM—using physiological data from smart wearable devices. By analyzing heart rate variability, movement patterns, and oxygen levels, we aim to provide a non-invasive, accessible alternative for sleep monitoring. Our findings indicate that wearable-based models can offer valuable insights into sleep quality, potentially empowering individuals to improve their sleep health and long-term well-being.

Introduction

Context & Motivation:

- **Essential Role of Sleep:** Supports cognitive function, physical health, and overall well-being
- **Clinical Limitations:** Polysomnography (PSG) is accurate but expensive and inconvenient (labor-intensive, clinical setting)
- **Wearable Potential:** Smartwatches and non-invasive sensors offer continuous, real-world data
- **Data Gaps:** Lack of publicly accessible, large-scale datasets for validating sleep staging algorithms

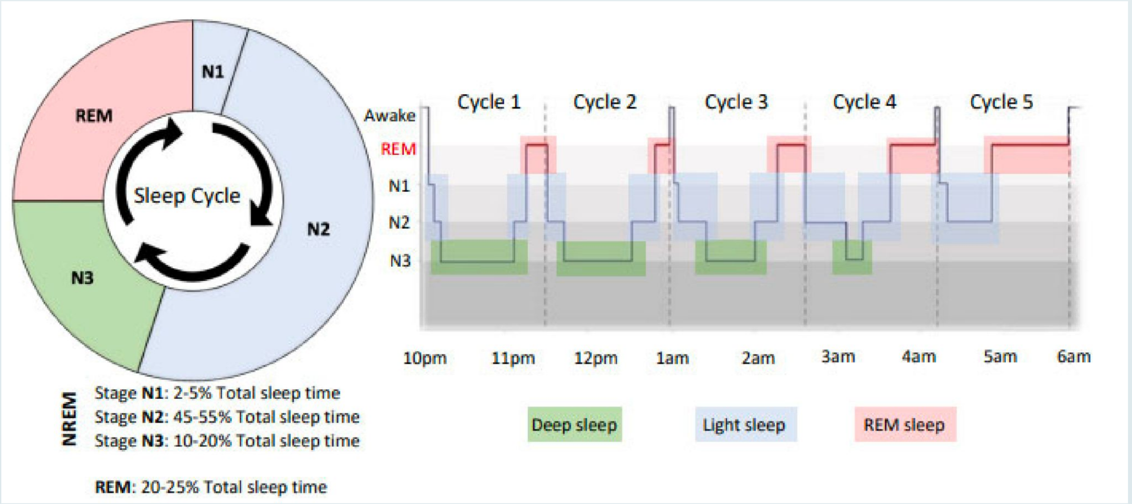

Figure 1: Hypnogram Depicting Sleep Stage Distribution Across Cycles [2]

Goals & Objectives:

- **Develop Explainable Machine Learning Models**: Classify sleep stages from wearable signals with known algorithm and transparent data use
- **Improvable Accessibility:** Provide a cost-effective, user-friendly approach to sleep monitoring and personal health insights, outside clinical labs

Understanding Sleep Cycles

- Typical sleep cycles last ~90 minutes, repeating 4–6 times per night, and include both Non-REM and REM stages.
- Deep sleep (N3) is most common earlier in the night, while REM sleep becomes more frequent toward the morning—both are essential for physical recovery and memory consolidation.

Figure 2: Empatica E4 Wristband with Core Specifications [1]

Dataset Overview:

- **DREAMT:** Dataset for Real-time sleep stage EstimAtion using Multisensor wearable Technology [3]
  - **100 individuals** recruited from Duke University Health System Sleep Disorder Lab
  - **Wearable (Empatica E4) signals:**
    - Raw: Blood Volume Pulse (BVP), Accelerometry (ACC_X,Y,Z), Electrodermal Activity (EDA), Skin Temperature (TEMP)
    - Derived: Heart Rate (HR), Inter-Beat Interval (IBI)
    - Sleep Stage Labels: Technician-annotated (PSG) every 30 seconds (W, N1, N2, N3, R)

Methodology

Exploratory Data Analysis (EDA) Insights:

- Sufficient Data Per Participant: Most participants have ~6.5-7 hours of data with minor variation
- Natural Class Imbalance: N2 (Light Sleep) is the most frequent stage, while N3 (Deep Sleep) is underrepresented, potentially impacting model performance.

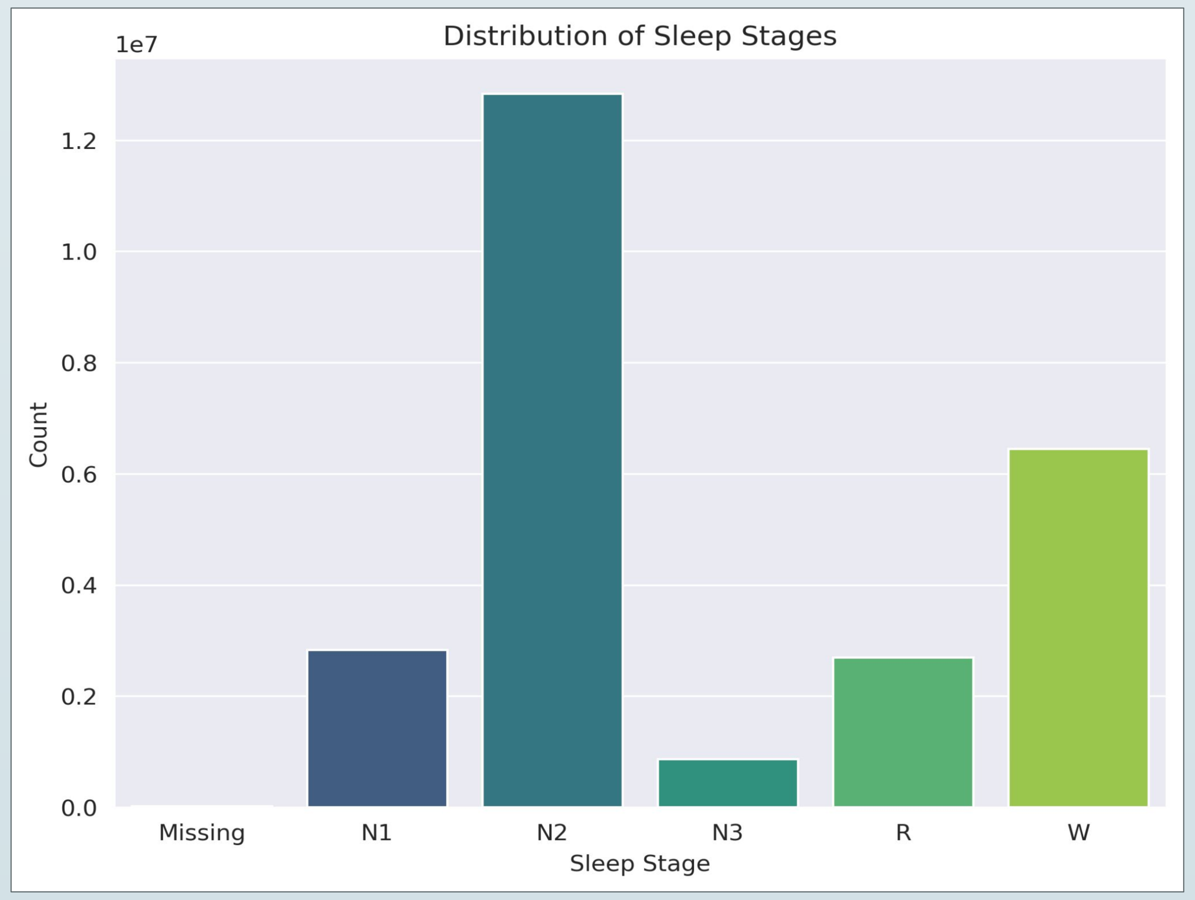

Figure 3: Distribution of Sleep Stage Frequency

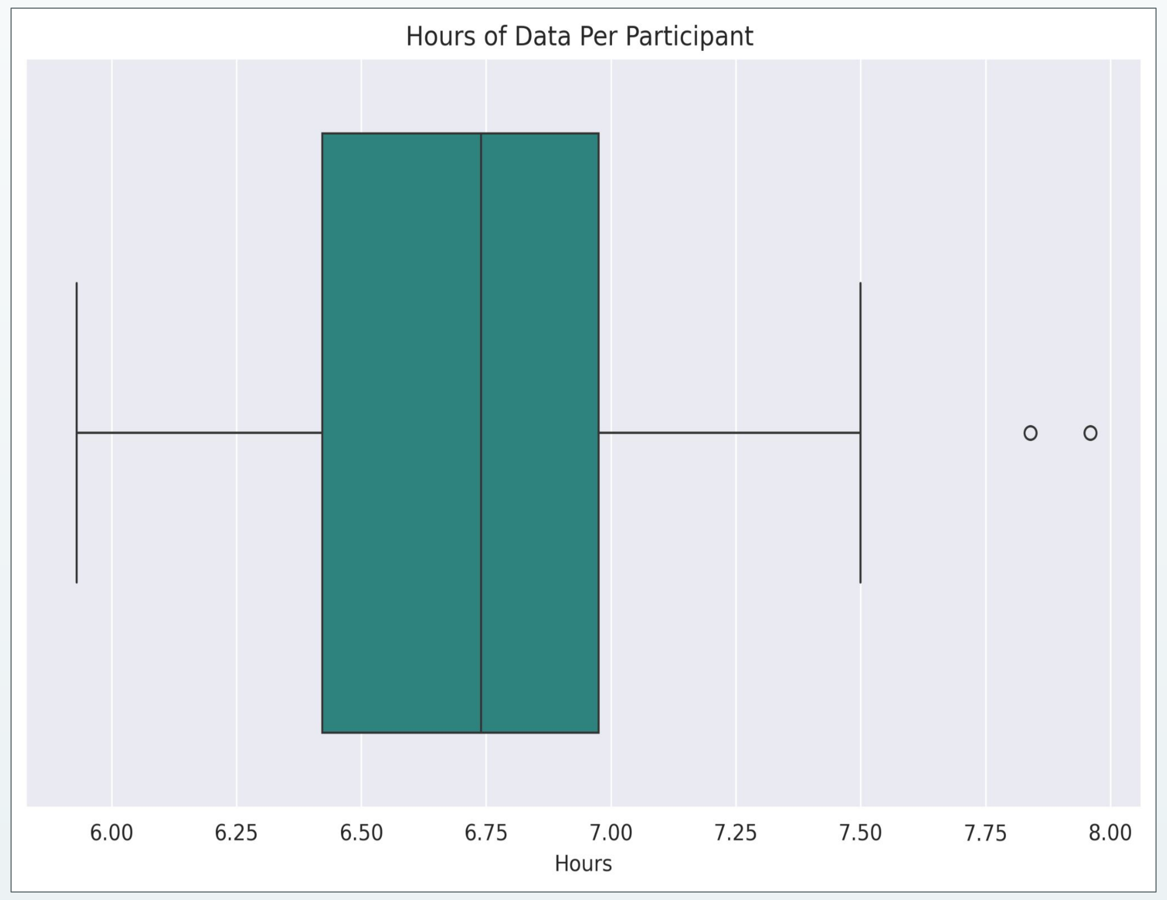

Figure 4: Hours of Data Per Participant

Data Preprocessing:

- **Filter** out preparation (P) stages
- **Raw signals** collected at 64 HZ:
  - Downsample to 10 Hz for full model- Data recorded 10 times/sec.
  - Downsample to 1 Hz for reduced model - Data recorded 1 time/sec.
- **Filtering & Noise Reduction:**
  - Applied 4th-order Butterworth low-pass filter to BVP, ACC, EDA, and TEMP to remove high-frequency noise
  - Used moving average to stabilize sensitive motion data
- **Sleep Stage Transition Preservation:**
  - Used majority voting with smoothing over downsampled windows to maintain sleep stage transitions

Neural Network Development & Training:

- **Long-Short Term Memory (LSTM)**
  - Reliable and frequently used in current literature
  - Designed to work with time-series data, where the past affects the present and future data
- **Training Setup:**
  - Train-Test Split of Data: 70% train, 15% validation, 15% test
  - Loss Function: Focal Loss
    - To address class imbalance by focusing on harder-to-classify sleep stages
  - Class Imbalance Handling:
    - Applied class weighting to reduce bias towards majority classes (e.g. N2, and Wake)
  - **Full-Feature Model:**
    - Utilizes all training features available from dataset
      - BVP, ACC_X,Y,Z, EDA, TEMP, HR, IBI
  - **Reduced-Feature Model:**
    - Utilizes training features most common to consumer-grade wearables
      - HR, IBI, ACC_X,Y,Z

Results

LSTM Performance:

- **Full-Feature Model:**
  - Overall Accuracy: 99%
  - Trained over 40 epochs: Best Epoch = 38
  - Sequences of 10 seconds at 10 Hz
- **Reduced-Feature Model:**
  - Overall Accuracy: 92%
  - Trained over 60 epocs: Best Epoch = 58
  - Sequences of 30 seconds at 1 Hz

	Precision	Recall	F1-Score
Wake	0.99	0.99	0.99
N1 (Non-REM 1, Light Sleep)	0.98	0.99	0.98
N2 (Non-REM 2, Light Sleep)	1.00	1.00	1.00
N3 (Non-REM 3, Deep Sleep)	0.90	0.96	0.93
REM	1.00	1.00	1.00

Table 1: Classification Report of LSTM, Full-Feature Model

	Precision	Recall	F1-Score
Wake	0.99	0.93	0.96
N1 (Non-REM 1, Light Sleep)	0.88	0.89	0.88
N2 (Non-REM 2, Light Sleep)	0.93	0.93	0.93
N3 (Non-REM 3, Deep Sleep)	0.59	0.84	0.70
REM	0.95	0.91	0.93

Table 2: Classification Report of LSTM, Reduced Model

Conclusion & Future Work

- This study demonstrates the potential of wearable-based sleep monitoring using deep learning models trained on physiological signals
- Our methodology effectively classifies sleep stages (R, N1, N2, N3, Wake) using smartwatch sensor data, achieving promising classification performance
- These findings contribute to the development of more accessible, non-invasive sleep tracking solutions, reducing reliance on expensive clinical sleep studies

Future Work:

- Explore domain adaptation techniques to improve robustness across different wearable devices
- Add noise and time-shifting into data to evaluate robustness of LSTM performance under real-world conditions

References

[1] M. W. Driller *et al*, "Pyjamas, Polysomnography and Professional Athletes: The Role of Sleep Tracking Technology in Sport," *Sports*, vol. 11, no. 1, Art. no. 1, Jan. 2023, doi: 10.3390/sports11010014.
[2] "E4 wristband technical specifications," *Empatica Support*, 2025. https://support.empatica.com/hc/en-us/articles/202581999-E4-wristband-technical-specifications (accessed Mar. 11, 2025).
[3] K. Wang, J. Yang, A. Shetty, and J. Dunn, "DREAMT: Dataset for Real-time sleep stage EstimAtion using Multisensor wearable Technology," Physionet.org, Feb. 05, 2025. https://physionet.org/content/dreamt/2.0.0/ (accessed Mar. 26, 2025).