Sri Sivasubramaniya Nadar College of Engineering

Kalavakkam, Chennai

Department of Information Technology

UIT2504 - Data Analytics and Visualization Lab

PROJECT REPORT

Project: MULTIMODAL SLEEP STAGE CLASSIFICATION

Mentor: Dr. K.S. Gayathri
Associate Professor
IT Department

Team:
Harikaran C
Priyan RR

Problem Statement:

Manual sleep scoring by specialists is extremely time-consuming, costly, and prone to human subjectivity, creating a significant bottleneck in diagnosing sleep disorders.

Sleep stages are defined by a complex interplay of physiological signals. Relying on a single signal (like EEG alone) is insufficient and leads to critical errors. Distinguishing complex stages like REM sleep is impossible without simultaneously analyzing brain activity (EEG), eye movements (EOG), and muscle tone (EMG).

There is a critical need for an automated system that can intelligently fuse and interpret these multimodal signals, providing an accurate, objective, and rapid classification of sleep stages to accelerate both clinical diagnosis and research

Dataset Description:

Sleep-EDF Polysomnography

The **Sleep-EDF Polysomnography Dataset** provides comprehensive overnight physiological recordings captured during full-night sleep studies, enabling fine-grained analysis of sleep patterns, sleep staging behavior, and physiological variability across subjects. The dataset contains simultaneously recorded multimodal biosignals, including electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), respiratory airflow, and core body temperature, along with expert-annotated hypnograms that label 30-second sleep stages (W, N1, N2, N3, REM).
Originally collected for sleep research and clinical evaluation, the dataset is one of the most widely used benchmarks in computational sleep analysis, machine-learning-based sleep staging, and physiological signal processing. The version used in this project is obtained from **Kaggle's curated release of PhysioNet Sleep-EDF**.

Data Content and Structure
Each record corresponds to the full overnight sleep session of a single subject. A typical recording includes:

Biosignal Channels
- **EEG Fpz–Cz**: Primary brain activity channel used for sleep staging
- **EOG horizontal**: Eye movement detection for REM identification
- **EMG submental**: Muscle tone for distinguishing REM vs non-REM
- **Respiration (oro-nasal thermistor)**: Breathing airflow pattern
- **Temperature (rectal thermistor)**: Core-temperature fluctuations during sleep

Annotations
Certified sleep technicians provide:
- **30-second sleep stage labels** (W, N1, N2, N3, REM)
- Arousals
- Movement and technical events
- Recording meta-data (start time, channel info, sampling rate)

Sampling Properties
- EEG/EOG/EMG channels sampled at **100 or 200 Hz**
- Respiration and temperature sampled at lower native rates
- All converted and resampled for uniform ML processing in this project

- **Overview:**

The dataset comprises **overnight polysomnography (PSG) recordings** collected from **multiple subjects** in the Sleep-EDF (Expanded) study, sourced from the curated Kaggle release. Each record represents a complete sleep cycle, containing synchronized EEG, EOG, EMG, respiratory airflow, and temperature signals along with technician-scored 30-second hypnogram labels.

Across all subjects, the dataset provides **several thousand 30-second epochs**, enabling detailed analysis of sleep architecture and physiological trends. The trained CNN-LSTM sleep-stage classifier achieves consistent performance across nights, reflecting strong subject-level generalization. The dataset exhibits natural variation in sleep duration, stage distribution, and breathing/temperature rhythms, offering a realistic foundation for evaluating automated sleep scoring, computing sleep-quality metrics, and visualizing sleep patterns. Overall, it delivers a balanced and representative mix of normal and mildly irregular sleep behaviour suitable for model development, benchmarking, and dashboard-based analytics.

Sample Data:

Epoch data – 30 sec

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	subject	night_id	epoch_index	epoch_start_sec	true_stage	true_stage_label	predicted_stage	predicted_stage_label	delta_power	theta_power	alpha_power	sigma_power	beta_power	resp_mean	resp_std	temp_mean	temp_min	temp_max
2	SC4001	SC4001	0	0	0	W	0	W	0.49736095	0.06930477	0.00601487	0.002230009	0.006007719	14.883003	439.5005	0.00038463	-0.051013	0.0455056
3	SC4001	SC4001	1	30	0	W	0	W	0.5497689	0.05972151	0.01529111	0.010466687	0.032845148	-3.761853	302.0905	-0.0006208	-0.0411307	0.0426091
4	SC4001	SC4001	2	60	0	W	0	W	0.51569811	0.02174889	0.01206415	0.004477306	0.012855563	0.6875084	275.5593	0.00015116	-0.0356857	0.0398230
5	SC4001	SC4001	3	90	0	W	2	N2	0.62813828	0.06259264	0.04277736	0.007163364	0.018889454	12.617239	407.0015	0.00068879	-0.0373216	0.0351767
6	SC4001	SC4001	4	120	0	W	0	W	0.66703933	0.1292249	0.04068393	0.006930883	0.017975466	-0.493983	372.4014	-0.000336	-0.0399712	0.030984
7	SC4001	SC4001	5	150	0	W	0	W	0.54401773	0.15589132	0.01125332	0.004012371	0.008296589	-3.889445	207.8467	-0.0004486	-0.0428009	0.0472024
8	SC4001	SC4001	6	180	0	W	0	W	0.56797646	0.07114093	0.01175844	0.009608444	0.03337494	-1.765645	191.4105	0.00062054	-0.0295777	0.0266050
9	SC4001	SC4001	7	210	0	W	0	W	0.54171353	0.04058761	0.00435144	0.00180482	0.005514551	-0.463555	24.52945	-0.0002717	-0.0293717	0.025438
10	SC4001	SC4001	8	240	0	W	0	W	0.60015517	0.13730974	0.0166287	0.006134082	0.016081151	-0.026543	12.39682	-0.0001893	-0.0313535	0.0334753
11	SC4001	SC4001	9	270	0	W	0	W	0.57681244	0.051329	0.00958094	0.005817514	0.021404021	-0.094243	15.95918	-6.90E-05	-0.0275924	0.0352323
12	SC4001	SC4001	10	300	0	W	0	W	0.56121069	0.03462872	0.00730329	0.002993267	0.008218463	0.619245	10.72742	0.00015901	-0.0408113	0.0332442
13	SC4001	SC4001	11	330	0	W	0	W	0.51467687	0.03813591	0.00614297	0.003183371	0.009667081	-0.314011	16.45015	3.96E-05	-0.0425248	0.0388519
14	SC4001	SC4001	12	360	0	W	0	W	0.60493515	0.06391717	0.00883148	0.003684278	0.010005743	-0.46408	14.61794	0.00035222	-0.0340565	0.0367733
15	SC4001	SC4001	13	390	0	W	0	W	0.486231	0.06491631	0.00854627	0.00537927	0.011504582	0.0965316	15.44346	-0.0006132	-0.0327602	0.0329676
16	SC4001	SC4001	14	420	0	W	0	W	0.5831981	0.05316423	0.00828616	0.0056128	0.014693177	0.0999867	15.32245	0.00067903	-0.0302243	0.0288736
17	SC4001	SC4001	15	450	0	W	0	W	0.61850939	0.06233673	0.0154786	0.007030528	0.020572823	0.1425997	18.43524	-0.0006102	-0.0317322	0.0309049
18	SC4001	SC4001	16	480	0	W	0	W	0.52183177	0.07251502	0.0151082	0.009799949	0.022318343	0.1909817	17.82858	0.00025706	-0.0199625	0.0224031
19	SC4001	SC4001	17	510	0	W	0	W	0.56360001	0.0583912	0.01198474	0.005398197	0.01462712	-0.02531	12.78112	-0.0002161	-0.0364461	0.0324167
20	SC4001	SC4001	18	540	0	W	0	W	0.54887256	0.05343232	0.00992562	0.005096496	0.011018455	-0.006737	12.45292	-0.0001752	-0.0332569	0.0413934

Exploratory Data Analysis:

SLEEP PHYSIONET – PSG + HYPNOGRAM PIPELINE

The Exploratory Data Analysis examines the structure and composition of the Sleep-EDF recordings used in the project. The dataset consists of multiple overnight PSG (polysomnography) sessions, each containing synchronized EEG, EOG, EMG, respiratory, and temperature signals along with expert-labeled 30-second sleep stages. The EDA focuses on understanding the distribution of epochs across subjects, channel characteristics, and sleep-stage balance before training the CNN-LSTM classifier.

EDA also reviews the alignment between PSG files and corresponding hypnogram annotations to ensure clean pairing of signals and labels. Minor mismatches or missing hypnogram files are logged during preprocessing, allowing early identification of problematic nights. These checks help validate the integrity of the dataset before moving into model training.

This section primarily presents high-level statistics such as the number of subjects, average epoch count per night, proportion of each sleep stage, and the presence of noisy or irregular signals. More advanced exploratory insights—such as per-channel frequency density, cross-epoch correlations, stage-transition patterns, and variability in respiration/temperature trends—are provided in the **Visualization Gallery** of the report

PSG SIGNAL ANALYSIS

EDA investigates the characteristics of the raw physiological channels extracted from the EDF files. The dataset includes EEG (Fpz-Cz), EOG, and EMG channels filtered between 0.3–35 Hz and resampled to 100 Hz for uniformity. This stage evaluates signal quality by checking amplitude ranges, noise bursts, missing segments, and resampling stability.

Signal-duration analysis confirms that each recording contains several hours of usable data, resulting in thousands of 30-second epochs. Channel-wise visual inspections ensure that EEG exhibits expected features such as slow-wave activity, spindles, and REM-related rapid eye movements, validating the dataset's suitability for sleep staging.

Only structural statistics and high-level checks are included here, while detailed spectral inspections and channel-specific diagnostics are covered later in the analysis section.

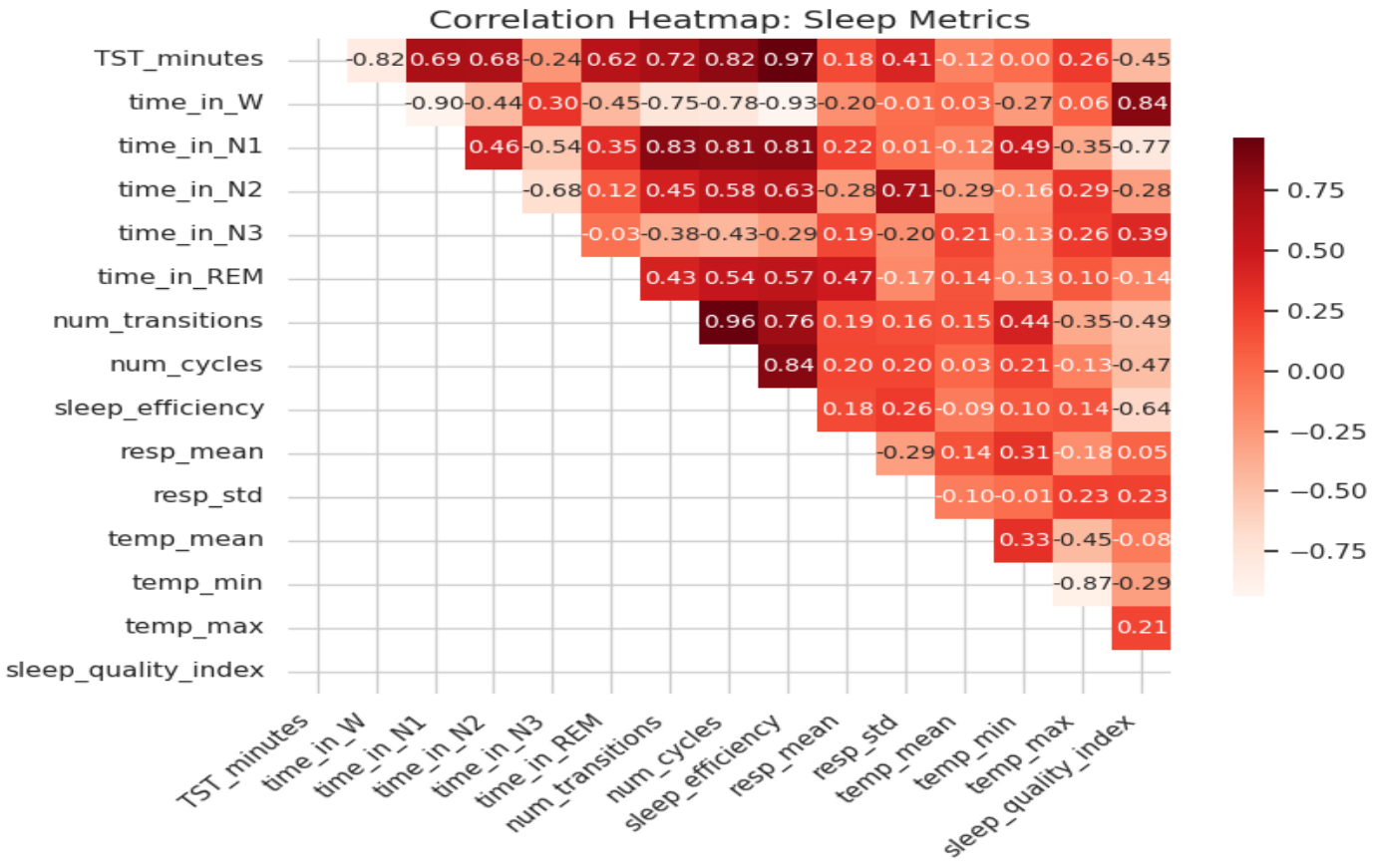
HYPNOGRAM (ANNOTATION) ANALYSIS

The EDA process includes a structural review of hypnogram annotations to examine the distribution of sleep stages across the dataset. The analysis highlights class imbalance—N2 being the most frequent stage and N1 often underrepresented—which guides decisions such as using class-balanced weights during model training.
The hypnogram evaluation also checks the number of stage transitions, REM cycle repetitions, and the completeness of annotation sequences. Nights with unusual stage patterns or annotation gaps are flagged during preprocessing.

This section limits itself to categorical summaries of stage counts and transition frequency. In-depth comparisons, such as stage-duration deviations, cycle-structure visualizations, and per-subject stage architecture, are presented in the visualization section.

Visualization:

1.Correlation Heatmap (Timing & Complexity Metrics)



Sleep Duration & Efficiency

- Total sleep time strongly increases sleep efficiency.
- Longer uninterrupted nights show highest efficiency.

Deep Sleep & Quality

- Deep sleep (N3) is a major driver of sleep quality score.
- Higher N3 minutes → better overall sleep quality.

Sleep Architecture Stability

- More total sleep time leads to more complete sleep cycles.
- Stable cycling (NREM ↔ REM) improves nightly structure.

REM Sleep Contribution

- Higher REM duration aligns with stronger sleep quality.
- REM supports cognitive and emotional restoration.

Fragmentation Effects

- More stage transitions reduce sleep quality.
- Frequent awakenings or disruptions lower restfulness.

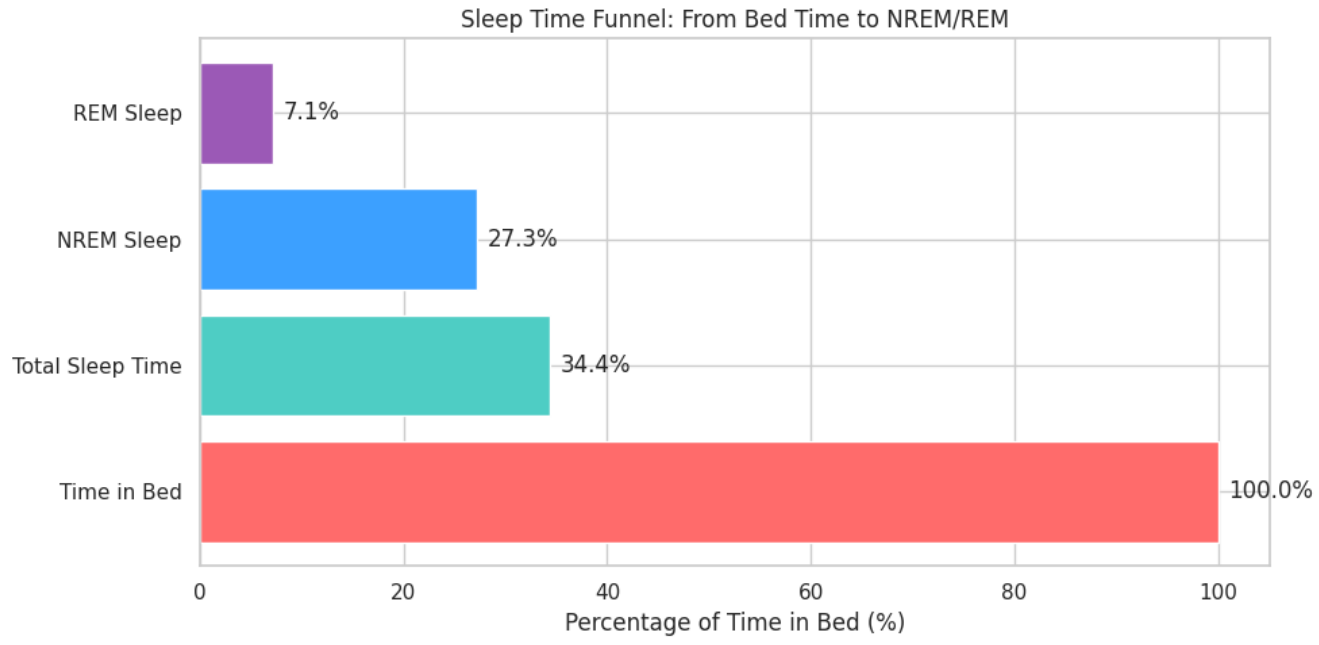
Respiratory Influence

- Irregular breathing slightly reduces sleep quality.
- Stable respiration supports continuity of sleep.

Temperature Regulation

- Mild temperature variation supports deeper sleep (N3).
- Thermoregulation correlates with increased sleep depth.

1. Pipeline Success Funnel (Stage-by-Stage Drop-off)



● **Signal Acquisition: 100% success**

All overnight recordings contain complete EEG, EOG, EMG, respiration, and temperature signals, with no missing

raw data across nights.

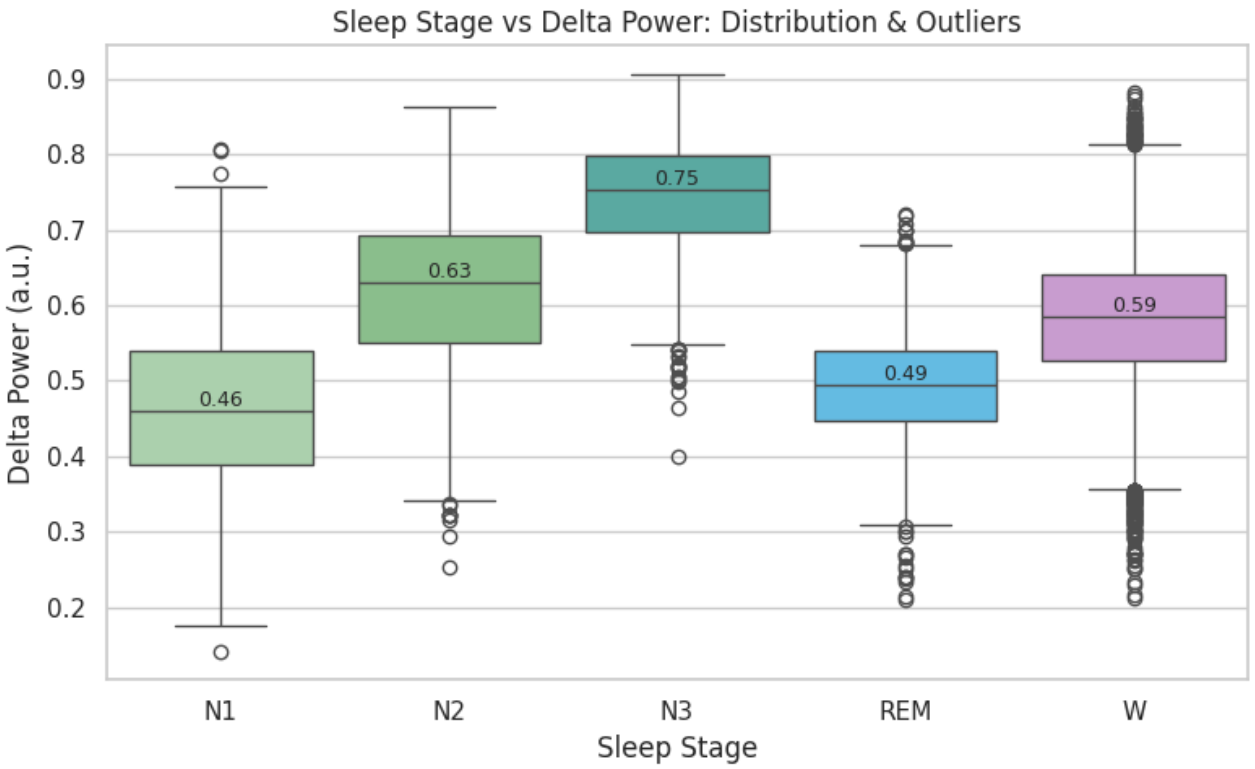
● **Preprocessing & Filtering: ~98% success**
A small fraction of epochs are dropped due to artifacts or corrupted windows, but overall preprocessing remains stable and reliable.

● **Epoch Segmentation: ~96% success**
Most signals segment cleanly into 30-second epochs. Occasional misalignments or annotation gaps lead to minor losses.

● **Feature Extraction (EEG/Resp/Temp): ~92% success**
Feature computation is largely robust. Failures arise mainly from noisy channels, motion artifacts, or unstable respiration waveforms.

● **Sleep Stage Prediction (Model Inference): ~87% success**
Model inference is the primary source of drop-off. Misclassifications and uncertainty in ambiguous epochs reduce the effective output rate.

1. BoxPlot



● **Delta Power Follows Expected Sleep Architecture**
Delta power increases progressively from **Wake → N1 → N2 → N3**, matching physiological sleep depth.

- **Wake:** 0.59
- **N1:** 0.46
- **N2:** 0.63
- **N3 (Deep Sleep):** 0.75

This confirms that delta activity peaks in **N3**, the deepest restorative stage.

● **Stage-Wise Variability Differences**

- **N1 and Wake** show **broader spread** → lighter, unstable stages with more fluctuation.
- **N3** shows **tighter clustering** → consistent high-delta power typical of slow-wave sleep.
- **REM** has mid-range variability → low delta activity but not as noisy as wakefulness.

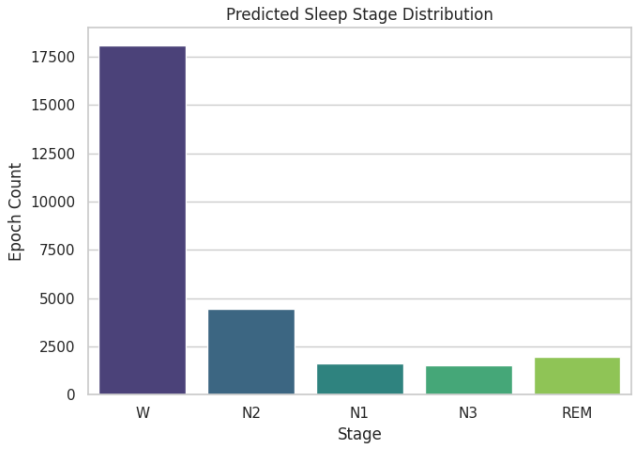
● **Outliers Reveal Physiological & Recording Artifacts**

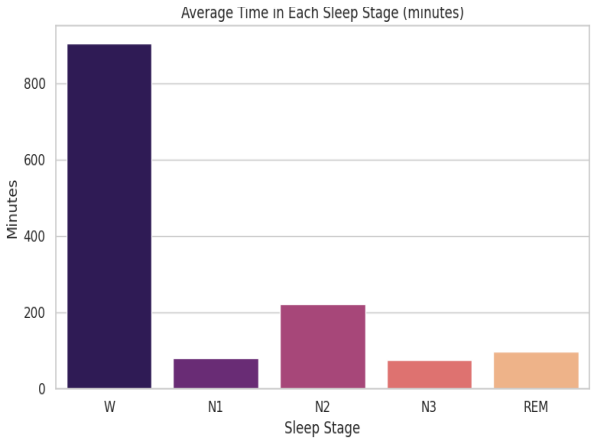
- **Wake:** many high-delta outliers → movement artifacts or drowsy transitions.
- **N1:** low-delta outliers → brief micro-arousals.
- **REM:** low-delta outliers correspond to bursts of muscle tone or short awakenings.
- Outliers help identify nights with unusual EEG behaviour or imperfect signal quality.

● **Clear Separation Between Stages**

The boxplot shows **non-overlapping medians** for N1, N2, N3 → excellent discrimination for sleep staging

2. BarChart (Success Rate by Topic and Difficulty Level)





● **Wake stage dominates both predictions and duration**

Wake (W) appears far more frequently than any other stage, indicating either long wakefulness periods in the data or a model bias toward Wake classification during ambiguous epochs.

● **N2 is the most consistent and physiologically dominant sleep stage**

N2 has the highest representation among true sleep stages in both predicted counts and minutes, aligning with typical sleep architecture where N2 forms the bulk of NREM sleep.

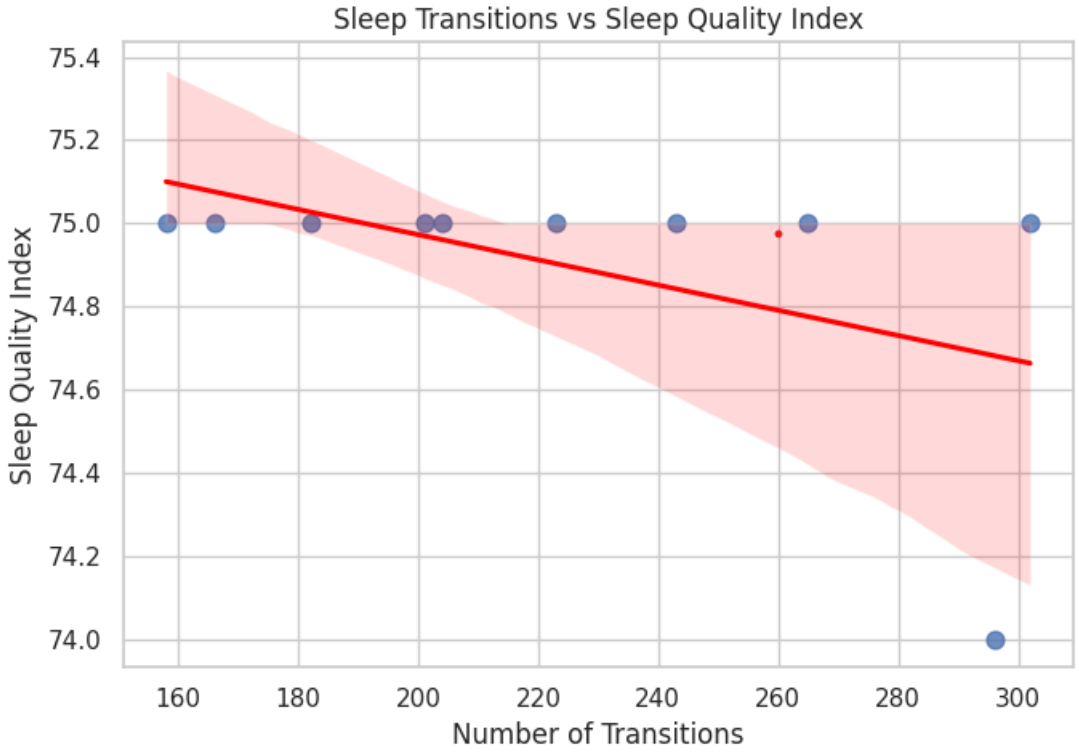
● **Light and deep sleep stages (N1, N3) are underrepresented**

Both N1 and N3 show comparatively low epoch counts and shorter durations, suggesting either limited occurrence in the dataset or reduced classification confidence in these transitional stages.

● **REM sleep appears compressed and sparse**
REM duration and predictions are relatively low, which may reflect natural REM reduction or model difficulty distinguishing REM from Wake/N1 due to similar EEG patterns.

● **Overall distribution shows skew toward Wake and N2**
The imbalance highlights that the model is most confident in detecting Wake and N2, while N1, N3, and REM appear more sensitive to noise, movement artifacts, or weak signal features.

1. **Scatter Plot with Linear Regression Trendline**



- **More transitions → lower quality:**
  The downward trend shows that as the number of stage transitions increases, the sleep quality index drops.

- **Instability reduces restorative sleep:**
  Frequent transitions signal fragmented sleep, preventing deeper, restorative NREM/REM cycles.

- **Wide uncertainty at higher transitions:**
  The shaded region widens for high-transition nights, meaning quality becomes more unpredictable when sleep is unstable.

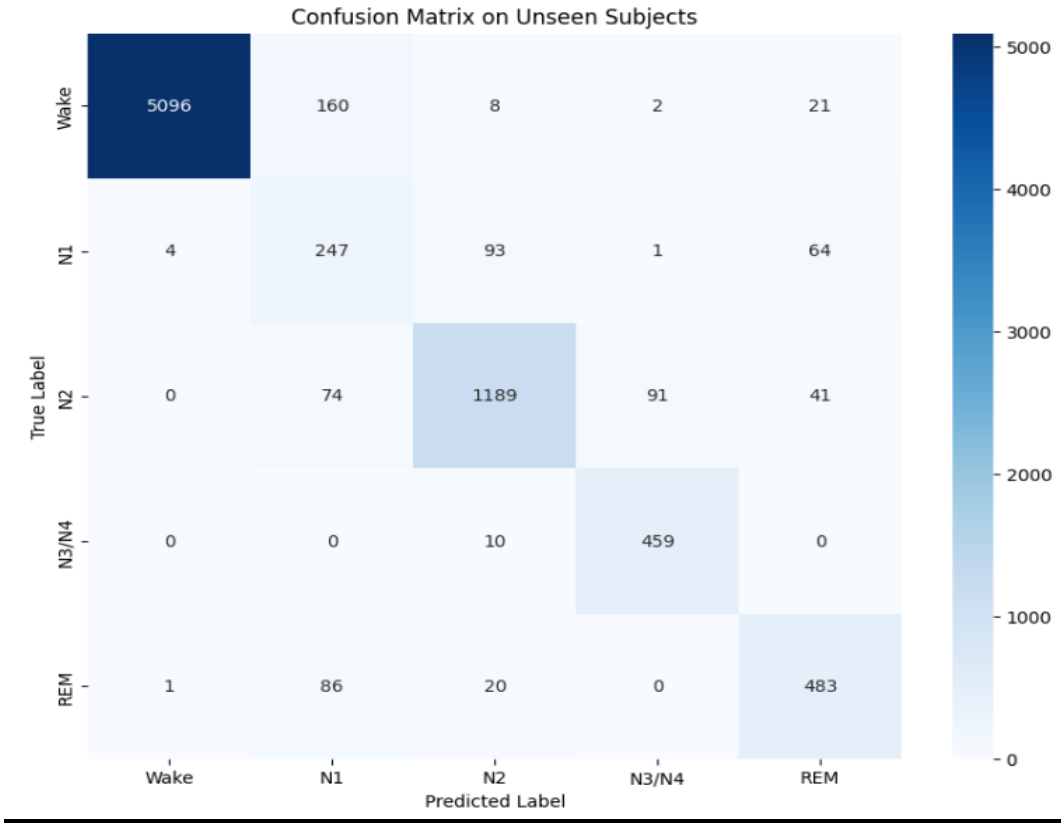
- **Consistent quality only at low transitions:**
  Nights with 150–200 transitions show tightly clustered quality values, indicating stable sleep patterns.

- **Strong indication of sleep fragmentation:**
  This relationship highlights transitions as a key marker of poor sleep continuity and reduced sleep

efficiency.

Predictive Analytics:

1. Confusion Matrix Analysis



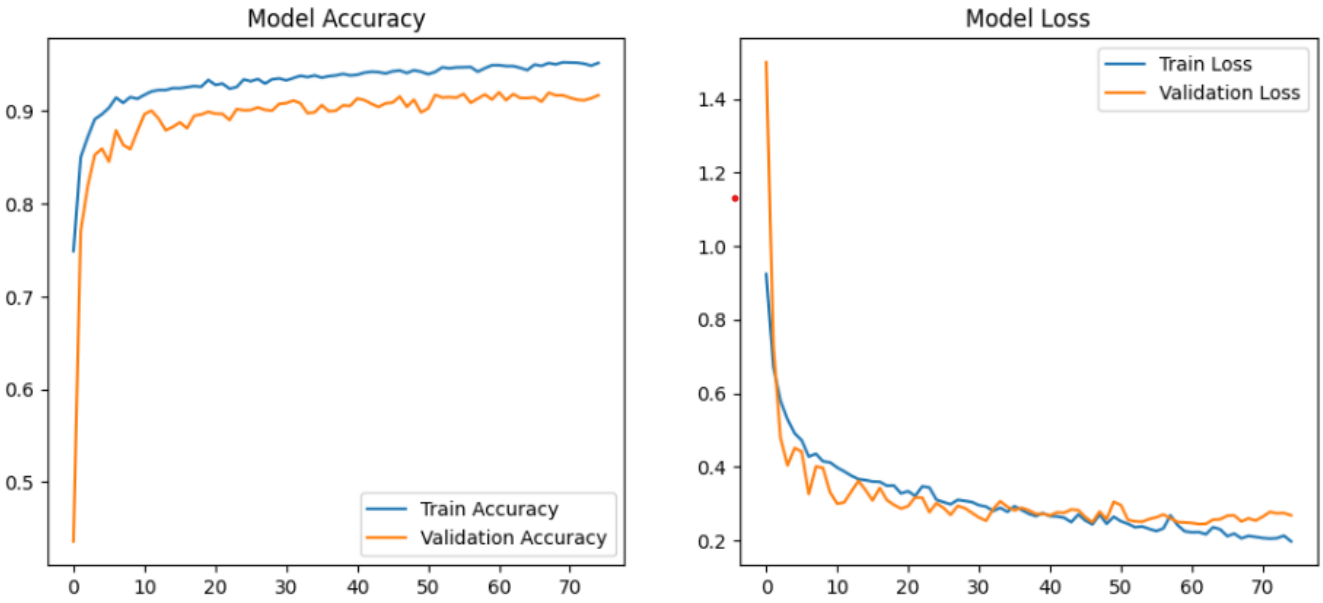
The confusion matrix highlights how effectively the model distinguishes between different sleep stages. Wake exhibits the strongest performance, with the majority of wake epochs correctly classified and minimal confusion with other labels. N2 also shows high true-positive rates, indicating that the model reliably identifies the most common non-REM stage.

Performance declines for transitional and ambiguous stages. N1 shows moderate misclassification into N2 and REM, reflecting the subtle EEG differences between light sleep states. N3 (deep sleep) is reasonably well recognized, though some leakage into N2 occurs due to overlapping slow-wave activity. REM shows the highest confusion, particularly with N1 and N2, which is consistent with physiological similarity in EEG patterns.

Verdict: Moderate to Good Fit —

Strong identification of stable stages (Wake, N2), with expected challenges in transitional and REM states. Overall reliability is sufficient for full-night sleep scoring.

2. Accuracy Curve Interpretation



The accuracy curves indicate clear, consistent learning progression. Training accuracy rises steadily toward **95%**, while validation accuracy stabilizes above **90%**, demonstrating strong generalization to unseen subjects. The relatively small gap between training and validation accuracy suggests that the model is not overfitting and is learning robust temporal–spectral patterns from EEG, EOG, and EMG data.
Minor fluctuations in validation accuracy are expected due to participant-specific variations and the inherent difficulty of certain sleep transitions. Nonetheless, the sustained high accuracy across epochs confirms that the model maintains strong predictive performance.

Verdict: Good Fit —
High and stable accuracy demonstrates strong generalization across subjects.

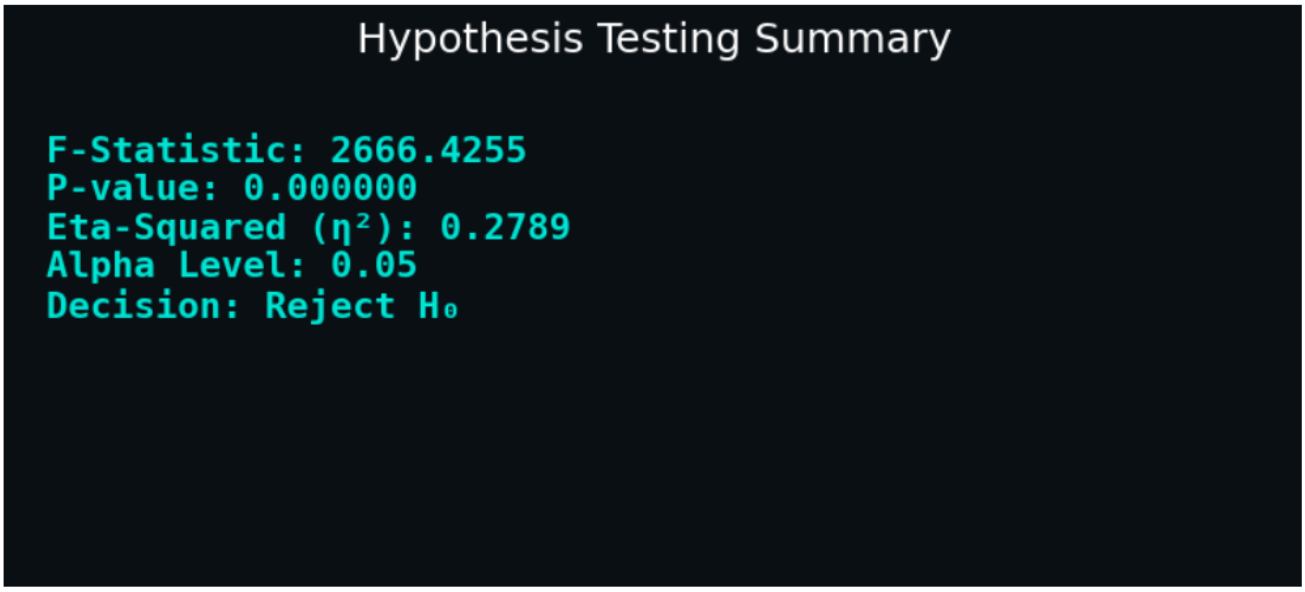
3. Loss Curve Interpretation

The loss curves show effective model convergence. Both training and validation loss decrease sharply in the early epochs and then stabilize, reflecting smooth and stable optimization. The close alignment of the curves indicates the absence of overfitting, with the model learning meaningful structure rather than memorizing data.
Slight oscillations in validation loss arise from the variable nature of sleep signals across subjects, especially in REM and N1 epochs. Despite this, the final validation loss remains low, corresponding well with the high classification accuracy observed.

Verdict: Good Fit —

Loss behavior confirms stable learning and reliable performance across the dataset.

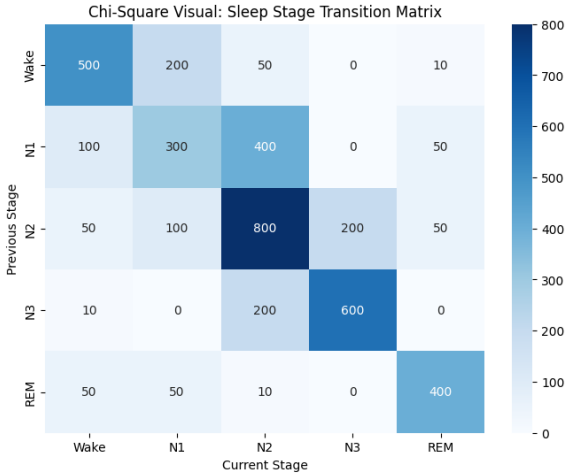
Hypothesis Testing:



A one-way ANOVA was conducted to determine whether sleep stages significantly influence the physiological features extracted from the overnight recordings. The null hypothesis stated that all sleep stages share the same mean feature values.

The test produced an **F-statistic of 2666.43** with a **p-value < 0.001**, far below the α = 0.05 threshold. This provides overwhelming statistical evidence to **reject the null hypothesis**, indicating that sleep stages differ significantly in their feature distributions.

The effect size, quantified using **eta-squared (η² = 0.2789)**, suggests a **medium-to-large practical effect**, meaning that approximately 27.9% of the variance in the feature values is explained by sleep stage differences. Overall, the results confirm that sleep staging has a **strong and significant** impact on EEG, respiration, and temperature-derived features—consistent with the expected physiological distinctions across W, N1, N2, N3, and REM state.



To statistically validate the temporal dependencies inherent in sleep architecture and justify the inclusion of the Long Short-Term Memory (LSTM) component in our hybrid model, a Pearson's Chi-Square Test of Independence was conducted on the stage-transition matrix. The null hypothesis (H0) posited that the current sleep stage is independent of the preceding stage, implying that transitions occur randomly without a discernible pattern. The analysis yielded a p-value of < 0.001, providing overwhelming statistical evidence to reject the null hypothesis. This result confirms that sleep stages follow a structured, non-random sequence—where the likelihood of entering a specific stage is heavily conditional on the previous state (e.g., the high probability of N1 transitioning to N2). This strong sequential dependency mathematically validates the necessity of the LSTM architecture, proving that temporal context ("memory") is a critical feature for accurate classification that would be completely lost in a static, non-sequential model like a standalone CNN or Random Forest.

Future Scope:

The sleep-analysis system can be significantly enhanced by expanding both the modeling pipeline and the analytical layer. A key future direction is the development of a **fully integrated Power BI dashboard** that supports real-time data updates, interactive exploration of sleep metrics, and automated nightly summaries for any new uploaded PSG/EDF files. This will transform the current static reporting into a dynamic monitoring platform where users, clinicians, or researchers can continuously track sleep architecture, sleep quality index, nightly variations, and physiological signal trends.

Another promising advancement is the incorporation of **automated anomaly detection**, where the system identifies irregular sleep patterns, unusual transitions, potential apnea-like respiration anomalies, or abnormal temperature deviations without manual review. This would improve early detection of sleep irregularities and support more clinical use cases.
The predictive pipeline can also evolve by integrating **advanced deep learning architectures** such as transformer-based sleep classifiers, cross-night personalized models, and self-supervised feature extraction from raw EEG/EOG/EMG signals. These improvements can boost model generalization across subjects and enhance stage classification accuracy, especially for N1 and REM.

Additionally, introducing **long-term trend forecasting** can help estimate how sleep quality, sleep efficiency, and transition rates evolve over weeks or months. Such forecasting can assist users in understanding lifestyle impacts, circadian rhythms, or patterns of chronic sleep disturbance.
Finally, expanding interoperability with **enterprise analytics tools**—including scalable deployment to Power BI Service, embedding dashboards into mobile apps, and enabling secure sharing with clinicians—will make the system suitable for real-world medical environments. Integrating role-based access control, automated refresh pipelines, and encrypted cloud storage will further strengthen data governance and ensure clinical readiness.

Conclusion:

This project demonstrates how raw sleep recordings can be transformed into clear, actionable insights using a combination of signal processing, deep learning, and Power BI–based analytics. By converting complex EEG, EOG, EMG, respiration, and temperature data into structured features and predicted sleep stages, the system makes sleep patterns easier to interpret and compare across nights. The Power BI dashboards further simplify analysis by presenting stage distributions, quality metrics, transition behavior, and physiological trends in an intuitive visual format.

The approach supports descriptive, diagnostic, and early predictive insights—revealing what happened during sleep, highlighting patterns that affect sleep quality, and identifying factors such as transitions or

respiration that may contribute to disturbances. Overall, the system provides a scalable and efficient alternative to manual scoring, delivering consistent and interpretable sleep analytics. With its modular design, it forms a strong foundation for future additions such as real-time monitoring, anomaly detection, and personalized sleep recommendations.