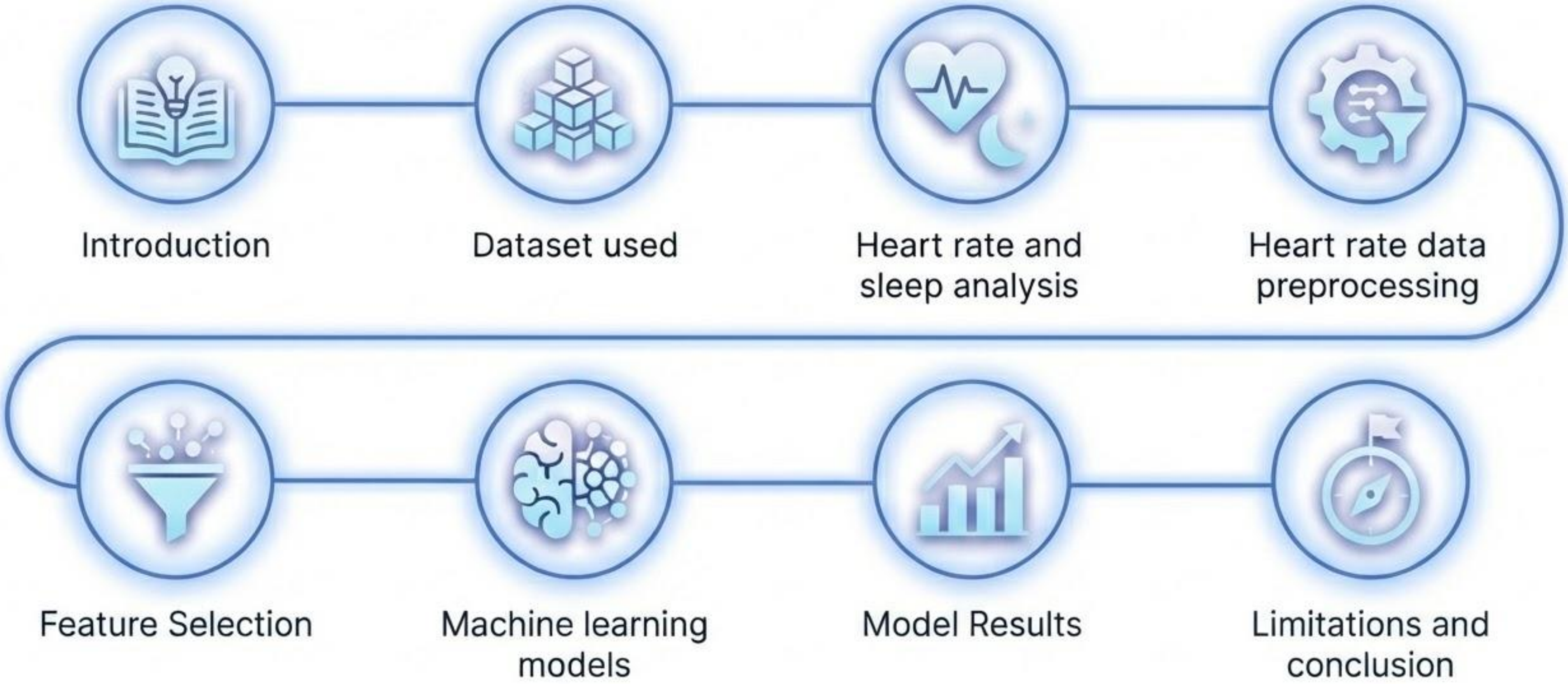
Analyzing Sleep Data in Sleeping Beauty Syndrome

Rajani Jagarwal
Tarikwa Bedane

Outlines


Introduction
Dataset used
Heart rate and sleep analysis
Heart rate data preprocessing
Feature Selection
Machine learning models
Model Results
Limitations and conclusion

What is KLS or Sleeping Beauty Syndrome?

Kleine–Levin Syndrome (KLS) is a rare disorder where people:

Sleep excessively for days or weeks

Have changes in behavior and thinking

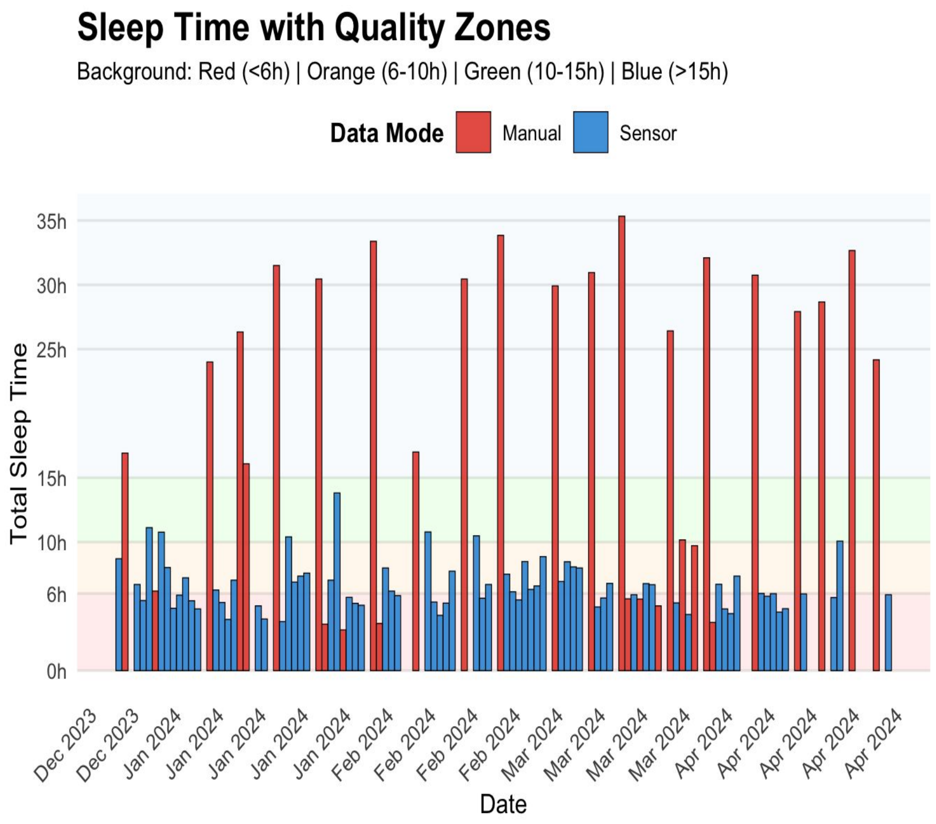
Have episodes that come and go

PROBLEM: UNPREDICTABILITY

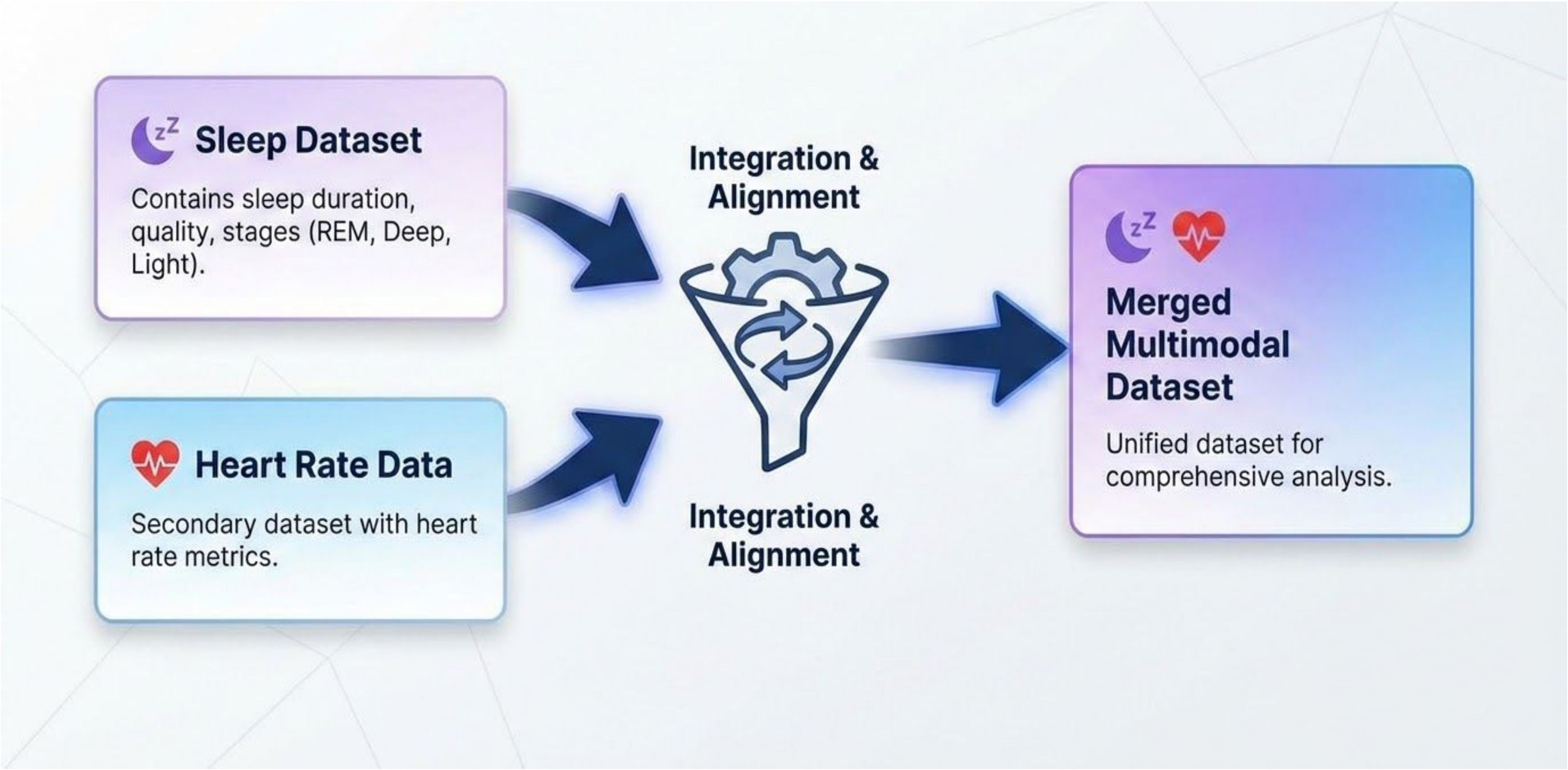
It is very hard to predict when an episode will happen.

Dataset Overview

- ~100 days of sleep data from one KLS patient
- Data collected daily as a **time-series**
- Contains two data modes:
  - **Sensor data** (normal nights)
  - **Manual data** (often during hypersomnia episodes)
  - Around 20 attack days



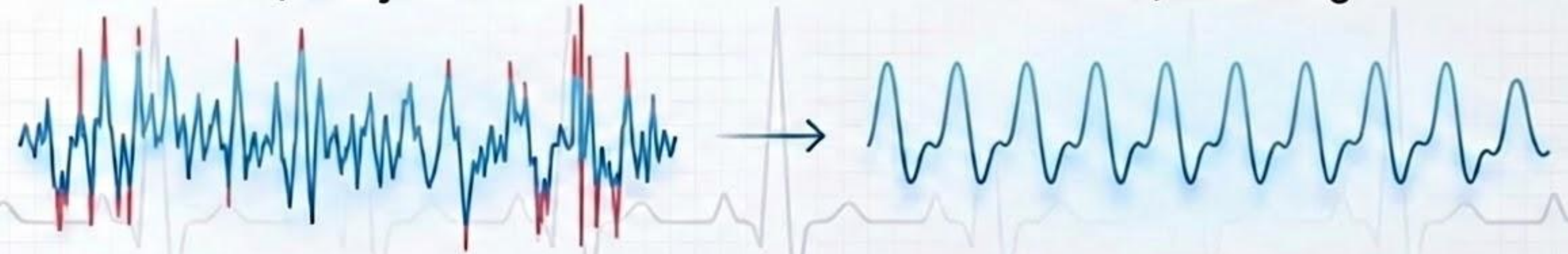
Dataset Used



Cleaning Heart Rate Data

We cleaned it by:

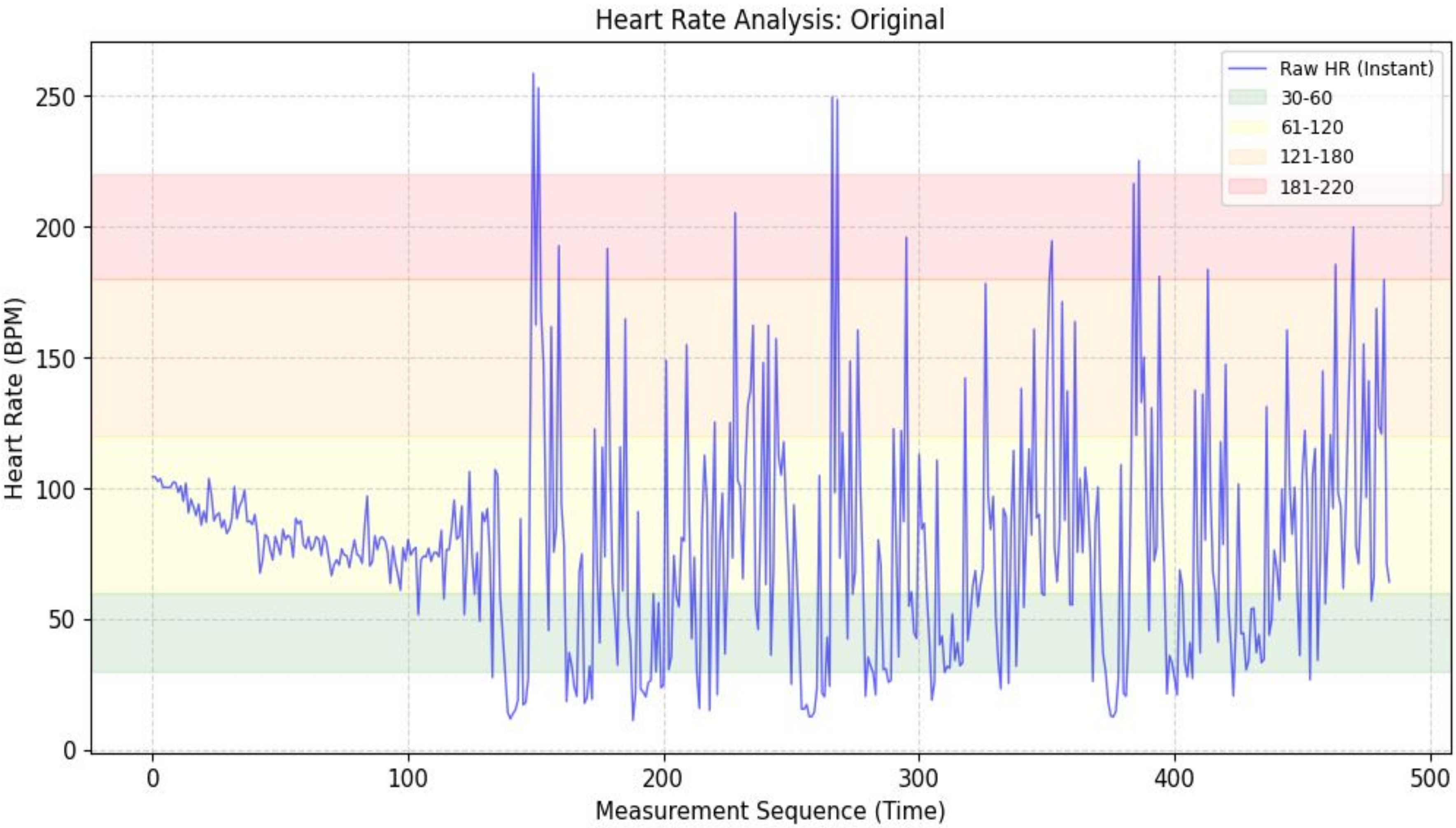
Raw, Noisy Data → **Cleaned, Stable Signal**

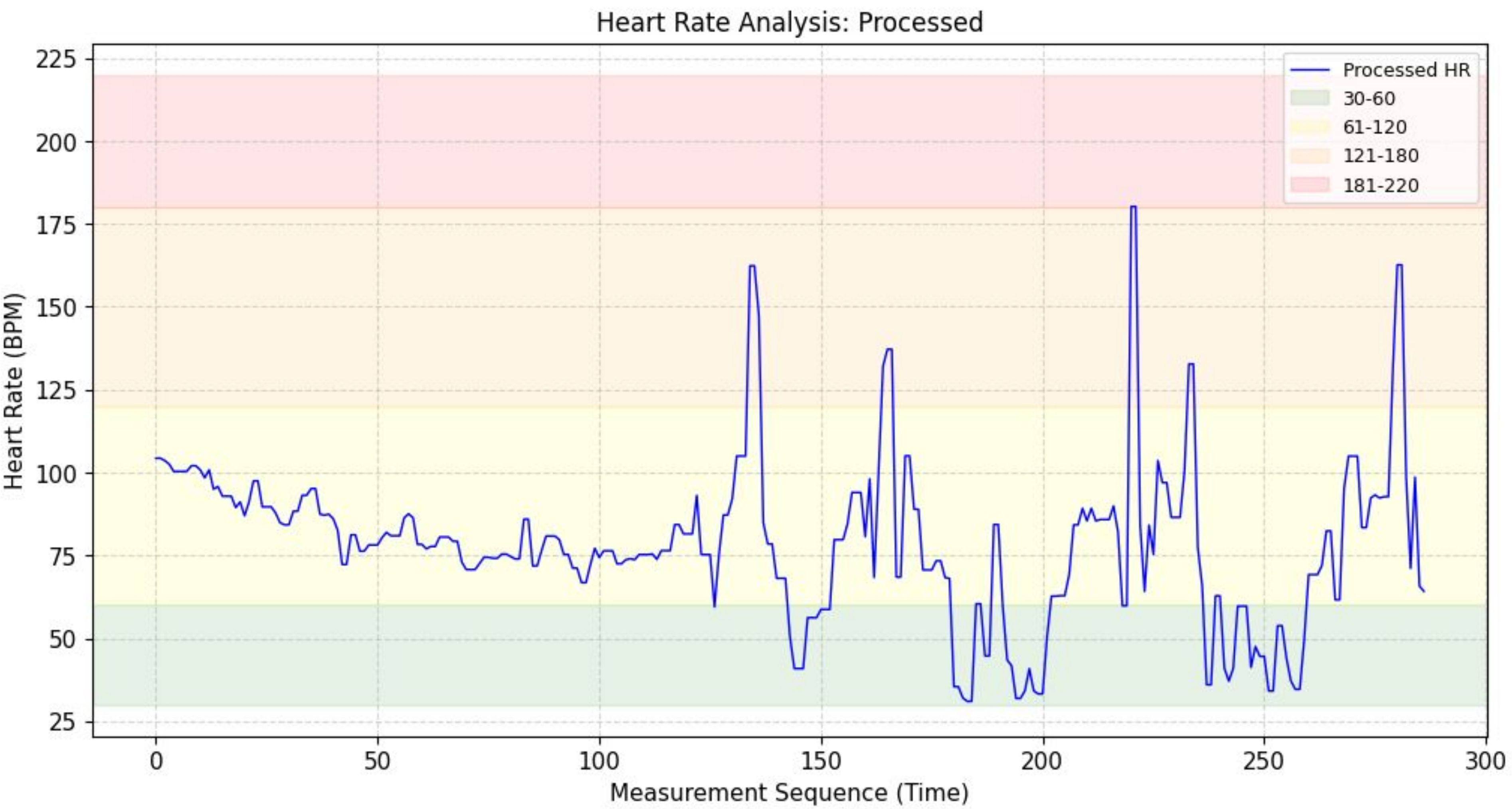


Raw heart rate data is noisy.

To obtain stable and physiologically meaningful heart rate signals.

1. safe parsing (Removing invalid values)
2. Physiological filtering (Keeping only realistic heart rate values 30-220 BMP)
3. Artifact removal (sudden jumps) diff>40BMP
4. Median filtering (Smoothing the signal to reduce noise)


Heart Rate Analysis: Original
Raw HR (Instant)
30-60
61-120
121-180
181-220
Heart Rate (BPM)
Measurement Sequence (Time)
0
50
100
150
200
250
0
100
200
300
400
500



Heart Rate Analysis: Processed
Processed HR
30-60
61-120
121-180
181-220
Heart Rate (BPM)
225
200
175
150
125
100
75
50
25
0
50
100
150
200
250
300
Measurement Sequence (Time)

Heart Rate Bins

Heart rate values were grouped into predefined ranges:

- 30–60 BPM
- 60–120 BPM
- 120–180 BPM
- 180–220 BPM

These features describe the distribution of heart rate levels during sleep.

Feature Importance Analysis

To identify features that contribute most to predicting KLS episodes.

Feature importance was assessed using:

1. Random Forest feature importance
2. Correlation analysis between features and future attacks

- Sleep duration :
  - Sleep_total_time
  - Sleep_interruption
  - Sleep_deviation
  - sleep_score
- Sleep stage percent :
  - REMPercent
  - lightPercent
- HR metrics:
  - Hr_mean
  - hr_bin_30_60_pct
  - Hr_bin_120_180_pct
- Weekday

Time-Lagged Feature Analysis

- Captures features from **t-1, t-2, t-3** days.
- Adds new feature with appending string **“_lag<day>”**
- Uses **Random forest** to list down most important/contributing features.

Features included sample:

- Stat_total_sleep_time_minutes
- Stat_total_sleep_time_minutes_lag1
- stat_total_sleep_time_minutes_lag2
- lightPercent_lag2
- Hr_bin_30_60_pct_lag1
- hr_bin_30_60_pct_lag2
- REMPercent
- REMPercent_lag2
- NREMPercent_lag2

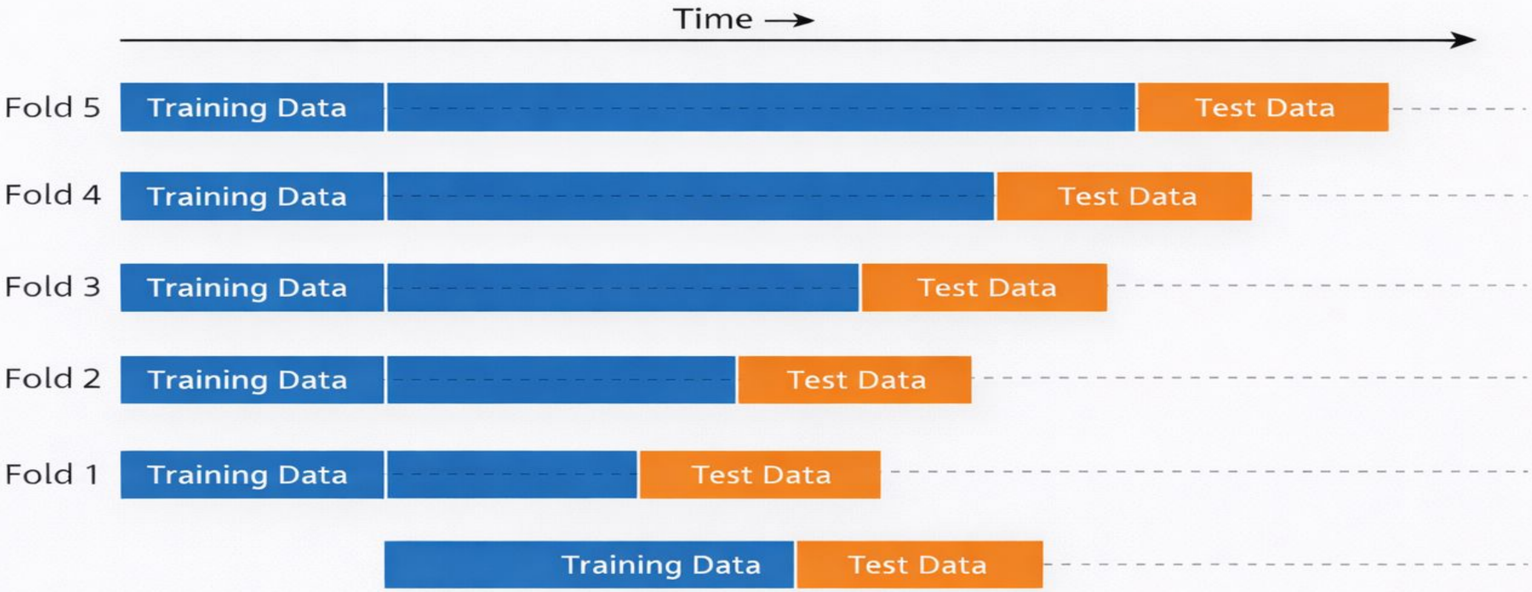
What are we predicting ?

- If the attack will happen next day or not.
- **Target** : **attack_in_1_day**
- Classification Problem (attack_in_1_day = 0 , attack_in_1_day = 1)

Time Series Cross Validation

- We used time-series cross-validation with 5 folds.
- In each fold, models were trained on past data and tested on future data.

This approach avoids data leakage and reflects real-world prediction scenarios.



Machine Learning Models

Models Evaluated:

- Logistic Regression
- Random Forest
- XGboost(Gradient Boosting)
- Neural Network

Temporal Models:

Also RNN and LSTM were tested for temporal pattern learning but we rejected them because data scarcity and rare KLS episode caused unreliable and poor performance.

Model Evaluation Metrics

Because KLS episodes are rare events, accuracy alone is not a reliable measure of model performance.

Therefore, we focus on metrics that reflect the model's ability to detect rare attacks:

- **Recall** (Sensitivity): ability to detect KLS episodes (TP/TP+FN)
- **Precision**: reliability of attack predictions (TP/TP+FP)
- **F1-score**: balance between recall and precision
- **PR-AUC**: performance under class imbalance

Logistic Regression

Effective for binary prediction with limited and imbalanced data.

Model setup:

Balanced class weighting + 1000 iterations

Important Features

- Sleep_score_averaged
- Stat_total_sleep_time_minutes_lag1
- hr_bin_120_180_pct
- Stat_sleep_interruption_minutes
- stat_total_sleep_time_minutes_lag2

Confusion Matrix

[[15 1]
 [0 4]]

Accuracy	: 0.950
Precision	: 0.800
Recall	: 1.000
F1-score	: 0.889
PR AUC	: 0.800

Random Forest

Combines multiple decision trees to improve stability and capture complex sleep–heart rate patterns.

Model setup

Ensemble of depth-limited trees (max_depth=8) + 300 n_estimators, balanced class weighting

Important Features

- stat_total_sleep_time_minutes,
- sleep_score_averaged,
- stat_bed_time_deviation_minutes,
- sleep_score,
- stat_total_sleep_time_minutes_lag2

Confusion Matrix

[[16 0]
 [0 4]]

Accuracy	: 1.000
Precision	: 1.000
Recall	: 1.000
F1-score	: 1.000
AUC	: 1.000

XGBoost

Advanced boosting algorithm designed to learn complex patterns and handle class imbalance.

Model setup:

Gradient-boosted trees(max_depth=8),
learning_rate=0.01, + imbalance-aware weighting

Important Features

- Sleep_score,
- NREMPercent,
- hr_bin_30_60_pct,
- Stat_total_sleep_time_minutes
- sleep_interruption_count

Confusion Matrix

[[16 0]
 [0 4]]

Accuracy : 1.000
Precision : 1.000
Recall : 1.000
F1-score : 1.000
AUC : 1.000

Neural Network

Designed to learn complex, non-linear relationships in physiological data.

Model setup

Architecture

Input features → Hidden layers (64, 32, ReLU) → Sigmoid output

Optimization

- Dropout = 0.3
- Adam(lr = 0.001)
- Binary cross-entropy

Training

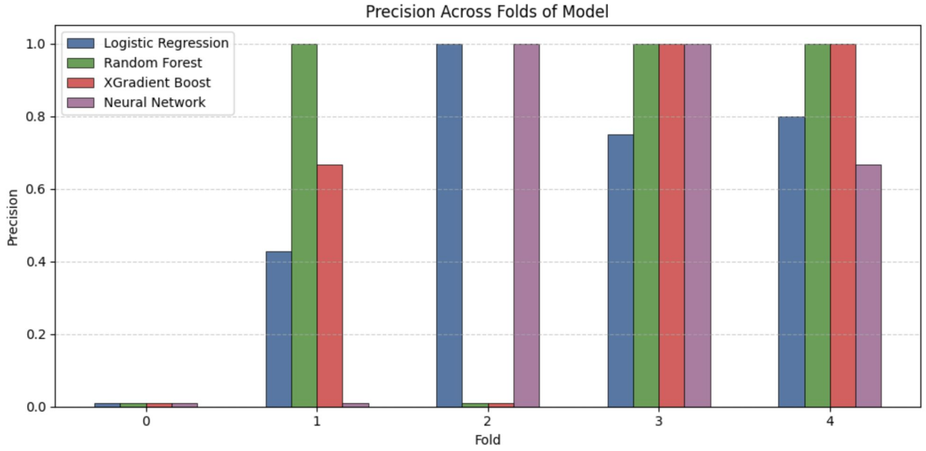
- Epochs = 30
- Batch size = 32, Threshold = 0.5

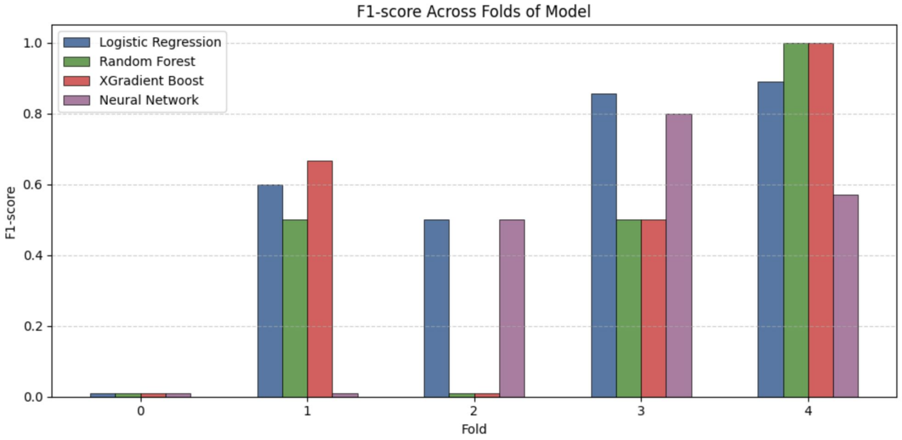
Confusion Matrix

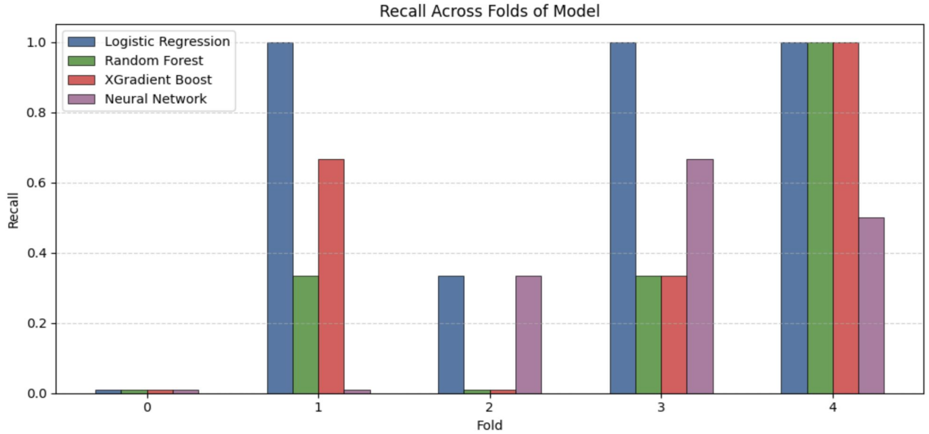
[[15 1]
 [2 2]]

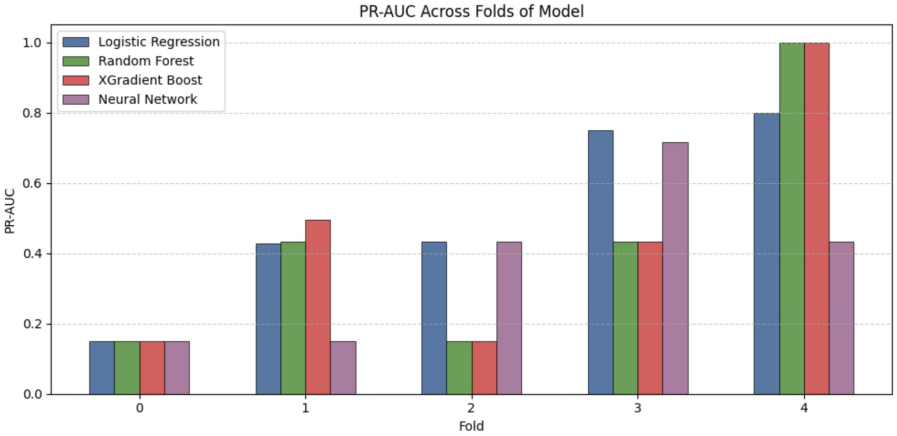
Accuracy	: 0.850
Precision	: 0.667
Recall	: 0.500
F1-score	: 0.571
PR AUC	: 0.433

Model Performance Comparison



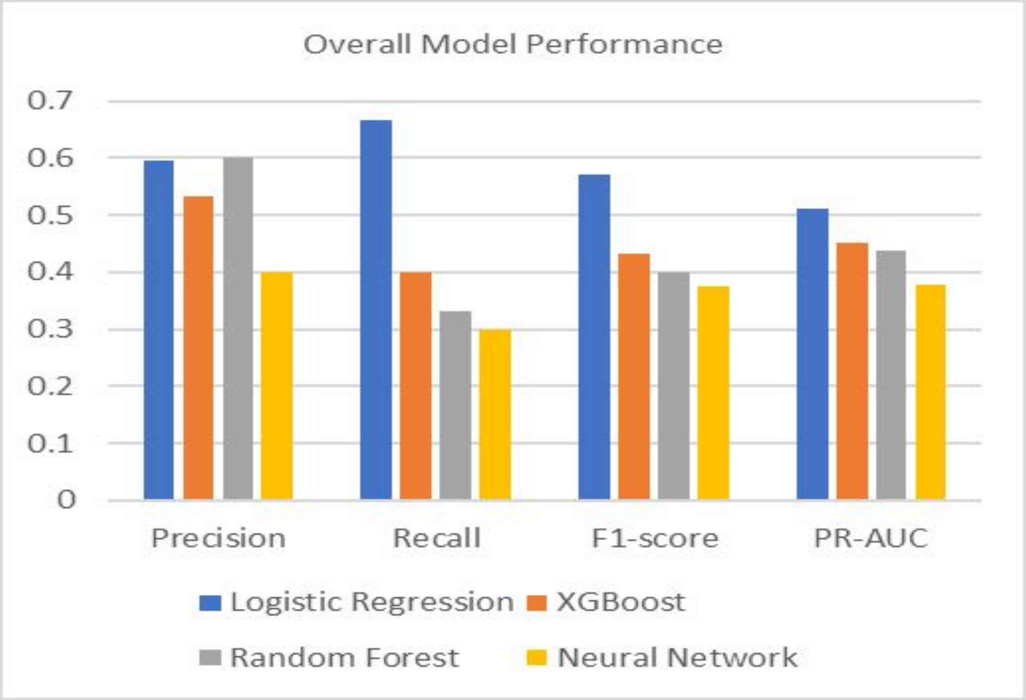






Model Performance Comparison

- **Stability Across Time** (Folds) - Logistic Regression
- **Recall Performance** -
  The **Tree-based models** miss positive cases in initial folds
- **Random Forest and XGradient** Boost outperforms overall in later folds - 4 and 5
- **Neural Network** is not performing well and F1-score, PR-AUC reduces in later folds



Limitations

Key limitations include:

- Limited number of KLS episodes in the dataset
- Potential uncertainty in attack labeling
- Noise and missing values in wearable data
- Class imbalance affecting model training
- Limited generalizability beyond the studied subject(s)

Conclusion

This study demonstrates that:

- Sleep score, duration, stage NREM/REM percent and HR metrics can be integrated to model physiological changes associated with KLS episodes.
- Adding features with lag of 1 and 2 days improved performance of models.
- Machine learning methods can identify patterns preceding KLS attacks.(Also few models perform better than others)
- However, robust prediction requires larger datasets and clinical validation.

THANK YOU!

Any Questions?