Integrating Hidden Markov Models for Polysomnographic Data Analysis

Likhitha Marrapu
Department of Data Science

Kent State University
Kent OH 44240, USA
lmarrapu@kent.edu

***Abstract*—This study explores the application of a Hidden Markov Model (HMM) for classifying respiratory events and sleep stages, specifically in the context of sleep apnea detection. By leveraging physiological signals as features, the model was trained to predict events such as apneas and hypopneas, while simultaneously decoding sleep stages using an innovative HMM approach. The implementation of SMOTE (Synthetic Minority Over-sampling Technique) helped address class imbalance, enhancing the model's ability to accurately classify less-represented events. Despite challenges in sleep stage classification, the model achieved notable results in respiratory event detection, showcasing its potential for real-world healthcare applications. The findings highlight the importance of continuous model refinement and the exploration of alternative techniques to further improve performance in sleep disorder diagnostics.**

I. Introduction

Polysomnography is a comprehensive diagnostic tool used in the assessment of sleep disorders, capturing physiological signals such as brain activity, eye movement, muscle tone, and respiratory events. The analysis of polysomnographic data is crucial for diagnosing conditions like sleep apnea, hypopnea, and various stages of sleep, which are vital for understanding sleep health. The MIT-BIH Polysomnographic Database provides a valuable collection of annotated data, representing a wide range of sleep stages and respiration events. By leveraging this dataset, researchers can gain insights into the intricate patterns of human sleep and respiration, leading to better diagnostic tools and treatment plans for sleep-related disorders.

The Hidden Markov Model (HMM) is a statistical model that is particularly effective in time-series analysis, making it ideal for tasks involving sequential data such as sleep stage classification and respiration event detection. In the context of polysomnography, HMM can be used to model the underlying hidden states, such as sleep stages and apnea events, based on observable physiological signals. The sequential nature of HMMs allows for the prediction and classification of sleep stages and respiration events, which are essential for the accurate diagnosis of sleep disorders. This approach promises to enhance the automation and accuracy of sleep analysis, aiding in the timely identification of abnormal sleep patterns and improving patient care.

II. Methodology

A. *Dataset*

The MIT-BIH Polysomnographic Database is a comprehensive collection of polysomnographic recordings obtained from subjects monitored at Beth Israel Hospital Sleep Laboratory in Boston. This dataset is designed to support research on sleep-related disorders, particularly chronic obstructive sleep apnea syndrome, and the effectiveness of therapeutic interventions like continuous positive airway pressure (CPAP). It encompasses multi-channel recordings, including electrocardiogram (ECG), electroencephalogram (EEG), and respiration signals, which provide a detailed representation of physiological activity during sleep. These signals are meticulously annotated for sleep stages, respiratory events, and cardiac activity, making the dataset a valuable resource for advancing the understanding and analysis of sleep health.

The database consists of 18 individual records, each including four key files: sleep and apnea annotations (.st), ECG beat annotations (.ecg), signal data (.dat), and header files (.hea). The annotations classify various sleep stages—such as wakefulness (W), REM (R), and non-REM stages (1-4)—as well as respiratory events like hypopneas, apneas, and their variants. This well-structured dataset, comprising over 80 hours of data, serves as a reliable foundation for developing models aimed at detecting sleep disorders and classifying complex respiratory and sleep-stage events with precision.

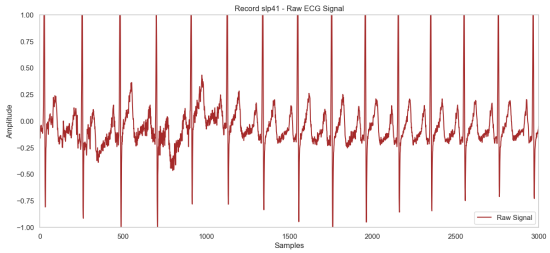
B. *Signal Processing and Feature Extraction*

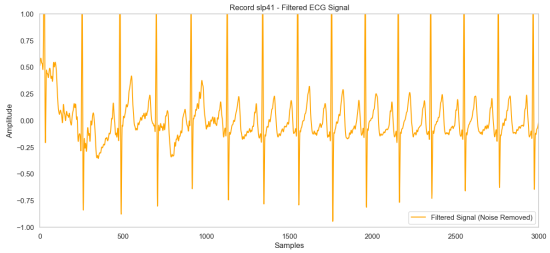
This study employs a systematic approach to analyze ECG signals, specifically focusing on P-wave detection and classification from the MIT-BIH Arrhythmia Database. The methodology encompasses several key steps, including Annotation Parsing and Event Processing,noise reduction, Signal Processing and Windowing, feature extraction:

*1) **Annotation Parsing and Event Processing**:* The implementation of the annotation parser leverages key concepts in data parsing, sequence mapping, and signal processing. Sleep stages are categorized into predefined classes (e.g., Awake, REM, and non-REM stages) while respiratory events are classified into categories such as apnea, hypopnea, and no apnea based on annotations provided in the dataset. The

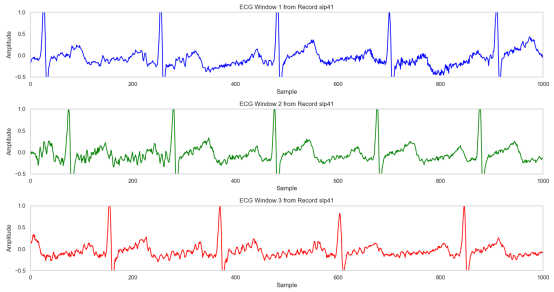
parser processes event sequences with a strict mapping to these categories and ensures proper synchronization of time-stamped annotations with signal data. A robust mechanism for handling unknown stages and events fills gaps in annotations using forward and backward filling techniques, improving continuity in the data. Additionally, windowing techniques are applied to segment the data into fixed durations with overlapping intervals, allowing for efficient analysis of time-series signals. These concepts enable a structured approach to extracting meaningful insights from complex polysomnographic data, ensuring accuracy in the classification of sleep stages and respiratory events.

*2) **Noise Handling Using Butterworth Bandpass Filtering**:* To ensure accurate analysis of polysomnographic signals, a Butterworth bandpass filter is employed for effective noise suppression. The filter eliminates frequencies outside the desired range, retaining only the relevant signal components for analysis. By defining low and high cutoff frequencies relative to the Nyquist frequency, the filter removes low-frequency drift and high-frequency noise. This approach preserves the integrity of physiological signals such as EEG, ECG, and respiration while minimizing artifacts, thereby enhancing the reliability of feature extraction and subsequent classification processes.





*3) **Signal Processing and Windowing**:* To process the ECG signals, a Butterworth bandpass filter is applied to remove unwanted noise and retain the relevant frequencies, particularly in the range of 0.5 to 50 Hz. The filtered signal is then divided into overlapping windows of fixed duration, which are essential for capturing temporal dynamics. Each window is analyzed for various features, ensuring that the signal's characteristics are well-represented for classification tasks. Additionally, sleep stage and event annotations are mapped to these windows, aligning physiological signals with labeled sleep stages and respiratory events. The use of windowing facilitates the analysis of non-stationary signals and prepares the data for machine learning models that can classify sleep stages and respiration events based on the extracted features.



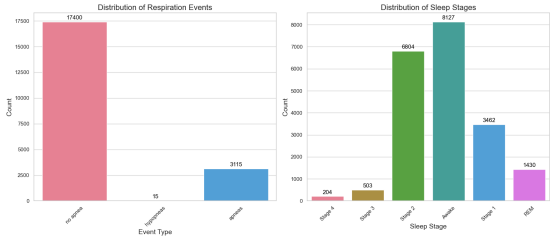
*4) **Feature Extraction**:* The feature extraction process involves calculating a variety of statistical, morphological, and frequency-domain features from segmented ECG windows. These include central tendency measures (mean), variability (variance), and amplitude-based features (peak-to-peak range, RMS value). In addition, statistical features such as skewness and kurtosis provide insights into the distribution of signal values, while morphological features like zero crossings capture the signal's shape. The frequency-domain analysis, using the Welch method, computes the power spectral density, highlighting the energy distribution across frequencies. Furthermore, complexity measures such as the mean absolute deviation (MAD) and Shannon entropy quantify the signal's unpredictability and informational content. These extracted features are combined with event and stage annotations to enrich the dataset and enable more accurate modeling of sleep and respiration events.

C. *Data Pre-processing*

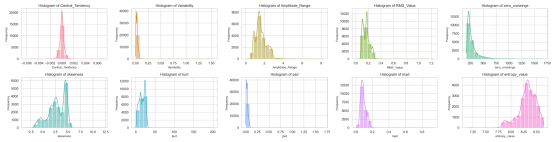
Data preprocessing is a critical step to ensure the dataset is clean, consistent, and ready for analysis. This process involved checking for and handling missing values to avoid potential biases or inaccuracies in the results. Data type conversions were performed to ensure compatibility with analytical and machine learning tools, such as converting categorical labels to numerical formats for efficient processing. Additionally, any anomalies or inconsistencies in the data were identified and addressed, ensuring that the dataset accurately represents the physiological signals and event annotations. These preprocessing steps laid the foundation for reliable feature extraction and model training.

D. *Exploratory Data Analysis (EDA)*

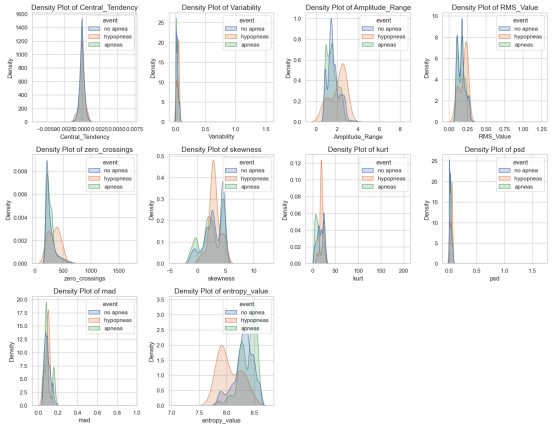
Exploratory Data Analysis (EDA) plays a crucial role in understanding the structure and distribution of the dataset before applying any advanced modeling techniques. In this project, EDA was performed to investigate the relationships between sleep stages, respiratory events, and the extracted features. By visualizing the distribution of features like central tendency, variability, and frequency-domain measures, we gained insights into their patterns and variability across different sleep stages and event types. The analysis also included examining correlations between physiological signals and event annotations, providing a deeper understanding of the data's inherent structure. EDA helped patterns, and trends in the data, guiding subsequent feature selection and preprocessing steps, ensuring the preparation of clean and meaningful data for model training.



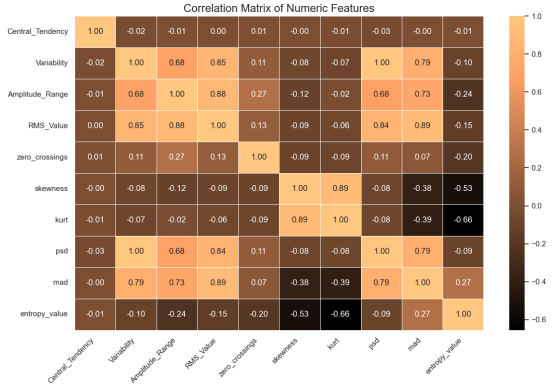
*1) **Univariate Analysis**:* The univariate analysis of the extracted features reveals distinct patterns that provide valuable insights into the characteristics of the physiological signals. Features such as Central Tendency and RMS Value exhibit tight clustering around lower values, indicating consistent signal behavior with minimal intensity variation. In contrast, features like Variability, Amplitude Range, and Zero Crossings display skewed distributions, highlighting the presence of generally stable signals punctuated by occasional larger fluctuations. The analysis of Skewness and Kurtosis uncovers asymmetry and the presence of sharp outliers, reflecting the non-normal nature of the signals. Additionally, the Power Spectral Density (PSD) is predominantly concentrated at lower frequencies, suggesting that the majority of the signal's energy is found in these regions. Measures like Mean Absolute Deviation (MAD) and Entropy point to moderate signal complexity, with some instances of higher randomness and deviation, offering key insights into the intricate dynamics of the sleep-related physiological signals.



*2) **Bivariate Analysis**:* Bivariate analysis was conducted to explore the relationship between numerical features and the target variable, 'event,' which classifies data into categories such as 'no apnea,' 'hypopneas,' and 'apneas.' Density plots of features like Central Tendency, Variability, and RMS Value reveal clear distribution shifts across the event classes, with 'apneas' displaying more widespread densities compared to other categories. Additionally, Amplitude Range and zero crossings exhibit distinct patterns, underscoring the varying behaviors of each class. Statistical features such as skewness, kurtosis, and entropy value show prominent peaks for 'no apnea,' indicating a strong separation between classes. Furthermore, the analysis of features like MAD and PSD unveils specific density concentrations for 'apneas,' highlighting their potential as important predictors. These insights from bivariate analysis provide a deeper understanding of how individual features relate to apnea events and inform feature selection for further model development.



*3) **Correlation Analysis**:* The correlation analysis reveals several significant relationships between the extracted features, providing valuable insights into their interdependencies. A strong positive correlation is observed between features that capture signal amplitude and variability, indicating that as one feature increases, the other tends to follow a similar trend. Additionally, features related to the distribution shape, such as skewness and kurtosis, exhibit a notable correlation, suggesting that the data's asymmetry and peakedness are closely related. Some features, like those measuring signal amplitude and variability, show high inter-correlation, implying that they may share common information about the underlying signal characteristics. On the other hand, certain features such as zero crossings and entropy display weaker correlations with the rest of the features, indicating that they contribute unique information with less overlap. These correlations offer a deeper understanding of the relationships within the data, which is essential for identifying key features for model development and enhancing the accuracy of predictions in subsequent stages of analysis.



III. Model Building And Evaluation

A. Hidden Markov Model (HMM) Architecture and Description

This section provides an overview of the HMM framework used for classifying sleep stages and respiratory events. Hidden Markov Models (HMMs) are statistical models ideal for sequential data, where observations depend on unobserved states. In this study, two separate Gaussian HMMs were designed: one for sleep stage classification and another for respiratory event classification. Each model incorporates key features extracted from the ECG signal, leveraging statistical, frequency, and morphological characteristics to identify latent states effectively. The model parameters, including transition probabilities, initial state probabilities, and emission distributions, were initialized using KMeans clustering to ensure a robust starting point for learning.

B. Training Process for Classifying Respiratory Events Using HMM

*1) **Data Preprocessing and Feature Engineering**:* Before training the HMM, the input data undergoes preprocessing steps to ensure quality and consistency. This includes handling missing values, scaling features using a standard scaler, and clipping extreme outliers to improve the stability of training. Important features such as central tendency, variability, and entropy are extracted to capture the characteristics of respiratory patterns. Labels for respiratory events (e.g., apnea, hypopnea) are encoded using a label encoder to facilitate HMM training.

*2) **Class Balancing Using SMOTE**:* Class imbalance in respiratory event data is addressed using Synthetic Minority Oversampling Technique (SMOTE). This technique generates synthetic samples for underrepresented classes, ensuring a balanced dataset. Balanced data improves the HMM's ability to generalize and perform accurately across all respiratory event categories.

*3) **Model Initialization and Training**:* The Gaussian HMM for respiratory events was initialized with 3 hidden states. Cluster centers obtained from KMeans served as initial emission means, and uniform probabilities were set for state transitions. The HMM's covariance matrices were adjusted to account for feature variance. The model was then trained using the Expectation-Maximization (EM) algorithm, iteratively refining parameters to maximize the likelihood of observed respiratory event sequences.

C. Evaluation Process for Respiratory Event Classification

*1) **Performance Metrics**:* The trained HMM's performance was evaluated using standard metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive understanding of the model's ability to classify respiratory events. Weighted averages were calculated to account for class imbalances, and confusion matrices were plotted to visualize misclassification patterns.

D. *Decoding Process for Predicting Sleep Stage Sequences*

*1) **Sequence Decoding with HMM**:* The HMM for sleep stage classification utilized 6 hidden states, corresponding to known sleep stages (Awake, Stage 1-4, REM). After training, the model decoded sequences of sleep stages from observed ECG features. The Viterbi algorithm was employed to find the most likely sequence of hidden states, offering a robust interpretation of sleep dynamics.

*2) **Classifying Sleep Stages and Respiratory Events**:* Decoded sleep stages were aligned with corresponding respiratory event predictions to provide a comprehensive view of sleep and breathing patterns. This integration enables detailed classification of sleep states and their association with respiratory irregularities, offering potential insights for diagnosing sleep disorders.

*3) **Visualization and Interpretation**:* The decoded sequences and classification results were visualized using heatmaps and confusion matrices. These visualizations highlighted the transitions between sleep stages and their impact on respiratory events, showcasing the interdependence of sleep and respiration.

IV. Results

A. *Sleep stages classification results*

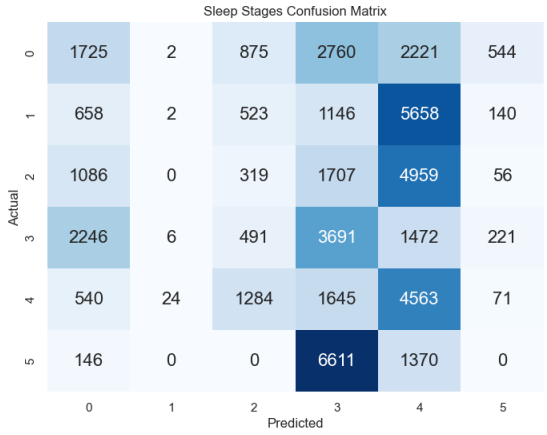
The classification of sleep stages using the Hidden Markov Model (HMM) yielded a low performance, with an overall accuracy of approximately 21.12%. The model struggled with distinguishing between the different stages, as indicated by the low precision and recall values for most stages, particularly for ”REM” and ”Stage 4,” where both metrics were almost zero. While ”Stage 3” showed a relatively better recall of 0.56, the model’s overall ability to correctly classify sleep stages was limited. The weighted F1 score of 0.15 further highlights the challenges faced in correctly classifying the sleep stages, indicating the need for improved feature engineering, model tuning, or the incorporation of more complex models for better accuracy and classification performance.The confusion matrix for sleep stage classification reveals varying levels of performance across classes. While Stage 3 demonstrates relatively strong classification accuracy, several stages, including Stage 5, show significant misclassifications into other stages, particularly Stage 3. This indicates potential issues with overlapping features or insufficient differentiation between sleep stages in the feature space. Improving feature representation and addressing possible class imbalances could enhance the model’s ability to distinguish between similar sleep stages more effectively

Class	Precision	Recall	F1-Score	Support
Awake	0.27	0.21	0.24	8127
REM	0.06	0.00	0.00	8127
Stage 1	0.09	0.04	0.05	8127
Stage 2	0.21	0.45	0.29	8127
Stage 3	0.23	0.56	0.32	8127
Stage 4	0.00	0.00	0.00	8127
Accuracy		0.21		48762
Macro Avg	0.14	0.21	0.15	48762
Weighted Avg	0.14	0.21	0.15	48762

TABLE I
CLASSIFICATION REPORT FOR SLEEP STAGES

Metric	Value
Accuracy	0.2112
Precision	0.1425
Recall	0.2112
F1 Score	0.1503

TABLE II
EVALUATION METRICS FOR SLEEP STAGES CLASSIFICATION



B. *Respiratory Event classification results*

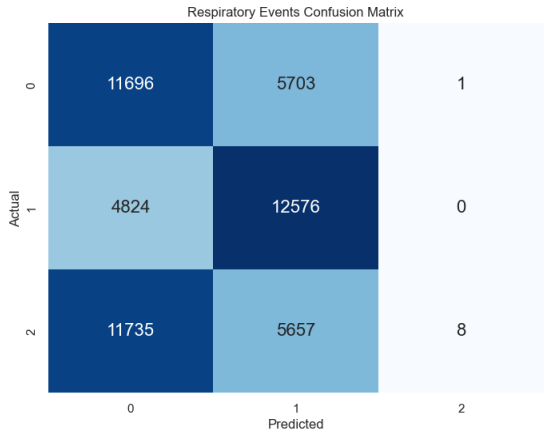
The classification results indicate moderate performance in distinguishing between respiratory events using the HMM-based approach. Hypopneas achieved the highest recall (72%) and a satisfactory f1-score (61%), followed by apneas with a lower f1-score of 51%. However, the model struggled significantly with classifying ”no apnea” events, resulting in a recall of 0%, which drastically affected the overall accuracy. The weighted accuracy of 46.51% suggests the need for improved feature selection or model tuning to enhance the detection of ”no apnea” events and balance the classification across all event types. These findings highlight the model’s current limitations and areas for optimization.The confusion matrix indicates that the model performs well in predicting Class 1, with the majority of instances correctly classified. However, Class 0 shows moderate accuracy, with a significant number of misclassifications into Class 1. Class 2 performs poorly, with very few correct predictions and a large number of samples misclassified as Class 0 or Class 1. This suggests potential challenges such as class imbalance or overlapping feature representations, affecting the model’s ability to distinguish between certain respiratory events effectively.

Class	Precision	Recall	F1-Score	Support
Apneas	0.41	0.67	0.51	17400
Hypopneas	0.53	0.72	0.61	17400
No Apnea	0.89	0.00	0.00	17400
Accuracy		0.47		52200
Macro Avg	0.61	0.47	0.37	52200
Weighted Avg	0.61	0.47	0.37	52200

TABLE III
CLASSIFICATION REPORT FOR RESPIRATORY EVENT CLASSIFICATION

Metric	Value
Accuracy	0.4651
Precision	0.6094
Recall	0.4651
F1 Score	0.3739

TABLE IV
EVALUATION METRICS FOR RESPIRATORY EVENT CLASSIFICATION



V. DISCUSSION

In evaluating the performance of the Hidden Markov Model (HMM) for classifying sleep stages and respiratory events, the results indicate that the model has achieved moderate performance in both areas, with particularly low accuracy and precision for sleep stages. The accuracy for sleep stages is 0.211, and the precision stands at 0.142, which suggests that the model struggles to correctly predict sleep stages. Conversely, for respiratory events, the accuracy is higher at 0.465, and precision improves significantly at 0.609, showing a better performance for this classification task. These metrics indicate that the model is more successful in identifying respiratory events than sleep stages, which may be attributed to the inherent complexity and variability of sleep stage classification.

To address potential class imbalances in the dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was utilized to oversample the minority classes. This technique helped balance the distribution of classes, ensuring that the model was trained on a more representative set of data. However, despite this adjustment, the HMM model's overall performance in sleep stage prediction remains limited. This could be due to the nature of the Hidden Markov Model itself, which may not fully capture the complex temporal dependencies present in the sleep stages. Further model optimization, possibly by integrating more advanced techniques such as deep learning, may help improve the prediction accuracy for sleep stages while maintaining performance for respiratory events.

Metric	Sleep Stages	Respiratory Events
Accuracy	0.211230	0.465134
Precision	0.142545	0.609411
Recall	0.211230	0.465134
F1 Score	0.150322	0.373920

TABLE V
EVALUATION METRICS FOR SLEEP STAGES AND RESPIRATORY EVENTS

VI. LIMITATIONS

While the HMM-based approach demonstrated potential in classifying respiratory events and sleep stages, there are several limitations that must be addressed. One of the key challenges was the relatively low accuracy and precision observed in sleep stage classification, which may be due to the complexity of sleep patterns and the quality of the available data. Additionally, although SMOTE was used to mitigate class imbalance, the model still faced challenges with overfitting, especially in less-represented classes. Furthermore, the model's performance could be further impacted by noise and inconsistencies in physiological signal data, highlighting the need for better data preprocessing and feature engineering. Future work should focus on refining the model architecture, exploring other advanced sampling techniques, and incorporating higher-quality datasets to improve overall performance and robustness.

VII. CONCLUSION

In conclusion, the Hidden Markov Model (HMM) demonstrated promising results for classifying respiratory events, with moderate accuracy and precision. The model's ability to accurately predict respiratory events highlights its potential in real-world applications such as sleep apnea detection. The use of SMOTE to address class imbalances contributed to the model's ability to better classify these events by enhancing the representation of minority classes, thus improving its overall performance. However, the performance for sleep stage classification was less than ideal, with lower accuracy and precision, indicating the challenges in accurately predicting such intricate sequences.

While the current model offers valuable insights into respiratory event classification, further improvements are necessary, particularly for sleep stage prediction. Exploring alternative approaches, such as deep learning models, could help capture the complex temporal relationships inherent in sleep data. Overall, this study underscores the importance of model optimization and further experimentation to enhance the capabilities of HMMs and other machine learning techniques in the realm of sleep health and diagnostics.