# Predictive Modeling of Long-Term CPAP Non-Adherence in OSA patients from Post-Initiation Treatment Telemonitoring Data

**Benedetta Giachetti** [1 2] **Simone Costantini** [3 4] **Elena Mugellini** [2] **Arnaud Prigent** [5 6]

## Abstract

Long-term adherence to continuous positive airway pressure (CPAP) therapy remains a major challenge in the management of obstructive sleep apnea, despite its well-established clinical benefits. The objective of this study is to predict long-term CPAP non-adherence one year after a baseline month that does not correspond to treatment initiation but can occur within a time window ranging from the fourth to the twenty-fourth month of therapy. We propose and compare a machine learning (ML) pipeline and a deep learning (DL) approach that leverage daily CPAP telemonitoring time series and electronic health record (EHR) variables. Both pipelines are evaluated on a held-out test set: the ML model achieved a macro F1-score of 0.83, while the DL model achieved 0.81, indicating comparable and robust predictive performance. These results suggest that CPAP usage patterns observed during an intermediate treatment phase remain highly informative for identifying patients at risk of future non-adherence and could support targeted long-term telemonitoring strategies, serving as a data-driven second opinion to assist clinicians in decision-making.

## 1. Introduction

Obstructive Sleep Apnea (OSA) is one of the most frequent chronic diseases, affecting nearly one billion people worldwide (Benjafield et al., 2019). Continuous Positive Airway Pressure (CPAP) is the first-line therapy for OSA (Epstein et al., 2009), significantly improving symptoms and quality of life if well adhered to (Lévy et al., 2015). Despite the potential benefits of CPAP therapy, adherence to treatment remains paramount. A recent study showed that overall CPAP termination rates after 1, 2, and 3 years were 23.1%, 37.1% and 47.7%, respectively (Pépin et al., 2021). Factors such as the severity of OSA, psychological considerations, and the management of CPAP-related side effects have all been linked to patient adherence to this treatment (Pépin et al., 2021).

Although several studies have focused on identifying patients at risk of poor adherence to CPAP therapy during the initial months of treatment, little is known about how to effectively telemonitor patients beyond this early phase. In clinical practice, once the first four months have passed, healthcare teams often face the challenge of determining how to continue supporting patients over time. Currently, no standardized approach exists to guide the telemonitoring of patients who remain under CPAP treatment after the initial adaptation period, particularly those who were initially predicted to be at high risk of low adherence.

The objective of this study is to develop a predictive model using longitudinal time series data derived from CPAP telemonitoring during a baseline month that is between the fourth and the 24th month of CPAP treatment, capable of estimating long-term CPAP adherence one year after. This approach aims to support clinicians in identifying patients at risk of future non-adherence and to improve long-term treatment management.

### 1.1. Related works

Based on the current scientific literature, there is no single clinical gold standard for predicting adherence to CPAP treatment in patients with OSA (Weaver & Grunstein, 2008). Instead, a combination of factors has been shown to be predictive of long-term use. Most studies focus on a restricted set of predictors, primarily the mean adherence during the first days or months of treatment (Kribbs et al., 1993; Budhiraja et al., 2007; Qiao et al., 2023) or patients' perceived treatment benefit (e.g., (Salepci et al., 2013; Weaver & Grun-

[1]Department of Computer Science, University of Fribourg, Fribourg, Switzerland [2]Haute école spécialisée de Suisse occidentale, Institute of HumanTech, Fribourg, Switzerland [3]Department of Electronics Information and Bioengineering, Politecnico di Milano, Milan, Italy [4]Scientific Institute, IRCCS "E. Medea", Bosisio Parini, Italy [5]Polyclinique Saint-Laurent, Groupe Medical de Pneumologie, Rennes, France [6]Polyclinique Saint-Laurent, Centre du sommeil, Rennes, France. Correspondence to: Benedetta Giachetti <benedett.giachett@hes-so.ch>.

*Proceedings of the Swiss AI Days 2026*, Martigny, VS & Fribourg, FR, Switzerland PMLR 309, 2026. 

stein, 2008)). Only a limited number of works (Scioscia et al., 2022; Eguchi et al., 2022; Araujo et al., 2018) explore a more integrated perspective by considering multiple patient and device-related features. However, these approaches typically rely on summary statistics such as average nightly use and focus on short-term or mid-term adherence (e.g., 3–6 months from treatment initiation). These studies have investigated CPAP adherence using data-driven approaches, either by characterizing adherence patterns in an unsupervised manner or by predicting future adherence using supervised learning models.

#### 1.1.1. Unsupervised modeling of adherence patterns

Rodrigues et al. (Rodrigues et al., 2022) proposed an unsupervised framework to characterize CPAP adherence behaviors using daily telemonitoring data of CPAP usage hours collected from a large real-world cohort of 2,381 patients. Their method relies on extracting recurrent 30-day usage patterns (snippets) from long CPAP time series and clustering them using Gaussian Mixture Models, resulting in eight clinically interpretable adherence profiles ranging from struggling to ideal usage. Importantly, the study focuses on pattern discovery and behavioral characterization, rather than predicting future adherence outcomes, and deliberately avoids aggregating time series into single summary statistics in order to preserve temporal dynamics. Similarly, Dielesen et al. (Dielesen et al., 2025) investigated early CPAP adherence trajectories using an unsupervised approach in 1,000 patients, with external validation on an independent cohort of 200 patients. Using daily CPAP usage data collected during the first 14 days of treatment, they identified distinct latent adherence patterns through growth mixture modeling, which were strongly associated with adherence status at 3 months. It demonstrated that early usage trajectories explain a large proportion of the variability in later adherences.

#### 1.1.2. Supervised prediction of CPAP adherence

Araujo et al. (Araujo et al., 2018) analyzed data from 3,588 patients, combining daily CPAP telemonitoring variables with electronic health record (EHR) data, and used the first 13 to 30 days of treatment to predict adherence at 6 months. Their models outperformed rule-based clinical baselines, demonstrating that early telemonitoring data contain predictive information well before standard follow-up visits. Eguchi et al. (Eguchi et al., 2022) conducted a single-center pilot study on 219 patients, using weekly aggregated features derived from daily CPAP data to predict adherence over a 12-week horizon. Adherence prediction was performed using logistic regression and learn-to-rank machine, achieving an F1-score of 0.864 and an AUC of 0.763. The most influential predictors identified by the models included the average daily usage duration within a week, the total weekly usage duration, the standard deviation of daily usage duration in a week, the standard deviation of daily average mask pressure in a week (from auto mode only), use of auto CPAP mode, and female sex. Scioscia et al. (Scioscia et al., 2022) investigated long-term CPAP adherence in a smaller cohort of 86 patients, using baseline clinical characteristics and follow-up CPAP parameters collected at 3, 6, and 12 months. Support vector machines achieved the best performance (68.6% accuracy and 72.9% AUC), with age, sex, mask interface, and air leakage emerging as relevant factors.

Prior work has demonstrated the feasibility of predicting CPAP adherence using early telemonitoring data and machine learning. However, existing studies either focus on short- or mid-term prediction horizons or emphasize unsupervised pattern discovery without explicit outcome prediction. In contrast, the present study targets one-year adherence prediction using daily time series data collected over an entire baseline month not corresponding to the first month of treatment, and it compares feature-based machine learning models with a deep learning approach.

## 2. Methodology

### 2.1. Data recordings and population

The e-QUALISAS study analyzed one-month de-identified CPAP telemonitoring data obtained from a single homecare provider database (ELIA Medical) at three different time points: January, June, and December 2021. The dataset included CPAP adherence expressed in hours per day, device-reported residual apnea–hypopnea index ($AHI_{PAP}$) expressed in events per hour, and the 95th percentile of unintentional leaks expressed in liters per minute.

All participants were adults aged over 18 years who initiated CPAP treatment before September 2020 and had been treated for at least four months prior to the beginning of the study period. Data were collected using the CPAP software AirSense 10 (ResMed, Australia). CPAP usage was defined as device use during a 24-hour period.

Age and sex were also available in the database. All included participants provided informed consent for data collection and anonymization. The study was registered on the Health Data Hub platform (No.F20220715144543).

### 2.2. Data preprocessing

#### 2.2.1. Data engineering

Three preprocessed datasets were used:

- **Demographics dataset**: containing patient-level demographic, clinical, and device-related information (e.g., age, sex, CPAP usage duration, mask type, device model).

- **Time series dataset**: daily CPAP telemonitoring data, including nightly adherence (hours of use), apnea–hypopnea index ($AHI_{PAP}$), and leak rate (l/min), indexed by treatment day. This dataset corresponds to CPAP usage recorded during **January 2021** and **December 2021**. It was used to extract input features from the whole month of January and adherence labels (adherence: 0 and non-adherence:1) from the whole month of December. The label adherence in December is defined as usage of the CPAP device for more than 4 hours per night for more than 70% of the nights in December (Rezaie et al., 2018).

#### 2.2.2. Cohort selection

A stepwise exclusion pipeline was applied to construct a high-quality analytical cohort. The sequence of inclusion and exclusion criteria, together with the corresponding patient counts at each stage, is illustrated in the Cohort selection panel of Figure 1 (only the steps with patient eliminations are shown). At each step, only patients satisfying all previous criteria were retained:

1. **Multi-source availability**: patients were required to be present simultaneously in the demographics, time-series, and labels datasets.
2. **Demographic validity**: patients with missing age or undefined/ambiguous sex information were excluded.
3. **Mask information completeness**: only patients with a valid CPAP mask type recorded were retained.
4. **Device homogeneity**: the analysis was restricted to patients using ResMed S10 CPAP devices to reduce device-related variability.
5. **Seniority constraint**: only patients with a CPAP usage duration of $\leq$2 years were included.
6. **Sensor plausibility checks**: patients exhibiting adherence values exceeding 24 hours were excluded as these values are physiologically implausible and indicative of sensor or recording errors.
7. **Time series completeness**: patients were required to have exactly 31 consecutive days with no missing values for adherence, $AHI_{PAP}$, or leak rate.
8. **Variance integrity**: patients whose time series showed zero variance in adherence, $AHI_{PAP}$, or leak rate were excluded to remove non-informative signals.

#### 2.2.3. Data parsing and splitting

Following cohort selection, data parsing and splitting were performed as illustrated in the corrisponding panel of Figure 1. For each patient in the final cohort, baseline daily CPAP telemonitoring features extracted (see Section 2.3.1) and time series, were used as input for machine learning (ML) and deep learning (DL), respectively, including adherence (hours of use), $AHI_{PAP}$, and leak rate. Time series features and time series were combined with static demographic and clinical variables extracted from the EHR, forming the complete input representation. The December adherence label was used as the classification target.

To enable unbiased model evaluation, the dataset was split at the patient level into a training–validation set (80%) and a held-out test set (20%) using stratified sampling based on the adherence label. This strategy preserved class proportions across splits while preventing information leakage between training and evaluation phases. The resulting partitions were subsequently used for both the machine learning and deep learning pipelines described in the following sections.

### 2.3. Machine learning

The complete processing ML pipeline is illustrated in Figure 1, in its corresponding panel.

#### 2.3.1. Features pre-processing

For each patient, January daily CPAP adherence time series were processed to extract both custom clinical and statistical features.

**Custom adherence-related features** The following five clinically motivated proportions were computed: percentage of nights with zero CPAP usage, percentage of nights with $\geq$4 hours of CPAP use, percentage of nights with $\geq$5 hours of CPAP use, percentage of nights with $AHI_{PAP} < 5$ events/hour, percentage of nights with leak rate $< 24$ L/min.

**Statistical and temporal features** For each January time serie (adherence, $AHI_{PAP}$, leak rate), ten descriptive features were extracted: mean, median, minimum, maximum, standard deviation, interquartile range, skewness, kurtosis, mean absolute day-to-day difference, linear trend (slope over time).

**Demographic and clinical features** Five EHR variables were included: age, sex, years of CPAP usage (seniority), clinical agency, mask type.

The final feature matrix combined:

- 5 custom time series features,
- 30 statistical time series features,
- 5 EHR features,

resulting in a total of 40 features per patient.

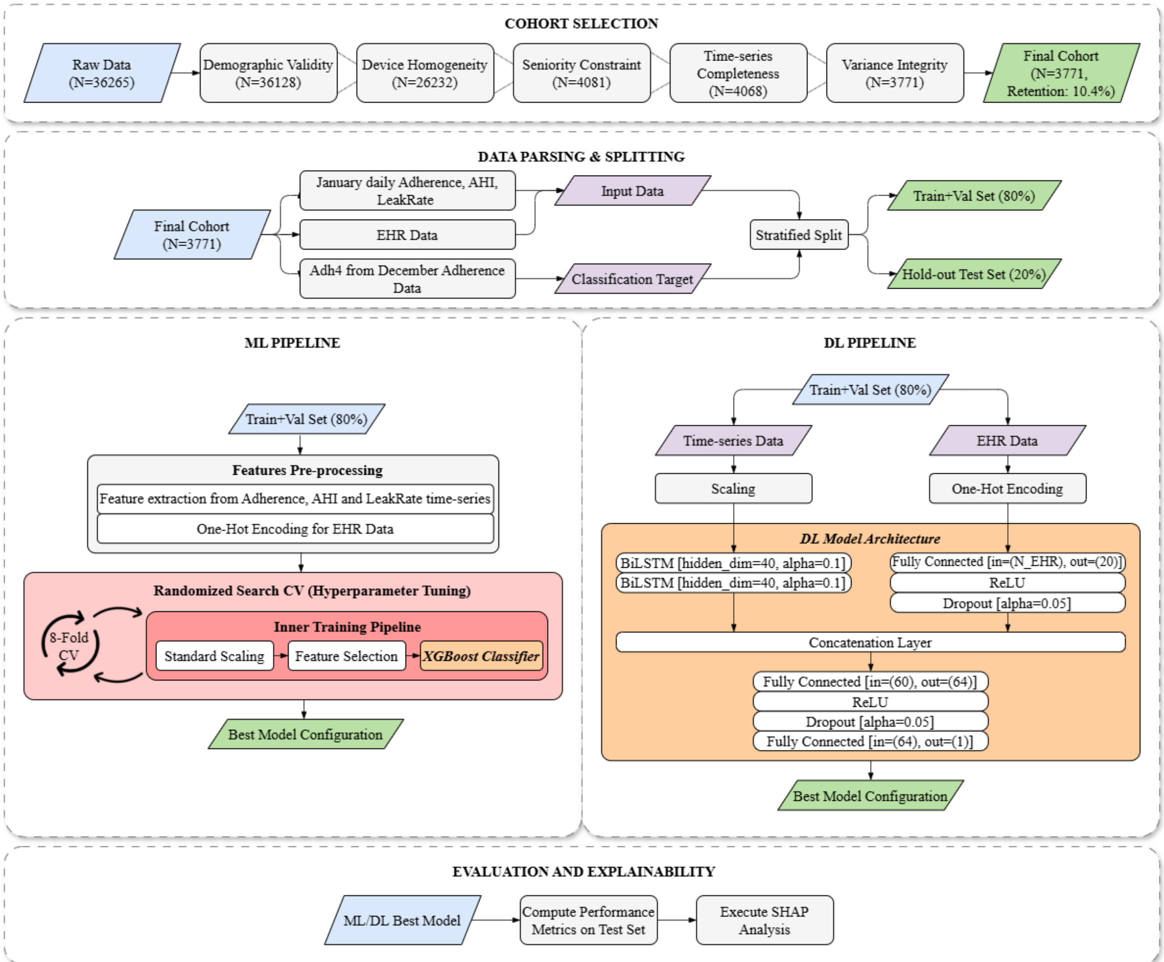


*Figure 1.* Overview of the proposed machine learning and deep learning pipeline. The figure illustrates cohort selection, feature extraction from daily CPAP time series, one-hot encoding of categorical EHR variables, model training, and post-hoc explainability analysis using SHAP.

One-hot encoding was applied to the categorical EHR variables.

#### 2.3.2. TRAINING

Numerical features were standardized using a z-score normalization.

Feature selection was embedded in the pipeline using Mutual Information (MI). This approach allows identification of features that are most informative for adherence prediction while remaining computationally efficient. The number of features selected was one of the hyperparameter fine-tuned during cross-validation.

Data were split into training (80%) and validation (20%) sets. The Extreme Gradient Boosting (XGBoost) classifier was evaluated.

Class imbalance was addressed through scale-positive-weight adjustment.

Hyperparameters were optimized via randomized search with 8-fold stratified cross-validation and a fixed number of random hyperparameter configurations (100 iterations) were sampled, using macro average F1-score as the primary optimization metric.

### 2.4. Deep learning approach

The complete processing DL pipeline is illustrated in Figure 1, in its corresponding panel.

#### 2.4.1. MODEL ARCHITECTURE

The following deep neural network architecture has been designed to integrate time series with EHR features.

The temporal component of the model is based on a two-layerbidirectional Long Short-Term Memory (BiLSTM) network, which is specifically employed to capture the sequential dynamics of daily CPAP usage, $AHI_{PAP}$, and leak rate over time, with a hidden size of 40. The final hidden states from the forward and backward LSTM directions are concatenated to form a fixed-length temporal representation.

In parallel, EHR variables are processed through a fully connected neural network with ReLU (Rectified Linear Unit) activation and dropout layer with dropout probability set to 0.05 for regularization.

The latent representations learned by the temporal and EHR data branches are subsequently concatenated and passed to a fully connected classification head. The final output layer produces a single logit representing the probability of adherence.

#### 2.4.2. Training

Within the training+validation cohort, data were further split at the patient level into training and validation sets maintaining a 7:1 ratio corresponding to 70% and 10% of the total dataset, respectively. The model was trained using binary weighted cross-entropy loss with logits to address imbalance, the Adam optimizer, and learning-rate scheduling. Early stopping based on validation macro-averaged F1-score was applied to mitigate overfitting.

### 2.5. Evaluation

The optimized ML and DL models were finally evaluated on the held-out test set.

### 2.6. Explainability

To improve model interpretability, explainability analyses were conducted using SHAP (SHapley Additive exPlanations) GradientExplainer for both the feature-based machine learning models and the deep learning architecture.

Although the proposed DL model learns high-level latent representations of CPAP usage dynamics, these internal features are inherently difficult to interpret. To address this limitation, explainability analysis was conducted at the input level, focusing on the contribution of individual days and input signals rather than on the latent representations learned by the BiLSTM. This approach enables identification of which days within the 31-day sequence had the greatest influence on the predicted probability of adherence, as well as which CPAP-derived signals were most influential on those days.

## 3. Results

### 3.1. Cohort selection

The initial cohort comprised 36,265 patients with available telemonitoring and clinical data. After excluding patients with missing demographic information, 36,128 patients remained. Restricting the analysis to users of ResMed S10 devices reduced the cohort to 26,232 patients, while applying the treatment seniority criterion resulted in 4,081 patients. Patients with incomplete time series data over the observation window (13 patients) were subsequently excluded, followed by the removal of 297 patients exhibiting zero variance in at least one monitored signal. The final cohort consisted of 3,771 patients, corresponding to an overall retention rate of 10.40%.

### 3.2. ML pipeline

Hyperparameter optimization via randomized search (100 candidates, 8-fold CV) selected a relatively shallow ensemble (193 trees, maximum depth of 4) with a very low learning rate ($\eta = 0.0038$), moderate regularization ($\alpha = 0.0014$, $\lambda = 0.4569$), and subsampling (`subsample`=0.50, `colsample_bytree`=0.8139). Feature selection based on mutual information retained $k = 25$ features out of 40 input variables.

In Figure 2, the top-ranked adherence-related features are shown. The variables *adherence_4hrs_perc* and *adherence_5hrs_perc* quantify the proportion of nights with CPAP usage exceeding 4 and 5 hours/night, respectively. Central tendency is captured by *adherence_mean* and *adherence_median*, while *adherence_min* represents extreme low-use episodes. Variability and irregularity are summarized by *adherence_std*, *adherence_iqr*, *adherence_mean_abs_diff*, and *adherence_skew*. Beyond the top 9, the MI-selected set also includes additional adherence descriptors (e.g., *adherence_null_perc*, *adherence_max*, *adherence_kurt*, *adherence_slope*) as well as respiratory and leakage statistics from the baseline window ($AHI_{PAP}$: median/min/std/iqr/skew/kurt/mean absolute difference; leak: mean/median/min/std/iqr).
The SHAP summary in Figure 2 highlights both direction and magnitude of each feature's contribution. Points to the right of zero increase the predicted probability of long-term non-adherence, whereas points to the left indicate a protective effect. For usage-related features (e.g., *adherence_4hrs_perc*, *adherence_5hrs_perc*, *adherence_mean*, *adherence_median*), high values (red) are predominantly associated with negative SHAP values, consistent with a reduced risk of non-adherence. Conversely, higher dispersion/instability (e.g., *adherence_std*, *adherence_iqr*, *adherence_mean_abs_diff*) tends to yield positive SHAP values.

The classification performance of the ML pipeline on the

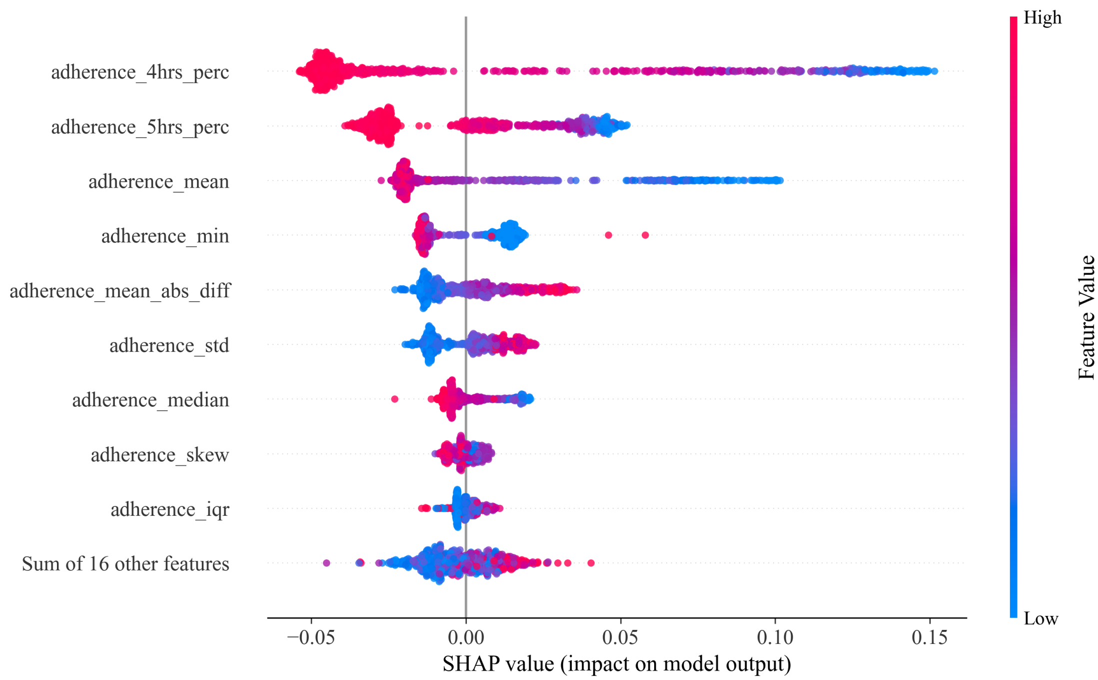


*Figure 2.* SHAP summary plot for the XGBoost-based machine learning model. Each point represents a patient, with the x-axis indicating the SHAP value (impact on the model output) and color encoding the feature value from low (blue) to high (red). Features are ranked by mean absolute SHAP value. The plot displays the top 9 features; the remaining 16 selected features are aggregated in the *Sum of 16 other features* row (total selected: 25). Higher CPAP usage (e.g., *adherence_4hrs_perc*, *adherence_5hrs_perc*, *adherence_mean*) is associated with negative SHAP values (lower predicted probability of non-adherence), whereas lower usage and higher instability (e.g., *adherence_std*, *adherence_mean_abs_diff*) shift predictions toward non-adherence.

*Table 1.* Classification performance comparison between the ML and DL pipelines on the test set.

| Model | Class | Precision | Recall | F1-score |
|---|---|---|---|---|
| ML | Adherent | 0.94 | 0.87 | 0.90 |
| | Non-Adherent | 0.68 | 0.84 | 0.75 |
| | Macro average | 0.81 | 0.85 | 0.83 |
| | Weighted average | 0.87 | 0.86 | 0.86 |
| DL | Adherent | 0.90 | 0.91 | 0.91 |
| | Non-Adherent | 0.73 | 0.72 | 0.72 |
| | Macro average | 0.82 | 0.81 | 0.81 |
| | Weighted average | 0.86 | 0.86 | 0.86 |

held-out test set is reported in Table 1. Overall, the model shows a strong and well-balanced predictive performance, with a balanced accuracy of 0.85.
For the Adherent class, the model achieves a high precision (0.94) and recall (0.87), resulting in an F1-score of 0.90. This indicates that most patients predicted as adherent are correctly classified, while only a limited fraction of true adherent patients are missed.
For the Non-Adherent class, the model attains a precision of 0.68 and a notably high recall of 0.84, yielding an F1-score of 0.75. The high recall reflects the model's ability to correctly identify the majority of patients at risk of long-term non-adherence, although at the cost of a moderate number of false positives.

### 3.3. DL pipeline

The classification performance of the DL pipeline on the test set is also summarized in Table 1. The model achieves an overall balanced accuracy of 0.81, indicating a solid but slightly lower discriminative ability compared to the ML approach.
For the Adherent class, the DL model shows a precision of 0.90 and a recall of 0.91, resulting in an F1-score of 0.91. This reflects a well-balanced performance, with both a low false-positive rate and a low false-negative rate for adherent patients.
For the Non-Adherent class, the model reaches a precision of 0.73 and a recall of 0.72, corresponding to an F1-score of 0.72. Compared to the ML model, the DL approach exhibits a more symmetric trade-off between precision and recall for non-adherence, but with a reduced sensitivity to high-risk patients. The macro-averaged F1-score (0.81) and

weighted F1-score (0.86) confirm consistent performance across classes, with slightly lower class-balanced effectiveness relative to the ML pipeline.
Overall, the DL model provides robust adherence predictions, with strong performance in identifying adherent patients, while offering a more conservative detection of non-adherence compared to the classical ML approach.

Figure 3 illustrates the temporal feature importance obtained from the SHAP analysis applied to the DL model. The heatmap reports the mean absolute SHAP values across the 31-day observation window for each input signal, highlighting the contribution of the day-to-day temporal data to the final prediction.
The results indicate that daily CPAP adherence is by far the most influential input signal across the entire temporal window, whereas $AHI_{PAP}$ and leak rate exhibit consistently negligible contributions. From a temporal perspective, the contribution of adherence is not uniform across the observation period. Higher importance is observed during the final portion of the monitoring window, with pronounced peaks around treatment days 21, 23, 24, 30. In contrast, the earliest days of treatment show comparatively lower influence on the model's output.

## 4. Discussion

This study investigated the prediction of long-term CPAP adherence using telemonitoring data collected during January 2021. By integrating demographic and clinical data with daily CPAP-derived time series, we developed and evaluated both feature-based machine learning models and a deep learning approach designed to capture temporal adherence dynamics.

A result emerging from this study is that the predictive performance of feature-based machine learning models was comparable to that achieved by the deep learning architecture. Both approaches achieved comparable macro F1 scores and class-wise performance on the held-out test set. This finding suggests that the handcrafted features extracted from CPAP telemonitoring data, when combined with appropriate feature selection, already capture most of the information required to predict long-term non-adherence. In this context, well-designed feature extraction and selection pipelines provide a favorable trade-off between predictive performance, computational efficiency, and interpretability. The macro-averaged F1-score (0.83) confirms a balanced performance across classes, while the weighted F1-score (0.86) reflects the strong overall predictive ability when accounting for class prevalence.
The ML model provides reliable adherence risk stratification, with particularly strong sensitivity to non-adherent patients.
The SHAP analysis of the ML model further supports the interpretation of clear behavioral differences between adherent and non-adherent patients.
Patients predicted as non-adherent are characterized by lower average nightly usage, a reduced proportion of nights exceeding clinically relevant thresholds (4–5 hours), and increased day-to-day variability, as reflected by higher standard deviation and mean absolute differences in adherence. This suggests that irregular nightly use and abrupt day-to-day changes are risk factors for poor long-term treatment persistence. In contrast, adherent patients tend to exhibit more stable and consistent usage patterns in the treatment course.
Regular and consistent CPAP use may indicate patients with better sleep routines and more structured habits. These traits not only facilitate better CPAP adherence but also extend adherence to other treatments and medications. It is well established from both cardiovascular (Granger et al., 2005) and non-cardiovascular (Curtis et al., 2011) clinical trials that adherence per se, including adherence to placebo, is associated with markedly improved health outcomes, an effect that is often substantially larger than that of active therapy. On the other hand, irregular CPAP use may reflect poor sleep routine, frequent awakenings, or irregular sleep schedules, which increase opportunities to discontinue the device. Frequent disruptions during sleep may make it harder for patients to adapt to CPAP therapy, further reinforcing irregular patterns of use (Somiah et al., 2012).

While the DL model is inherently more computationally demanding and less transparent, it offers the advantage of modeling day-to-day temporal dependencies that are not explicitly accessible to feature-based approaches, which instead summarize the entire observation window into aggregate descriptors. This analysis revealed that the importance of daily CPAP adherence is not uniformly distributed across time, with specific final days exhibiting disproportionately higher contributions to the model output. The contribution of adherence appears to stabilize or slightly decrease toward the end of the observation period. One possible interpretation is that the BiLSTM progressively saturates while moving backwards as sufficient behavioral information is accumulated.

Consistent with the ML explainability results, the DL SHAP analysis also indicates negligible contributions from $AHI_{PAP}$ and leak rate across the entire temporal window suggesting that the prediction is primarily driven by the autoregressive nature of the adherence signal, where behavior acts as the strongest proxy for long-term adherence.

In a previous study (**?**) based on the same database, the objective was primarily descriptive and stratificative. The methodology relied on the analysis of 2 features only extracted from the same baseline month of CPAP usage:

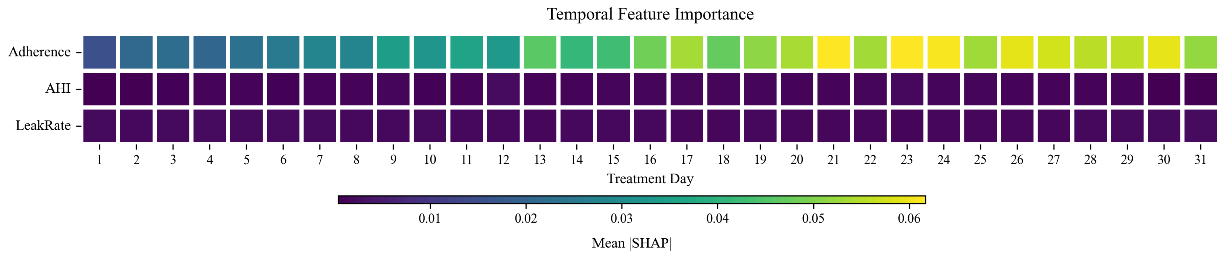


*Figure 3.* Temporal feature importance obtained from SHAP analysis applied to the BiLSTM-based deep learning model. The heatmap shows the mean absolute SHAP values for each input signal (Adherence, $AHI_{PAP}$, and LeakRate) across the 31-day observation window. Warmer colors indicate a stronger contribution to the predicted probability of long-term non-adherence.

monthly usage mean and standard deviation. Patients were grouped using thresholds (4 h/night for mean and 1.76 h for standard deviation), resulting in four clinically interpretable adherence profiles. Descriptive statistical analyses were then used to compare these groups at 6 and 12 months, without building a predictive model in the strict sense. The approach was transparent, and closely aligned with clinical reasoning, aiming to support patient monitoring prioritization rather than individual-level prediction.

Both studies confirm the importance of the monthly mean and standard deviation of CPAP usage. A high monthly usage mean is a well-established predictor of long-term adherence (Sabil et al., 2021; Budhiraja et al., 2007; Van Ryswyk et al., 2019), while usage variability has only been recognized as a relevant factor in more recent studies, (Eguchi et al., 2022) found that standard deviation of daily usage duration in a week possibly correlated to poor CPAP adherence. The present study further extends these findings by showing that additional features characterizing monthly usage patterns also contribute meaningfully to adherence prediction, even if those features have a less transparent explanation. Eguchi et al.(Eguchi et al., 2022) adopted a ML approach for early detection of non-adherence, using CPAP data from the first month of treatment to predict mid-term adherence at 6 months. Their ML pipeline which incorporated the XGB model and EHR variables, achieved F1, F1(-), precision, precision(-), sensitivity, and specificity values of 0.85, 0.61, 0.83, 0.65, 0.87, and 0.57, respectively. In contrast, the present work focuses on adherence classification at a later stage of device use. Our model achieves comparable, and in some metrics superior, performance, with a weighted F1-score of 0.86, an F1-score of 0.75 for the non-adherent class, a weighted precision of 0.87, and a precision of 0.68 for non-adherent patients. These results extend the findings of Eguchi et al., demonstrating that CPAP usage patterns remain informative for adherence prediction even at later stages of treatment, beyond the early initiation period.

The proposed predictive framework is well aligned with real-world CPAP telemonitoring workflows and has the potential to support clinical decision-making in routine practice. By leveraging data collected during the treatment, the model enables identification of patients at high risk of long-term non-adherence, well before treatment discontinuation becomes clinically evident. It could be integrated into existing telemonitoring platforms to prioritize clinical attention and tailor follow-up strategies. Rather than increasing the overall clinical workload, the model could help clinicians focus their efforts on a subset of patients who are most likely to benefit from targeted interventions, such as adherence counseling, mask refitting, or behavioral support.

## 5. Conclusions

In this work, we studied whether CPAP telemonitoring data can be used to predict long-term treatment adherence by comparing a feature-based machine learning approach with a deep learning model. The two approaches achieved very similar predictive performance, showing that simple and well-designed features extracted from the baseline time window already contain most of the information needed to identify patients at risk of non-adherence.

Explainability analyses consistently showed that nightly CPAP usage patterns are the main drivers of prediction. Both machine learning and deep learning models relied primarily on usage level and variability, while $AHI_{PAP}$ and leak rate played a negligible role.

From a practical point of view, these results indicate that feature-based machine learning models offer a good balance between performance, interpretability, and computational efficiency, making them well suited for deployment in clinical telemonitoring systems.

## Limitations

One limitation of this study is the lack of data regarding patients' initial clinical characteristics, including baseline apnea–hypopnea index (AHI), disease severity, sleepiness, psychological considerations, and body mass index (BMI). Additionally, CPAP device settings, such as applied pressure, which may have heterogeneous effects on treatment adherence, were not available in the dataset.

A further limitation is that the baseline month considered in this analysis corresponds to a treatment period between the third and the 24th month of CPAP therapy. Consequently, the adherence patterns identified in this study may not generalize to patients with CPAP treatment seniority outside this time window. As the analysis focuses on a specific stage of therapy, model performance and applicability may differ in patients at earlier or more advanced stages of treatment. Future studies should therefore validate these findings across broader treatment durations to improve external validity.

Although the CPAP initiation date was available, it remains unclear whether some patients had previously been treated with CPAP and discontinued therapy before restarting, which may have influenced adherence behavior.

Another limitation concerns the preprocessing strategy adopted to address missing data. To avoid biased interpretation, patients with missing data in at least one of the considered periods were excluded from the analysis. Potential causes of missing data include technical transmission failures, resulting in the absence of usage information, or complete discontinuation of CPAP use by the patient.

## Ethical Statement

All experiments were performed on a local workstation equipped with an Intel Core i7 processor and an NVIDIA GeForce GTX 1650 Ti GPU. Consequently, the energy consumption and associated CO2 emissions of this study are minimal when compared to those of large-scale deep learning training workloads.
The development of this work was based exclusively on open-source software tools, including Python, PyTorch, SciPy, and scikit-learn, all used in accordance with their respective licensing terms. In addition, Large Language Models (ChatGPT, GPT-5.2, OpenAI) were employed solely to improve the linguistic clarity and readability of the manuscript. All scientific content, data analyses, and interpretations were produced by the authors, who carefully reviewed and validated all AI-assisted outputs.

## References

Araujo, M., Bhojwani, R., Srivastava, J., Kazaglis, L., and Iber, C. ML Approach for Early Detection of Sleep Apnea Treatment Abandonment: A Case Study. In *Proceedings of the 2018 International Conference on Digital Health*, pp. 75–79, Lyon France, April 2018. ACM. ISBN 978-1-4503-6493-5. doi: 10.1145/3194658.3194681.

Benjafield, A. V., Ayas, N. T., Eastwood, P. R., Heinzer, R., Ip, M. S. M., Morrell, M. J., Nunez, C. M., Patel, S. R., Penzel, T., Pépin, J.-L., Peppard, P. E., Sinha, S., Tufik, S., Valentine, K., and Malhotra, A. Estimation of the global prevalence and burden of obstructive sleep apnoea: A literature-based analysis. *The Lancet Respiratory Medicine*, 7(8):687–698, August 2019. ISSN 22132600. doi: 10.1016/S2213-2600(19)30198-5.

Budhiraja, R., Parthasarathy, S., Drake, C. L., Roth, T., Sharief, I., Budhiraja, P., Saunders, V., and Hudgel, D. W. Early CPAP use identifies subsequent adherence to CPAP therapy. *Sleep*, 30(3):320–324, March 2007. ISSN 0161-8105.

Curtis, J. R., Larson, J. C., Delzell, E., Brookhart, M. A., Cadarette, S. M., Chlebowski, R., Judd, S., Safford, M., Solomon, D. H., and Lacroix, A. Z. Placebo adherence, clinical outcomes, and mortality in the women's health initiative randomized hormone therapy trials. *Medical Care*, 49(5):427–435, May 2011. ISSN 1537-1948. doi: 10.1097/MLR.0b013e318207ed9e.

Dielesen, J., Ledwaba-Chapman, L. J., Kasetti, P., Husain, N. F., Skinner, T. C., Pengo, M. F., Whiteman, T., Asimakopoulou, K., Merritt, S., Jones, D., Dickel, P., Pulakal, S., Ward, N. R., Pepperell, J., Steier, J., and Sathyapala, S. A. Six early CPAP-usage behavioural patterns determine peak CPAP adherence and permit tailored intervention, in patients with obstructive sleep apnoea. *Thorax*, 80(5):300–308, April 2025. ISSN 1468-3296. doi: 10.1136/thorax-2024-221763.

Eguchi, K., Yabuuchi, T., Nambu, M., Takeyama, H., Azuma, S., Chin, K., and Kuroda, T. Investigation on factors related to poor CPAP adherence using machine learning: A pilot study. *Scientific Reports*, 12 (1):19563, November 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-21932-8.

Epstein, L. J., Kristo, D., Strollo, P. J., Friedman, N., Malhotra, A., Patil, S. P., Ramar, K., Rogers, R., Schwab, R. J., Weaver, E. M., Weinstein, M. D., and Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 5(3):263–276, June 2009. ISSN 1550-9389.

Granger, B. B., Swedberg, K., Ekman, I., Granger, C. B., Olofsson, B., McMurray, J. J. V., Yusuf, S., Michelson, E. L., Pfeffer, M. A., and CHARM investigators. Adherence to candesartan and placebo and outcomes in chronic heart failure in the CHARM programme: Double-blind, randomised, controlled clinical trial. *Lancet*, 366(9502):

2005–2011, December 2005. ISSN 1474-547X. doi: 10.1016/S0140-6736(05)67760-4.

Kribbs, N. B., Pack, A. I., Kline, L. R., Smith, P. L., Schwartz, A. R., Schubert, N. M., Redline, S., Henry, J. N., Getsy, J. E., and Dinges, D. F. Objective Measurement of Patterns of Nasal CPAP Use by Patients with Obstructive Sleep Apnea. *American Review of Respiratory Disease*, 147(4):887–895, April 1993. ISSN 0003-0805. doi: 10.1164/ajrccm/147.4.887.

Lévy, P., Kohler, M., McNicholas, W. T., Barbé, F., McEvoy, R. D., Somers, V. K., Lavie, L., and Pépin, J.-L. Obstructive sleep apnoea syndrome. *Nature Reviews Disease Primers*, 1(1):15015, June 2015. ISSN 2056-676X. doi: 10.1038/nrdp.2015.15.

Pépin, J.-L., Bailly, S., Rinder, P., Adler, D., Szeftel, D., Malhotra, A., Cistulli, P. A., Benjafield, A., Lavergne, F., Josseran, A., Tamisier, R., Hornus, P., and On Behalf Of The medXcloud Group, n. CPAP Therapy Termination Rates by OSA Phenotype: A French Nationwide Database Analysis. *Journal of Clinical Medicine*, 10 (5):936, March 2021. ISSN 2077-0383. doi: 10.3390/jcm10050936.

Qiao, M., Xie, Y., Wolff, A., and Kwon, J. Long term adherence to continuous positive Airway pressure in mild obstructive sleep apnea. *BMC Pulmonary Medicine*, 23 (1):320, September 2023. ISSN 1471-2466. doi: 10.1186/s12890-023-02612-3.

Rezaie, L., Phillips, D., and Khazaie, H. Barriers to acceptance and adherence to continuous positive airway pressure therapy in patients with obstructive sleep apnea: A report from Kermanshah province, western Iran. *Patient Preference and Adherence*, 12:1299–1304, 2018. ISSN 1177-889X. doi: 10.2147/PPA.S165905.

Rodrigues, J. F., Bailly, S., Pepin, J.-L., Goeuriot, L., Spadon, G., and Amer-Yahia, S. CPAP Adherence Assessment via Gaussian Mixture Modeling of Telemonitored Apnea Therapy. *Applied Sciences*, 12(15):7618, July 2022. ISSN 2076-3417. doi: 10.3390/app12157618.

Sabil, A., Le Vaillant, M., Stitt, C., Goupil, F., Pigeanne, T., Leclair-Visonneau, L., Masson, P., Bizieux-Thaminy, A., Humeau, M.-P., Meslier, N., and Gagnadoux, F. A CPAP data-based algorithm for automatic early prediction of therapy adherence. *Sleep & Breathing = Schlaf & Atmung*, 25(2):957–962, June 2021. ISSN 1522-1709. doi: 10.1007/s11325-020-02186-y.

Salepci, B., Caglayan, B., Kiral, N., Parmaksiz, E. T., Comert, S. S., Sarac, G., Fidan, A., and Gungor, G. A. CPAP Adherence of Patients With Obstructive Sleep Apnea. *Respiratory Care*, 58(9):1467–1473, September 2013. ISSN 0020-1324. doi: 10.4187/respcare.02139.

Scioscia, G., Tondo, P., Foschino Barbaro, M. P., Sabato, R., Gallo, C., Maci, F., and Lacedonia, D. Machine learning-based prediction of adherence to continuous positive airway pressure (CPAP) in obstructive sleep apnea (OSA). *Informatics for Health & Social Care*, 47(3):274–282, July 2022. ISSN 1753-8165. doi: 10.1080/17538157.2021.1990300.

Somiah, M., Taxin, Z., Keating, J., Mooney, A. M., Norman, R. G., Rapoport, D. M., and Ayappa, I. Sleep Quality, Short-Term and Long-Term CPAP Adherence. *Journal of Clinical Sleep Medicine*, 8(5):489–500, October 2012. ISSN 1550-9389, 1550-9397. doi: 10.5664/jcsm.2138.

Van Ryswyk, E., Anderson, C. S., Antic, N. A., Barbe, F., Bittencourt, L., Freed, R., Heeley, E., Liu, Z., Loffler, K. A., Lorenzi-Filho, G., Luo, Y., Margalef, M. J. M., McEvoy, R. D., Mediano, O., Mukherjee, S., Ou, Q., Woodman, R., Zhang, X., and Chai-Coetzer, C. L. Predictors of long-term adherence to continuous positive airway pressure in patients with obstructive sleep apnea and cardiovascular disease. *Sleep*, 42(10):zsz152, October 2019. ISSN 1550-9109. doi: 10.1093/sleep/zsz152.

Weaver, T. E. and Grunstein, R. R. Adherence to Continuous Positive Airway Pressure Therapy: The Challenge to Effective Treatment. *Proceedings of the American Thoracic Society*, 5(2):173–178, February 2008. ISSN 1546-3222, 1943-5665. doi: 10.1513/pats.200708-119MG.