Explainable Profiling of Sleep Disorders to Support Trustworthy Clinical Interventions

Sifeddine Sellami*, Juba Agoun†, Lamia Yessad‡, Auday Berro§, Louenas Bounia¶
* Université Lumière Lyon 2, Laboratoire ERIC, Lyon, France
ks_sellami@esi.dz
†Université Lumière Lyon 2, Laboratoire ERIC, Lyon, France
juba.agoun1@univ-lyon2.fr
‡École Nationale Supérieure d'Informatique (ESI), Alger, Algérie
l_yessad@esi.dz
§Université de Lyon 1, LIRIS, UMR CNRS 5205, France
audayberro@gmail.com
¶Université Sorbonne Paris Nord, LIPN, UMR CNRS 7030, France
louenas.bounia@univ-paris13.fr

***Abstract*—Sleep disorders present significant diagnostic challenges due to symptom heterogeneity, profoundly impacting public health. To address this, we leverage real-world data from the KANOPEE mobile application—a validated digital tool for managing sleep complaints—to develop a trustworthy model for disorder detection and patient profiling using explainable artificial intelligence (XAI). Guided by clinical expertise, our analysis focuses on key variables including demographic characteristics, clinical assessment scores (e.g., Insomnia Severity Index, anxiety, and depression scores), and physiological metrics. Our dataset comprises approximately *1k* complete patient records. Our dual methodological approach involves: (1) regression models to predict treatment success metrics, enabling the identification of modifiable behavioral factors influencing mental health and sleep outcomes; and (2) k-means clustering to segment patients into three or more distinct profiles, differentiating subtypes of sleepers (*e.g.,* young professionals with mild insomnia and anxiety, older individuals with regular sleep and low depression, and middle-aged patients with high daytime sleepiness and irregular sleep patterns). XAI techniques, including SHAP and formal explanations, elucidate key drivers like sleep regularity and anxiety, validated against clinical thresholds and expert feedback from sleep specialists. This framework aligns algorithmic insights with medical reasoning, enhancing interpretability and supporting personalized interventions in digital sleep medicine. Future extensions will incorporate application completion rates for broader efficacy assessment.**

***Index Terms*—Sleep Disorder Profiling, Explainable AI (XAI), K-means Clustering, Inter-Cluster Statistical Analysis, Real-World Data Analysis**

I. Introduction

Sleep disorders, a pressing public health challenge, impair patients' health, well-being, and daily functioning [1]. Their diagnosis is complex due to diverse symptoms and the lack of standardized pathways, leading to frequent underdiagnosis or inadequate treatment [3], [17]. The advent of digital health tools, such as the KANOPEE smartphone application, has generated extensive real-world data, offering new opportunities for analysis [4], [5]. Yet, these raw, often unlabeled datasets pose challenges for direct medical use, necessitating advanced data mining and explainable artificial intelligence (XAI) techniques to extract interpretable patterns for clinicians.

In this study, we analyze approximately $\approx$ 58,000 sleep-related records from KANOPEE, refined to 933 complete instances after rigorous preprocessing, focusing on clinically relevant variables (e.g., age, gender, ISI, ESS, ANX, DEP, sleep duration, and regularity). Collaborating with sleep medicine specialists, we apply data mining and unsupervised learning to uncover patterns, validated through coherent clinical interpretations. Our primary contribution is a systematic XAI framework that extends expert insights, enhancing trust in AI-driven sleep disorder detection.

A key challenge is bridging algorithmic data analysis with clinical reasoning. It lies in creating effective interfaces between two distinct forms of information processing: the algorithmic approaches used by AI systems to extract statistical regularities from data, and the intuitive, experiential processes through which humans naturally interpret and make sense of information [11]. Interpretability and explainability are the fundamental concepts involved in addressing this challenge and building the first step of trustworthy AI [2]. While interpretability remains focused on bridging the gap between what machines discover and what humans can meaningfully comprehend, explainability involves providing humans with meaningful information about the computational logic involved, as well as the significance of specific features that contributed to the final decision outcome.

We focus on two approaches: (1) regression models to predict treatment outcomes, such as improvements in insomnia and depression, identifying behavioral factors like sleep regularity; and (2) K-means clustering to segment patients into three or more profiles (e.g., young professionals with mild insomnia vs. older adults with stable sleep). These methods reveal clinically meaningful patterns, validated by experts, enhancing diagnostic and intervention strategies.

This paper is organized as follows. Section II introduces the

2375-9259/25/$31.00 ©2025 IEEE
DOI 10.1109/ICDMW69685.2025.00275

KANOPEE dataset used for analysis. Section III describes our regression-based approach with SHAP interpretation for predicting clinical MO. Section IV details the K-means clustering methodology and patient profile characterization. Section V presents experimental results, and Section VI reports initial expert feedback. Section VII outlines the refined clustering analysis incorporating expert recommendations. Section VIII reviews related work, positioning our contributions. Section IX summarizes findings and discusses future directions.

II. KANOPEE Dataset

This study analyzes data collected by Philip et al. [5] using *KANOPEE*, a smartphone application designed to reduce sleep complaints over a 17-day period through repeated interactions with a virtual assistant. KANOPEE is France's first virtual companion app dedicated to the follow-up of patients with sleep disorders [1]. The application was developed and clinically validated by Professor Pierre Philip's team at *CHU de Bordeaux* and *CNRS SANPSY unit*.

KANOPEE provides digital support to patients suffering from sleep disorders through two virtual assistants (*Louise* and *Jeanne*). These agents conduct personalized assessments during three scheduled visits: **Day 0 - Visit 0 (D0-V0)**, **Day 7 - Visit 1 (D7-V1)**, and **Day 17 - Visit 2 (D17-V2)**[2]. The questionnaires administered during these visits generate validated clinical indicators, including the **PHQ-9** (Patient Health Questionnaire-9 for depression assessment) and the **ISI** (Insomnia Severity Index).

The dataset encompasses information from three key interactions [5]:

- **Screening Interview (V0)** on **Day 0 (D0)**: Initial assessment where users complete the Insomnia Severity Index (ISI) and provide baseline information.
- **Follow-up Interview (V1)** on **Day 7 (D7)**: Mid-point evaluation including a summary of sleep diary entries, second ISI assessment, and personalized sleep recommendations.
- **Final Interview (V2)** on **Day 17 (D17)**: Concluding assessment where users report adherence to recommendations and complete the ISI again. Based on the final ISI score, users are either advised to continue using KANOPEE autonomously (ISI ≤ 21) or referred to a sleep specialist (ISI > 21).

To facilitate analysis, the dataset variables are organized into three categories:

1) **Baseline demographics and clinical data**: age, gender, BMI, geographic location (department, region), socio-professional category, education level, NOSAS score (sleep apnea risk assessment), presence of OSA (Obstructive Sleep Apnea) or RLS (Restless Legs Syndrome).
2) **Longitudinal measurements across visits (D0, D7, D17)**: standardized scores for anxiety (ANX), depression (DEP), insomnia (ISI), excessive daytime sleepiness (ESS), and corresponding visit timestamps.
3) **Derived sleep parameters**: statistical measures (means and standard deviations) of sleep duration, bedtime, and wake time calculated for two periods: D0→D7 and D7→D17.

A significant challenge in this dataset was the high rate of missing values, as many users did not complete all three visits. To ensure analytical reliability in the medical context, only complete records with data from all three visits were retained. This stringent filtering resulted in a final dataset of **145 patients** with **37 features**, providing a robust foundation for our clustering analysis.

III. Regression-Based Approach with SHAP Interpretation

Our first methodological approach aims to identify key variables that explain inter-patient variations in sleep duration. The objective is to determine which factors significantly impact sleep duration by identifying variables that may increase or decrease this critical sleep parameter, thereby providing insights into sleep disorder mechanisms. We developed a regression-based predictive model and employed Explainable AI (XAI) techniques to interpret the results using a representative set of patient instances. The methodology, illustrated in Figure 1, comprises four sequential steps.

A. Dataset Preparation and Target Variable Selection

We selected average sleep duration as the target variable for our regression models. Given that the original dataset contains sleep duration measurements across two distinct temporal periods (D0–D7 and D7–D17)[3], we constructed two independent datasets by excluding the sleep duration column corresponding to the alternative period. This temporal separation enables the training of period-specific models, allowing us to capture potential variations in sleep patterns and influencing factors across different phases of the intervention.

B. Model Training and Performance Evaluation

Following dataset preparation, categorical variables were encoded using label encoding to ensure compatibility with machine learning algorithms. Four regression algorithms were evaluated : Linear Regression, Decision Tree, Random Forest, and XGBoost. Each model was trained using an 80%–20% train-test split strategy. Model performance was assessed using two complementary metrics : Mean Squared Error (MSE) and Mean Absolute Error in minutes (MAE_{min}). The MAE in minutes is formally defined as :

$$\text{MAE}_{\text{min}} = \frac{1}{N} \sum_{i=1}^{N} |(y_i - \hat{y}_i) \times 60|$$

[1]CHU de Bordeaux: Launch of Kanopée, the 1st virtual companion application to help with sleep issues, addiction, and stress related to lockdown, accessed: 2025-05-23.

[2]D: Day since registration, V: Visit number

[3]D0–D7 and D7–D17 refer to Days 0–7 and Days 7–17 post-registration, respectively.

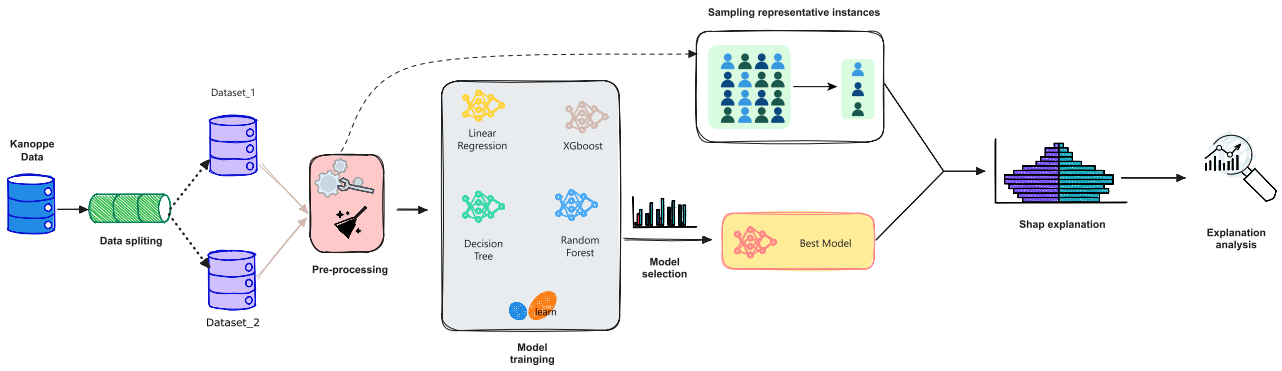

Fig. 1. Methodology for identifying variables influencing sleep duration using predictive models and SHAP interpretation.

where y_i represents the true sleep duration, $\hat{y}_i$ is the predicted value for the i-th instance, N denotes the total number of test instances, and the factor 60 converts the error from hours to minutes for clinical interpretability. The best-performing model was selected for subsequent explanation analysis.

C. Representative Instance Selection and SHAP Analysis

To identify the most influential variables affecting sleep duration, we selected six representative instances from each dataset : two patients with average sleep duration, two with notably low values, and two with high values. These instances were systematically excluded from the training datasets to prevent data leakage. Using the optimally performing model, we generated predictions for these representative cases and applied SHapley Additive exPlanations (SHAP) to decompose each prediction into individual feature contributions, thereby identifying variables that most significantly influenced the predicted sleep duration for each patient profile.

D. Interpretation and Clinical Insights

SHAP values and summary plots were analyzed to derive clinical insights regarding factors that positively or negatively impact sleep duration. This analysis provides a quantitative foundation for understanding which patient characteristics and clinical measurements are most predictive of sleep duration variations.

E. Limitations and Motivation for Clustering Approach

While this regression-based approach successfully identifies key factors influencing sleep duration at a population level, it inherently assumes homogeneous relationships between variables across all patients. However, sleep disorders manifest heterogeneously across patient populations due to variations in lifestyle, comorbidities, and sociodemographic characteristics. This limitation motivated the development of our complementary clustering approach, detailed in Section IV, which segments patients into distinct subgroups to enable more personalized and clinically relevant interpretations that better capture population heterogeneity.

IV. Clustering and Characterization for Patient Profile Construction

To construct patient profiles based on their sleep patterns, we employed the K-means clustering algorithm to group 145 patient instances. The optimal number of clusters ($K = 2$) was determined by maximizing the silhouette score, which evaluates clustering quality by measuring intra-cluster cohesion and inter-cluster separation. Principal Component Analysis (PCA) further validated this choice by visually confirming the clear separability of the two clusters.

To characterize the clusters and identify the most influential variables, we applied three complementary methods:

- Statistical comparison tests between clusters to highlight significant differences [6].
- A distance-based analysis (*in-pattern/out-pattern*) to assess cluster distinctiveness [7].
- An explainable artificial intelligence (XAI) approach to interpret cluster boundaries, adapted from [8].

Figure 2 provides an overview of our methodological approach, with details described in the subsequent subsections.

a) K-means Clustering.: Prior to applying the K-means algorithm, we preprocessed the data by encoding categorical variables using label encoding and standardizing numerical features. Standardization was performed using the formula $x' = \frac{x-\mu}{\sigma}$, where μ is the feature mean and σ is the standard deviation. This ensured all features contributed equally to the clustering process.

The K-means algorithm ($K = 2$) successfully partitioned the patients into two distinct groups. Subsequently, the three complementary methods listed above were applied to characterize the clusters and identify the key variables driving the separation of each group.

A. Comparative Statistical Tests

This characterization method, adapted from [6], systematically identifies variables that significantly differentiate between patient clusters. For numerical variables, we calculate cluster-specific means and apply appropriate statistical tests to evaluate inter-cluster differences. The testing procedure begins with an *F-test* to assess variance homogeneity between

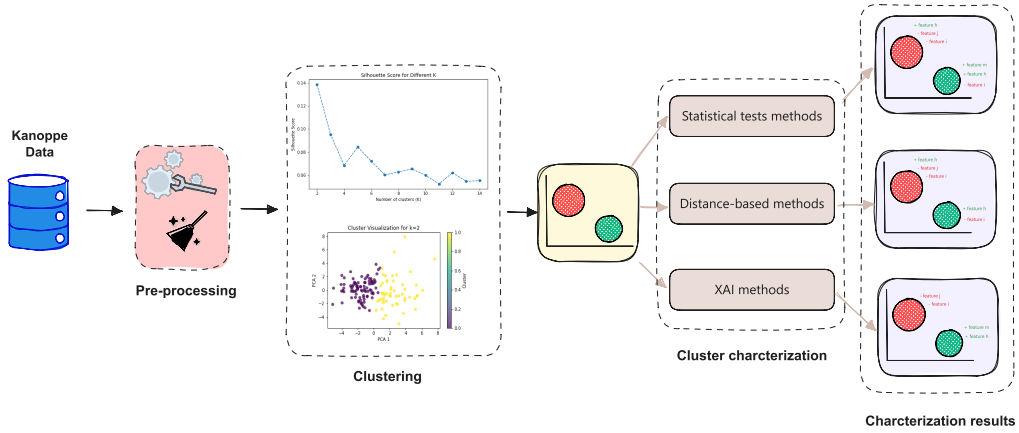

Fig. 2. Illustrative diagram of the patient profile construction process.

clusters. Based on this result, we apply either **Student's** t-test (equal variances) or **Welch** t-test (unequal variances). Variables achieving $p \leq 0.05$ are retained as cluster-discriminating characteristics. For categorical variables, we analyze category-specific proportions within each cluster using contingency table methods. Binary variables are evaluated using **Fisher's exact** test, while multi-category variables employ the **Chi-square** test. Again, only variables with $p \leq 0.05$ are considered statistically significant.

We chose the p-value threshold of 0.01 to adopt a more cautious approach, as recommended in [26]. Lowering the significance level from the conventional 0.05 to 0.01 or 0.001 helps reduce false positives and enhances the robustness of findings.

Following variable identification[4], we interpret cluster-specific characteristics through targeted analysis. For standardized clinical measures (`ANX, DEP, ISI, ESS, BMI`), we reference established normative ranges to provide absolute clinical interpretations. For remaining numerical variables, we conduct relative inter-cluster comparisons to identify directional differences. Categorical variables are analyzed by examining within-cluster and between-cluster percentage distributions. This systematic approach enables comprehensive characterization of each patient subgroup.

B. In-pattern/Out-pattern Method

The In-pattern/Out-pattern methodology, adapted from [7], segments cluster observations into core and peripheral subgroups to identify distinctive variables and their behavioral patterns within each cluster. This distance-based approach reveals whether variables exhibit characteristically high or low values for each patient group. The method follows a systematic six-step procedure:

1) **Centroid computation**: Calculate X_k as the geometric center (mean) of all instances in cluster k.
2) **Distance-based classification**: For each instance i, compute its Euclidean distance d_i to centroid X_k. Instances with $d_i \in [\mu_k \pm z \cdot \sigma_k]$ are classified as *in-pattern* (core cluster members), while those outside this range are labeled *out-pattern* (peripheral members), where μ_k and σ_k represent the mean and standard deviation of distances in cluster k, and z is a user-defined threshold parameter.
3) **Subgroup characterization**: Compute means $\mu_{\text{in}}(k, v)$ and $\mu_{\text{out}}(k, v)$ for each cluster k and variable v across in-pattern and out-pattern subgroups.
4) **Differentiation factor computation**: Calculate the relative difference between subgroup means for each variable v in cluster k:
$$df(k, v) = \frac{\mu_{\text{in}}(k, v) - \mu_{\text{out}}(k, v)}{\mu_{\text{out}}(k, v)}$$
This factor quantifies how distinctively a variable behaves within the cluster core compared to its periphery.
5) **Statistical characterization**: Calculate descriptive statistics (μ, σ) for each variable v within cluster k to understand its overall distribution pattern.
6) **Salient variable identification**: A variable v is deemed salient in cluster k when its differentiation factor falls outside the expected range:
$$df(k, v) \leq \mu_{df}(k) - z\sigma_{df}(k)$$
$$\text{or} \quad df(k, v) \geq \mu_{df}(k) + z\sigma_{df}(k)$$
Positive factors indicate that variable v exhibits characteristically high values in cluster k, while negative factors signify predominantly low values.

This distance-based analysis identifies cluster-specific characteristic variables and their directional tendencies, providing interpretable insights into each patient group's defining features.

C. XAI-based Cluster Interpretation

Following the framework established by [8], this method employs Explainable AI techniques to analyze cluster struc-

[4]Variables showing significant differences between clusters

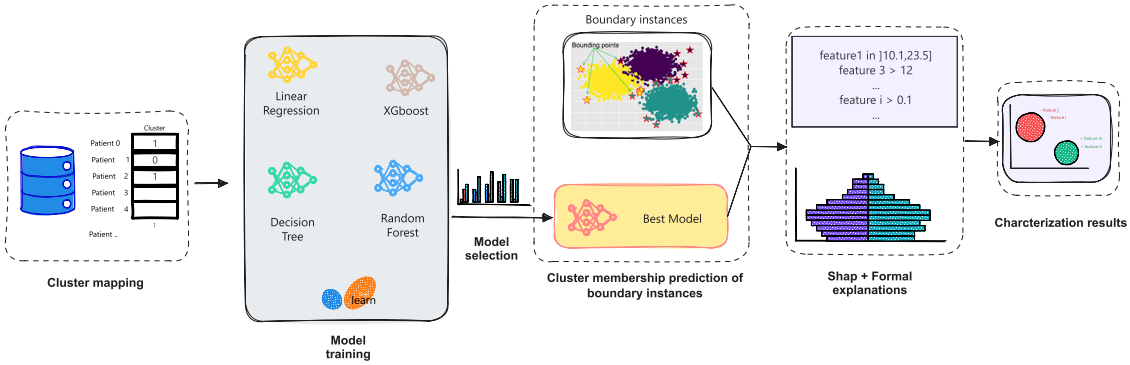

Fig. 3. Methodological framework for XAI-based cluster characterization and variable extraction.

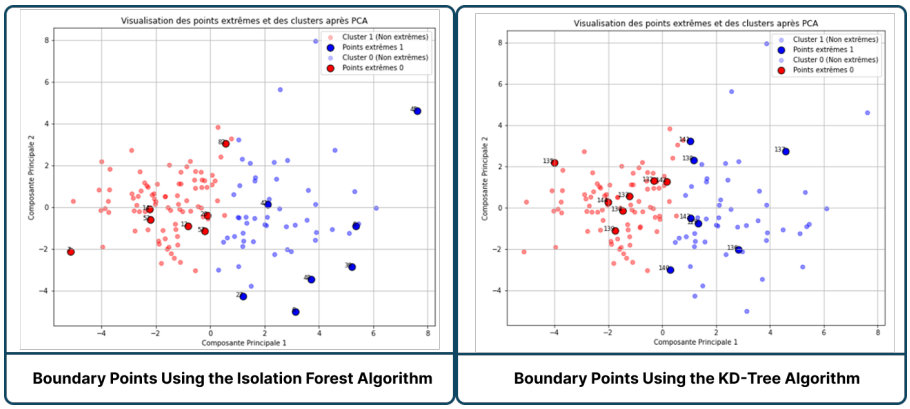

Fig. 4. Cluster boundary instances identified using KD-Tree and Isolation Forest algorithms.

tures and generate interpretable patient profiles. The six-step methodology, illustrated in Figure 3, integrates clustering, classification, and explanation generation:

(a) Cluster labeling: Apply clustering algorithm and assign each instance a cluster label, treating cluster membership as the target variable (`Cluster`).

(b) Classifier training: Train multiple classification models (**Logistic Regression**, **Decision Tree**, **Random Forest**, **XGBoost**) using a 75%-25% train-test split on the labeled dataset. Select the best-performing model for each cluster based on accuracy and F_1 score metrics.

(c) Boundary instance identification: Following [8], focus explanation efforts on instances near cluster decision boundaries, as these provide the most informative insights into cluster-distinguishing characteristics. Apply *KD-Tree* and *Isolation Forest* algorithms to detect boundary instances within each cluster. Visualize these critical points using PCA projection (Figure 4). Select the top 5% of instances as representative boundary cases.

(d) Hybrid explanation generation: For boundary instances, generate explanations using a combined approach integrating **SHAP** [9] and formal rule-based explanations via PyXAI[5] [14]. This hybrid methodology (detailed in Section IV-D) produces interpretable rules explaining why each boundary instance is predicted to belong to cluster 0 or 1, highlighting the most discriminative features.

(e) Feature importance interpretation: For each explained instance, identify key variables driving cluster assignment predictions. Interpret these variables according to their clinical nature: reference established thresholds for standardized medical scores (ANX, DEP, ISI, ESS, BMI) or perform comparative analysis against cluster means for other variables.

(f) Profile generalization: Synthesize boundary instance explanations to derive generalizable, representative patient profiles that characterize each cluster's typical members.

Summary: This XAI-based approach elucidates cluster assignment rationale by analyzing decision boundaries, enabling the construction of interpretable and clinically meaningful patient profiles for each identified subgroup.

D. SHAP-Enhanced Formal Explanations for Cluster Analysis

We introduce a hybrid **SHAP + Formal** approach that combines SHAP-based feature attribution with formal abductive

[5]PyXAI documentation: www.cril.univ-artois.fr/pyxai/

explanations, addressing individual method limitations while leveraging their strengths for robust cluster interpretation.

1) Individual Method Limitations: The SHAP method [9] quantifies feature contributions through additive attribution but presents critical limitations: (1) it assumes feature independence, neglecting variable interactions prevalent in medical data, and (2) its reliability remains questionable in healthcare domains. Recent studies [10], [13], [16] demonstrate discrepancies between agnostic XAI methods and formal approaches. SHAP scores also often fail to provide clear cluster differentiation in sleep disorder contexts.

Formal explanations offer mathematical rigor [10], [14] through rule-based interpretations but tend to be complex, containing 7 ± 2 elements [15], [16] that exceed cognitive processing limits [18] and complicate cluster characterization.

2) Hybrid Methodology: Our approach combines SHAP's feature ranking with formal explanations' logical reliability through four steps:

1) **SHAP feature ranking**: Identify top 10 influential variables for boundary instances based on importance scores.
2) **Formal rule generation**: Generate formal explanations using PyXAI[6] [14].
3) **Feature intersection**: Compute intersection between SHAP-ranked variables and formal explanation features.
4) **Rule construction**: Generate concise rules using intersected variables, ensuring short, reliable, and interpretable explanations.

This hybrid approach maintains logical consistency while remaining cognitively accessible, providing reliable cluster characterization through variables validated by both statistical importance and logical consistency.

V. Experiments and Results

This section presents experimental findings from our dual methodological framework: (1) regression-based sleep duration prediction with XAI interpretation, and (2) clustering-based patient profile construction and characterization.

A. Regression-Based Sleep Duration Analysis

Following the methodology outlined in Section III, we evaluated four regression algorithms for sleep duration prediction across two temporal datasets: **Dataset_1** (Day 0 to Day 7) and **Dataset_2** (Day 7 to Day 17). Model performance metrics are summarized in Table I.

Random Forest achieved superior performance across both datasets, demonstrating the lowest MSE and MAE values. This model was subsequently selected for SHAP-based explanation of representative patient instances.

SHAP analysis was conducted on twelve carefully selected instances (six per dataset: two average, two low, and two high sleep duration cases). The feature attribution analysis, illustrated in Figure 5, identifies key determinants of sleep duration and their directional effects:

[6]PyXAI: www.cril.univ-artois.fr/pyxai/

TABLE I
Regression model performance comparison across temporal datasets.

Model	Dataset_1		Dataset_2	
	MSE	**MAE_{min}**	**MSE**	**MAE_{min}**
Linear Regression	0.66	42.11	0.54	36.38
Decision Tree	0.78	48.02	0.70	41.80
Random Forest	**0.58**	**40.73**	**0.53**	**35.85**
XGBoost	0.70	43.56	0.62	37.20

- **Later bedtime correlates with increased sleep duration** (MIDMOY)
- **Sleep regularity promotes longer sleep duration** (SRIV)
- **Clinical factors negatively impact sleep**: insomnia severity (ISIV), elevated BMI, and excessive daytime sleepiness (ESS) consistently reduce sleep duration
- **Age demonstrates non-linear effects**: middle-aged patients exhibit shorter sleep duration compared to both younger and older adults

B. Clustering-Based Patient Segmentation

Cluster optimization using silhouette scores (Figure 2) and PCA visualization (Figure 6) for the top three k values confirmed $k = 2$ as optimal. K-means clustering partitioned the dataset into `Cluster 0` (90 instances, 62.07%) and `Cluster 1` (55 instances, 37.93%). Three complementary characterization methods were subsequently applied to identify discriminative variables.

Statistical Test Characterization. Comparative statistical analysis (Section IV-A) revealed no significant categorical variables, while several numerical variables demonstrated substantial inter-cluster differences. The most discriminative variables are presented in Table II.

Clinical interpretation according to established thresholds reveals distinct patient profiles. **Cluster 1** comprises younger patients exhibiting excessive daytime sleepiness (ESS), moderate insomnia (ISI), moderate depression (DEP) with major depressive episode risk (ANX), and irregular sleep patterns (reduced SRIV, elevated TSTISD variability). **Cluster 0** encompasses older patients with normal sleepiness, mild subclinical insomnia, minimal depression without significant anxiety, and preserved sleep regularity.

Clinical Summary. Statistical analysis identifies a clear dichotomy: Cluster 1 represents younger patients with comorbid sleep-mood disorders characterized by clinical depression, moderate insomnia, and sleep dysregulation, suggesting complex psychopathological presentations. Cluster 0 represents older patients with maintained sleep architecture and mild mood symptoms, indicating more stable clinical profiles.

*** Distance-based characterization.** We implemented the methodology described in Section IV-B, utilizing the calculation of in-patterns and out-patterns to reproduce the approach developed by [19]. This technique enabled us to identify the characteristic variables defining each cluster by determining

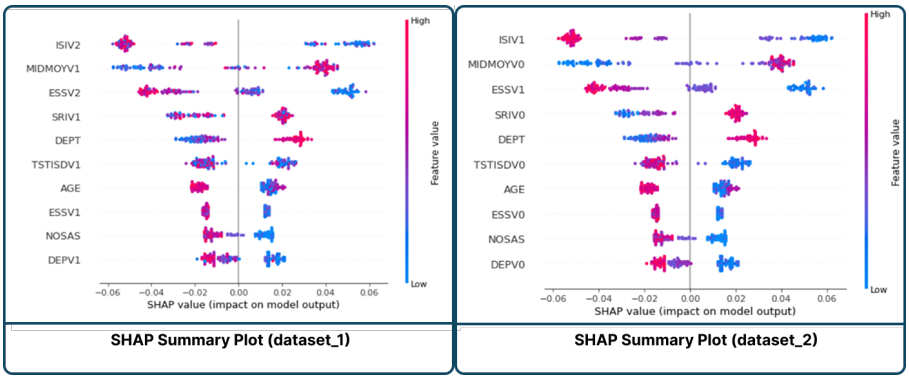

Fig. 5. SHAP summary plots revealing variable importance and directional effects on sleep duration prediction across both temporal datasets.

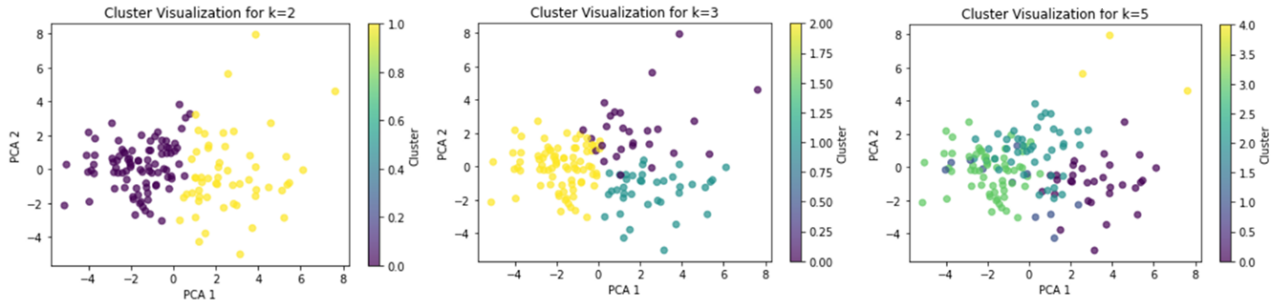

Fig. 6. PCA-based cluster visualization demonstrating optimal separation for varying cluster numbers.

TABLE II
STATISTICALLY SIGNIFICANT VARIABLES DISTINGUISHING CLUSTERS (MEAN ± STANDARD DEVIATION). POSITIVE DIFFERENCES INDICATE HIGHER VALUES IN CLUSTER 1.

Variable	Cluster 0	Cluster 1	Difference	p-value
Age	52.04 ± 12.13	47.14 ± 12.95	−4.90	2.29×10^{-2}
ESSV0	6.68 ± 4.27	10.25 ± 5.03	3.40	2.61×10^{-5}
ISIV0	13.12 ± 4.32	17.44 ± 3.68	4.31	6.84×10^{-9}
DEPV0	6.07 ± 3.37	13.25 ± 4.24	7.19	1.77×10^{-21}
ANXV0	2.52 ± 1.63	4.51 ± 1.35	1.99	3.69×10^{-12}
TSTISDV0	2.30 ± 0.92	2.87 ± 1.03	0.58	6.44×10^{-4}
SRIV0	85.59 ± 5.94	81.98 ± 7.73	−3.61	3.80×10^{-3}

whether their values exhibited consistently elevated or diminished patterns within each patient group.

Cluster 0. This patient group demonstrates predominantly elevated values for *Age*, *SRIV0*, and *SRIV1* variables, indicating an older demographic with enhanced sleep regularity patterns. Conversely, this cluster exhibits predominantly diminished values across multiple clinical indicators including *ISIV1*, *ISIV2*, *DEPV0*, *DEPV1*, *DEPV2*, *ANXV1*, and *ANXV2*, suggesting reduced severity levels across insomnia, depression, and anxiety symptom domains.

Cluster 1. This group manifests elevated values across several clinical variables: *ISIV1*, *ISIV2*, *DEPV0*, *DEPV1*, *DEPV2*, *ANXV1*, *ANXV2*, and *CRP*, indicating substantial prevalence of insomnia symptoms, depressive manifestations, anxiety disorders, and potential inflammatory processes. In contrast, *Age* presents reduced values, characterizing a younger demographic cohort.

Summary. The distance-based analysis reveals two distinct patient phenotypes: Cluster 1 encompasses younger patients presenting a symptomatic triad consisting of clinical depression, sleep disturbances, and occupational engagement, suggesting a potential association with work-related psychological stress. Cluster 0 corresponds to older patients exhibiting milder depressive symptoms and preserved sleep pattern regularity, reflecting greater psychological stability.

*** XAI-based characterization.** Following cluster identification and the addition of the *Cluster* variable to our dataset (detailed in Section IV-C), we trained four classification algo-

TABLE III
PERFORMANCE COMPARISON OF CLASSIFICATION MODELS FOR DISTINGUISHING CLUSTERS C_0 AND C_1.

Model	Accuracy	F_1 (C_0)	F_1 (C_1)
Logistic Regression	97.29%	98%	96%
Decision Tree	83.78%	88%	75%
Random Forest	**97.30%**	**98%**	**96%**
XGBoost	94.59%	96%	92%

rithms. Their performance metrics, optimized using **Optuna**[7], are presented in Table III. The Random Forest model demonstrated superior classification performance and was therefore selected for subsequent analytical procedures.

To characterize cluster boundaries, we employed the KD-Tree [8] and Isolation Forest [9] algorithms, selecting 5% of the dataset (corresponding to 7 instances per cluster) as representative boundary points. The spatial distribution of these critical instances was previously visualized in Figure 4.

After identifying these boundary points, the *SHAP+Formal* explanation methodology (described in Section IV-D) revealed the key variables influencing cluster assignment predictions for these critical boundary instances. Variable interpretation followed established protocols: clinical variables corresponding to standardized medical scores were evaluated against recognized clinical thresholds [3], while remaining variables were assessed through comparative analysis against cluster-specific mean values. Representative examples of boundary instance explanations from each cluster are provided below.

- Cluster 0 (Instance 142). The predictive variables for this boundary case include: DEPV0, DEPV1, DEPV2, ANXV0, ANXV1, ANXV2, ISIV0, ISIV1, TSTISDV0, and AGE. This patient demonstrates minimal depressive symptomatology (DEP), absence of significant anxiety manifestations (ANX), mild subclinical insomnia severity (ISI), reduced sleep duration variability (TSTISDV0), and advanced age relative to the population mean.
- Cluster 1 (Instance 143). The influential variables for this boundary prediction encompass: DEPV0, DEPV1, ISIV1, ANXV1, TSTISDV1, ESSV0, ESSV1, and AGE. This patient exhibits moderate depressive symptomatology (DEP), probable major depressive episode indicators (ANX), pronounced excessive daytime sleepiness (ESS), moderate clinical insomnia severity (ISI), elevated sleep duration variability (TSTISDV1), and younger age compared to the population average.

Summary. The interpretation of boundary instances between clusters reveals distinct patient characteristics. Individuals in Cluster 0 are typically older in age, exhibit reduced depressive symptomatology, and maintain consistent sleep behavioral patterns. Conversely, patients in Cluster 1 demonstrate younger age demographics, experience significant depression and insomnia symptoms, and present irregular and insufficient sleep patterns.

[7]Optuna is a hyperparameter optimization framework for automated machine learning model tuning

[8]The KD-Tree algorithm organizes data points within a k-dimensional space for efficient spatial partitioning (see [20]).

[9]The Isolation Forest algorithm computes anomaly scores by measuring the path length required to isolate individual instances.

VI. FEEDBACK FROM EXPERTS

Feedback from sleep medicine specialists at CHU de Bordeaux guided refinements to our methodology and dataset, enhancing clinical relevance and interpretability for trustworthy interventions in digital sleep medicine.

First, the experts recommended refining the KANOPEE dataset to ensure consistent patient identifiers across both application versions, improving data quality and preprocessing reliability.

Second, they proposed focusing on clinically relevant variables for clustering: age, gender, BMI, education level, socio-professional category (CSP), ISI (V0, V1, V2), ESS (V0, V1, V2), ANX (V0, V1, V2), DEP (V0, V1, V2), TSTMOY, MIDMOY, and SRIV. This increased the sample size to 1,000 complete records, boosting statistical power. Variables like ID, dates, region (REG), department (DEPT), NOSAS, OSA, and RLS were excluded as clinically irrelevant or imprecise. CSP and education level were identified as categorical, requiring label encoding.

Third, the experts suggested using at least three clusters to capture nuanced patient profiles. The initial two-cluster solution oversimplified sleep disorder heterogeneity into "good" versus "poor" sleepers. Adopting $k = 3$, as shown in Figure 7, revealed subtypes like young professionals with mild insomnia, older adults with stable sleep, and middle-aged patients with excessive sleepiness and irregular sleep.

Fourth, for regression, the experts advised shifting from predicting sleep duration (TSTMOY), a modifiable parameter, to outcomes reflecting KANOPEE's success, such as Delta ISI (ISI V2 - ISI V0) and Delta PHQ-9 (DEP V2 - DEP V0). Application completion rates (Visit 1 and Visit 2, Yes/No) were also suggested as engagement indicators.

Finally, the experts emphasized predicting sleep-related (ISI, ESS, TSTMOY, SRIV) or mental health (ANX, DEP) outcomes, excluding irrelevant predictions like age or gender. This focused analysis on actionable insights for personalized interventions.

These changes were integrated into the refined approach (Section VII), yielding a 933-instance dataset, a three-cluster solution, and clinically relevant predictions. Ongoing collaboration with CHU de Bordeaux supports further data updates and validation, ensuring alignment with clinical expertise.

VII. REFINED CLUSTERING ANALYSIS

Based on expert feedback, we enhanced our K-means clustering analysis by using a larger, refined dataset. We excluded clinically irrelevant variables, such as dates, geographic region (REG), department (DEPT), NOSAS score, obstructive sleep apnea (OSA), and restless legs syndrome (RLS), retaining 933

TABLE IV
CLASSIFICATION MODEL PERFORMANCE FOR CLUSTERS C_0, C_1, AND C_2.

Model	Accuracy	F_1 (C_0)	F_1 (C_1)	F_1 (C_2)
Logistic Regression	**97.86%**	**97%**	**98%**	**98%**
Decision Tree	80.77%	74%	89%	77%
Random Forest	90.17%	86%	95%	89%
XGBoost	89.32%	83%	92%	88%

complete records (an increase of 888 instances over the initial 145-patient dataset). Applying K-means clustering with $k = 3$, optimized using silhouette scores and validated through PCA visualization, we identified three distinct patient profiles, as illustrated in Figure 7.

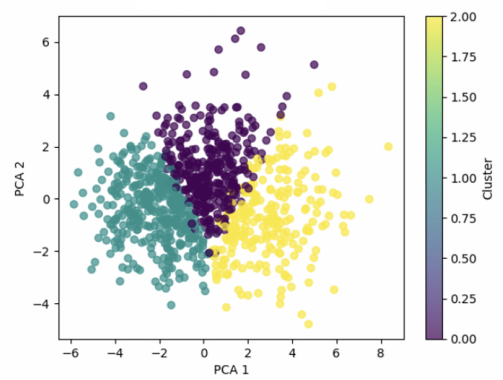

Fig. 7. PCA-Based Visualization of Three Clusters ($k = 3$) with 933 Instances.

The resulting clusters, labeled Cluster 0, Cluster 1, and Cluster 2, contain 325, 368, and 240 instances, respectively.

A. Cluster Characterization

We characterized the three clusters using three complementary methods: statistical tests, in-pattern/out-pattern analysis, and SHAP-based XAI. These approaches consistently identified distinct patient profiles:

- **Cluster 0**: Young patients, often in higher intellectual professions, with adequate sleep duration, mild insomnia, mild depression, a probable risk of major depressive episodes, and later bedtimes.
- **Cluster 1**: Older patients, predominantly male, with average but improvable sleep quality, minimal depression, mild insomnia, and low anxiety levels.
- **Cluster 2**: Middle-aged patients with excessive, potentially pathological daytime sleepiness, moderate insomnia, moderate depression, and irregular sleep patterns.

For the XAI-based characterization, we trained four classification models: Logistic Regression, Decision Tree, Random Forest, and XGBoost, and selected Logistic Regression for its superior performance and inherent interpretability (Table IV).

Logistic Regression's coefficients were used to assess feature importance for cluster assignment. We calculated the relative importance of each feature x_j as:

$$\text{Importance}(x_j) = \frac{|\beta_j|}{\sum_{k=1}^{p} |\beta_k|}$$

where β_j is the coefficient for feature x_j. A positive coefficient ($\beta_j > 0$) indicates that higher feature values increase the likelihood of belonging to a cluster, while a negative coefficient ($\beta_j < 0$) decreases it.

Cluster 0: Features with positive contributions, increasing the probability of assignment to Cluster 0, include *MIDMOY* (sleep midpoint), *TSTMOY* (average sleep duration), *CSP* (socio-professional category), *ANX* (anxiety), and *ISI* (insomnia severity). This suggests patients in higher intellectual professions with later bedtimes, longer sleep duration, and elevated anxiety and insomnia levels. Features with negative contributions, reducing the probability, include *ESS* (daytime sleepiness), *DEP* (depression), *SRIV* (sleep regularity), and *AGE*, indicating younger patients with lower depression, less regular sleep, and reduced sleepiness.

Cluster 1: Positive contributors include *SEXE* (gender), *AGE*, *ETUDE* (education level), and *SRIV*, characterizing older, mostly male patients with higher education and regular sleep patterns. Negative contributors include *DEP*, *ANX*, *MIDMOY*, *ISI*, *TSTMOY*, and *ESS*, reflecting minimal depression, low anxiety, earlier bedtimes, shorter sleep duration, and reduced insomnia and sleepiness.

Cluster 2: Positive contributors include *ANX*, *DEP*, *ESS*, *ISI*, and *AGE*, indicating middle-aged patients with higher insomnia, anxiety, depression, and daytime sleepiness. Negative contributors include *TSTMOY*, *MIDMOY*, and *SRIV*, suggesting shorter, earlier, and less regular sleep.

These interpretable profiles align with clinical expectations, supporting personalized interventions in digital sleep medicine.

VIII. RELATED WORK

Comparison with Prior Studies. Recent studies have applied machine learning (ML) to analyze sleep behaviors and diagnose sleep disorders using data from wearables, mobile apps, and clinical cohorts, employing supervised and unsupervised methods.

Supervised approaches include Trujillo et al. [21], who used regression to predict sleep duration from 30-day physical activity and sleep data, highlighting exercise and sleep regularity as key factors. Al-Mamun et al. [22] applied a CatBoost classifier to detect abnormal sleep duration in students based on demographic and psychological data. These studies, however, often overlook model interpretability, limiting clinical applicability.

Unsupervised methods, such as those by Park et al. [23] and Ferreira-Santos and Rodrigues [24], used clustering to identify subtypes in insomnia or sleep apnea patients. Gool et al. [25] applied hierarchical clustering to define hypersomnolence subtypes. While these approaches offer population-level insights, they lack individualized, interpretable profiles critical for clinical use.

Positioning of Our Approach. Our study advances prior work by integrating supervised regression and unsupervised clustering with explainable AI (XAI) on KANOPEE app data. We focus on clinically relevant outcomes like Delta ISI and Delta PHQ-9, as advised by experts. Our three-cluster K-means solution identifies nuanced profiles (e.g., young professionals with mild insomnia, older adults with stable sleep, middle-aged patients with excessive sleepiness), validated via statistical tests, in-pattern/out-pattern analysis, and XAI. Using Logistic Regression and a SHAP+Formal explanation framework, we ensure global and individual-level interpretability, enhancing usability for clinicians and supporting personalized interventions in digital sleep medicine.

IX. Conclusion

Our regression analysis showed that later bedtimes and higher sleep regularity predict longer sleep, whereas elevated BMI, insomnia severity, and daytime sleepiness are associated with shorter sleep. Additionally, middle-aged patients tend to sleep less compared to younger and older individuals, highlighting non-linear age effects. The optimized K-means clustering, refined to three clusters with expert feedback, revealed distinct patient profiles: (*i*) younger, anxious individuals with delayed sleep; (*ii*) older, low-anxiety males with average sleep; and (*iii*) middle-aged patients characterized by daytime sleepiness and irregular sleep schedules.

Our XAI-based approach highlights the critical role of integrating expert medical knowledge into the AI development lifecycle. As argued by [27], simply making a model interpretable for data scientists is insufficient; true interpretability must be built with concepts and expertise from relevant domains. The explanations derived from our experiments align with established clinical knowledge about sleep and mental health. This fosters trust in the clustering process, aligning algorithmic outputs with clinical reasoning for personalized interventions in digital sleep medicine. By co-developing these models with sleep medicine specialists, we ensure that the explanations are not just computationally sound but also clinically meaningful.

Future research directions will involve improving missing value handling. Our explainable framework could be extended to incorporate imputation techniques while preserving interpretability. This would allow models to remain transparent and reliable in more realistic and incomplete data contexts. Furthermore, to support reproducibility and advance research, the source code for this study will be publicly available on GitHub under an open-source license upon publication.

References

[1] L. C. Markun and A. Sampat, "Aperçu et développements axés sur le clinicien en polysomnographie," Sleep Med., vol. 67, 2020.

[2] M. Ennab and H. Mcheick, "Designing an interpretability-based model to explain the artificial intelligence algorithms in healthcare," *Diagnostics*, vol. 12, no. 7, pp. 1557, 2022.

[3] S. Xu, O. Faust, S. Seoni, S. Chakraborty, P. D. Barua, H. W. Loh, H. Elphick, F. Molinari, and U. R. Acharya, "A review of automated sleep disorder detection," Comput. Biol. Med., vol. 150, 2022.

[4] S. Shajari, K. Kuruvinashetti, A. Komeili, and U. Sundararaj, "The emergence of AI-based wearable sensors for digital health technology: A review", Sensors, vol. 23, art. 9498, 2023.

[5] P. Philip, L. Dupuy, P. Sagaspe, E. de Sevin, M. Auriacombe, J. Taillard, J.-A. Micoulaud Franchi, and C. M. Morin, "Efficacy of a smartphone-based virtual companion to treat insomniac complaints in the general population: Sleep diary monitoring versus an internet autonomous intervention," J. Clin. Med., vol. 11, no. 15, art. 4387, 2022.

[6] J. Kim, B. T. Keenan, D. C. Lim, S. K. Lee, A. I. Pack, and C. Shin, "Symptom-based subgroups of Koreans with obstructive sleep apnea," J. Clin. Sleep Med., vol. 14, no. 3, pp. 437–443, 2018.

[7] M. R. Khoie, T. S. Tabrizi, E. S. Khorasani, S. Rahimi, and N. Marhamati, "A hospital recommendation system based on patient satisfaction survey," Appl. Sci., vol. 7, no. 10, art. 966, 2017.

[8] S. Bobek, M. Kuk, M. Szelą.zek, and G. J. Nalepa, "Enhancing cluster analysis with explainable AI and multidimensional cluster prototypes," *IEEE Access*, vol. 10, pp. 24984–24997, 2022.

[9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, vol. 30, Long Beach, CA, USA, 2017, pp. 4765–4774.

[10] J. Marques-Silva, "Logic-based explainability in machine learning," arXiv preprint arXiv:2301.02622, 2023.

[11] A. Vellido, "The importance of interpretability and visualization in machine learning for applications in medicine and health care," *Neural computing and applications*, vol. 32, no. 24, pp. 18069–18083, 2020.

[12] G. Audemard, S. Bellart, L. Bounia, F. Koriche, J.-M. Lagniez, and P. Marquis, "On the computational intelligibility of Boolean classifiers," in Proc. of KR'21, 2021.

[13] G. Audemard, S. Bellart, L. Bounia, F. Koriche, J.-M. Lagniez, and P. Marquis, "On preferred abductive explanations for decision trees and random forests," in Proceedings of IJCAI-22, 2022.

[14] G. Audemard, J.-M. Lagniez, P. Marquis, and N. Szczepanski, "PyXAI: an XAI library for tree-based models," in Proc. of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24), Demo Track, Jeju, South Korea, August 2024, pp. 8601–8605.

[15] Y. Izza, X. Huang, A. Ignatiev, N. Narodytska, M. C. Cooper, and J. Marques-Silva, "On computing probabilistic abductive explanations," Int. J. Approx. Reason., vol. 159, art. 108939, Aug. 2023.

[16] L. Bounia and I. Setitra, "Enhancing the intelligibility of decision trees with concise and reliable probabilistic explanations," Data & Knowl. Eng., vol. 156, art. 102394, Mar. 2025.

[17] N. Watson, K. Yu, D. Campbell, L. Bloudek, J. Yee, J. Cronin, A. Sanchez Azofra, and N. Boskovic, "Prevalence and Unmet Need of Obstructive Sleep Apnea in the United States," 2025.

[18] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," Psychol. Rev. 1956.

[19] J. Agoun, Y. Bouallouche, and M.-S. Hacid, "OptiClust4Rec: Unsupervised data-driven methodology for quality of life recommendations during a medical therapy," in Proc. Int. Conf. Cooperative Inf. Syst. (CoopIS), Groningen, Netherlands, Oct. 2023.

[20] J. L. Bentley, "Multidimensional binary search trees used for associative searching," Commun. ACM, vol. 18, no. 9, pp. 509–517, Sept. 1975.

[21] R. Trujillo, E. Zhang, J. M. Templeton, and C. Poellabauer, "Predicting long-term sleep deprivation using wearable sensors and health surveys," Comput. Biol. Med., 2024.

[22] F. Al-Mamun, M. A. Mamun, M. E. Hasan, M. M. ALmerab, and D. Gozal, "Exploring sleep duration and insomnia among prospective university students: A study with geographical data and machine learning techniques," Nat. Sci. Sleep, vol. 16, pp. 1235–1251, Aug. 2024.

[23] S. Park, S. W. Lee, S. Han, and M. Cha, "Clustering insomnia patterns by data from wearable devices: Algorithm development and validation study," JMIR mHealth uHealth, vol. 7, no. 12, art.

[24] D. Ferreira-Santos and P. P. Rodrigues, "Obstructive sleep apnea: a categorical cluster analysis and visualization," Pulmonology, vol. 29, no. 3, pp. 207–213, May–Jun. 2023, doi: 10.1016/j.pulmoe.2021.10.003.

[25] J. K. Gool, Z. Zhang, M. S. S. L. Oei, S. Mathias, Y. Dauvilliers, G. Mayer, G. Plazzi, R. del Rio-Villegas, J. S. Cano, K. Šonka, et al., "Data-driven phenotyping of central disorders of hypersomnolence with unsupervised clustering," Neurology, 2022.

[26] B. Vidgen and T. Yasseri, "P-values: misunderstood and misused," Frontiers in Physics, vol. 4, art. 6, 2016, doi: 10.3389/fphy.2016.00006.

[27] T. Miller, P. Howe, and L. Sonenberg, Explainable AI: Beware of inmates running the asylum or: How I learnt to stop worrying and love the social and behavioural sciences, *arXiv preprint arXiv:1712.00547*, 2017.