Received: 3 July 2024 | Revised: 20 August 2024 | Accepted: 3 September 2024

DOI: 10.1111/jsr.14349

RESEARCH ARTICLE

A comparative analysis of unsupervised machine-learning methods in PSG-related phenotyping

Mohammadreza Ghorvei[1,2] | Tuomas Karhu[1,2] | Salla Hietakoste[1,2] | Daniela Ferreira-Santos[1,3] | Harald Hrubos-Strøm[4,5] | Anna Sigridur Islind[6] | Luka Biedebach[6] | Sami Nikkonen[1,2] | Timo Leppänen[1,2,7] | Matias Rusanen[1,2,8]

[1]Department of Technical Physics, University of Eastern Finland, Kuopio, Finland

[2]Diagnostic Imaging Center, Kuopio University Hospital, Kuopio, Finland

[3]INESC TEC - Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal

[4]Department of ear-nose and throat, Akershus University Hospital, Lørenskog, Norway

[5]Campus Akershus University Hospital, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[6]Department of Computer Science, Reykjavik University, Reykjavik, Iceland

[7]School of Electrical Engineering and Computer Science, The University of Queensland, Brisbane, Queensland, Australia

[8]HP2 Laboratory, INSERM U1300, Grenoble Alpes University, Grenoble Alpes University Hospital, Grenoble, France

Correspondence
Mohammadreza Ghorvei, Department of Technical Physics, University of Eastern Finland, Kuopio, FI-70211, Finland.
Email: mohammadreza.ghorvei@uef.fi

Funding information
Respiratory Foundation of Kuopio Region; The State Research Funding for university-level health research, Kuopio University Hospital, Wellbeing Service County of North Savo, Grant/Award Numbers: 5041828, 5041790, 5041794, 5041798, 5041809, 5041797; Tampere Tuberculosis Foundation; European Union's Horizon 2020 Research and Innovation Programme, Grant/Award Number: 965417; Orion Research Foundation; Research Foundation of the pulmonary diseases; French National Research Agency, MIAI@Grenoble Alpes, Grant/Award Number: ANR-15-IDEX-02: ANR-19-P3IA-0003

Summary

Obstructive sleep apnea is a heterogeneous sleep disorder with varying phenotypes. Several studies have already performed cluster analyses to discover various obstructive sleep apnea phenotypic clusters. However, the selection of the clustering method might affect the outputs. Consequently, it is unclear whether similar obstructive sleep apnea clusters can be reproduced using different clustering methods. In this study, we applied four well-known clustering methods: Agglomerative Hierarchical Clustering; K-means; Fuzzy C-means; and Gaussian Mixture Model to a population of 865 suspected obstructive sleep apnea patients. By creating five clusters with each method, we examined the effect of clustering methods on forming obstructive sleep apnea clusters and the differences in their physiological characteristics. We utilized a visualization technique to indicate the cluster formations, Cohen's kappa statistics to find the similarity and agreement between clustering methods, and performance evaluation to compare the clustering performance. As a result, two out of five clusters were distinctly different with all four methods, while three other clusters exhibited overlapping features across all methods. In terms of agreement, Fuzzy C-means and K-means had the strongest ($\kappa = 0.87$), and Agglomerative hierarchical clustering and Gaussian Mixture Model had the weakest agreement ($\kappa = 0.51$) between each other. The K-means showed the best clustering performance, followed by the Fuzzy C-

means in most evaluation criteria. Moreover, Fuzzy C-means showed the greatest potential in handling overlapping clusters compared with other methods. In conclusion, we revealed a direct impact of clustering method selection on the formation and physiological characteristics of obstructive sleep apnea clusters. In addition, we highlighted the capability of soft clustering methods, particularly Fuzzy C-means, in the application of obstructive sleep apnea phenotyping.

KEYWORDS
hard clustering, polysomnography, sleep disorders, soft clustering, unsupervised machine-learning methods

1 | INTRODUCTION

Obstructive sleep apnea (OSA) is recognized as a multifactorial and heterogeneous sleep disorder with a high prevalence (de Araujo Dantas et al., 2023). At present, the diagnosis and assessment of OSA severity predominantly rely on a single polysomnographic (PSG) index, the apnea–hypopnea index (AHI). However, the AHI only reflects the frequency of respiratory events, thus ignoring a vast amount of physiological information available on PSGs, as well as leaving out essential information on metabolic, cardiovascular and neural physiology that characterize the heterogeneity of OSA (Malhotra et al., 2021). Also, the first-line treatment for OSA has centered around continuous positive airway pressure (CPAP; Semelka et al., 2016), which heavily relies on the AHI, and thus it might not be the optimal choice for all patients with OSA (Randerath et al., 2021; Sutherland et al., 2018). To address these shortcomings, several studies have employed cluster analyses to discover distinct groups of patients with similar characteristics (i.e. OSA phenotypes), aiming to fully capture the heterogeneity of OSA reflected in PSGs (Zinchuk et al., 2017; Zinchuk & Yaggi, 2020). Such methods have also shown great potential in personalized OSA management and treatment (Bonsignore et al., 2017; de Chazal et al., 2020; McNicholas & Korkalainen, 2023; Randerath et al., 2022; Verbraecken et al., 2022).

Unsupervised machine-learning methods have played a key role in the main core of cluster analysis, enabling the analysis of complex and multidimensional health data, like PSG, without requiring prior label knowledge (Pitafi et al., 2023). In addition, unsupervised machine-learning methods can enable the integration of the PSG-based variables with anthropometrics and demographic characteristics, symptoms, comorbidity profiles, as well as questionnaire data (Zinchuk et al., 2017). Therefore, these methods are well suited for discovering various OSA phenotypes.

However, the choice of clustering methods is a recognized challenge of cluster analysis in various research fields (Jaeger & Banks, 2023). This has also been quite controversial in the field of OSA phenotyping as previous studies have reported that the clustering findings heavily depend on the chosen clustering method (Kim et al., 2020; Zhang et al., 2022; Zhu et al., 2023); clustering outputs might change by utilizing different clustering methods and settings. In this regard, follow-up analyses and the development of targeted treatment strategies might be significantly biased, leading to incorrect OSA treatment and unsatisfied patient outcomes.

In this comparative study, we aimed to investigate how the clustering method selection affects the cluster formations and physiological outcomes by utilizing a large number of individuals with suspected OSA. We applied assorted unsupervised machine-learning methods and compared the clustering outputs across all methods. In addition, we evaluated the agreements and similarities between the clustering methods and examined the clustering performance of applied methods through internal cluster validation indices (Ran et al., 2022).

2 | MATERIALS AND METHODS

2.1 | Clinical PSG recordings

We conducted a retrospective study for 912 consecutive individuals referred for PSG at the Princess Alexandra Hospital, Brisbane, Australia due to OSA suspicion. PSGs were recorded between 2015 and 2017 using the Compumedics Grael 4 K PSG system (Abbotsford, Australia). All PSGs were conducted, adjusted and analysed by medical experts following the newest guidelines available at the time of analysis, set by the American Academy of Sleep Medicine (AASM; Troester et al., 2023). As OSA has been linked with a wide range of physiological outcomes, particularly neurocognitive impairment (Tam et al., 2014), we used a psychomotor vigilance task (PVT) to compare the vigilance levels of individuals within clusters. In our study, all individuals completed a PVT in the evening before the PSG. In PVT, a total of 120 visual stimuli were presented on a touchpad screen, and individuals were instructed to react by pressing a screen with their thumb or index finger as fast as possible. If the reaction time exceeded 500 ms, it was counted as a lapse. The Metro South Human Research Ethics Committee (Brisbane, Australia) accepted the retrospective data collection and research use of the data (LNR/2019/QMS/54313).

For further analysis, individuals with incomplete PSG recording or demographic information ($n = 40$) or a total sleep time (TST) of less than 1 hr ($n = 7$) were excluded. After exclusion, a total of 865 individuals with suspected OSA were included in the final analysis.

2.2 | Data preprocessing

We selected 42 clustering variables after careful review and examination. We used only continuous variables, and thus no categorical variables like sex were included. The selected variables included nine domains: heart rate variability (HRV); sleep architecture; breathing disturbance; blood oxygen saturation (SpO_2)-based variables; snoring; periodic limb movement (PLM); arousals; PVT outcomes; and a domain consisting of questionnaire, demographic and anthropometric data (Table 1).

In addition to conventional SpO_2 variables (i.e. oxygen desaturation index and percentage of time spent below 90% oxygen saturation [T90]), we calculated several more comprehensive SpO_2 variables (Figure 1) by using the ABOSA software (Karhu et al., 2022). The included variables primarily focus on the depth, area and duration of saturation and resaturation oxygen events. They have also been shown to more accurately characterize the hypoxic load and indicate the severity of OSA compared with conventional variables (Kulkas et al., 2013). These variables included desaturation severity (the sum of desaturation event areas divided by TST), desaturation duration (the sum of desaturation event durations divided by TST), recovery severity (the sum of recovery event areas divided by TST), recovery duration (the sum of recovery event durations divided by TST) and total transient drop severity (the sum of desaturation and recovery areas from the 100% reference). Moreover, we calculated median ratio parameters between desaturation and recovery event depths, areas, durations and slopes. In addition, we calculated the obstruction severity (Kulkas et al., 2013) with the following formula:

$$\text{Obstruction Severity}(s\%) = \frac{\sum_{n=1}^{L}(\text{HypDur}_n \times \text{DesatArea}_n) + \sum_{n=1}^{L}(\text{ApDur}_n \times \text{DesatArea}_n)}{\text{TST}} \quad (1)$$

where L is the number of desaturation events, $\text{DesatArea}\,(s\%)$ is a desaturation area of an individual event (Figure 1), and $\text{HypDur}\,(s)$ and $\text{ApDur}\,(s)$ are the durations of individual hypopnea and apnea events, respectively. In Equation (1), each hypopnea and apnea event were linked to a desaturation event.

Also, for conventional overnight (long-term) time- and frequency domain HRV metrics, we detected the R peaks from electrocardiographic signals and corrected the peaks using Kubios HRV Premium 3.4.1 (Kubios Oy, Kuopio, Finland) software with default settings (Tarvainen et al., 2014). The average of RR intervals (average RR) and standard deviation of RR intervals (SDRR) were calculated to reflect the overall HRV. The percentage of RR intervals differing by more than 50 ms (pRR50) and root mean square of successive differences (RMSSD) were calculated to reflect the parasympathetic activity. In addition, we calculated the power spectral densities (PSDs) from each 5-min segment of the detrended RR series with Welch's method (eight sections, 50% overlap, Hamming window). From the PSDs, the low-frequency to high-frequency band power ratio (LF/HF) ratio was

TABLE 1 Domains of selected variables.

Variable category	Variables
Questionnaire, demographic and anthropometric data	ESS (points)
	BMI (kg m^{-2})
	Age (years)
Sleep architecture	WASO (min)
	TST (min)
	Sleep efficiency (%)
	Percentage of stage N1 (% of TST)
	Percentage of stage N2 (% of TST)
	Percentage of stage N3 (% of TST)
	Percentage of REM sleep (% of TST)
Breathing disturbance	AHI (events per hr)
	Hypopnea index (events per hr)
	Apnea index (events per hr)
	Central apnea index (events per hr)
	REM AHI (events per hr)
	NREM AHI (events per hr)
	Supine AHI (events per hr)
	Non-supine AHI (events per hr)
	Mean apnea duration (s)
	Mean hypopnea duration (s)
HRV	Average RR (s)
	pRR50 (%)
	RMSSD (s)
	SDRR (s)
	LF/HF ratio
SpO_2-based variables	Oxygen desaturation index (events per hr)
	Desaturation severity (%-unit)
	Desaturation duration (%-unit)
	Total transient drop severity (%-unit)
	Recovery severity (%-unit)
	Recovery duration (%-unit)
	Obstruction severity (s%)
	Percentage of time spent below 90% oxygen saturation (% of TST)
	Median duration ratio
	Median depth ratio
	Median area ratio
	Median slope ratio
Snoring	Snoring time (% of TST)
Arousals	Arousal index (events per hr)
PLM	PLM index (events per hr)
PVT outcomes	Median reaction time (ms)
	Number of lapses

AHI, apnea–hypopnea index; average RR, average of RR intervals; BMI, body mass index; ESS, Epworth Sleepiness Scale; HRV, heart rate variability; LF/HF, low-frequency to high-frequency ratio; NREM, non-rapid eye movement; PLM, periodic limb movement; pRR50, percentage of RR intervals differing by more than 50 ms; PVT, psychomotor vigilance task; REM, rapid eye movement; RMSSD, root mean square of successive differences; SDRR, standard deviation of RR intervals; SpO_2, blood oxygen saturation; TST, total sleep time; WASO, wake after sleep onset.

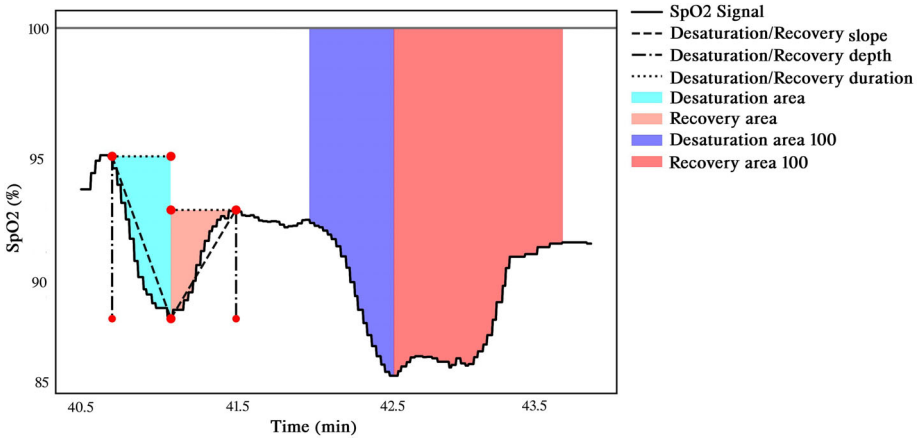

FIGURE 1 The visualization of SpO_2-based variables.

calculated to illustrate the sympathovagal balance (Shaffer & Ginsberg, 2017).

Before clustering, the selected variables were normalized by utilizing the *Z*-score technique (Schaffer & Green, 1996). The normality of the cluster distributions was examined using the Kolmogorov–Smirnov test. Also, the difference between clustering variables across all clusters in each clustering method was examined using a one-way analysis of variance or Kruskal–Wallis as appropriate. Moreover, all the variables were pairwise compared using a post-hoc Dunn's test with a Bonferroni correction for multiple comparisons. The limit for statistical significance was set at $p < 0.05$.

2.3 | Unsupervised machine-learning methods

We selected a diverse range of unsupervised machine-learning methods utilized in the medical research and application of OSA phenotyping, i.e. Agglomerative Hierarchical Clustering (AHC; Johnson, 1967), K-means (MacQueen, 1967), Gaussian Mixture Model (GMM; Cai et al., 2019) and Fuzzy C-means (FCM; Bezdek et al., 1984). These methods encompass both hard and soft clustering methods (Pitafi et al., 2023). The hard clustering algorithms (i.e. AHC and K-means) assign each data point to a single cluster, allowing straightforward categorization where distinct and non-overlapping clusters are presented. On the other hand, soft clustering methods (i.e. FCM and GMM) can assign each data point to multiple groups with different membership probabilities (Pitafi et al., 2023).

We employed diverse model selection criteria to obtain the optimal number of clusters. However, the recommended number of clusters differed across the clustering methods. Also, these selection methods do not ensure that the recommended number of clusters implies the natural clusters in the data (Alashwal et al., 2019). We determined the number of clusters to be five for this study based on a consensus from multiple methods, the clinical interpretability of the clusters, and the necessity for a balanced and fair comparison across methods.

2.3.1 | Agglomerative Hierarchical Clustering

The AHC starts by creating sub-clusters that include only one observation each, and continues by combining sub-clusters from bottom to top until there is only one large cluster. The sub-clusters are merged through different similarity measures (e.g. based on distance, the nearest neighbour, and the partition's quality evaluation after merging; Govender & Sivakumar, 2020). In this study, the five clusters were made using Ward's method as the similarity measure to merge sub-clusters, and Euclidean distance was used as the similarity metric. The Lance–Williams updating formula and the corresponding coefficients were employed for updating similarity measures. The stats (version 3.6.2) package in R was used to implement the AHC method (Flynt & Dean, 2016).

2.3.2 | K-means

The K-means assigns observations to a single cluster with the smallest distance to the centroid of a cluster (Pitafi et al., 2023). In this study, the Euclidean distance was chosen for computing the distance between cluster centroids and observations. Also, to ensure that the algorithm has fully converged, the maximum number of iterations was set to 2000. The initial centroids were randomly chosen. However, the number of starts was set at 25; this means that the algorithm was executed 25 times with different random initial centroids, and the final cluster was returned by the one with the lowest cluster sum of squares. The implementation was done by the stats (version 3.6.2) package in R (Flynt & Dean, 2016).

2.3.3 | Fuzzy C-means

The FCM generalizes K-means by assigning each observation to all clusters based on different membership probabilities. It iteratively updates the cluster centroids vector and membership matrix until the convergence of the objective function is achieved or the maximum

iteration is reached (Askari, 2021). We used Euclidian distance as the distance metric in the objective function. The initial centroid vectors and membership matrix were chosen based on the K-means++ initialization (Bahmani et al., 2012) and random sampling function (Cebeci, n.d.). Similar to K-means, the number of starts and the maximum number of iterations were set to 25 and 2000, respectively. To avoid leading to suboptimal clusters, we set the conservative threshold of 1e-9 in the process of converging the objective function. The fuzzy membership degree was set to 1.1. The "ppclust" (version 1.1.0) package in R was used for the implementation of the FCM (Cebeci et al., 2019).

2.3.4 | Gaussian Mixture Model

The GMM is a soft parametric clustering technique that uses multivariate Gaussian distributions to group data points according to their maximum likelihood under the distributions (Torkkola, 2003). In our implementation, the expectation–maximization (EM) algorithm was used to maximize the complete likelihood of the objective function. The performance of different EM models was measured using the Bayesian information criterion (Jung & Wickrama, 2008). The "mclust" (version 6.0.1) package in R was used to implement the GMM method (Jung & Wickrama, 2008).

2.4 | Comparison of clustering results

We utilized the t-distributed stochastic neighbour embedding (t-SNE) technique to visually compare the clustering methods and understand how clusters are formed. This method effectively visualizes patterns between data by converting high dimensions to lower dimensions, while maintaining local structures and emphasizing dissimilarities (Van der Maaten & Hinton, 2008).

We also evaluated the agreement between clustering methods, and measured their similarity and dissimilarity. We utilized Cohen's kappa statistics to assess the pairwise agreement between clustering methods. For this purpose, we adjusted the same labels for two corresponding clusters derived from two different clustering methods if they exhibited approximately similar physiological characteristics.

Furthermore, we evaluated clustering performance using validation indices. There are two ways to assess the clustering performance: external and internal validation methods. External validation methods focus on comparing the clustering results with true membership labels. In contrast, internal validation methods emphasize internal properties such as the separateness and compactness of the clusters. In this study, due to the lack of access to true membership labels, we only compared clustering methods using internal validation methods. For this purpose, we utilized the Calinski–Harabasz index (CHI), Davies–Bouldin index (DBI), Dunn index (DI), and Silhouette coefficient (SC; Ran et al., 2022).

3 | RESULTS

3.1 | Physiological characteristics of individuals

The demographics and PSG characteristics of the study population are presented in Table 2. The heatmap of various clustering variables and the summary of cluster characteristics are presented in Figures 2 and 3, respectively. In addition, the detailed parameter values and statistical analyses, as well as the distribution of OSA severity in each cluster, are reported in Tables S1–S4 and Figure S1, respectively (Supplementary Data S1).

To summarize the findings, the following results were consistent between all clusters.

Cluster 1 was the least populated cluster, comprising 2.19% ($n = 19$) of the population. Individuals within Cluster 1 were young, non-obese and did not snore (Tables S1–S4). They had the lowest AHI (i.e. no-OSA) and no hypoxic load during sleep. They also indicated a higher level of vigilance compared with other clusters. However, this result did not reach statistical significance ($p > 0.05$). In terms of the duration of respiratory events, these clusters had the shortest apnea duration.

Cluster 2 was the most densely populated group. This cluster included individuals with mild OSA who most often reported excessive daytime sleepiness (EDS; median ESS ≥ 11; Tables S1–S4). The individuals in this cluster had good sleep quality, as evidenced by the greatest sleep efficiency, longest TST and shortest wake after sleep onset (WASO). In addition, the individuals in this cluster generally had low hypoxic load.

Cluster 3 corresponded to individuals who had the highest number of PLM events during the night (Tables S1–S4). Also, individuals in this cluster had a significantly higher number of respiratory events and more severe hypoxic load in comparison with individuals within Cluster 2 ($p < 0.05$). However, they experienced fewer central apnea compared with Cluster 2.

Cluster 4 comprised individuals with severe OSA, having longer breathing disturbances, higher hypoxic load, and more snoring compared with individuals in Clusters 2 and 3 (Tables S1–S4). Furthermore, although individuals in Cluster 4 had a significantly higher frequency of arousals compared with those in Cluster 3, they had significantly better sleep quality in most clustering methods (i.e. the case of AHC, K-means and FCM).

Cluster 5 corresponded to individuals with extremely severe OSA (median AHIs ≥ 73.5; Figure S1) and hypoxic load. They were obese, snorers, and had the lowest number of PLM events (Tables S1–S4). They exhibited poor sleep quality, characterized by a high percentage of N1 sleep, and very low percentages of N3 and REM sleep. Individuals within this cluster also experienced the highest frequency of arousal during sleep ($p < 0.05$). In terms of HRV parameters, individuals in Cluster 5 tended to have the highest values of the LF/HF ratio with the shortest average of RR intervals.

TABLE 2 Demographic, clinical and PSG characteristics of the applied dataset are reported as medians (interquartile range) or counts (percentage).

Characteristic	All ($n = 865$)	No OSA (AHI < 5) ($n = 135$)	Mild OSA (5 ≤ AHI < 15) ($n = 245$)	Moderate OSA (15 ≤ AHI < 30) ($n = 206$)	Severe OSA (AHI ≥ 30) ($n = 279$)
Female (%)	392 (45.3)	89 (65.9)	133 (54.3)	87 (42.2)	83 (29.7)
Age (years)	56.0 (45.0–66.0)	45.0 (32.0–59.0)	55.0 (45.0–64.5)	56.0 (48.0–66.0)	59.0 (48.0–69.0)
BMI (kg m^{-2})	34.0 (29.0–40.0)	29.4 (25.2–35.2)	33.6 (28.4–38.9)	34.0 (30.4–39.8)	36.5 (31.9–43.4)
TST (min)	309.0 (254.0–360.0)	344.0 (281.0–382.0)	324.0 (265.0–374.0)	309.0 (260.0–356.0)	284.0 (222.0–332.0)
Sleep efficiency (%)	72.0 (60.0–83.0)	79.0 (67.0–87.0)	76.0 (63.0–85.0)	70.0 (60.0–81.0)	66.0 (54.0–76.0)
WASO (min)	96.0 (57.0–142.0)	64.5 (42.5–106.0)	77.5 (46.0–119.0)	102.0 (61.0–137.1)	124.5 (86.5–170.5)
Percentage of stage N1 (% of TST)	11.0 (7.0–19.0)	6.0 (3.0–9.0)	9.0 (5.0–13.0)	11.0 (7.0–17.0)	20.0 (12.0–32.0)
Percentage of stage N2 (% of TST)	48.0 (41.0–56.0)	50.0 (42.0–57.0)	50.0 (43.0–57.0)	49.0 (42.0–57.0)	46.0 (38.0–54.0)
Percentage of stage N3 (% of TST)	18.0 (10.0–27.0)	22.0 (14.0–31.0)	21.0 (13.0–30.0)	19.0 (11.0–28.0)	13.0 (5.0–22.0)
Percentage of stage REM (% of TST)	17.0 (12.0–22.0)	18.0 (14.0–25.0)	19.0 (14.0–23.0)	17.0 (12.0–22.0)	14.0 (10.0–20.0)
AHI (events per hr)	18.0 (8.0–38.0)	2.5 (1.3–3.6)	9.6 (7.3–12.1)	21.2 (17.6–25.5)	51.5 (39.0–73.1)
Oxygen desaturation index (events per hr)	13.0 (3.0–32.0)	1.2 (0.4–2.5)	4.9 (2.1–9.3)	14.2 (9.1–21.2)	44.4 (30.5–65.7)

AHI, apnea–hypopnea index; BMI, body mass index; OSA, obstructive sleep apnea; REM, rapid eye movement; TST, total sleep time; WASO, wake after sleep onset.

In addition to similar physiological features that were consistently discovered between clustering methods, several physiological characteristics differed across the clustering methods. For instance, the individuals within Cluster 4 discovered by AHC exhibited low sympathetic activity and generally good HRV, while other clustering methods did not find similar characteristics in Cluster 4. In terms of the duration of respiratory events, inconsistencies were found between clustering methods; in most of the cases, Cluster 1 had the shortest, and Cluster 5 had the longest apnea durations, whereas the hypopnea durations differed across clustering methods.

3.2 | Agreements between the clustering methods

The highest agreement in the clustering of the individuals into the same clusters was achieved between FCM and K-means with Cohen's kappa reaching an excellent value of $\kappa = 0.87$ (Table 3). Additionally, a good agreement was obtained between GMM and K-means ($\kappa = 0.70$). On the other hand, AHC demonstrated moderate agreement with other clustering methods. The weakest agreement was observed between the GMM and AHC methods with $\kappa = 0.51$.

3.3 | Internal validation of the clustering methods

In the internal validation of the clustering methods, K-means outperformed both soft clustering methods (i.e. FCM and GMM) as well as AHC with a SC score of 0.090, a DBI score of 2.386, and a CHI score of 102.037 (Table 4). According to the DI, AHC had the best performance with a score of 0.097. When comparing soft clustering methods, FCM consistently outperformed GMM in all criteria (SC = 0.081 versus 0.060, DBI = 2.510 versus 2.838, CHI = 100.989 versus 92.226, and DI = 0.093 versus 0.071). In summary, although AHC had the best performance based on the DI, K-means demonstrated the best performance followed by FCM in the other three remaining criteria.

3.4 | Visualization of the clusters

To visually represent the cluster formations, we reduced 42 dimensions of input data into two-dimensional embeddings using t-SNE and visualized the cluster formations using the output labels of each method (Figure 4). Additionally, in Figure 4(e), we utilized traditional OSA severities as labels to represent the diversity of OSA severity among individuals within each cluster across all methods.

As seen in Figure 4, Cluster 1 was completely separated from the other clusters with all clustering methods. In addition, individuals in Cluster 5 were identified highly similarly across all clustering methods. However, the formation of Clusters 2–4 varied among different clustering methods. That is, individuals within these clusters showed inconsistencies in clustering outputs and physiological characteristics (Figures 2 and 3). According to membership probabilities and cluster formations, Cluster 3 was surrounded by Clusters 2 and 4, and had

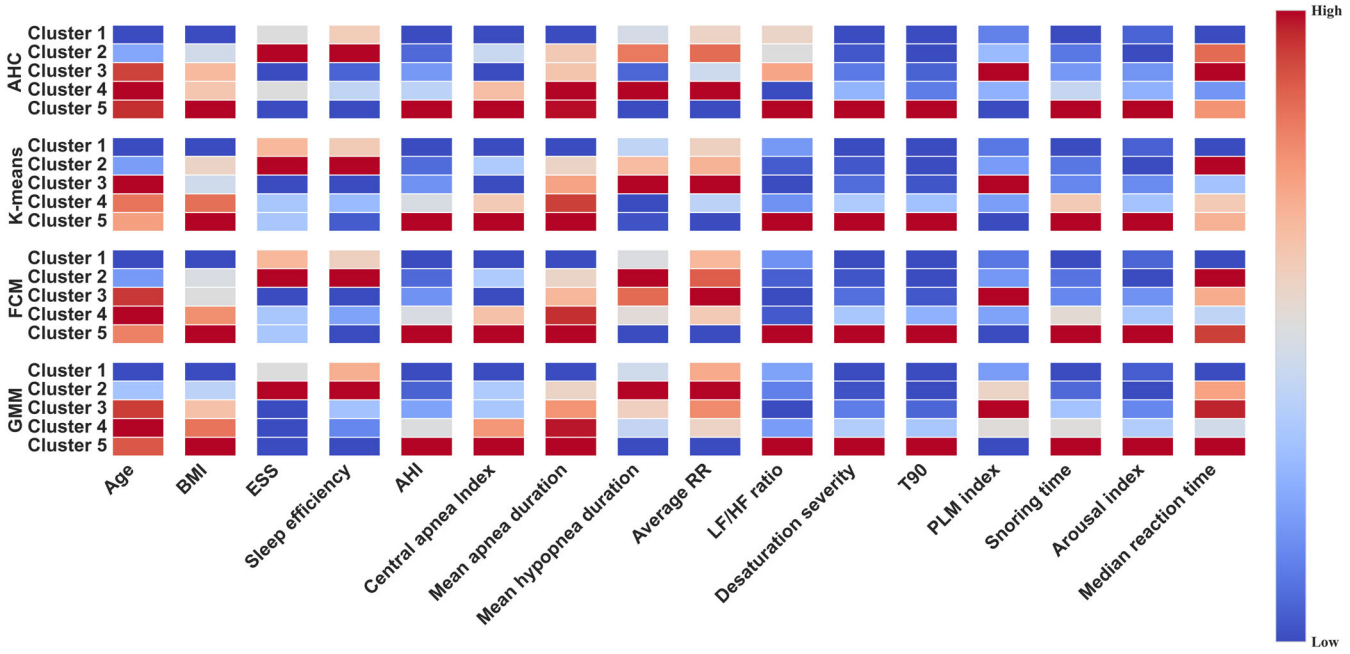

FIGURE 2 Illustration of a subset of clustering variables with heatmap. The absolute values for all clustering variables with all clustering methods are presented in Tables S1–S4. AHI, apnea–hypopnea index; average RR, average of RR intervals; BMI, body mass index; ESS, Epworth Sleepiness Scale; LF/HF ratio, low-frequency to high-frequency ratio; PLM, periodic limb movement; T90, percentage of time spent below 90% oxygen saturation.

complex mixed boundaries with them. Thus, we summarized the membership probability of individuals within Cluster 3 categorized by FCM and GMM (Table 5). For instance, those individuals who were categorized to Cluster 3 by FCM had probabilities of 0.224 and 0.112 to belong to Clusters 2 and 4, respectively. As another example, individuals with no OSA within Cluster 3 showed a membership probability of 0.334 for Cluster 2 and a membership probability of 0.053 for Cluster 1. GMM tended to produce high posterior probabilities.

4 | DISCUSSION

In this study, we applied four clustering methods to examine how different clustering methods affect cluster formation and physiological outcomes in a highly obese (BMI ≥ 30 kg m^{-2}, 72.37%) population of 865 individuals suspected of having OSA. The comparison revealed that clearly separated observations were clustered highly similarly across all methods (Figure 4). However, with overlapping clusters, the agreement between methods decreased. This may suggest that, regardless of the selected set of variables, the agreement between clustering methods is likely to decrease when dealing with overlapping clusters within the data. In such scenarios, the choice of clustering method can affect the cluster formations and physiological characteristics (Figures 2 and 3; Tables S1–S4).

In our study, Cohen's kappa values indicated the weakest agreement between AHC and the other three clustering methods (i.e. the case of K-means, FCM and GMM; Table 3). AHC is a subgroup of hierarchical clustering algorithms, while other methods utilized in this study, like K-means, are partitional clustering algorithms (Pitafi et al., 2023). Therefore, this weak agreement is likely due to the different clustering mechanisms and structures. It should be noted that AHC and K-means have been primarily utilized in the latest applications of OSA phenotyping (Bazoukis et al., 2023). This indicates that clustering outputs can be highly sensitive to the methods derived by different clustering mechanisms. Thus, it would be beneficial to validate the reproducibility of the discovered clusters by using methods with varying mechanisms. On the other hand, K-means and FCM were in excellent agreement, suggesting that the formed clusters are highly similar (Table 3). This discrepancy in agreements between different clustering methods is something to be aware of when comparing the results between different publications. However, it has to be stated that our study does not provide an answer to which of the methods is the best from the clinical point of view. Further studies are therefore required to link different clusters and clustering methods to OSA-related symptoms, co-morbidities and other longitudinal outcomes.

We also evaluated the performances of clustering methods using internal evaluation indices. K-means demonstrated the best performance in discovering well-separated and cohesive groups of individuals. FCM also demonstrated satisfactory performance, outperforming GMM in all performance indices and surpassing AHC in most of the indices (Table 4). FCM also showed great potential to handle overlapping clusters by providing membership probabilities for all individuals (Table 5). As FCM exhibited more flexibility compared with other methods, FCM has great potential to be considered in the current application of OSA phenotyping. However, as stated above,

	AHC	K-means	FCM	GMM
Cluster 1	N:19 Age:43.7 BMI:26.7 ESS:10.0 AHI:2.4 — young, non-obese, no-OSA, no hypoxic load, vigilant, non-snorers, shortest apnea durations	N:19 Age:43.7 BMI:26.7 ESS:10.0 AHI:2.4 — young, non-obese, no-OSA, no hypoxic load, vigilant, non-snorers, shortest apnea durations	N:19 Age:43.7 BMI:26.7 ESS:10.0 AHI:2.4 — young, non-obese, no-OSA, no hypoxic load, vigilant, non-snorers, shortest apnea durations	N:19 Age:43.7 BMI:26.7 ESS:10.0 AHI:2.4 — young, non-obese, no-OSA, no hypoxic load, vigilant, non-snorers, shortest apnea durations
Cluster 2	N:260 Age:47.3 BMI:32.3 ESS:11.0 AHI:7.4 — EDS, well-structured sleep, mild OSA, lowest hypoxic load, lowest arousal frequency	N:315 Age:47.3 BMI:33.6 ESS:11.0 AHI:8.1 — EDS, well-structured sleep, mild OSA, lowest hypoxic load, non-vigilant, lowest arousal frequency	N:303 Age:47.1 BMI:32.7 ESS:11.0 AHI:7.6 — EDS, well-structured sleep, mild OSA, lowest hypoxic load, non-vigilant, lowest arousal frequency	N:301 Age:48.9 BMI:31.4 ESS:11.0 AHI:6.7 — EDS, well-structured sleep, mild OSA, lowest hypoxic load, lowest arousal frequency
Cluster 3	N:236 Age:58.7 BMI:34.8 ESS:9.0 AHI:15.4 — moderate OSA, lowest central apnea index, highest number of PLM events, non-vigilant	N:240 Age:62.1 BMI:32.2 ESS:8.0 AHI:16.0 — old, moderate OSA, lowest central apnea index, longest hypopnea durations, highest number of PLM events	N:240 Age:60.6 BMI:32.9 ESS:8.0 AHI:15.6 — moderate OSA, highest number of PLM events	N:263 Age:58.9 BMI:34.4 ESS:9.0 AHI:19.5 — moderate OSA, highest number of PLM events, non-vigilant
Cluster 4	N:217 Age:59.9 BMI:34.4 ESS:10.0 AHI:30.0 — old, severe OSA, severe hypoxic load, longest apnea and hypopnea durations, highest average RR, SDRR, RMSSD, and pRR50, lowest LF/HF ratio	N:202 Age:59.0 BMI:37.0 ESS:9.0 AHI:41.2 — severe OSA, severe hypoxic load, shortest hypopnea durations	N:203 Age:61.6 BMI:36.3 ESS:9.0 AHI:40.1 — old, severe OSA, severe hypoxic load	N:186 Age:59.9 BMI:36.8 ESS:9.0 AHI:42.1 — old, severe OSA, severe hypoxic load
Cluster 5	N:133 Age:59.2 BMI:39.1 ESS:9.0 AHI:73.4 — obese, poor sleep quality, severe OSA, severe hypoxic load, shortest hypopnea durations, highest arousal frequency, lowest number of PLM events, snorers, lowest average RR, highest LF/HF ratio	N:89 Age:57.2 BMI:39.0 ESS:9.0 AHI:83.3 — obese, poor sleep quality, severe OSA, severe hypoxic load, longest apnea durations, highest arousal frequency, lowest number of PLM events, snorers, lowest average RR, highest LF/HF ratio	N:100 Age:58.1 BMI:39.0 ESS:9.0 AHI:80.6 — obese, poor sleep quality, severe OSA, severe hypoxic load, longest apnea and shortest hypopnea durations, highest arousal frequency, lowest number of PLM events, snorers, lowest average RR, highest LF/HF ratio	N:96 Age:58.1 BMI:38.8 ESS:9.0 AHI:82.7 — obese, poor sleep quality, severe OSA, severe hypoxic load, shortest hypopnea durations, longest apnea durations, highest arousal frequency, lowest number of PLM events, snorers, lowest average RR, highest LF/HF ratio

FIGURE 3 Typical characteristics of individuals in different clusters identified with the four compared methods. AHI, apnea–hypopnea index; average RR, average of RR intervals; BMI, body mass index; EDS, excessive daytime sleepiness; ESS, Epworth Sleepiness Scale; LF/HF ratio, low-frequency to high-frequency ratio; *N*, number of individuals within clusters; OSA, obstructive sleep apnea; PLM, periodic limb movement; pRR50, percentage of RR intervals differing by more than 50 ms; REM, rapid eye movement; RMSSD, root mean square of successive differences; SDRR, standard deviation of RR intervals, All values are based on Median.

this has to be confirmed by linking the FCM-defined clusters to OSA-related adverse health consequences.

We incorporated new features into common OSA phenotyping variables, including detailed SpO_2-based variables, HRV parameters and PVT outcomes. The objective of the utilization of such parameters was to provide more detailed information and better characterization of hypoxic load, autonomic nervous system and individual level of vigilance, allowing us to elucidate new phenotypic expressions of OSA. For instance, the included SpO_2 variables, such as desaturation and recovery severities and duration indexes, demonstrated that individuals in clusters with severe OSA characteristics experienced more frequent and longer desaturations, coupled with shorter and less effective recovery durations.

TABLE 3 Cohen's kappa values for pairwise comparisons of clustering methods.

Clustering methods (number of clusters = 5)	Cohen's kappa values among 865 suspected OSA individuals			
	AHC	K-means	FCM ($m = 1.1$)	GMM
AHC	1	–	–	–
K-means	0.56	1	–	–
FCM ($m = 1.1$)	0.60	0.87	1	–
GMM	0.51	0.70	0.71	1

AHC, Agglomerative Hierarchical Clustering; FCM, Fuzzy c-means; GMM, Gaussian Mixture Model; OSA, obstructive sleep apnea.

TABLE 4 Performance evaluation of clustering methods.

Clustering methods (number of clusters = 5)	Cluster validation metrics (index scores) among 865 suspected OSA individuals			
	CHI (↑)	DBI (↓)	DI (↑)	SC (↑)
AHC	88.36	2.673	**0.097**	0.062
K-means	**102.037**	**2.386**	0.092	**0.090**
FCM ($m = 1.1$)	100.989	2.510	0.093	0.081
GMM	92.226	2.838	0.071	0.060

Note: Bold indicates the best index values.
(↑) and (↓) indicate whether higher or lower values will result in improved performance, respectively.
AHC, Agglomerative Hierarchical Clustering; CHI, Calinski–Harabasz index; DBI, Davies–Bouldin index; DI, Dunn index; FCM, Fuzzy c-means; GMM, Gaussian Mixture Model; OSA, obstructive sleep apnea; SC, Silhouette coefficient.

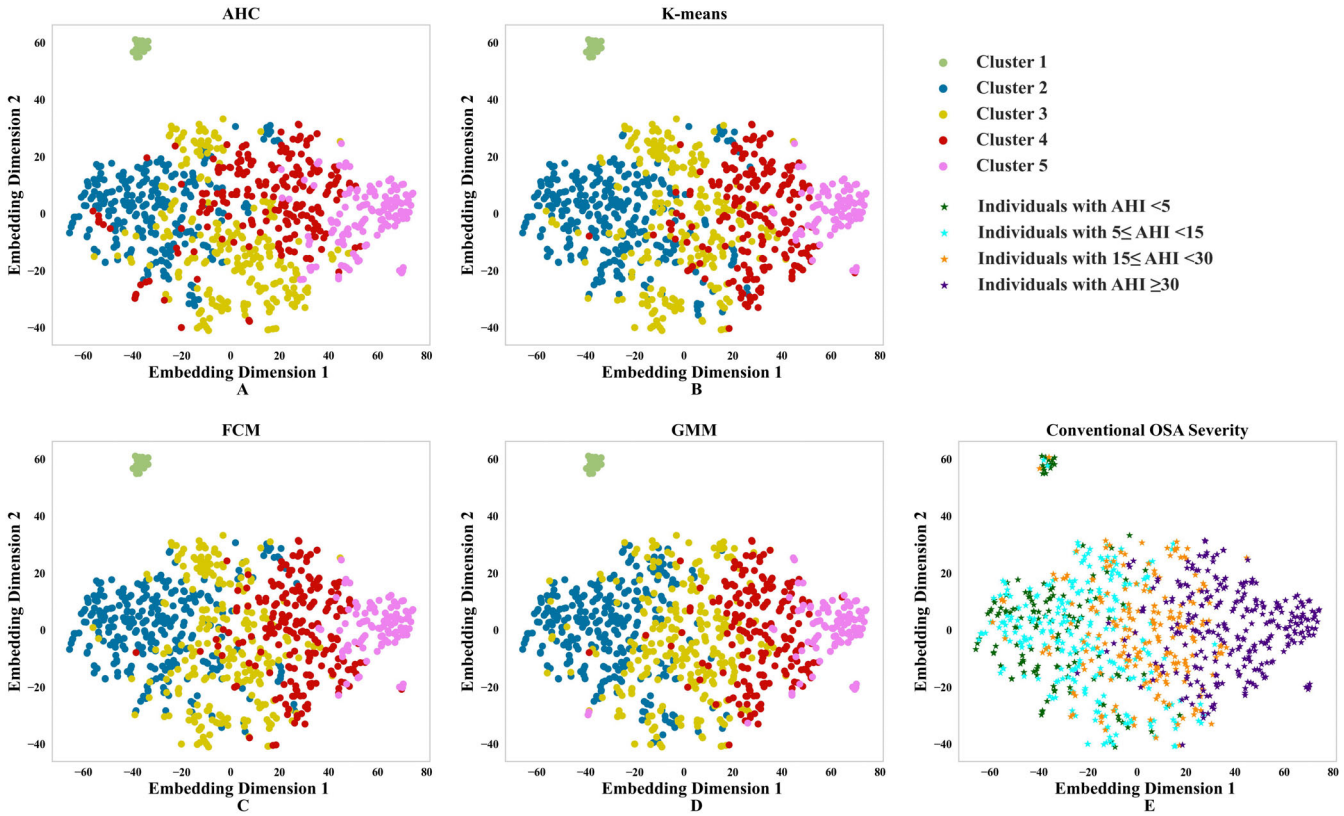

FIGURE 4 Visualization of cluster formations using clustering labels and conventional OSA severity categorization. AHC, agglomerative hierarchical clustering; AHI, apnea-hypopnea index; FCM, Fuzzy c-means; GMM, Gaussian Mixture Model; OSA, obstructive sleep apnea.

In this study, we identified a broad spectrum of phenotypic characteristics that were in line with previous literature. Cluster 1 primarily included the mixture of individuals with no OSA (i.e. AHI < 5, $n = 15$, 79.0%) and a few individuals with a higher frequency of respiratory events (i.e. 5 ≤ AHI < 15, $n = 2$, 10.5%; and 15 ≤ AHI < 30, $n = 2$, 10.5%; Figure 4). One common characteristic among these individuals was the absence of hypoxemia, despite experiencing few hypopneas but basically no apneas (Tables S1–S4). In addition, several physiological characteristics of individuals within Cluster 1 are highly similar to those described in "Cluster 1" by Eun-Yeol et al. (Ma et al., 2021), where the individuals were young and non-obese, with good sleep quality and few respiratory and desaturation events.

Individuals in Cluster 2 had the best sleep quality, marked by the greatest sleep efficiency, longest TST and shortest WASO (Tables S1–S4). However, these individuals reported the worst daytime sleepiness based on the Epworth Sleepiness Scale (ESS) scores. Monitoring and managing EDS is essential as it increases the risk of motor vehicle and occupational accidents as well as decreases the overall quality of life (Garbarino et al., 2016; Lal et al., 2021; Tregear et al., 2009). Furthermore, several previous studies related to OSA phenotyping have demonstrated daytime sleepiness to be one of the clustering characteristics (Bailly et al., 2021; Gasa et al., 2023; Ida et al., 2022). However, the main characteristics of EDS clusters differ between studies. Our findings in Cluster 2 are the most closely in line with "Cluster 2" described by Gasa et al. (2023), where individuals had mild OSA, good sleep quality, yet reported symptoms of daytime sleepiness.

Cluster 3 comprised individuals with traditional OSA severities ranging from no-OSA to severe (Figure S1), and the highest number of PLM events (Tables S1–S4). Furthermore, Cluster 3, located between Clusters 2 and 4 with heavily mixed boundaries (Figure 4) showing high uncertainty and a wide range of membership probabilities (Table 5). This indicates that individuals in Cluster 3 can be easily mixed with adjacent clusters. Moreover, it is possible that the night-to-night variability, in terms of the AHI, plays an important role in Cluster 3 individuals as the majority of them have mild-to-moderate OSA and are more susceptible to such variation compared with those with severe OSA (Stöberl et al., 2017). Therefore, by considering all these aspects, it is likely that the identification and correct classification of Cluster 3 individuals is the most complex and challenging task among the found clusters. Based on these findings, it is possible to further speculate that planning, tailoring and managing effective treatment for these individuals would be a challenge.

Cluster 4 contained individuals with severe OSA. Although this cluster was more extreme than Cluster 3 regarding most of the physiological domains, they had better sleep quality in most cases. The individuals in Cluster 5 also exhibited severe OSA, but on more extreme scales as the PSG parameters were significantly worse compared with Cluster 4 (Tables S1–S4). In our study, the main characteristics of Clusters 4 and 5 are highly similar to "Cluster 3, moderate to severe OSA with hypopnea" and "Cluster 4, severe OSA with hypoxaemia" described by Kim et al. (2020), respectively. In Kim's study, "Cluster 4" was marked by very severe OSA, severe obesity, the highest AHI and the most oxygen desaturation events, while "Cluster 3" included older individuals with more PLM events and longer apnea–hypopnea durations compared with "Cluster 4".

In this study, the utilization of soft clustering methods enabled the identification of specific subgroups of clusters that presented multiple phenotypic characteristics (Table 5). For instance, we showed that individuals in Cluster 3 had various membership probabilities to multiple clusters, and this information can further assist in clinical decision-making. For example, there may be a necessity to closely monitor the worsening of symptoms and comorbidities in individuals within Cluster 3 who showed a considerable probability of being members of Cluster 4, where OSA was more severe in general.

Our study is not without limitations. In this study, we only utilized continuous variables derived from PSG and PVT, while other possible phenotypic variables such as sex, ethnicity, dentofacial characteristics,

TABLE 5 Example of membership probabilities for individuals belonging to Cluster 3.

Clusters	Soft clustering methods	Median membership probabilities of individuals within cluster 3				
		Individuals with AHI ≤ 5	Individuals with 5 ≤ AHI < 15	Individuals with 15 ≤ AHI < 30	Individuals with 30 ≤ AHI	ALL individuals
Cluster 1	FCM	0.005	0.005	0.001	0.002	0.003
	GMM	0.005	0.000	0.000	0.000	0.001
Cluster 2	FCM	0.334	0.267	0.170	0.117	0.224
	GMM	0.060	0.040	0.020	0.000	0.020
Cluster 3	FCM	0.575	0.605	0.580	0.520	0.582
	GMM	0.930	0.950	0.970	0.920	0.940
Cluster 4	FCM	0.045	0.065	0.210	0.310	0.112
	GMM	0.000	0.040	0.005	0.063	0.027
Cluster 5	FCM	0.000	0.000	0.000	0.000	0.002
	GMM	0.000	0.000	0.000	0.000	0.000

AHI, apnea–hypopnea index; FCM, Fuzzy C-means; GMM, Gaussian Mixture Model.

symptoms and comorbidities were not considered (Eckert et al., 2013; Ferreira-Santos & Rodrigues, 2023; Subramanian et al., 2011; Ye et al., 2009). Including these variables in PSG-based phenotyping could further help to understand the pathophysiology of OSA, and lead to the identification of optimal treatments and improvements in prognostics (Zinchuk & Yaggi, 2020). However, this selection was done with a purpose: being able to identify OSA phenotypes based on continuous parameters such as age, BMI, ESS, PSG metrics and PVT measurements could help in clinical decision-making without having information on OSA-related comorbidities. That is, if certain PSG phenotypes (i.e. OSA clusters) are related to high risks of, for example, cardiovascular consequences, this information can be used for planning further clinical considerations or treatments. We also set the number of clusters to five for this study to have clinically interpretable clusters and to fairly compare the outputs of different clustering methods. Considering a fixed number of clusters for each method might present a limitation in this study as different clustering algorithms operate using unique mechanisms and might include varying numbers of clusters based on different model selection criteria.

In conclusion, we compared the physiological outcomes of five clusters across assorted unsupervised machine-learning methods where Cluster 1 consisted of individuals with no OSA, Cluster 2 consisted of sleepy individuals with mild OSA and good sleep quality, Cluster 3 consisted of a mix of different OSA severities and a high number of PLM events, Cluster 4 consisted of individuals with traditional severe OSA, and Cluster 5 consisted of individuals with extremely severe OSA and poor sleep quality. However, some characteristics like the duration of respiratory events and HRV parameters differed across the clustering methods. The results of this study indeed highlight the effect of clustering methods on cluster formations and physiological outcomes. Also, our findings showed that FCM demonstrated satisfactory performance in terms of identifying distinct OSA clusters and handling complex and mixed structures within the data. By utilizing fuzzy clustering methods, we can identify subgroups of individuals within clusters who might indicate multiple phenotypic characteristics. This can significantly reduce the discrepancies between clustering methods, and enhance the robustness and reliability of the clustering findings.

AUTHOR CONTRIBUTIONS

Mohammadreza Ghorvei: Conceptualization; investigation; funding acquisition; writing – original draft; writing – review and editing; visualization; validation; software; formal analysis; resources; data curation; methodology. **Tuomas Karhu:** Writing – review and editing; methodology; visualization; software; data curation; resources; investigation. **Salla Hietakoste:** Resources; data curation; software; formal analysis; methodology; writing – review and editing; investigation. **Daniela Ferreira-Santos:** Writing – review and editing; project administration; supervision. **Harald Hrubos-Strøm:** Writing – review and editing. **Anna Sigridur Islind:** Writing – review and editing. **Luka Biedebach:** Writing – review and editing. **Sami Nikkonen:** Conceptualization; investigation; funding acquisition; writing – review and editing; validation; visualization; methodology; software; formal analysis; project administration; resources; supervision; data curation. **Timo Leppänen:** Data curation; supervision; resources; project administration; formal analysis; software; methodology; validation; visualization; writing – review and editing; funding acquisition; investigation; conceptualization. **Matias Rusanen:** Conceptualization; investigation; funding acquisition; writing – review and editing; visualization; validation; methodology; project administration; formal analysis; software; data curation; supervision; resources.

ACKNOWLEDGEMENTS

MG has received research funding from the Respiratory Foundation of Kuopio Region, Orion Research Foundation, Tampere Tuberculosis Foundation, and The State Research Funding for university-level health research, Kuopio University Hospital, Wellbeing Service County of North Savo (project number 5041828). TK has received research funding from The State Research Funding for university-level health research, Kuopio University Hospital, Wellbeing Service County of North Savo (project number 5041798). SH has received research funding from The State Research Funding for university-level health research, Kuopio University Hospital, Wellbeing Service County of North Savo (project number 5041790) and the Research Foundation of the Pulmonary Diseases. HHS has received research funding from the European Union's Horizon 2020 Research and Innovation Programme (decision number 965417). SN has received research funding from The State Research Funding for university-level health research, Kuopio University Hospital, Wellbeing Service County of North Savo (project number 5041809). TL has received research funding from the European Union's Horizon 2020 Research and Innovation Programme (decision number 965417), and The State Research Funding for university-level health research, Kuopio University Hospital, Wellbeing Service County of North Savo (project number 5041794). MR's work was supported by research funding from the French National Research Agency, MIAI@Grenoble Alpes (ANR-15-IDEX-02 & ANR-19-P3IA-0003).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest relevant to this study.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Mohammadreza Ghorvei https://orcid.org/0000-0002-3303-5969
Tuomas Karhu https://orcid.org/0000-0001-9464-447X
Salla Hietakoste https://orcid.org/0000-0001-5355-6203
Daniela Ferreira-Santos https://orcid.org/0000-0002-0390-9944
Harald Hrubos-Strøm https://orcid.org/0000-0003-0065-0145
Anna Sigridur Islind https://orcid.org/0000-0002-4563-0001
Luka Biedebach https://orcid.org/0000-0003-0974-202X
Sami Nikkonen https://orcid.org/0000-0003-0615-4118

Timo Leppänen https://orcid.org/0000-0003-4017-821X
Matias Rusanen https://orcid.org/0000-0001-9633-6016

REFERENCES

Alashwal, H., El Halaby, M., Crouse, J. J., Abdalla, A., & Moustafa, A. A. (2019). The application of unsupervised clustering methods to Alzheimer's disease. *Frontiers in Computational Neuroscience*, *13*, 31.

Askari, S. (2021). Fuzzy C-means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development. *Expert Systems with Applications*, *165*, 113856.

Bahmani, B., Moseley, B., Vattani, A., Kumar, R., & Vassilvitskii, S. (2012). Scalable k-means++. arXiv Preprint arXiv:1203.6402.

Bailly, S., Grote, L., Hedner, J., Schiza, S., McNicholas, W. T., Basoglu, O. K., Lombardi, C., Dogas, Z., Roisman, G., Pataka, A., Bonsignore, M. R., Pepin, J. L., & ESADA Study Group. (2021). Clusters of sleep apnoea phenotypes: A large pan-European study from the European sleep Apnoea database (ESADA). *Respirology*, *26*, 378–387.

Bazoukis, G., Bollepalli, S. C., Chung, C. T., Li, X., Tse, G., Bartley, B. L., Batool-Anwar, S., Quan, S. F., & Armoundas, A. A. (2023). Application of artificial intelligence in the diagnosis of sleep apnea. *Journal of Clinical Sleep Medicine*, *19*, 1337–1363.

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computational Geosciences*, *10*, 191–203.

Bonsignore, M. R., Giron, M. C. S., Marrone, O., Castrogiovanni, A., & Montserrat, J. M. (2017). Personalised medicine in sleep respiratory disorders: Focus on obstructive sleep apnoea diagnosis and treatment. *European Respiratory Review*, *26*, 170069.

Cai, T. T., Ma, J., & Zhang, L. (2019). CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality. *Annals of Statistics*, *47*, 1234–1267. https://doi.org/10.1214/18-AOS1711

Cebeci, Z., & Cebeci, C. (2018). Initialization Algorithms for Partitioning Cluster Analysis. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)* (pp. 1–5). IEEE. https://doi.org/10.1109/IDAP.2018.8620920

Cebeci, Z. (2019). Comparison of internal validity indices for fuzzy clustering. *Journal of Agricultural Informatics*, 1–14. https://doi.org/10.17700/jai.2019.10.2.537

de Araujo Dantas, A. B., Gonçalves, F. M., Martins, A. A., Alves, G. Â., Stechman-Neto, J., Corrêa, C. C., Santos, R. S., Nascimento, W. V., de Araujo, C. M., & Taveira, K. V. M. (2023). Worldwide prevalence and associated risk factors of obstructive sleep apnea: A meta-analysis and meta-regression. *Sleep and Breathing*, *1–27*, 2083–2109.

de Chazal, P., Sutherland, K., & Cistulli, P. A. (2020). Advanced polysomnographic analysis for OSA: A pathway to personalized management? *Respirology*, *25*, 251–258.

Eckert, D. J., White, D. P., Jordan, A. S., Malhotra, A., & Wellman, A. (2013). Defining phenotypic causes of obstructive sleep apnea. Identification of novel therapeutic targets. *American Journal of Respiratory and Critical Care Medicine*, *188*, 996–1004.

Ferreira-Santos, D., & Rodrigues, P. P. (2023). Obstructive sleep apnea: A categorical cluster analysis and visualization. *Pulmonology*, *29*, 207–213.

Flynt, A., & Dean, N. (2016). A survey of popular R packages for cluster analysis. *Journal of Educational and Behavioral Statistics*, *41*, 205–225.

Garbarino, S., Guglielmi, O., Sanna, A., Mancardi, G. L., & Magnavita, N. (2016). Risk of occupational accidents in workers with obstructive sleep apnea: Systematic review and meta-analysis. *Sleep*, *39*, 1211–1218.

Gasa, M., Salord, N., Fontanilles, E., Pérez Ramos, S., Prado, E., Pallarés, N., Santos Pérez, S., & Monasterio, C. (2023). Polysomnographic phenotypes of obstructive sleep apnea in a real-life cohort: A pathophysiological approach. *Archivos de Bronconeumología*, *59*, 638–644.

Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, *11*, 40–56.

Ida, H., Suga, T., Nishimura, M., & Aoki, Y. (2022). Unique clinical phenotypes of patients with obstructive sleep apnea in a Japanese population: A cluster analysis. *Journal of Clinical Sleep Medicine*, *18*, 895–902.

Jaeger, A., & Banks, D. (2023). Cluster analysis: A modern statistical review. *Wiley Interdisciplinary Reviews: Computational Statistics*, *15*, e1597.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*, 302–317.

Karhu, T., Leppänen, T., Töyräs, J., Oksenberg, A., Myllymaa, S., & Nikkonen, S. (2022). ABOSA–freely available automatic blood oxygen saturation signal analysis software: Structure and validation. *Computer Methods and Programs in Biomedicine*, *226*, 107120.

Kim, J.-W., Won, T. B., Rhee, C. S., Park, Y. M., Yoon, I. Y., & Cho, S. W. (2020). Polysomnographic phenotyping of obstructive sleep apnea and its implications in mortality in Korea. *Scientific Reports*, *10*, 13207.

Kulkas, A., Tiihonen, P., Julkunen, P., Mervaala, E., & Töyräs, J. (2013). Novel parameters indicate significant differences in severity of obstructive sleep apnea with patients having similar apnea–hypopnea index. *Medical & Biological Engineering & Computing*, *51*, 697–708.

Lal, C., Weaver, T. E., Bae, C. J., & Strohl, K. P. (2021). Excessive daytime sleepiness in obstructive sleep apnea. Mechanisms and clinical management. *Annals of the American Thoracic Society*, *18*, 757–768.

Ma, E.-Y., Kim, J. W., Lee, Y., Cho, S. W., Kim, H., & Kim, J. K. (2021). Combined unsupervised-supervised machine learning for phenotyping complex diseases with its application to obstructive sleep apnea. *Scientific Reports*, *11*, 4457.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297). University of California Press.

Malhotra, A., Ayappa, I., Ayas, N., Collop, N., Kirsch, D., Mcardle, N., Mehra, R., Pack, A. I., Punjabi, N., White, D. P., & Gottlieb, D. J. (2021). Metrics of sleep apnea severity: Beyond the apnea-hypopnea index. *Sleep*, *44*, zsab030.

McNicholas, W. T., & Korkalainen, H. (2023). Translation of obstructive sleep apnea pathophysiology and phenotypes to personalized treatment: A narrative review. *Frontiers in Neurology*, *14*, 1239016.

Pitafi, S., Anwar, T., & Sharif, Z. (2023). A taxonomy of machine learning clustering algorithms, challenges, and future realms. *Applied Sciences*, *13*, 3529.

Ran, X., Xi, Y., Lu, Y., Wang, X., & Lu, Z. (2022). Comprehensive survey on hierarchical clustering algorithms and the recent developments. *Artificial Intelligence Review*, *1–46*, 8219–8264.

Randerath, W., de Lange, J., Hedner, J., Ho, J. P. T. F., Marklund, M., Schiza, S., Steier, J., & Verbraecken, J. (2022). Current and novel treatment options for obstructive sleep apnoea. *ERJ Open Research*, *8*, 126–2022.

Randerath, W., Verbraecken, J., de Raaff, C. A. L., Hedner, J., Herkenrath, S., Hohenhorst, W., Jakob, T., Marrone, O., Marklund, M., McNicholas, W. T., Morgan, R. L., Pepin, J. L., Schiza, S., Skoetz, N., Smyth, D., Steier, J., Tonia, T., Trzepizur, W., van Mechelen, P. H., & Wijkstra, P. (2021). European Respiratory Society guideline on non-CPAP therapies for obstructive sleep apnoea. *European Respiratory Review*, *30*, 210200.

Schaffer, C. M., & Green, P. E. (1996). An empirical comparison of variable standardization methods in cluster analysis. *Multivariate Behavioral Research*, *31*, 149–167.

Semelka, M., Wilson, J., & Floyd, R. (2016). Diagnosis and treatment of obstructive sleep apnea in adults. *American Family Physician*, *94*, 355–360.

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, *5*, 258.

Stöberl, A. S., Schwarz, E. I., Haile, S. R., Turnbull, C. D., Rossi, V. A., Stradling, J. R., & Kohler, M. (2017). Night-to-night variability of obstructive sleep apnea. *Journal of Sleep Research*, *26*, 782–788.

Subramanian, S., Guntupalli, B., Murugan, T., Bopparaju, S., Chanamolu, S., Casturi, L., & Surani, S. (2011). Gender and ethnic differences in prevalence of self-reported insomnia among patients with obstructive sleep apnea. *Sleep and Breathing*, *15*, 711–715.

Sutherland, K., Kairaitis, K., Yee, B. J., & Cistulli, P. A. (2018). From CPAP to tailored therapy for obstructive sleep Apnoea. *Multidisciplinary Respiratory Medicine*, *13*, 1–13.

Tam, S., Woodson, B. T., & Rotenberg, B. (2014). Outcome measurements in obstructive sleep apnea: Beyond the apnea-hypopnea index. *Laryngoscope*, *124*, 337–343.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-Aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV–heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*, 210–220.

Torkkola, K. (2003). Feature extraction by non-parametric mutual information maximization. *Journal of Machine Learning Research*, *3*, 1415–1438.

Tregear, S., Reston, J., Schoelles, K., & Phillips, B. (2009). Obstructive sleep apnea and risk of motor vehicle crash: Systematic review and meta-analysis. *Journal of Clinical Sleep Medicine*, *5*, 573–581.

Troester, M., Quan, S., & Berry, R. (2023). The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications, Version 3. *Darien Illinois: American Academy of Sleep Medicine*.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 11.

Verbraecken, J., Dieltjens, M., op de Beeck, S., Vroegop, A., Braem, M., Vanderveken, O., & Randerath, W. (2022). Non-CPAP therapy for obstructive sleep apnoea. *Breathe*, *18*, 220164.

Ye, L., Pien, G. W., & Weaver, T. E. (2009). Gender differences in the clinical manifestation of obstructive sleep apnea. *Sleep Medicine*, *10*, 1075–1084.

Zhang, X. L., Zhang, L., Li, Y. M., Xiang, B. Y., Han, T., Wang, Y., & Wang, C. (2022). Multidimensional assessment and cluster analysis for OSA phenotyping. *Journal of Clinical Sleep Medicine*, *18*, 1779–1788.

Zhu, W., Xiang, L., Long, Y., Xun, Q., Kuang, J., & He, L. (2023). Cluster analysis of clinical phenotypic heterogeneity in obstructive sleep apnea assessed using photoplethysmography. *Sleep Medicine*, *102*, 134–141.

Zinchuk, A., & Yaggi, H. K. (2020). Phenotypic subtypes of OSA: A challenge and opportunity for precision medicine. *Chest*, *157*, 403–420.

Zinchuk, A. V., Gentry, M. J., Concato, J., & Yaggi, H. K. (2017). Phenotypes in obstructive sleep apnea: A definition, examples and evolution of approaches. *Sleep Medicine Reviews*, *35*, 113–123.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ghorvei, M., Karhu, T., Hietakoste, S., Ferreira-Santos, D., Hrubos-Strøm, H., Islind, A. S., Biedebach, L., Nikkonen, S., Leppänen, T., & Rusanen, M. (2025). A comparative analysis of unsupervised machine-learning methods in PSG-related phenotyping. *Journal of Sleep Research*, *34*(3), e14349. https://doi.org/10.1111/jsr.14349