Decoding the cardiac response to pediatric sleep apnea: a morphological explainability study based on SHAP analysis of nocturnal ECG signals

C. García-Vicente[1,2], G. C. Gutiérrez-Tobal[1,2], Lidia Enríquez-Meijome[1], C. de León-Morán[1], D. Gozal[3], R. Hornero[1,2]

[1]Biomedical Engineering Group, University of Valladolid, Valladolid, Spain

[2]CIBER de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III Valladolid, Spain

[3]Office of The Dean and Department of Pediatrics, Joan C. Edwards School of Medicine, Marshall University, 1600 Medical Center Dr, Huntington, WV 25701, United States

Abstract

Pediatric obstructive sleep apnea (OSA) is a prevalent breathing disorder linked to cardiovascular diseases. While polysomnography is the diagnostic gold standard, its complexity, high cost, and limited accessibility often lead to underdiagnosis. Therefore, automated methods based on simplified signals, such as single-lead electrocardiogram (ECG), have been developed. While deep learning (DL) models using ECG signals have demonstrated high diagnostic performance, their "black-box" nature hinders clinical adoption. To address this challenge, we propose a novel quantitative explainable artificial intelligence (XAI) framework based on SHAP (SHapley Additive exPlanations) to characterize the cardiac response to pediatric OSA at the morphological level. Building upon an interpretable DL stacking model evaluated on 3,320 nocturnal pediatric ECG recordings, we computed a quantitative SHAP analysis at the ECG segment level using the test subset. Specifically, we calculated Shapley values for ECG signals from the test subset and mapped them to specific waveform components, including QRS complex, ST segment, TP segment, and QT interval. This enabled quantifying the contribution of each component to the estimated apnea hypopnea index (AHI) across the four OSA severities. Results demonstrated a severity-dependent increase in the contribution of the QRS complex and QT and ST intervals to model predictions (p<0.01), supporting their role as key cardiac biomarkers in pediatric OSA. Importantly, an analysis of model misclassifications revealed that wrong predictions in the healthy group exhibited significantly higher Shapley values in the QRS and QT regions (p<0.05), suggesting the detection of subclinical cardiac stress unrecognized by standard AHI criteria. Conversely, misclassified mild OSA cases showed significantly lower values (p<0.05), suggesting an absence of cardiac abnormalities despite apneic events or that the cardiac patterns are similar in both groups. This segment-level SHAP framework could translate DL predictions into clinically meaningful cardiac biomarkers, paving the way for personalized risk stratification in pediatric OSA.

1. Introduction

Pediatric obstructive sleep apnea (OSA) is a prevalent breathing disorder that leads to changes in the cardiovascular system and increases cardiac risk [1]. The standard diagnostic method, polysomnography (PSG), is costly, complex, and uncomfortable, leading to underdiagnosis of children [2]. To solve this problem, studies have focused on the development of automated diagnostic methods based on simplified signals, such as the single-lead electrocardiogram (ECG) [3].

In recent years, deep learning (DL) has shown promising results in estimating pediatric OSA through cardiorespiratory signals [4], [5]. However, DL models are frequently criticized for their lack of transparency. This limitation hinders their integration into clinical practice. In this context, explainable artificial intelligence (XAI) methods have emerged as a promising solution to provide insights into the decision-making processes of DL models [6]. While XAI has been explored in many contexts [6], its application in pediatric OSA remains limited. Moreover, existing studies predominantly rely on qualitative interpretations, which, although useful for visualization, fail to provide clinically applicable knowledge. Specifically, there is a lack of quantitative frameworks capable of translating explainability outputs into meaningful clinical biomarkers.

To overcome these limitations, previous studies from our group applied XAI techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) to interpret DL models based on cardiorespiratory signals, including ECG alone and ECG combined with SpO_2 [5], [7]. These works identified relevant cardiorespiratory patterns associated with OSA and demonstrated that the models leveraged physiologically coherent information. Particularly, our latest study [7] introduced an interpretable stacking DL model using ECG and SpO_2 signals, and SHAP analysis quantitatively explained the contribution of both signals to the model's predictions.

Despite these advances, our previous work focused on assessing interpretability at the signal level, quantifying the global contribution of ECG and SpO_2 to model performance. While that approach was useful for validating the model's reliability, it did not allow for a quantitative understanding of the specific cardiac mechanisms associated with OSA. For this reason, this proposal introduces a quantitative SHAP framework that characterizes the contribution of specific ECG waveform components to pediatric OSA detection. By decomposing the ECG signal into clinically recognized segments (QRS complex, ST segment, TP interval, and QT interval), we move from global interpretability to perform a quantitative analysis of the cardiac system in pediatric OSA.

Accordingly, the objective of this study is to quantitatively characterize the cardiac response to pediatric OSA by leveraging SHAP interpretability for ECG-derived physiological components. Through this approach, we aim to transform DL explainability into clinically meaningful cardiac biomarkers, facilitating both improved understanding of pediatric OSA mechanisms and increased trust in automated diagnostic systems.

2. Methods and Materials

2.1. Signals and subjects

This study used the Childhood Adenotonsillectomy Trial (CHAT) [8] and Pediatric Adenotonsillectomy Trial Snoring (PATS) public datasets [9], and one private dataset from the University of Chicago (UofC). A total of 3,320 valid ECG

Table 1: Demographic and clinical variables of the study population.

Variables	CHAT Training set	PATS Training set	CHAT Validation set	PATS Validation set	CHAT Test set	PATS Test set	UofC Test set
Subjects (*n*)	987 (61.34)	426(58.28)	323 (20.07)	152 (20.79)	299 (18.58)	153 (20.93)	980 (100)
Age (years)	7.00 [2.00]	7.69 [0.00]	7.00 [2.00]	7.50 [0.00]	6.90 [2.00]	7.56 [0.00]	6.00 [6.00]
Males (*n*)	510 (51.67)	221(51.88)	164 (50.77)	71(46.71%)	161 (53.85)	86 (56.21)	379 (38.67)
AHI<1 e/h[(1)] (*n*)	212 (21.48)	163(38.26)	67 (20.74)	72 (47.37)	65 (21.74)	75 (49.02)	173 (17.65)
1≤AHI<5 e/h[(2)](*n*)	487 (49.34)	190(44.60)	167 (51.70)	62 (40.79)	144 (48.16)	53 (34.64)	401 (40.92)
5≤AHI<10e/h[(3)](*n*)	159 (16.11)	29 (6.81)	44 (13.62)	11 (7.24)	49 (16.39)	14 (9.15)	177 (18.06)
AHI≥10 e/h[(4)] (*n*)	129 (13.07)	44 (10.33)	45 (13.93)	7 (4.61)	41 (13.71)	11 (7.19)	229 (23.37)

Data are presented as median [interquartile range] or n (%). BMI: body mass index; AHI: apnea-hypopnea index; e/h: events/hour.
AHI<1 e/h[(1)]: No OSA; 1≤AHI<5 e/h[(2)]: mild OSA; 5≤AHI<10 e/h[(3)]: moderate OSA; AHI≥10 e/h[(4)]: severe OSA.

recordings collected during nocturnal PSG studies of children were analyzed. Recordings were randomly assigned to approximately 60% for training, 20% for validation, and 20% for testing (see Table 1). Each subject was uniquely assigned to a single set to prevent duplication.

2.2. ECG delineation

ECG waveform delineation was performed using an adaptive algorithm for ECG signals at a sampling rate of 100 Hz. QRS complex detection was based on the Pan-Tompkins framework [10]. Additionally, we incorporated regional polarity analysis via local voting (analyzing a window of 9 adjacent beats) to ensure accurate identification of the R-peak even in the presence of signal inversions or extreme noise. Once the R-peak was established, the peaks of the remaining waves (P, Q, S, T) were identified by searching for local extrema within temporal windows based on standard cardiac physiology [11], [12]. Finally, using these reference points, the QRS complex, the ST, PQ, and QT intervals were defined.

2.3. Deep Learning algorithm: Stacking-based CNNs

The selected DL model was based on an algorithm previously developed and evaluated by our group using ECG and SpO_2 signals [7]. The algorithm consisted of a stacking-based regression model built upon convolutional neural networks (CNNs) designed to detect pediatric OSA using these two signals. The network took as input the full overnight recordings of both signals, which were processed independently through CNNs. The model's output was an estimate of the apnea hypopnea index (AHI) at the subject level. The hyperparameter configuration was optimized using the validation dataset, as described in previous studies [5], [13]. The stacking model demonstrated high diagnostic performance in the CHAT test subset (n = 299), achieving a 4-class Cohen's kappa of 0.549, a 4-class accuracy of 70.23%, and a 2-class accuracy of 81.61% for 1 e/h, 92.64% for 5 e/h, and 94.98% for 10 e/h.

2.4. Interpretability using quantitative SHAP

In this study, we have applied the SHAP method to quantitatively interpret the model's decisions on the test subset of the CHAT dataset (n=299). The objective was to assess how specific ECG morphological components contribute to the model's output. For this purpose, we used the DeepExplainer method. DeepExplainer is based on the Deep Learning Important Features algorithm, which assigns contribution scores to individual neurons by propagating activation differences relative to a reference baseline [14]. Specifically, we computed Shapley values for each ECG signal sample to gain insight into how different waveform morphologies contribute to the model's decision-making process. This analysis enabled a detailed characterization of the importance (Shapley values) of distinct morphological ECG patterns and their contribution to the model's detection of OSA severities. The locations of the intervals obtained through ECG delineation (PQ, QRS, ST, and QT segments) were used to index the Shapley values, transforming the sample-level Shapley values into an accumulative quantification. In this way, the contribution of each phase of the cardiac cycle to the estimation of the AHI was evaluated by analyzing the Shapley values at the positions of the detected segments. Finally, Shapley values for each interval (PQ, QRS, ST, and QT segments) were aggregated at the subject level and stratified according to clinical groupings, allowing comparison of morphological contribution patterns across OSA severities (no OSA, mild OSA, moderate OSA, and severe OSA).

2.5. Statistical Analysis by OSA severity

The subjects were categorized into four groups based on the estimated AHI: No OSA (AHI<1), mild OSA (1≤AHI<5), moderate OSA (5≤AHI<10), and severe OSA (AHI≥10). To evaluate the feature importance across these groups, we performed a Kruskal-Wallis test with pairwise comparisons corrected for multiple testing using the Benjamini–Hochberg method. Additionally, to assess misclassifications, Shapley values for each ECG segment were compared between model successes and failures within each severity group using the Mann-Whitney U test, corrected using the Benjamini–Hochberg method.

3. Results

Figure 1 presents boxplots showing the sum of the Shapley values assigned to the QRS complex, QT interval, and ST and TP segments for each subject across the four pediatric OSA severity groups. The results indicate that the total sum of Shapley values in the QRS complex and the QT interval increases with higher severity levels. Moreover, for all ECG components, the progressive increase of the upper quartile is shown as OSA severity increases across all intervals.

Statistical analysis using the Kruskal-Wallis test corrected using Benjamini–Hochberg showed significant differences in QRS complexes and QT intervals in all pairwise comparisons, except for the transition between healthy and mild OSA subjects and moderate and severe OSA (see Table 2). The ST segment differentiated healthy subjects from mild OSA.

Since the QRS complex and the QT interval were found to be the main ECG patterns influencing the model's decisions, we evaluated boxplots showing the sum of the Shapley values assigned to these segments, comparing correct and incorrect model predictions within each OSA severity group (see Figure 2). Following statistical analysis, Shapley values showed significant differences in the QRS complex and QT interval between successes and failures within both the healthy and mild OSA groups ($p<0.05$). Notably, in the healthy group, the model's failures exhibited higher importance in these ECG segments than successes. Conversely, in the mild OSA group, the failures relied less on these ECG segments compared to correctly classified mild OSA.

4. Discussion

This study presents a quantitative SHAP-driven framework to characterize the cardiac response to pediatric OSA at the level of ECG morphology. Moving beyond prior approaches focused on global signal attribution, we provide, for the first time, a quantitative analysis of Shapley values across clinically meaningful ECG segments (QRS, ST, TP, and QT). This enables a direct link between DL predictions and cardiac mechanisms, transforming model interpretability into potential cardiac biomarkers.

Only a few studies have explored XAI for pediatric OSA, primarily relying on engineered features or qualitative interpretations. Previous works from our group demonstrated the feasibility of using Grad-CAM and SHAP to interpret DL models based on ECG and ECG combined with SpO_2 signals, highlighting relevant patterns associated with pediatric OSA. However, these approaches were limited to explanations at the signal level. This proposal advances interpretability by quantifying specific ECG components, enabling the translation of model explainability into clinically applicable insights.

Figure 1: Boxplots showing the sum of the Shapley values assigned to the QRS complex, QT interval, ST segment, and TP interval for each subject, grouped by pediatric OSA severity groups.

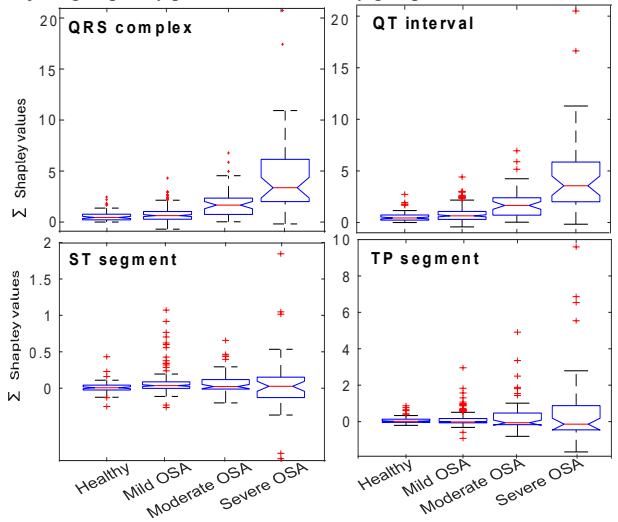
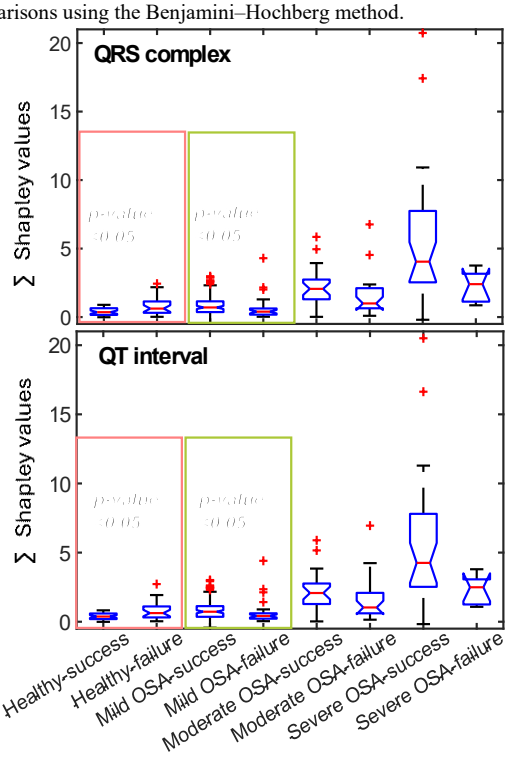


Table 2: *P*-values obtained with the comparison of Shapley values between the 4 OSA severities. The ST interval goes from the end of the QRS complex to the end of the T wave.

	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
QRS	n.s	4×10^{-7}	1×10^{-12}	3×10^{-5}	7×10^{-11}	n.s
QT	n.s.	2×10^{-7}	2×10^{-13}	5×10^{-5}	6×10^{-11}	n.s
ST_f	1×10^{-2}	n.s	n.s	n.s	n.s	n.s
TP	n.s	n.s	n.s	n.s	n.s	n.s

1: healthy group; 2: mild OSA group; 3: moderate OSA group; 4: severe OSA group; n.s: no significant differences. ST_f: from the end of the QRS complex to the end of the T wave.

Figure 2: Boxplots showing the sum of the Shapley values assigned to the QRS complex and QT interval, comparing correct and incorrect model predictions within each OSA severity group. P-values were calculated using the Mann-Whitney U test and corrected for multiple comparisons using the Benjamini–Hochberg method.



The results reveal that SHAP values capture consistent and physiologically meaningful ECG patterns across OSA severities. As shown in Figure 1, there is a progressive increase in Shapley values from healthy subjects to severe OSA, particularly for the QRS complex and QT interval. This trend supports the idea that the DL model is not relying on noise but rather identifying distinct cardiac patterns associated with OSA severities.

Among all analyzed ECG components, both the QRS complex and the QT interval emerge as the key cardiac markers related to pediatric OSA. Significant differences are observed between the healthy group and the higher-severity groups (moderate and severe OSA), as well as between the early-stage OSA group and the more advanced stages. From a pathophysiological perspective, the increasing reliance of the model on the QT interval aligns with delayed or heterogeneous ventricular repolarization, a known consequence of heightened sympathetic tone, which increases the risk of ventricular arrhythmias in severe pediatric OSA [15]. Simultaneously, the growing relevance of the QRS complex suggests that ventricular depolarization abnormalities, which could be linked to early structural remodeling, cyclic heart rate variability (bradycardia-tachycardia) [15], [16], or subclinical ventricular hypertrophy

[17], become progressively more pronounced as the disease advances. These mechanisms are also contributors to the cardiovascular risk in pediatric OSA [15].

The ST segment provides additional insight into the differences between healthy subjects and mild OSA, suggesting that this segment contributes more to AHI estimation in the early stages of the disease. Given that the ST segment reflects early ventricular repolarization, its modulation may be influenced by transient nocturnal hypoxemia and oxidative stress [18], which are known to affect myocardial oxygen demand.

Analysis of QRS complex and QT interval comparing correct and incorrect model predictions within each OSA severity group revealed significant differences in both the healthy and mild groups. In the healthy group, incorrect predictions exhibited higher median Shapley values, suggesting that the model could be detecting subclinical cardiac patterns that the AHI criteria fail to capture. In contrast, missclassifications in the mild OSA group showed lower medians, reflecting that there may be a patient phenotype where apneic events do not cause cardiac abnormalities.

Despite these promising results, several limitations should be acknowledged. First, SHAP analysis was conducted on a single dataset, and external validation across independent cohorts would be necessary to ensure generalizability. Second, the accuracy of ECG delineation directly impacts segment-level analysis. In this sense, improvements in waveform segmentation could further refine the results. Third, while SHAP provides robust attribution, it does not establish causality, and complementary physiological validation would strengthen clinical interpretation. Finally, extending this framework to identify distinct electrophysiological phenotypes of pediatric OSA could provide new insights into disease heterogeneity and personalized cardiac risk stratification.

This finding indicates that the model's predictions are based on cardiac impact, and not solely on the detection of apneic events. The combined analysis of Shapley values demonstrates that the QRS complex and QT interval contribute differentially across OSA severities, capturing significant cardiac changes associated with different OSA severities. This approach not only improves the interpretability of DL systems but also opens the door to using these models for personalized risk stratification, identifying which children require urgent intervention based on their cardiac phenotype rather than their AHI alone.

In conclusion, this study demonstrates that quantitative XAI applied at the ECG segment level enables the identification of clinically meaningful cardiac patterns associated with pediatric OSA. Thus, this proposal quantitatively corroborates our prior qualitative XAI findings, identifying the QRS complex and the QT and ST intervals as relevant cardiac biomarkers for the characterization of pediatric OSA. Overall, this approach advances the development of a clinically applicable and quantitatively interpretable automatic tool for characterizing pediatric OSA.

5. Acknowledgments

This work is part of the project PID2023-148895OB-I00 funded by MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU"/PRTR and was also co-funded by "CIBER-Consorcio Centro de Investigación Biomédica en Red" (CB19/01/00012) through "Instituto de Salud Carlos III". This research is also part of the project CPP2022-009735, funded by MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU"/PRTR, and was also co-funded by POCTEP 2021-2027 (0043_NET4SLEEP_2_E).

6. References

[1] R. Bhattacharjee, *et al.*, "Cardiovascular Complications of Obstructive Sleep Apnea Syndrome: Evidence from Children," *Prog. Cardiovasc. Dis.*, vol. 51, no. 5, pp. 416–433, Mar. 2009.

[2] P. Moridian *et al.*, "Automatic diagnosis of sleep apnea from biomedical signals using artificial intelligence techniques: Methods, challenges, and future works," *WIREs Data Min. Knowl. Discov.*, vol. 12, no. 6, Nov. 2022.

[3] B. Samadi, *et al.*, "Systematic Review of Detecting Sleep Apnea Using Artificial Intelligence: An Insight to Convolutional Neural Network Method," *Arch. Neurosci.*, vol. 11, no. 1, pp. 1–14, 2024.

[4] E. Mortazavi, *et al.*, "Deep learning approaches for assessing pediatric sleep apnea severity through SpO2 signals," *Sci. Rep.*, vol. 14, no. 1, p. 22696, Oct. 2024.

[5] C. García-Vicente, *et al.*, "SleepECG-Net: explainable deep learning approach with ECG for pediatric sleep apnea diagnosis," *IEEE J. Biomed. Heal. Informatics*, vol. 29, no. 2, pp. 1021–1034, 2025.

[6] A. Saranya and R. Subhashini, "A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends," *Decis. Anal. J.*, vol. 7, p. 100230, Jun. 2023.

[7] C. García-Vicente, *et al.,* "Combined explainable deep learning model to predict pediatric sleep apnea from ECG and SpO2," *Meas. J. Int. Meas. Confed.*, vol. 264, p. 120259, Mar. 2026.

[8] C. L. Marcus *et al.*, "A Randomized Trial of Adenotonsillectomy for Childhood Sleep Apnea," *N. Engl. J. Med.*, vol. 368, no. 25, pp. 2366–2376, Jun. 2013.

[9] R. Wang *et al.*, "Pediatric Adenotonsillectomy Trial for Snoring (PATS): protocol for a randomised controlled trial to evaluate the effect of adenotonsillectomy in treating mild obstructive sleep-disordered breathing," *BMJ Open*, vol. 10, no. 3, p. e033889, Mar. 2020.

[10] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[11] L. Sörnmo and P. Laguna, "The Electrocardiogram—A Brief Background," in *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, Elsevier, 2005, pp. 411–452.

[12] J. P. Martínez, *et al.*, "A Wavelet-Based ECG Delineator Evaluation on Standard Databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, Apr. 2004.

[13] F. Vaquerizo-Villar *et al.*, "A Convolutional Neural Network Architecture to Enhance Oximetry Ability to Diagnose Pediatric Obstructive Sleep Apnea," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 8, pp. 2906–2916, 2021.

[14] D. Rothman, *Hands-On Explainable AI (XAI) with Python: Interpret, visualize, explain, and integrate reliable AI for fair, secure, and trustworthy AI apps*. 2020.

[15] C. Guilleminault, *et al.*, "Cyclical variation of the heart rate in sleep apnoea syndrome. Mechanisms, and Usefulness of 24 h Electrocardiography as a Screening Technique," *Lancet*, vol. 323, no. 8369, pp. 126–131, Jan. 1984.

[16] C. L. Marcus, "Sleep-disordered Breathing in Children," *Am. J. Respir. Crit. Care Med.*, vol. 164, no. 1, pp. 16–30, Jul. 2001.

[17] R. S. Amin *et al.*, "Left Ventricular Hypertrophy and Abnormal Ventricular Geometry in Children and Adolescents with Obstructive Sleep Apnea," *Am. J. Respir. Crit. Care Med.*, vol. 165, no. 10, pp. 1395–1399, May 2002.

[18] S. Ebrahimian *et al.*, "Beat-to-beat cardiac repolarization lability increases during hypoxemia and arousals in obstructive sleep apnea patients," *Am. J. Physiol. Circ. Physiol.*, Mar. 2024.

Address for correspondence:
Clara García-Vicente.
Biomedical Engineering Group, University of Valladolid, Av. Ramón y Cajal 7, 47003, Valladolid, Spain. e-mail: clara.garciav@uva.es.