Explainable hybrid convolutional and transformer network for pediatric sleep apnea diagnosis using nocturnal oximetry

C. García-Vicente, *Student Member, IEEE,* G. C. Gutiérrez-Tobal, *Member, IEEE,* J. Gomez-Pilar, *Member, IEEE,* F. Vaquerizo-Villar, A. Martín-Montero, M. Domínguez-Guerrero, D. Gozal, and R. Hornero, *Senior Member, IEEE*

***Abstract*— Pediatric obstructive sleep apnea (OSA) is a breathing disorder marked by pauses in airflow (apneas) and reduced airflow (hypopneas), contributing to neurocognitive and behavioral impairments, cardiovascular complications, and other health issues in affected children. Polysomnography is the gold standard for diagnosis, but complexity, high cost, and limited accessibility issues often lead to underdiagnosis. To address these challenges, we propose a simplified diagnostic approach based on blood oxygen saturation (SpO_2) recordings from nocturnal oximetry. A total of 1,609 SpO_2 recordings from the Childhood Adenotonsillectomy Trial (CHAT) were analyzed. We developed an interpretable approach leveraging a convolutional-transformer network to estimate pediatric OSA severity. Furthermore, we evaluated the explainable artificial intelligence method Gradient-weighted Class Activation Mapping (Grad-CAM). The model achieved 4-class Cohen's kappa and accuracy of 0.529 and 68.56% in the test set, respectively. The proposed model demonstrated enhanced performance correlating with increasing disease severity, with accuracy values ranging from 82% to 95% at different severity cut-offs, thereby signaling improved diagnostic performance when compared to previous approaches. Furthermore, Grad-CAM identified key SpO_2 patterns linked to OSA, such as SpO_2 desaturations related to clusters of apneic events and desaturations occurring independently of events. This innovative approach represents a promising alternative for diagnosing OSA and provides valuable insights into respiratory abnormalities associated with pediatric OSA.**

***Clinical Relevance*—This study highlights the potential of an interpretable deep-learning approach using overnight oximetry for diagnosing pediatric obstructive sleep apnea. It effectively identifies clinically relevant desaturation patterns associated with the disease and supports its early, objective, and efficient detection in clinical practice.**

I. INTRODUCTION

Obstructive sleep apnea (OSA) in children is a respiratory disorder characterized by recurrent episodes of partial or complete upper airway obstruction during sleep, leading to intermittent hypoxemia and desaturation-reoxygenation patterns in oxygen saturation (SpO_2) [1]. These fluctuations in SpO_2 directly reflect the respiratory disturbances underlying OSA, including hypopneas and apneas, which impair tissue oxygenation and metabolic homeostasis [1]. The severity of these events is associated with systemic effects such as autonomic dysregulation, cardiovascular dysfunction, and an increased risk of long-term comorbidities, especially in severe OSA cases [2].

The gold standard for diagnosing OSA is overnight polysomnography (PSG), which monitors multiple biomedical signals, including SpO_2, to calculate the apnea-hypopnea index (AHI) and classify OSA severity (no OSA: AHI<1 events/hour, e/h; mild OSA: 1≤AHI<5 e/h; moderate OSA: 5≤AHI<10 e/h; and severe OSA: AHI≥10 e/h) [3]. While PSG is highly effective, its widespread use is limited by high costs, restricted accessibility, and the discomfort of the procedure, leading to significant underdiagnosis in children [4].

In this context, the SpO_2 signal has emerged as a promising alternative for evaluating OSA, as it is non-invasive and directly reflects the hypoxemia and desaturation-reoxygenation patterns characteristic of the disorder [5]. In recent years, deep learning (DL) methods have demonstrated significant potential for the automated analysis of SpO_2, easing OSA detection without necessarily requiring PSG. However, in the pediatric population, the development of DL models for OSA prediction using SpO_2 remains limited. To date, the extant studies have implemented convolutional and/or recurrent neural networks (CNNs, RNNs) to analyze SpO_2 signals and estimate OSA severity [6], [7]. Although these studies have shown promising results, more advanced hybrid architectures, such as CNNs and transformer (TF) networks, which effectively capture both global and local relationships in temporal signals, have yet to be explored in this context.

On the other hand, DL interpretability remains a significant barrier to clinical adoption, including the sleep domain [8]. In this regard, explainable artificial intelligence (XAI) techniques are essential for providing transparency in identifying relevant SpO_2 patterns and understanding the relationship between oxygenation fluctuations and OSA severity. Accordingly, a recent adult study applied the Gradient-weighted Class Activation Mapping (Grad-CAM) XAI technique to interpret

This work is part of the projects PID2023-148895OB-I00, TED2021-131913B-I00, and CPP2022-009735, funded by MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU"/PRTR. This research was also co-funded by the European Union through the Interreg VI-A Spain-Portugal Program (POCTEP) 2021-2027 (0043_NET4SLEEP_2_E), and by "CIBER-Consorcio Centro de Investigación Biomédica en Red" (CB19/01/00012) through "Instituto de Salud Carlos III". C. García-Vicente was supported by 'Ayudas para contratos predoctorales para la Formación de Doctores' grant from the 'Ministerio de Ciencia, Innovación y Universidades' (PRE2021-100792). D. Gozal is supported in part by NIH grants HL166617 and HL169266.

C. García-Vicente, G. C. Gutiérrez-Tobal, J. Gomez-Pilar, F. Vaquerizo-Villar, A. Martín-Montero, and R. Hornero are with the Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain, and with the Biomedical Engineering Group, Universidad de Valladolid, Av. Ramón y Cajal 7, 47003, Valladolid, Spain (phone: +34 983 423000, ext. 4713) (e-mail: clara.garciav@uva.es). M. Domínguez-Guerrero is with the Biomedical Engineering Group.

D. Gozal is with the Office of The Dean and the Department of Pediatrics, Joan C. Edwards School of Medicine, Marshall University, 1600 Medical Center Dr, Huntington, WV 25701 (email: gozal@marshall.edu).

979-8-3315-8618-8/25/$31.00 ©2025 IEEE

DL models based on cardiorespiratory signals, including SpO_2, providing key patterns associated with OSA [9].

Based on the aforementioned considerations, we propose an explainable and hybrid DL approach based on a CNN and a TF algorithm to estimate pediatric OSA severity from nocturnal SpO_2 that uses Grad-CAM to identify patterns of this signal that drive the model to detect the disease. Thus, this study presents two key contributions. First, we propose a DL approach based on a CNN-TF architecture with transfer learning to estimate the severity of pediatric OSA using SpO_2 signals, aiming to enable automated and efficient diagnosis. Second, we propose the Grad-CAM method to identify critical patterns in SpO_2 signals and provide visual interpretations that contribute to a better understanding of the relationship between desaturation events and pediatric OSA.

II. MATERIALS AND METHODS

A. Signals and subjects

This study used the public dataset from the Childhood Adenotonsillectomy Trial (CHAT), accessible at https://sleepdata.org/datasets/chat [10]. A total of 1,609 valid SpO_2 recordings collected during nocturnal PSG studies of children aged 5 to 10 were analyzed. Recordings were randomly assigned to three independent subsets with approximately 60% for training, 20% for validation, and 20% for testing, verifying that age, sex, body mass index, and AHI showed no statistically significant differences ($p > 0.01$) between sets. Each subject was uniquely assigned to a single set to prevent duplication. Table 1 provides a summary of the clinical and sociodemographic data of the participants.

The single-channel SpO_2 signals from CHAT recordings were resampled to a frequency of 1 Hz [6]. Recordings were empirically set to 8 hours, as this value achieved the highest performance on the validation set. Shorter signals were padded with zeros at the start, while longer signals were cropped by removing data from the beginning, following previous studies [11]. Nocturnal SpO_2 data were divided into 24 segments, each lasting 20 minutes ($24 \times 1{,}200 \times 1 = 28{,}800$ samples). This segmentation approach was determined to be the most effective for training a previously evaluated CNN model for pediatric OSA diagnosis [6]. Additionally, this format facilitated the adaptation of CNN's optimal architecture for the model used in this study. For each subject, the output label was the AHI, as annotated in the CHAT dataset.

B. DL-based approach

A hybrid approach based on a CNN-TF was developed to capture both spatial structure and long-range relationships. The model received single-channel overnight SpO_2 signals as input. The convolutional part corresponded to the previously presented CNN model [6]. Subsequently, a time series adapted TF-based encoder was implemented using attentional mechanisms through the transfer learning technique to capture dependencies and contextual information throughout the night sequence [12]. The output of the model was the AHI for each subject.

C. AHI Estimation

It is important to note that the AHI value estimated by this model tends to underestimate the AHI from the original PSG, since our estimation uses the total recording time, which is longer than the total sleep time [13]. To address this discrepancy, we re-estimated the final AHI using a support vector regression model trained on the training group to correct for this bias, following previous studies [13].

TABLE I. DEMOGRAPHIC AND CLINICAL INFORMATION. THE DATA ARE PRESENTED AS N (%) OR MEDIAN [INTERQUARTILE RANGE].

Variables	Training	Validation	Test
Subjects (*n*)	987 (61,3%)	323 (20,1%)	299 (18,6%)
Age (*years*)	7,0[2,0]	7,0[2,0]	6,9[2,0]
Males (*%*)	51,7%	49,2%	46,1%
BMI (kg/m^2)	17,3[5,9]	17,1[6,3]	17,4[6,0]
AHI(events/h)	2,6 [4,8]	2,5[4,8]	2.3[5,1]
AHI≥1(e/h)*	487	167	144
AHI≥5(e/h)*	159	44	49
AHI≥10(e/h)*	129	45	41

*AHI≥1 (e/h): mild OSA; AHI≥5: moderate OSA; AHI≥10: severe OSA.

D. Interpretability using Grad-CAM

In this study, the Grad-CAM method was used to analyze and understand the internal mechanisms of the model concerning identifying apneic events and detecting respiratory patterns associated with pediatric OSA [14]. Grad-CAM calculations involved utilizing gradients from each specific convolutional layer to produce heatmaps corresponding to each layer. The final heatmap was then generated by averaging all the individual layer heatmaps [11].

E. Performance assessment

The diagnostic performance for pediatric OSA was assessed using confusion matrices, 4-class accuracy (Acc_4), and 4-class Cohen's kappa coefficient (k) across four severity groups: No OSA, mild, moderate, and severe OSA. Additionally, we calculated global accuracy (Acc), sensitivity (Se), specificity (Sp), positive and negative predictive values (PPV and NPV), as well as the positive likelihood ratio (LR^+) for AHI severity thresholds of 1, 5, and 10 e/h.

III. RESULTS

A. Optimal CNN-TF architecture

The optimal hyperparameter configuration was determined heuristically by training the model on the training set and evaluating its performance on the CHAT validation set. Cohen's kappa (k) was used as the evaluation metric, comparing the actual severity of OSA with the severity estimated by the model. The final model was optimized through transfer learning from the previously described CNN model [6], completed by the addition of a TF-based encoder architecture. The CNN-TF architecture included the previously presented 6 CNN blocks, each consisting of 64 filters [6], followed by 7 TF-based encoder layers. Within each TF module, the key dimension for multi-head attention was set to 16, with 8 attention heads to capture global relationships in the sequence. A dropout rate of 0.1 was applied within the TF, while the final dense layer consisted of 64 units. Finally, a global dropout rate of 0.1 was used. The Adam optimizer was used to update the weights with an initial learning rate of 10^{-5}. The optimal architecture was trained with a batch size of 150 and 100 epochs.

B. Pediatric OSA Diagnostic performance

Fig. 1 shows the confusion matrices after the classification of the OSA severity in the CHAT test set. The 4-class metrics obtained were *Acc4*=68.56% and k_4 = 0.529. Table 2 shows that the highest *Acc* (94.65%) is obtained for identifying the most severely affected children. This is particularly remarkable, as these children are the ones who will benefit the most from an accurate and timely diagnosis [15].

C. SpO_2 patterns

Fig. 2 shows the zooms of the heatmaps obtained using the Grad-CAM method on SpO_2 signals from a nocturnal recording. Fig. 2 (a) shows a SpO_2 signal in which Grad-CAM identifies clusters of desaturations (SpO_2 drops > 3%) by leveraging the temporal context of the events within the nocturnal time series. Fig. 2 (b) highlights how the model focuses its attention on SpO_2 desaturation regions with events, demonstrating its ability to learn SpO_2 patterns associated with apneic events. Additionally, the model distinguishes between oxygen drops related to OSA desaturations and those considered artifacts, prioritizing only those within pathophysiological ranges. Finally, in Fig. 2 (c), Grad-CAM highlights a signal region without apneic events but with desaturations that may not have been annotated by specialists or that could be linked to other pathologies, such as chronic obstructive pulmonary disease (COPD) [16].

IV. DISCUSSION AND CONCLUSIONS

This work presents the development of an interpretable hybrid DL model based on CNN-TF and nocturnal SpO_2, which has achieved high performance in diagnosing the severity of pediatric OSA. Grad-CAM was useful for interpreting the proposed model and enabled achieving a deeper understanding of the pathophysiological behavior of SpO_2 related to pediatric OSA.

According to the explainability obtained from Grad-CAM, the model's AHI prediction based on information from desaturation patterns in the SpO_2 signal reflects the physiologic response to apneic events, where frequent reductions in oxygen levels are key features [1]. Moreover, the model's focus on desaturation clusters associated with apneic events would help to detect the amplitude and duration of these desaturations and the oxygen recovery time, which may reveal the severity of the related desaturation and the degree of upper airway flow limitation along with declining pulmonary functional reserve. In addition, fast oscillations and signal fragmentation may also be linked to ventilatory instability, which can affect the cardiovascular system [17]. Lastly, the examination of the desaturation episodes in this study, together with the future inclusion of other signals such as electrocardiogram (ECG), could lead to a more precise assessment of the cardiovascular risk associated with OSA, especially in severe cases [18].

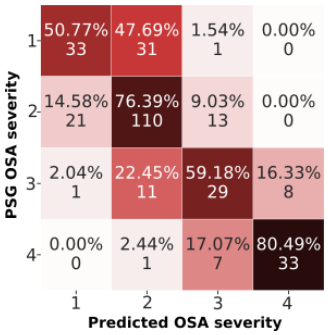

Figure 1. Confusion matrix for the 4 severity levels in the CHAT test set. 1: AHI<1 e/h, 2: 1≤AHI<5 e/h, 3: 5≤AHI<10 e/h, and 4: AHI≥10 e/h.

TABLE II. PERFORMANCE METRICS (%) FOR THE BINARY CLASSIFICATION OF THE CNN-TF APPROACH.

IAH	*Se(%)*	*Sp(%)*	*PPV(%)*	*NPV(%)*	LR^+	*Acc (%)*
1 e/h	90.60	50.77	86.89	60.00	1.84	81.94
5 e/h	85.56	93.30	84.62	93.75	12.77	90.97
10 e/h	80.49	96.90	80.49	96.90	25.96	94.65

When evaluating the performance of our model, notable advancements are evident compared to prior studies that also relied on SpO_2 for diagnosing pediatric OSA. For instance, Calderón et al. [19] used SpO_2 features for binary classification (5 e/h cut-off), achieving a *Se* of 62% vs. 85.6%, *Sp* of 96% vs. 93.3%, and an *Acc* of 79% vs. 91%. In comparison, the present model demonstrated a significantly improved balance between *Se* and *Sp* while increasing diagnostic *Acc*. Vaquerizo-Villar et al. [6] used a CNN based on SpO_2. When compared to this study, our model achieved a higher k_4 (0.529 vs. 0.510), obtaining a better performance in terms of agreement. Moreover, our CNN-TF demonstrated key advantages in *Se* and *Acc*, particularly at 1 e/h (90.6% vs. 71.2% and 81.9% vs. 77.6%, respectively), making it more effective for detecting the presence of OSA. At 5 and 10 e/h, our model achieved a better balance between *Se* and *Sp*, with slightly improved *Se* (85.6% vs. 83.7%) at 5 e/h. Moreover, the explanation of the model using XAI in this approach reflects the major difference and advantage. Mortazavi et al. [7] implemented a CNN-RNN model with attention using SpO_2 signals. In comparison, this study achieved a higher k_4 (0.610 vs. 0.529) by using a smaller and different subset of the CHAT dataset (n=844 vs. n=1609). Our approach, using the full database and coherently handling longitudinal records, offers an advantage in terms of representativeness and generalizability of the results. Additionally, Grad-CAM highlights the most relevant regions directly in SpO_2 input without losing the temporal relationship, allowing for a physiology-aligned representation.

Among the limitations of the study, it is worth mentioning that only one database was used, and therefore, it would be beneficial to validate the model with a larger number of SpO_2 recordings to assess its performance under different conditions and in various populations. Additionally, a future line of research could focus on evaluating the model using global XAI approaches, enabling the quantitative identification of relevant SpO_2 patterns in the context of heterogeneous pediatric phenotypes.

In conclusion, integrating an interpretable CNN-TF model in the analysis of nocturnal SpO_2 provides a reliable diagnosis of pediatric OSA. Furthermore, Grad-CAM helps to identify respiratory patterns related to the disease and suggests other patterns not previously annotated or that could be associated with other diseases [16]. This approach could serve as a foundation for future use of multiple simultaneous signals from PSG, such as ECG and SpO_2, assessing cardiovascular risk, a consequence that is particularly prevalent in severe pediatric OSA. Ultimately, our proposed methodology

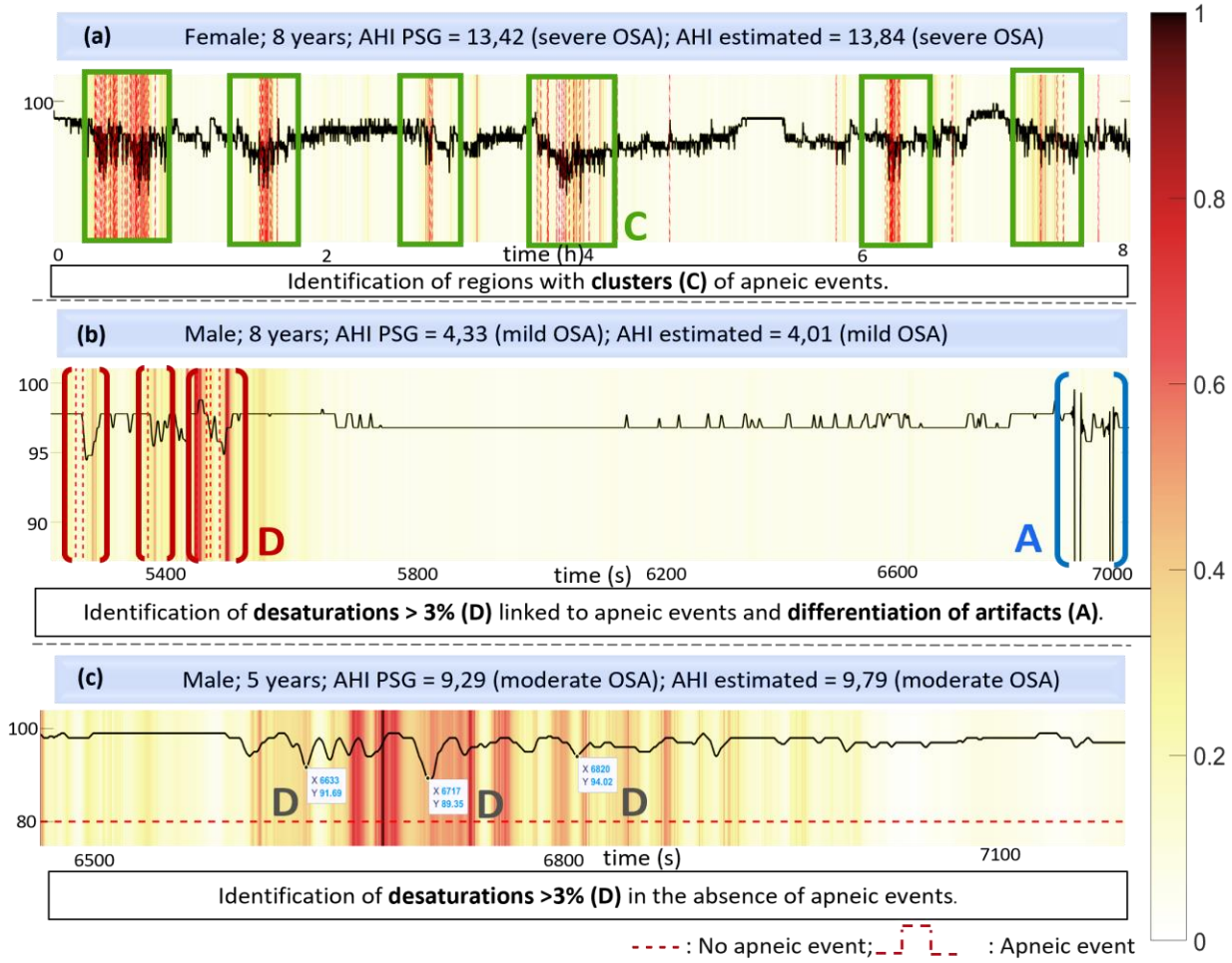

Figure 2: Grad-CAM displays some representative findings in SpO_2 signals from the CHAT test subset. Fig. 2 (a) shows the model's focus on clusters (C) of events. Fig. 2 (b) shows the model focusing on desaturations (D) associated with apneic events. Fig. 2 (c) shows the identification of desaturations (D) in the absence of apneic events. The color bar indicates at 0 (yellow) the areas of lower relevance and at 1 (brown) the areas of higher relevance.

presents a promising alternative to PSG, offering a simplified, fast, and objective method for diagnosing pediatric OSA.

REFERENCES

[1] C. L. Marcus *et al.*, "Diagnosis and Management of Childhood Obstructive Sleep Apnea Syndrome," *Pediatrics*, vol. 130, no. 3, pp. e714–e755, Sep. 2012.

[2] O. Vitelli *et al.*, "Autonomic imbalance during apneic episodes in pediatric obstructive sleep apnea," *Clin. Neurophysiol.*, vol. 127, no. 1, pp. 551–555, Jan. 2016.

[3] R. B. Berry *et al.*, "The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.6," *Am. Acad. Sleep Med.*, 2020.

[4] G. D. Church, "The Role of Polysomnography in Diagnosing and Treating Obstructive Sleep Apnea in Pediatric Patients," *Curr. Probl. Pediatr. Adolesc. Health Care*, vol. 42, no. 1, pp. 2–25, Jan. 2012.

[5] F. del Campo *et al.*, "Oximetry use in obstructive sleep apnea," *Expert Rev. Respir. Med.*, vol. 12, no. 8, pp. 665–681, 2018.

[6] 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.

[7] E. Mortazavi *et al.*, "Deep learning approaches for assessing pediatric sleep apnea severity through SpO2 signals," *Sci. Rep.*, pp. 1–21, 2024.

[8] A. Chaddad *et al.*, "Survey of Explainable AI Techniques in Healthcare," *Sensors*, vol. 23, no. 2, p. 634, Jan. 2023.

[9] Á. S. Alarcón *et al.*, "Obstructive sleep apnea event detection using explainable deep learning models for a portable monitor," *Front. Neurosci.*, vol. 17, no. July, Jul. 2023.

[10] 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.

[11] C. Garcia-Vicente *et al.*, "SleepECG-Net: explainable deep learning approach with ECG for pediatric sleep apnea diagnosis," *IEEE J. Biomed. Heal. Informatics*, pp. 1–14, 2024.

[12] A. Vaswani *et al.*, "Attention Is All You Need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, pp. 5999–6009, Jun. 2017.

[13] C. García-Vicente *et al.*, "ECG-based convolutional neural network in pediatric obstructive sleep apnea diagnosis," *Comput. Biol. Med.*, vol. 167, no. September, p. 107628, Dec. 2023.

[14] R. R. Selvaraju *et al.*, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020.

[15] G. C. Gutiérrez-Tobal *et al.*, "Reliability of machine learning to diagnose pediatric obstructive sleep apnea: Systematic review and meta-analysis," *Pediatr. Pulmonol.*, vol. 57, no. 8, pp. 1931–1943, Aug. 2022.

[16] W. T. McNicholas, *Impact of Other Respiratory Conditions and Disorders During Sleep*. Academic Press, 2013.

[17] H. H. Chuang *et al.*, "The 3% Oxygen Desaturation Index is an Independent Risk Factor for Hypertension Among Children with Obstructive Sleep Apnea," *Nat. Sci. Sleep*, vol. 14, pp. 1149–1164, Jun. 2022.

[18] A. Kulkas *et al.*, "Severity of desaturation events differs between hypopnea and obstructive apnea events and is modulated by their duration in obstructive sleep apnea," *Sleep Breath.*, vol. 21, no. 4, pp. 829–835, Dec. 2017.

[19] J. M. Calderón *et al.*, "Development of a minimally invasive screening tool to identify obese Pediatric population at risk of obstructive sleep Apnea/Hypopnea syndrome," *Bioengineering*, vol. 7, no. 4, pp. 1–13, Oct. 2020.