

IFMBE Proceedings 112

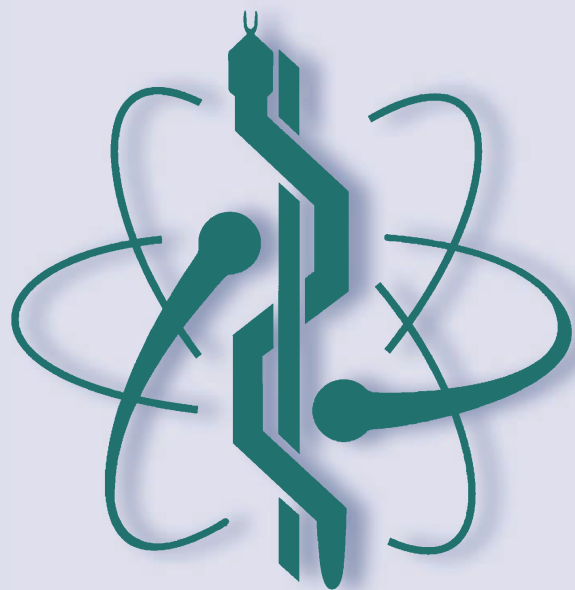
Tomaž Jarm

Rok Šmerc

Samo Mahnič-Kalamiza *Editors*

9th European Medical and Biological Engineering Conference

Proceedings of EMBEC 2024,
June 9–13, 2024, Portorož, Slovenia,
Volume 1



 Springer

Series Editor

Ratko Magjarević, *Faculty of Electrical Engineering and Computing, ZESOI, University of Zagreb, Zagreb, Croatia*

Associate Editors

Piotr Ładyżyński, *Warsaw, Poland*

Fatimah Ibrahim, *Department of Biomedical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur, Malaysia*

Igor Lackovic, *Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia*

Emilio Sacristan Rock, *Mexico DF, Mexico*

Tomaž Jarm · Rok Šmerc ·
Samo Mahnič-Kalamiza
Editors

9th European Medical and Biological Engineering Conference

Proceedings of EMBEC 2024, June 9–13, 2024,
Portorož, Slovenia, Volume 1

Editors

Tomaž Jarm
University of Ljubljana, Faculty of Electrical
Engineering
Ljubljana, Slovenia

Rok Šmerc
University of Ljubljana, Faculty of Electrical
Engineering
Ljubljana, Slovenia

Samo Mahnič-Kalamiza
University of Ljubljana, Faculty of Electrical
Engineering
Ljubljana, Slovenia

ISSN 1680-0737

ISSN 1433-9277 (electronic)

IFMBE Proceedings

ISBN 978-3-031-61624-2

ISBN 978-3-031-61625-9 (eBook)

<https://doi.org/10.1007/978-3-031-61625-9>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland








If disposing of this product, please recycle the paper.

Contents

A Machine Learning Approach for Predicting Electrophysiological Responses in Genetically Modified HEK Cells	1
<i>Jacopo Vitale, Martina Sassi, and Leandro Pecchia</i>	
A Machine Learning Framework for Gait and EMG Analysis for Post-stroke Motor Dysfunctions Assessment	15
<i>Francesco Romano, David Perpetuini, Daniela Cardone, and Arcangelo Merla</i>	
A Novel University Course on Medical Devices Design and Certification at the University of Siena	23
<i>Ernesto Iadanza</i>	
A Complex Spinal Surgery Lifting System for Prone Positioning	29
<i>Lutong Li, Stuart Watson, Glyn Smurthwaite, John Large, Andrew Weightman, and Glen Cooper</i>	
Accurate and Interpretable Deep Learning Model for Sleep Staging in Children with Sleep Apnea from Pulse Oximetry	38
<i>Fernando Vaquerizo-Villar, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, Adrián Martín-Montero, David Gozal, Eduardo Tamayo, and Roberto Hornero</i>	
Active Compensation for OPM-MEG Inside a Two-Layer Magnetically Shielded Room	48
<i>Michał Władziński, Anna Jodko-Władzińska, and Tilmann H. Sander</i>	
Alarms Early Detection in Dialytic Therapies via Machine Learning Models ...	55
<i>Alessia Nicosia, Nunzio Cancilla, Marco Siino, Michele Passerini, Francesca Sau, Ilenia Tinnirello, and Andrea Cipollina</i>	
An Everyday Hat for Detection of Eye Blinks and Forehead Clenching	67
<i>S. M. Musfequr Rahman, Henna Mattila, Asif Shaikh, Pasi Raunonen, and Johanna Virkki</i>	
An Innovative Solution for Efficient Workflow Management in Healthcare	77
<i>Alessio Luschi and Ernesto Iadanza</i>	



Accurate and Interpretable Deep Learning Model for Sleep Staging in Children with Sleep Apnea from Pulse Oximetry

Fernando Vaquerizo-Villar^{1,2,3} , Daniel Álvarez^{1,2} ,
Gonzalo C. Gutiérrez-Tobal^{1,2} , Adrián Martín-Montero^{1,2} , David Gozal⁴ ,
Eduardo Tamayo^{3,5,6} , and Roberto Hornero^{1,2} 

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

fernando.vaquerizo@gib.tel.uva.es

² Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Instituto de Salud Carlos III, Valladolid, Spain

³ Department of Anaesthesiology, Clinic University Hospital of Valladolid, Valladolid, Spain

⁴ Joan C. Edwards School of Medicine, Marshall University, Huntington, WV, USA

⁵ Group for Biomedical Research in Critical Care Medicine, BioCritic, Valladolid, Spain

⁶ CIBER de Enfermedades Infecciosas, Instituto de Salud Carlos III, Valladolid, Spain

Abstract. Identification of sleep stages is crucial in the diagnosis of sleep-related disorders but relies on the labor-intensive and manual scoring of overnight polysomnography (PSG) recordings. To simplify the sleep staging process, deep learning (DL) algorithms have been proposed to automatically analyze pulse rate (PR) and blood oxygen saturation (SpO₂) signals from pulse oximetry in children with obstructive sleep apnea (OSA). However, existing approaches are perceived as black boxes, limiting their implementation in clinical settings. Accordingly, we develop a DL architecture based on a U-Net to automatically perform 4-class sleep stage classification (wake, light sleep, deep sleep, and rapid-eye movement sleep) using entire-night PR and SpO₂ recordings. Furthermore, Semantic Segmentation via Gradient-Weighted Class Activation Mapping (Seg-Grad-CAM), an eXplainable Artificial Intelligence methodology, is proposed to provide an interpretation of the sleep scoring process. PR and SpO₂ from 1,633 PSG recordings obtained from the Childhood Adenotonsillectomy Trial database were used for these purposes. The U-Net model showed a high performance for the 4-stage classification procedure in an independent set, with 78.2% accuracy and 0.696 Cohen's kappa. The Seg-Grad-CAM heatmaps revealed that the PR signal has a higher contribution than SpO₂ towards sleep staging, while also showing the key roles of mean and variance in PR amplitude, along with changes in the content of PR spectral bands, in the sleep staging process. These findings suggest that an explainable DL model to analyze pulse oximetry signals could be integrated in the clinical environment for automatic sleep staging in abbreviated pediatric OSA tests.

Keywords: Deep learning (DL) · explainable artificial intelligence (XAI) · pediatric obstructive sleep apnea (OSA) · pulse oximetry · sleep staging

1 Introduction

Identification of sleep macro-structural changes (i.e., sleep stages) plays a crucial role in diagnosing sleep-related disorders [1]. According to the American Academy of Sleep Medicine (AASM) guidelines, each 30-s consecutive epoch of sleep must be scored as wake (W), three levels of non-Rapid Eye Movement (non-REM) sleep (N1, N2, and N3), or REM sleep [2]. Sleep technicians perform this scoring by manually examining electroencephalography, electrooculography, and electromyography signals, which are recorded along with cardiorespiratory signals during polysomnography (PSG) [2]. While PSG is considered the gold standard, its analysis for sleep scoring is expensive, complex, time-consuming, and often limited in availability, leading to delays in diagnosing sleep disorders [3]. Additionally, manual sleep scoring faces significant inter-rater variability [4], potentially impacting diagnostic accuracy.

With recent advances in deep-learning (DL) methodologies, various alternatives based on the automated analysis of a reduced set of signals have been proposed to address these PSG-related limitations [5]. Particularly, pulse oximetry devices have emerged as a simplified option for sleep scoring and diagnosing sleep disorders, as they record blood oxygen saturation (SpO_2) and pulse rate (PR) signals at patient's home through a non-invasive probe [3, 5]. The time-frequency characteristics of PR and SpO_2 signals exhibit changes during sleep stages, leading to several studies using DL algorithms for automatic sleep stage scoring from pulse oximetry signals [5].

A large proportion of these investigations has concentrated on sleep staging in individuals with obstructive sleep apnea (OSA), a highly prevalent sleep disorder affecting nearly 1 billion people globally [3]. OSA diagnosis relies on the apnea-hypopnea index (AHI: the number of apneas and hypopneas per hour of sleep), highlighting the importance of scoring sleep stages and calculating the total sleep time (TST) in this context [2]. Importantly, most of these studies focused on adult OSA populations [6–9], whereas only two conference papers developed by our own group have approached sleep staging from PR and SpO_2 signals in pediatric OSA subjects [10, 11]. This imbalance is not surprising, given that pediatric OSA involves distinct etiological, diagnostic, and therapeutic considerations compared to adult subjects.

In these preliminary studies, sleep staging was performed using a standard convolutional neural network (CNN) [10] and a CNN combined with a recurrent neural network (CNN-RNN) [11], respectively. Despite proving valuable in learning stage-related features from pulse oximetry signals, these DL architectures were trained on sequences of 10 and 100 30-s consecutive epochs [10, 11], respectively, limiting their performance by not considering the whole-night dynamics of PR and SpO_2 during sleep. Another significant limitation in all state-of-the-art studies is the lack of interpretability of the DL models [5], thus hindering broader acceptance in real clinical settings. On this note, a recent report from the European Union emphasizes the necessity of enhancing the reliability, transparency, and interpretability of artificial intelligence systems for their responsible and well-informed implementation in society [12].

To overcome these limitations of the state-of-the-art, this study presents two novel contributions. On the one hand, a DL architecture based on the U-Net framework is proposed to point-wise identify W, light sleep (N1 and N2), deep sleep (N3) and REM sleep stages in pediatric OSA patients from whole-night (subject-based) pulse oximetry (PR

and SpO₂) recordings. Conversely, an eXplainable Artificial Intelligence (XAI) analysis based on Semantic Segmentation via Gradient-Weighted Class Activation Mapping (Seg-Grad-CAM) is proposed to provide an interpretation of the features of PR and SpO₂ signals linked with W, light sleep, deep sleep, and REM sleep stages. We hypothesize that the application of subject-based DL and XAI algorithms to pulse oximetry signals can derive in high-performance, interpretable, and consequently, clinically applicable models for the automated detection of sleep stages in childhood OSA patients. Accordingly, our main objective is to assess an explainable DL model with PR and SpO₂ data aimed at accurately detecting sleep stages in pediatric OSA patients.

2 Materials and Methods

2.1 Subjects and Signals

The Childhood Adenotonsillectomy Trial (CHAT) database was used in this work. CHAT is a semi-public dataset comprising 1639 sleep studies conducted on pediatric subjects aged 5–10 years old who were assessed due to clinical suspicion of OSA. Within this dataset, there were 1633 PSG-derived PR and SpO₂ recordings [13]. Each sleep study in the database also includes annotations of sleep stages and apnea/hypopnea events, which were performed following the AASM 2007 rules [14].

Data was partitioned into three independent sets: training, validation, and test. To prevent potential biases arising from including PR and SpO₂ recordings from the same pediatric subject in multiple sets, the test set was composed of 858 PR and SpO₂ recordings from the baseline (453 subjects) and follow-up (405 subjects) groups of the CHAT database, whereas the subjects from the non-randomized (775 subjects) group of the CHAT dataset were randomly allocated into training (75%, 588 subjects) and validation sets (25%, 187 subjects). Table 1 presents the clinical and polysomnographic information of the population under study.

PR and SpO₂ recordings, originally acquired using sampling rates (fs) from 1 to 512 Hz, were resampled to a common fs of 1 Hz [10, 11]. Subsequently, subject-based standardization was implemented to normalize PR and SpO₂ baseline levels across different children. Finally, adhering to the input requirements of the DL model and the total recording time, all PR and SpO₂ recordings were either padded or truncated to a common length of 12 h ($L = 43200$ samples). Truncation or zero padding was exclusively applied to the start of each recording, and a value of 12 h was chosen to ensure that only initial wake periods were either removed or added. According to the sleep stage label of the corresponding 30-s epoch, every sample of the whole recordings is individually labeled as W, light sleep, deep sleep, or REM sleep.

2.2 Proposed Deep-Learning Architecture

Figure 1 shows the overall U-Net-based DL architecture employed in this study. U-Net is an encoder-decoder deep neural network originally designed for image segmentation that allows point-wise prediction. U-Net has already shown its usefulness for sleep scoring in adult OSA subjects [15]. The input section of the proposed U-Net consists

of 12-h of resampled PR and SpO₂ data ($2 \times L$ samples). This input is first processed through an encoder consisting of four layers, aimed at learning low-level stage-related features from PR and SpO₂ signals. Each layer consists of a convolutional block (conv block) followed by a max-pooling and a dropout layer. Each conv block comprises three stacked sub-blocks, being each sub-block composed of a 2D convolution operation with N_f filters and a kernel size of 2×3 , a Rectified Linear Unit (ReLU) activation, and batch normalization. As in the standard U-Net, N_f is initially set as 64 and its value is doubled with the level of depth (layer) of the network.

After the encoder, the extracted feature maps are further processed at the bottleneck of the network through one conv block prior to the decoder. Analogously to the encoder, the decoder consists of four layers intended to extract high-resolution feature maps, each of them comprising a dropout layer, a 2D transposed convolution (Conv2DTranspose), and a conv block. To preserve the low-level features, skip connections are added from the encoder layers to the corresponding decoder layers. Finally, the last layer of the U-Net is a softmax activation, which is used to generate a point-wise prediction as the output ($L \times 4$ samples), which describes the probability of each sample belonging to W, light, deep, and REM sleep stages.

For the sake of completeness, U-Net was not only trained using PR and SpO₂ data (U-Net_{PR-SpO2}), but also using single-channel PR (U-Net_{PR}) and SpO₂ (U-Net_{SpO2}) data.

Table 1. Clinical and polysomnographic data of the children in the CHAT database. TRT = total recording time. Data are presented as median [interquartile range] or n (%).

	All	Training set	Validation set	Test set
Subjects (n)	1633 (100)	588 (36.0)	187 (11.5)	858 (52.5)
Age (years)	7 [6–8]	7 [6–8]	7 [6–8]	7 [6–8]
Males (n)	776 (47.5)	278 (47.3)	82 (43.9)	416 (48.5)
AHI (e/h)	2.5 [1.1–5.9]	1.4 [0.8–2.8]	1.5 [0.7–3.0]	4.1 [2.2–8.0]
TRT (hours)	9.8 [9.1–10.8]	9.6 [8.9–10.6]	9.7 [9.1–10.8]	9.9 [9.2–10.9]
Wake (%)	22.8 [15.8–29.4]	21.9 [14.7–28.5]	23.3 [16.5–30.0]	23.3 [16.3–30.0]
Light (%)	38.7 [33.0–44.5]	39.2 [33.1–45.0]	37.6 [32.6–43.4]	38.4 [33.0–44.2]
Deep (%)	23.1 [19.5–27.8]	23.3 [19.6–28.2]	23.8 [19.4–28.2]	22.9 [19.5–27.4]
REM (%)	14.1 [11.6–16.8]	13.9 [11.1–16.9]	14.3 [11.5–16.8]	14.1 [11.7–16.7]

2.3 eXplainable Artificial Intelligence Analysis

In this study, XAI has been applied using the Seg-Grad-CAM method to understand the internal processes of the U-Net in relation to the prediction of sleep stages [16]. Seg-Grad-CAM is an extension of the well-known Grad-CAM method for semantic segmentation, which can produce heatmaps that explain the relevance for the decision of individual points or regions of interest (ROI) in the input data [16]. Given a target

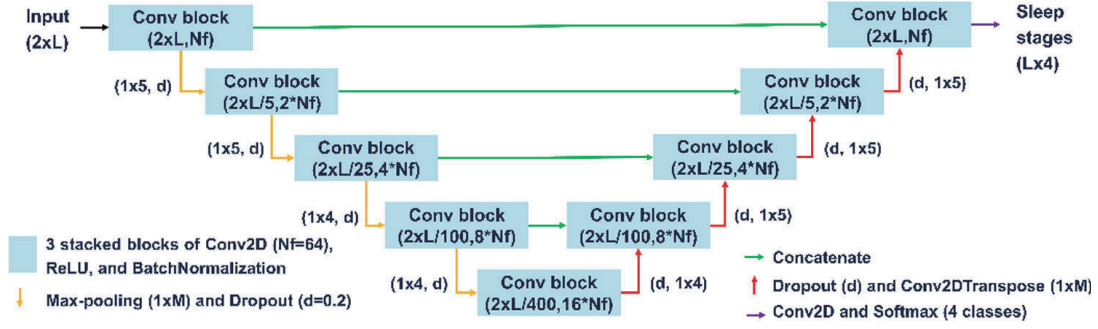


Fig. 1. Overview of the proposed U-Net-based DL architecture for sleep staging.

class c , a region of interest ROI^c , and the feature map A^k of a chosen 2D convolutional layer, Seg-Grad-CAM computes the heatmap L^c using the following expression:

$$L^c = ReLU\left(\sum_k \alpha_k^c \cdot A^k\right), \quad (1)$$

where α_k^c are the weights denoting the importance of each feature map for the output class c , y^c , in the region ROI^c :

$$\alpha_k^c = \frac{1}{Z} \sum_{u,v} \frac{\partial \sum_{i \in ROI^c} y_i^c}{\partial A_{u,v}^k}. \quad (2)$$

In this study, c is one out of the four sleep stages (W/Light/Deep/REM), ROI^c is the region of points predicted as c , and A^k are the feature maps at the last 2D convolution of the bottleneck layer [16]. Seg-Grad-CAM-derived heatmaps were finally normalized and resized for a joint visualization with the raw PR and SpO₂ data.

2.4 Statistical Analysis

The DL models output prediction probabilities for each of the 4 sleep stages at each point of the recordings, which were converted into sleep stage predictions by selecting the class with the highest probability. As manual sleep scoring is performed for each 30-s epoch, the per-epoch output label was obtained as the predicted sleep stage most represented within the 30-s epoch. All zero-padded regions were removed before calculating performance measures. The overall performance of the U-Net for automatic sleep staging was finally assessed by means of the 4-class accuracy (Acc), Cohen's kappa index (kappa), macro F1-score (MF1), and per-class F1-score (F1).

3 Results

3.1 Sleep Staging Performance

Table 2 shows the performance metrics of the three U-Net models (U-Net_{PR-SpO₂}, U-Net_{PR}, and U-Net_{SpO₂}) for the four-stage classification procedure (W/Light/Deep/REM) in the test set. As expected, the U-Net_{PR-SpO₂} model achieved a higher performance (Acc = 78.2%, kappa = 0.696, and MF1 = 78.3%) than the U-Net_{PR} (Acc = 75.5%, kappa = 0.659, and MF1 = 75.7%) and U-Net_{SpO₂} (Acc = 60.7%, kappa = 0.447, and MF1 = 58.9%) models.

Table 2. Sleep staging performance metrics for U-Net_{PR-SpO2}, U-Net_{PR}, and U-Net_{SpO2} models.

	Overall metrics			Per-class F1-score (F1) (%)			
	Acc (%)	kappa	MF1(%)	W	Light	Deep	REM
U-Net _{PR-SpO2}	78.2	0.696	78.3	86.7	75.1	76.3	74.9
U-Net _{PR}	75.5	0.659	75.7	85.7	71.9	73.9	71.3
U-Net _{SpO2}	60.7	0.447	58.9	76.6	59.1	59.6	40.2

3.2 Identification of PR and SpO₂ Patterns

Figure 2 and Fig. 3 show Seg-Grad-CAM visualizations of the PR and SpO₂ signals of a representative subject, respectively. These heatmaps were computed for samples predicted as W, Light, Deep, and REM sleep stages by the U-Net_{PR-SpO2} model. For each heatmap, a zoom in a relevant region of the heatmap is included at the right, together with the short-time Fourier transform, which better shows the time-frequency characteristics of the PR and SpO₂ patterns that the model is focusing on to make the prediction. Notice that the PR signal has a remarkably higher contribution than SpO₂ towards sleep staging, as derived from the heatmap amplitude in Figs. 2 and 3. In this respect, the heatmaps are highlighting well-known time-frequency PR characteristics (Fig. 2): mean and variance in PR, as well as predominance of very low frequency (VLF), low frequency (LF), and/or high frequency (HF) content. Despite having a lower importance for sleep staging, SpO₂ baseline amplitude and the presence and depth of oxygen drops (desaturations) are also emphasized by Seg-Grad-CAM (Fig. 3).

4 Discussion and Conclusions

To our knowledge, this is the first study combining DL and XAI approaches for the accurate and interpretable detection of sleep stages from pulse oximetry signals. The proposed DL model based on U-Net (U-Net_{PR-SpO2}) showed a remarkable performance, with 78.2% Acc, 0.696 kappa, and 78.3% MF1 for W/Light/Deep/REM sleep scoring. The superior performance of U-Net_{PR-SpO2} over U-Net_{PR}, and U-Net_{SpO2} models indicates that SpO₂ and PR show complementarity in sleep stage detection, as we also reported in [10]. The obtained kappa value (in the range 0.61–0.80) highlights substantial agreement with manual PSG scoring [10, 11], suggesting its potential for sleep staging in at-home pulse oximetry tests for childhood OSA screening.

Our results also showed that the proposed XAI approach based on Seg-Grad-CAM can identify time-frequency PR and SpO₂ features used by the U-Net model to point-wise predict each stage through the entire sleep recording. Regarding PR, the Seg-Grad-CAM heatmaps showed that PR amplitude is higher in W than in NREM (Light and Deep) sleep, which agrees with a recent work by Martín-Montero et al. [17] in the pediatric OSA context. We also showed the influence of: (i) HF activity during light and deep sleep, which is consistent with the higher content during NREM sleep found by Martín-Montero et al. [17] within a subject-adaptive respiratory band; (ii) LF activity during W

and deep sleep, which warrants further research; (iii) VLF activity in W, Light and REM sleep stages, which agrees with a macro-sleep disruptions band found by Martín-Montero et al. [17] within this range (0.001–0.005 Hz).

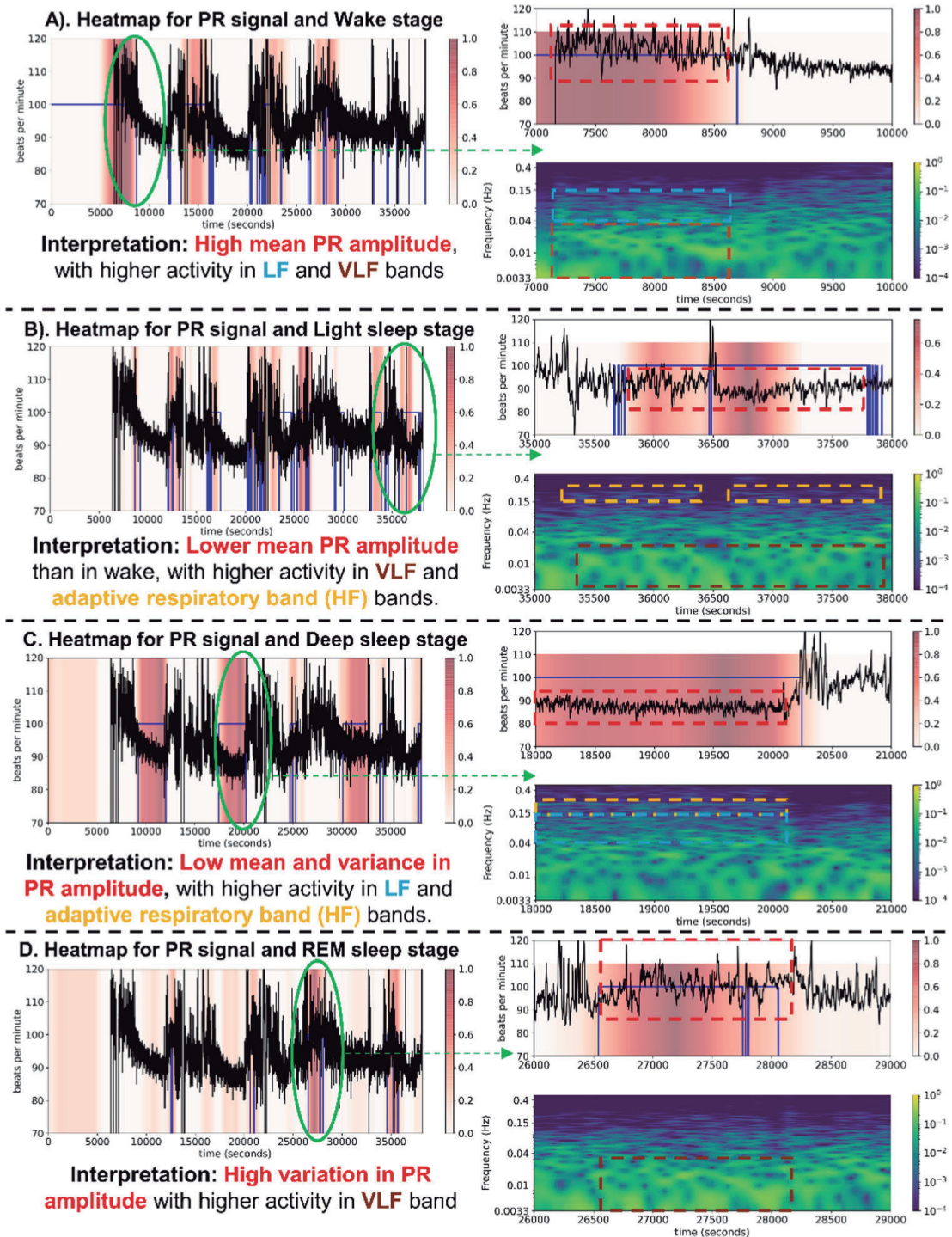


Fig. 2. Seg-Grad-CAM visualizations of the PR signal of a representative subject for: (A) Wake; (B) Light sleep; (C) Deep sleep; (D) REM sleep. Blue lines delineate the regions of interest containing samples predicted as the corresponding sleep stage.

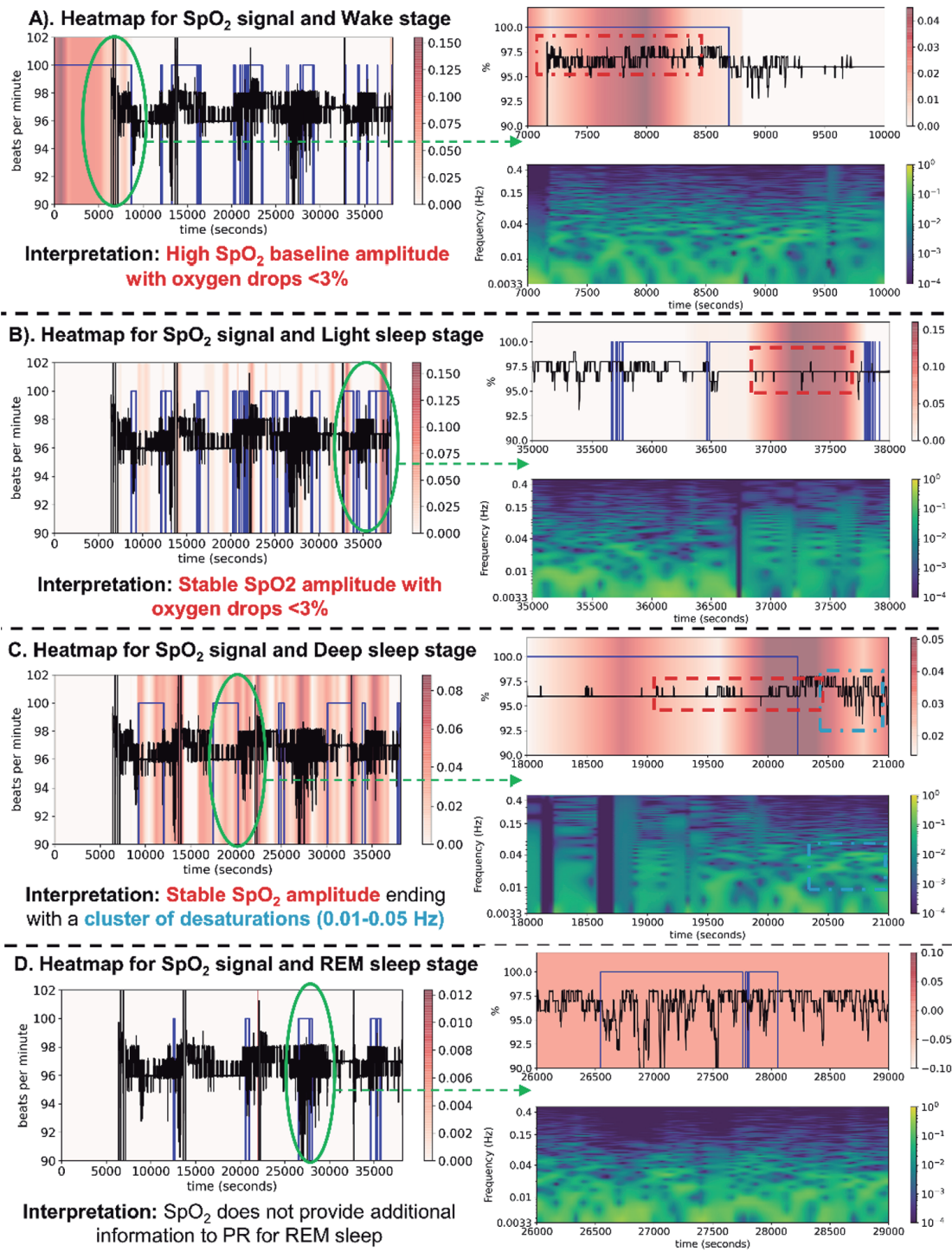


Fig. 3. Seg-Grad-CAM visualizations of the SpO₂ signal of a representative subject for: (A) Wake; (B) Light sleep; (C) Deep sleep; (D) REM sleep. Blue lines delineate the regions of interest containing samples predicted as the corresponding sleep stage.

Seg-Grad-CAM heatmaps also revealed some changes in SpO₂ baseline amplitude and oxygen drops during sleep stages, although their contribution towards sleep staging was considerably lower than PR patterns (see Figs. 2 and 3), which can be explained by the higher performance of U-Net_{PR} with respect to U-Net_{SpO₂}. In light of the reported findings, our XAI approach could facilitate: (i) the development of a novel guideline for

scoring sleep stages from pulse oximetry; (ii) its implementation in a remote processing server or portable devices to offer automatic and interpretable sleep stage predictions.

Recent studies have shown the usefulness of DL approaches applied to pulse oximetry signals for sleep staging in adult [6–9] and pediatric OSA patients [10, 11]. While prior research in adults reported kappa values between 0.64–0.75 for W/Light/Deep/REM detection [6–9], our study focuses on pediatric OSA, introducing an explainable DL model that obtained a kappa value within this range (0.696). In this respect, the differences in cardiorespiratory and neurophysiological activity during sleep in children compared to adults emphasize the need for specific automatic sleep scoring models for pediatric subjects [2]. Conversely, two preliminary works by our group reported kappa values of 0.680 [10] and 0.743 [11] for W/NREM/REM (3-stage) sleep detection in pediatric subjects, whereas the kappa value for 3-stage classification derived from our U-Net model is 0.787. This highlights the importance of considering the whole-night dynamics of PR and SpO₂ during sleep. The current study also contributes with a XAI analysis based on Seg-Grad-CAM that offers an interpretation of pulse oximetry patterns considered by the U-Net for sleep staging.

Despite these interesting results, some limitations remain. First, although the sample size is large, additional pediatric datasets would be desirable to improve model generalizability. Similarly, another potential future goal could be to design and assess a DL model tailored for sleep scoring across various age groups (children, adolescents, adults, and elderly patients). While the proposed interpretability and visualization approach based on Seg-Grad-CAM has successfully identify PR and SpO₂ patterns influencing stage predictions, it has been claimed that current XAI approaches may present challenges for individual patient decision-making [18]. Thus, further research exploring alternative XAI and visualization techniques is warranted for a comprehensive analysis.

In summary, we obtained an accurate U-Net-based DL model to automatically score sleep stages in childhood OSA patients from PR and SpO₂, with a higher performance than the reported by previous studies. Furthermore, our XAI approach based on Seg-Grad-CAM allowed to identify those PR and SpO₂ features contributing to the detection of wake, light sleep, deep sleep, and REM sleep stages. In particular, the following time-frequency patterns of the PR signal have been here highlighted as important for sleep stage detection: mean and variance in PR amplitude, as well as changes in the spectral content of PR within VLF, LF, and HF bands. These results suggest that combining DL and XAI analysis can be useful to accurately perform automatic sleep staging in pulse oximetry tests for pediatric OSA diagnosis.

Acknowledgements. This work was supported by 'Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación/1 0.13039/501100011033/', ERDF A way of making Europe, and NextGenerationEU/PRTR under projects PID2020-115468RB-I00 and PDC2021-120775-I00, and by 'CIBER -Consortio Centro de Investigación Biomédica en Red-' (CB19/01/00012 and CB21/13/00051) through 'Instituto de Salud Carlos III (ISCIII)'. Fernando Vaquerizo-Villar was supported by a 'Sara Borrell' grant (CD23/00031) from the ISCIII cofunded by the European Social Fund (ESF). Daniel Álvarez was supported by a "Ramón y Cajal" Grant (RYC2019-028566-I) funded by MCIN/AEI/1 0.13039/501100011033 and by ESF Investing in your future.

Conflicts of Interest Statement. The authors declare that they have no conflict of interest.

References

1. Sateia, M.J.: International classification of sleep disorders-third edition. *Chest* **146**, 1387–1394 (2014)
2. Berry, R.B., et al.: The AASM manual for the scoring of sleep and associated events. *Am. Acad. Sleep Med.* **53**, 1689–1699 (2018)
3. Chang, J.L., et al.: International consensus statement on obstructive sleep apnea. *Int. Forum Allergy Rhinol.* **13**, 1061–1482 (2023)
4. Stepnowsky, C., et al.: Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters. *Sleep Med.* **14**, 1199–1207 (2013)
5. Gaiduk, M., et al.: Current status and prospects of automatic sleep stages scoring: Review. *Biomed. Eng. Lett.* **13**, 247–272 (2023)
6. Huttunen, R., et al.: Assessment of obstructive sleep apnea-related sleep fragmentation utilizing deep learning-based sleep staging from photoplethysmography. *Sleep* **44**, 1–10 (2021)
7. Sridhar, N., et al.: Deep learning for automated sleep staging using instantaneous heart rate. *npj Digit. Med.* **3**, 106 (2020)
8. Radha, M., et al.: A deep transfer learning approach for wearable sleep stage classification with photoplethysmography. *npj Digit. Med.* **4**, 1–11 (2021)
9. Kotzen, K., et al.: SleepPPG-Net: a deep learning algorithm for robust sleep staging from continuous photoplethysmography. *IEEE J. Biomed. Heal. Informat.* **27**, 924–932 (2023)
10. Vaquerizo-Villar, F., et al.: A convolutional neural network to classify sleep stages in pediatric sleep apnea from pulse oximetry signals. *MELECON 2022 - IEEE Mediterr. Electrotech. Conf. Proc.* 108–113 (2022). <https://doi.org/10.1109/MELECON53508.2022.9842917>
11. Vaquerizo-Villar, F., et al.: A deep learning model based on the combination of convolutional and recurrent neural networks to enhance pulse oximetry ability to classify sleep stages in children with sleep apnea. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1–4. IEEE (2023). <https://doi.org/10.1109/EMBC40787.2023.10341100>
12. Hamon, R., Junklewitz, H., Sanchez, I.: Robustness and explainability of artificial intelligence. *Publ. Off. Eur. Union* (2020)
13. Marcus, C.L., et al.: A randomized trial of adenotonsillectomy for childhood sleep apnea. *N. Engl. J. Med.* **368**, 2366–2376 (2013)
14. Iber, C., et al.: The AASM Manual for the scoring of sleep and associated events: rules, terminology and technical specification. *J. Clin. Sleep Med.* **3**, 752 (2007)
15. Huttunen, R., et al.: A comparison of signal combinations for deep learning-based simultaneous sleep staging and respiratory event detection. *IEEE Trans. Biomed. Eng.* **70**, 1704–1714 (2022)
16. Vinogradova, K., Dibrov, A., Myers, G.: Towards interpretable semantic segmentation via gradient-weighted class activation mapping (student abstract). *Proceedings of the AAAI conference on artificial intelligence* **34**, 13943–13944 (2020)
17. Martín-Montero, A., et al.: Pediatric sleep apnea: Characterization of apneic events and sleep stages using heart rate variability. *Comput. Biol. Med.* **154**, 106549 (2023)
18. Ghassemi, M., et al.: The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit. Heal.* **3**, e745–e750 (2021)

Author Index

A

Álvarez, Daniel 38
Adolf, Jindřich 115
Adorno, Bruno V. 298
Agostino, Accardo 308
Ajčević, Miloš 308
Alabdah, Fahad 227
Aldana Palomino, Ana Cristina 346
Alshammari, Adel 227
Amato, Francesco 217, 373
Angelone, Francesca 217, 373
Arienzo, Arianna 373
Arredondo, María Teresa 197
Arroyo, Peña 197

B

Bini, Fabiano 256
Bliznakov, Zhivko 287
Bliznakova, Kristina 287
Bocchi, Leonardo 95
Bodenstorfer, Mirjam 266
Bohn, Martin 276
Boscarino, Tamara 356
Botis, George 363
Bromis, Konstantinos 363

C

Čuljak, Ivana 151
Cabrera, María Fernanda 197
Calcagno, Andrea 373
Cancilla, Nunzio 55
Cardenas Caceres, Pablo 346
Cardone, Daniela 15
Caricato, Marco 130
Carlson, Jim Ingebretsen 197
Castella, Lorenzo 373
Castillo, Andrés 197
Cieza Huane, Leslie Yessenia 346
Cifrek, Mario 151, 160

Cipollina, Andrea 55
Cooper, Glen 29, 227, 298
Cordella, Fulvio 373
Cortes, Jaime Barrio 197
Cruces Chanchhuaña, André Jesús 346

D

D'Ambrosio, Antonio 356
D'Antoni, Federico 316
de Brito Martins, Ayana 373
Dianderas Jorge, Angel Eduardo 346
Dlakić, Velid 178
Doležal, Jaromír 115
Dukov, Nikolay 287
Dužă, Ștefana 186

F

Fico, Giuseppe 197
Finti, Alessia 256
Florea, Bogdan Cristian 106
Fournier, Guillaume 373
Francia, Piergiorgio 95
Franzo, Michela 256

G

Gao, Yueming 160, 207, 336
García, María 85
Gozal, David 38
Grassi, Roberto 217
Gutiérrez-Tobal, Gonzalo C. 38, 85

H

Hanafusa, Akihiko 276
Haritou, Maria 363
Havlík, Jan 248
Heieck, Frieder 276
Herrero-Tudela, María 85
Hornero, Roberto 38, 85
Hren, Rok 122

I

Iadanza, Ernesto 23, 77
 Iskra, Katerina 308

J

Jagličić, Zvonko 170
 Jazbinšek, Vojko 170
 Jodko-Władzińska, Anna 48
 Jurčová, Kateřina 115

K

Kačerová, Ilona 115
 Kakkos, Ioannis 363
 Kim, Joohyeong 327
 Klaić, Luka 207
 Kouris, Ioannis 363

L

Large, John 29
 Ler, Daria 178
 Lhotská, Lenka 115
 Li, Dongming 336
 Li, Hungchun 336
 Li, Lutong 29
 Liebert, Adam 170
 Lipšová, Vladimíra 115
 Liu, Jinfen 238
 Liu, Jinlong 238
 Lombardi, Sara 95
 Longo, Isabel 316
 Lopez, María I. 85
 Lupiañez, Francisco 197
 Luschi, Alessio 77

M

Mahmutović, Lejla 178
 Maierová, Lenka 248
 Mak, Peng Un 336
 Manea, Ionut 106
 Marhl, Urban 170
 Marinozzi, Franco 256
 Martín-Montero, Adrián 38
 Martynenko, Alexander 141
 Mascianà, Gianluca 130
 Matarrese, Margherita A. G. 130
 Matsopoulos, George 363
 Mattila, Henna 67
 McGrath, Brendan 298

Merdović, Nejra 178
 Merino-Barbancho, Beatriz 197
 Merla, Arcangelo 15
 Merlo, Marco 308
 Merone, Mario 316
 Miladinović, Aleksandar 308
 Milanič, Matija 122
 Miloulis, Stavros-Theofanis 363
 Minev, Teodor 287
 Mohamaddan, Shahrol 276
 Moreno Elescano, Sergio Enrique 346
 Mou, Pedro Antonio 336
 Mrđanović, Emina 178
 Munaretto, Laura 308

N

Neagu, Georgeta Mihaela 106
 Neagu, Georgeta-Mihaela 186
 Nicosia, Alessia 55

P

Pagnano, Maria Elisabetta 130
 Palermo, Andrea 356
 Parrish, Amy 373
 Passerini, Michele 55
 Pastor, Xavier 141
 Pecchia, Leandro 1, 130, 316, 356
 Perpetuini, David 15
 Petrosino, Lorenzo 316
 Piccolo, Salvatore 373
 Piemonte, Vincenzo 316, 356
 Pokvić, Lejla Gurbeta 178
 Ponsiglione, Alfonso Maria 217, 373
 Preoteasa, Rareș-Marin 106
 Pun, Sio Hang 336

R

Rahman, S. M. Musfequr 67
 Raumonon, Pasi 67
 Rizzi, Jacopo Giulio 308
 Roca-Umbert, Ana 197
 Rodríguez, Ana Isabel Villimar 197
 Roglić, Matija 207
 Romano, Francesco 15
 Romano, Maria 373
 Romero-Oraá, Roberto 85
 Rujas, Miguel 197

S

Sander, Tilmann H. 48
Sander, Tilmann 170
Sansone, Mario 217
Sassi, Martina 1
Sau, Francesca 55
Sawosz, Piotr 170
Serša, Gregor 122
Shaikh, Asif 67
Siino, Marco 55
Simončič, Urban 122
Smurthwaite, Glyn 29
Softić, Adna 178
Song, Lipeng 227
Sparaco, Rossella 373
Stergar, Jošt 122
Sultana, Alina Elena 186
Sun, Qi 238

T

Țarălungă, Dragoș Daniel 106
Tamayo, Eduardo 38
Tang, Yuan 298
Tarousi, Marilena 363
Tavernise, Federica 95
Thein, Marc 276
Tinnirello, Ilenia 55

V

Vai, Mang I 336
Vaquerizo-Villar, Fernando 38
Vasić, Željka Lučev 151, 160, 207
Vidović, Domagoj 151
Virkki, Johanna 67
Vitale, Jacopo 1, 130
Vitale, Vincenzo Norman 373

W

Wang, Han 160
Wang, Jiamei 160, 336
Watanabe, Hayato 327
Watson, Stuart 29
Wei, Ziliang 160, 207
Weightman, Andrew 29, 298
Władziński, Michal 48
Wojtkiewicz, Stanislaw 170

X

Xiong, Jiwen 238
Xu, Zhimeng 160

Y

Yang, Jiejie 160
Yang, Shuang 160
Yin, Yadong 336
Yokosawa, Koichi 327