IEEE Region 8

IEEE Italy Section

Università degli Studi di Palermo

# IEEE MELECON 2022

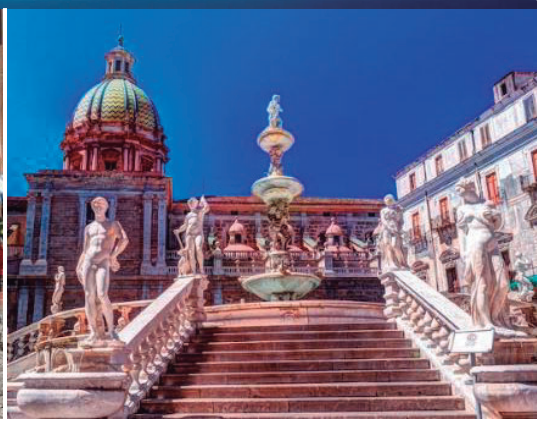
## PALERMO, ITALY / JUNE 14-16, 2022

## IEEE MEDITERRANEAN ELECTROTECHNICAL CONFERENCE

# PROCEEDINGS

**For further information, visit the website**

www.melecon2022.org





IEEE Catalogue Number: CFP22MEL-ART
ISBN: 978-1-6654-4280-0

Proceedings of 2022 IEEE Mediterranean Electrotechnical Conference

# IEEE MELECON 2022



IEEE Catalog Number: CFP22MEL-ART
ISBN: 978-1-6654-4280-0

---

## *SESSION S3A - Services, Applications and Solutions in Smart Healthcare*

**Room: Room 4 - University of Palermo**

**Chairs:** Maria Gabriella Signorini, *Politecnico di Milano, Italy*
Fernando Vaquerizo-Villar, *Institute for Bioengineering of Catalonia, Spain*
Maurizio Naldi, *LUMSA University, Italy*

126 **The Uncertain Times of COVID Mass Vaccine Deliveries: From Start-Up to Steady-State**
*Maurizio Naldi, LUMSA University, Italy*
*Gaia Nicosia, Università Roma Tre, Italy*
*Andrea Pacifici, Università di Roma "Tor Vergata", Italy*

131 **A Model to Improve the Quality of Low-Dose CT Scan Images**
*Francesca Chircop, University of Malta, Malta*
*Carl J. Debono, University of Malta, Malta*
*Paul Bezzina, University of Malta, Malta*
*Francis Zarb, University of Malta, Malta*

137 **PDapp: A Companion Mobile Application With Appcessories for Continuous Follow-Up of Parkinson's Disease Patients**
*Duarte Dias, INESC TEC, Portugal*
*Joana M Silva, Previous Member of INESC Technology and Science, Portugal*
*Nuno Oliveira, Dengun, Portugal*
*João Massano, Centro Hospitalar Universitário de São João, University of Porto, Portugal*
*João Paulo S. Cunha, INESC TEC, University of Porto, Portugal*

142 **Typical External Loads That Require to Be Resisted by a Non-Dominant Prosthetic Hand**
*Yesenia Aquilina, University of Malta, Malta*
*Michael A. Saliba, University of Malta, Malta*
*Jesmond Attard, University of Malta, Malta*
*Alexiei Dingli, University of Malta, Malta*

---

## *SESSION S4A - 3D Networks*

**Room: Room 5 - University of Palermo**
**Chairs:** Barbara Masini, *CNR-IEIIT, Italy*
Franco Davoli, *University of Genoa and CNIT S2N National Lab, Italy*
Mario Marchese, *University of Genoa and CNIT S2N National Lab, Italy*
Ana Perez-Neira, *Centre Tecnològic Telecomunicacions Catalunya, Spain*

148 **Power and Bandwidth Allocation Based on Age of Information Metrics in Satellite Uplink Channels**
*Jorge Torres Gomez, TU Berlin, Germany*
*Máximo Morales-Céspedes, Universidad Carlos III de Madrid, Spain*
*Musbah Shaat, Centre Tecnològic de Telecomunicacions de Catalunya, Spain*
*Ana Isabel Pérez Neira, UPC/CTTC, Spain*
*Ana Garcia Armada, Universidad Carlos III de Madrid, Spain*

154 **NS-3-Based 5G Satellite-Terrestrial Integrated Network Simulator**
*Nour Badini, University of Genoa, Italy*
*Mario Marchese, University of Genoa, Italy*
*Fabio Patrone, University of Genoa, Italy*

160 **Over the Air Computing for Satellite Networks in 6G**
*Marc M. Gost, Universitat Politecnica de Catalunya, Centre Tecnologic de Telecomunicacions de Catalunya, Spain*
*Ana Pérez-Neira, , Universitat Politecnica de Catalunya, Centre Tecnologic de Telecomunicacions de Catalunya, Spain*

165 **Federated Semi-Supervised Classification of Multimedia Flows for 3D Networks**
*Saira Bano, University of Pisa, ISTI-CNR, Italy*
*Achilles Machumilane, University of Pisa, ISTI-CNR, Italy*
*Lorenzo Valerio, CNR, Institute of Informatics and Telematics, Italy*
*Pietro Cassarà, CNR, Institute of Information Science and Technologies, Italy*
*Alberto Gotta, CNR, Institute of Information Science and Technologies, Italy*

---

## *SESSION S1D - Electrical Machines and Drivers for Industry*

**Room: Room 6 - University of Palermo**
**Chairs:** Giorgio Sulligoi, *University of Trieste, Italy*

# A convolutional neural network to classify sleep stages in pediatric sleep apnea from pulse oximetry signals

Fernando Vaquerizo-Villar,
Biomedical Engineering Group,
Universidad de Valladolid, Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, (CIBER-BBN),
Valladolid, Spain
fernando.vaquerizo@gib.tel.uva.es

Daniel Álvarez
Biomedical Engineering Group,
Universidad de Valladolid, Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, (CIBER-BBN),
Valladolid, Spain
daniel.alvarez@uva.es

Gonzalo C. Gutiérrez-Tobal,
Biomedical Engineering Group,
Universidad de Valladolid, Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, (CIBER-BBN),
Valladolid, Spain
gonzalo.gutierrez@gib.tel.uva.es

Félix del Campo,
Hospital Universitario Río Hortega,
Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, (CIBER-BBN),
Valladolid, Spain
fsas@telefonica.net

Leila Kheirandish-Gozal
Department of Child Health, The University of Missouri School of Medicine,
Columbia, Missouri, USA
gozall@health.missuori.edu

David Gozal
Department of Child Health, The University of Missouri School of Medicine,
Columbia, Missouri, USA
gozald@health.missuori.edu

Thomas Penzel
Interdisciplinary Center of Sleep Medicine, Charité-Universitätsmedizin Berlin,
Berlin, Germany
thomas.penzel@charite.de

Roberto Hornero
Biomedical Engineering Group,
Universidad de Valladolid, Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina, (CIBER-BBN),
Valladolid, Spain
robhor@tel.uva.es

***Abstract*— Characterization of the sleep and wake stages is essential in the diagnosis of pediatric obstructive sleep apnea (OSA). The onerous requirements and limitations of overnight polysomnography (PSG), the gold standard, have led to the search for simplified sleep scoring systems. Accordingly, the aim of this study was to assess the usefulness of a convolutional neural network (CNN)-based deep-learning architecture fed with pulse oximetry signals to automatically classify sleep stages in symptomatic children at risk of OSA. Nocturnal pulse rate (PR) and blood oxygen saturation ($SpO_2$) from 429 pediatric OSA patients were employed for this purpose. A 2D CNN architecture was trained using 30-s epochs from PR and $SpO_2$ signals for the automatic classification of the three main sleep stages: wake (W), non-Rapid Eye Movement (non-REM), and REM sleep. The proposed 2D CNN model showed a promising diagnostic performance for the three-stage classification procedure (W/NREM/REM) in an independent test set, with 83.1% accuracy and 0.680 Cohen's kappa, outperforming 1D CNN models trained using PR or $SpO_2$ signals alone. Furthermore, the total sleep time calculated for each subject using the 2D CNN model showed high agreement with the manually scored from PSG (intra-class correlation coefficient of 0.677). These results were superior to previous studies focused on the automated detection of sleep stages in pediatric OSA patients using photoplethysmography and PR signals derived from pulse oximetry. Therefore, joint analysis of PR and $SpO_2$ signals using CNNs can be helpful detect sleep stages in at-home pulse oximetry tests for pediatric OSA diagnosis.**

***Keywords*— *Convolutional neural networks (CNN), deep learning, pediatric obstructive sleep apnea (OSA), pulse oximetry, sleep staging.***

## I. Introduction

Obstructive sleep apnea (OSA) is a highly prevalent condition among children (1.2%-5.7%) [1], which presents differentiating etiological, diagnostic, and therapeutic considerations when compared to adults [1], [2]. Childhood OSA is characterized by recurrent apneas (cessation of airflow) and hypopneas (airflow reductions) during sleep [1], leading to oxygen desaturations, elevations in carbon dioxide, and recurrent arousals, i.e., sleep fragmentation [1], [3]. As a result, affected children are at increased risk of suffering from long-term neurobehavioral abnormalities and diminished cardiovascular and metabolic functions [1], [2], which impose a negative impact on their overall health and quality of life [1], [2]. Thus, it is essential to accelerate the diagnosis of pediatric OSA, which requires assessments during sleep and thus, identification of sleep and wake stages to characterize the sleep architecture.

Overnight polysomnography (PSG) is the gold standard to diagnose childhood OSA [1]. PSG consists in monitoring children's sleep in a specialized laboratory by recording a wide range of cardiorespiratory and neurophysiological signals [4], [5]. After the test, sleep stages and respiratory events are scored from PSG recordings following the guidelines of the American Academy of Sleep Medicine (AASM) [5]. In this respect, electroencephalogram (EEG), electrooculogram (EOG), and submental electromyogram (EMG) signals are visually inspected by a technician to identify the sleep stage corresponding to each non-overlapping 30-s epoch: wake (W), three levels of non-Rapid Eye Movement (non-REM) sleep (N1, N2, and N3) and REM sleep [6]. Conversely, technicians manually score apneic (apnea+hypopnea) events and subsequently obtain the apnea-

hypopnea index (AHI: number of apneas and hypopneas per sleep hour), which is one of the major parameters used to diagnose childhood OSA [5].

Despite its widespread acceptance, PSG is technically complex, costly, highly intrusive, and relatively unavailable, thus delaying the diagnosis and treatment of the affected children [4], [7]. Furthermore, sleep scoring suffers from a considerable inter-rater variability in OSA patients [8], which may alter the accuracy of the diagnosis [8].

Alternatives predicated on the automated analysis of a reduced set of signals have been proposed to overcome these PSG-related limitations. EEG, EOG, electrocardiogram (ECG), airflow, and pulse oximetry (photoplethysmography, pulse rate, and blood oxygen saturation) signals have been used, among others, for the automated detection of sleep stages and/or the diagnosis of OSA and its severity both in adult and pediatric patients [6], [9], [10]. In this regard, pulse oximetry signals have been frequently advocated. These signals can be obtained in a non-invasive way at children's home with a portable pulse-oximeter, typically located on the finger, thus being a simplified alternative to PSG [9].

Multiple studies have demonstrated the utility of the automated analysis of peripheral blood oxygen saturation ($SpO_2$) and pulse rate (PR) from nocturnal oximetry for the screening of pediatric OSA [9], [11]. Particularly, $SpO_2$ and PR signals allow to quantify oxygen desaturations and heart rate responses related to apneic events, respectively [5]. On the other hand, the automated detection of sleep stages in pediatric OSA patients has been marginally explored. In this respect, the PR signal derived from photoplethysmography (PPG) is a surrogate of the heart rate variability (HRV) signal [6], thus containing information regarding changes in the autonomic nervous system (ANS) activity related to sleep stages [12]. Furthermore, the $SpO_2$ baseline value is slightly higher during wakefulness, whereas oxygen desaturations associated to apneic events predominantly occur in REM sleep [13].

Only two preliminary studies have conducted automated sleep staging in pediatric OSA patients using pulse oximetry signals [14], [15]. Dehkordi *et al*. [14] performed sleep staging using common HRV time and spectral features extracted from the PR signal. Recently, a conference paper developed by our own group evaluated a deep-learning methodology based on a convolutional neural network (CNN) architecture to detect sleep stages from raw PPG data [15]. However, no studies have included the $SpO_2$ signal for automatic sleep staging. In this respect, an automated system for the detection of sleep stages from PR and $SpO_2$ could be easily integrated into simplified sleep monitoring devices.

The main novelty of this study is the use of a CNN architecture to automatically classify W, NREM, and REM sleep stages from raw PR and $SpO_2$ signals. As a deep-learning technique, CNNs can automatically learn intricate patterns from the data [16]. We hypothesized that a 2D CNN architecture fed with raw PR and $SpO_2$ data could enhance pulse oximetry ability to detect sleep stages in children suffering from OSA. Thus, our main goal was to evaluate the usefulness of CNNs to classify W, NREM, and REM stages from PR and $SpO_2$ signals in childhood OSA patients. In addition, the secondary goal of this work was to evaluate the usefulness of the 2D CNN model to estimate the total sleep time (TST), which is involved in the calculation of the AHI.

## II. Material and Methods

### A. Subjects and signals

The baseline dataset from the public multicenter Childhood Adenotonsillectomy Trial (CHAT) database was used in this study [17], [18]. The clinical trial is registered as NCT00560859 and its full protocol is provided in the supplementary material of Marcus *et al*. [17]. The baseline dataset is composed of PSG studies of 453 pediatric subjects (5-10 years) suffering from OSA, who were randomized to early adenotonsillectomy or a strategy of watchful waiting [18]. Sleep stages and apneic events were annotated using the AASM 2007 rules in order to derive the hypnogram and the AHI for a definitive diagnosis [5].

Complete PR and $SpO_2$ signals from PSG-derived pulse oximetry were downloaded for 429 pediatric subjects and used different sampling rates ranging from 1 to 512 Hz. These signals were resampled to a common sample frequency of 1 Hz in order to homogenize the frequency, as well as to reduce the computational load [19], [20]. Then, both signals were standardized to normalize heart rate and oxygen saturation baseline levels among different subjects. Finally, PR and $SpO_2$ signals were divided into 30-second non-overlapping epochs, being each epoch labeled as W, NREM, or REM using the sleep stages annotations provided by the clinicians [14], [15].

This dataset was split into three groups: training set (257 first subjects, 60%), used to train the CNN architecture; validation set (86 following subjects, 20%), employed to monitor the convergence of the CNN model; and test set (last 86 subjects, 20%), employed to assess the proposed methodology in an independent group. Table I shows clinical and polysomnographic data from the children under study.

### B. Proposed CNN architecture

Figure 1 shows the overall CNN-based deep-learning architecture used in this study. In order to assess whether PR and $SpO_2$ provide complementary information, the input signal of the CNN is: (*i*) PR signal; (*ii*) $SpO_2$ signal; (*iii*) PR and $SpO_2$ signals. For each of these configurations, the input section of the network consists of the input signal(s) for the 30-s epoch (i.e., 30 samples) to be classified, concatenated with the five preceding and the four following epochs [15], thus having a 10-epochs length 1D (300x1: PR or $SpO_2$) or 2D (300x2: PR+$SpO_2$) input vector. Using preceding and following epochs allows for a better modeling of the temporal context used by technicians for scoring sleep stages [6], [15].

The input is processed by the CNN architecture using six convolutional blocks (conv block), each one consisting of:

- Convolutional layer (conv). This layer extracts feature maps from the data using convolutional filters (kernels) [16]. In this work, each conv layer had 64 filters with a kernel size of: (*i*) 5x1 when the input vector of the CNN is 1D (PR or $SpO_2$); (*ii*) 5x2, when the input vector is 2D (PR+$SpO_2$).
- Batch normalization (BN). BN is employed to normalize the feature maps extracted by the convolutional layer [16].
- Rectified Linear Unit (ReLU). ReLU is the standard activation function for deep neural networks. It introduces a nonlinearity to the normalized feature maps, which allows the CNN to learn complex patterns from the input data [16].

TABLE I. DEMOGRAPHIC AND POLYSOMNOGRAPHIC DATA OF THE SUBJECTS IN THE STUDY

| | All | Training set | Validation set | Test set |
|---|---|---|---|---|
| **Subjects (n)** | 429 | 257 | 86 | 86 |
| **Age (years)** | 6 [5, 8] | 6 [5, 8] | 6 [5, 7] | 6 [5, 7] |
| **Males (n)** | 208 (48.5%) | 126 (49.2%) | 35 (40.7%) | 46 (53.5%) |
| **BMI (kg/m2)** | 17.2 [15.4, 22.0] | 17.1 [15.6, 21.6] | 18.8 [15.2, 23.4] | 16.5 [15.1, 22.5] |
| **AHI (e/h)** | 4.8 [2.7, 8.7] | 4.6 [2.6, 8.8] | 4.6 [2.5, 8.5] | 5.1 [3.2, 9.5] |
| **Wake (n)** | 133891 (25.4%) | 79814 (25.1%) | 27899 (26.9%) | 26178 (25.0%) |
| **NREM (n)** | 319038 (60.6%) | 193547 (60.9%) | 62153 (60.0%) | 63338 (60.6%) |
| **REM (n)** | 73405 (14.0%) | 44724 (14.1%) | 13661 (13.2%) | 15020 (14.4%) |
| **TRT (min)** | 608 [557, 658] | 618 [563, 661] | 589 [540, 651] | 608 [562, 641] |
| **TST (min)** | 466 [429, 494] | 472 [440, 497] | 447 [420, 482] | 461 [423, 500] |

Data are presented as median [interquartile range], n or %. BMI: Body Mass Index; AHI: Apnea-Hypopnea Index; e/h: events per hour; REM: Rapid Eye Movement; NREM: Non-REM; TRT: Total Recording Time; TST: Total Sleep Time

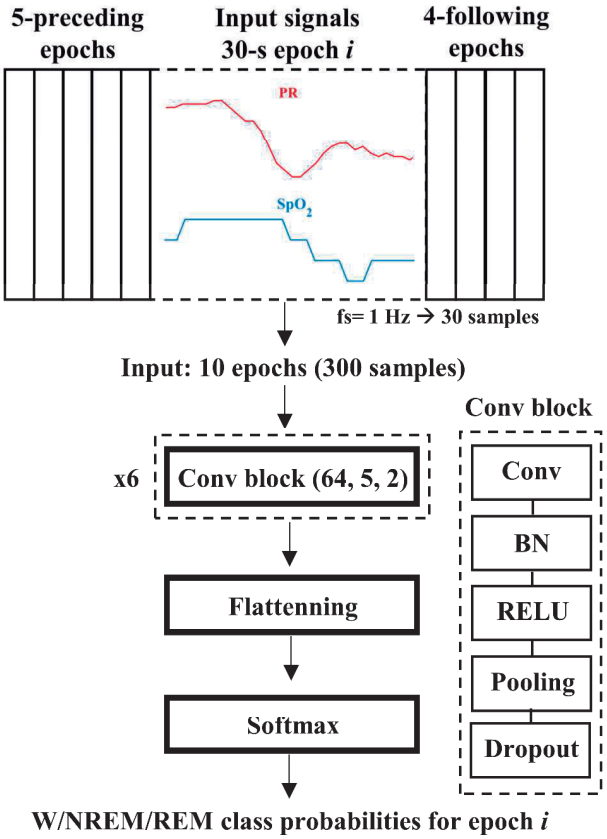



Fig. 1. Overview of the proposed convolutional neural network (CNN) architecture. Each convolutional block (conv block) includes a convolutional layer (conv), batch normalization (BN), a rectified linear unit (RELU) activation function, pooling, and dropout.

- Pooling. After the ReLU function, a max-pooling operation with a pool factor of 2 was applied to reduce dimensionality, while the most relevant feature maps are kept [16].
- Dropout. As the last step in each convolutional block, dropout operation was included to reduce overfitting. It randomly removes node connections with a probability $p$ during the training phase [16]. In this work, a dropout probability $p$=0.1 was chosen [20].

Following the last convolutional block, a flattening layer is used to reshape the feature maps into a 1-D series [16]. Finally, the output layer of the network is a softmax activation function, which yields the probability of belonging to W, NREM, and REM sleep stages for the input 30-s epoch.

The CNN training process was performed using the following configuration: He-normal method for the initialization of weights and biases of the network; categorical cross entropy as the loss function to minimize during training; adaptive moment estimation (Adam) algorithm with an initial learning rate of 0.0001 to update network weights at each iteration; and a batch size of 100 with a random shuffling strategy to accelerate the convergence of the Adam algorithm [15]. During the training process, the loss function was calculated in the validation set to monitor the convergence of the network. Accordingly, the learning rate was reduced by a factor of 2 after 10 epochs with no reduction of the validation loss, and the training was stopped (early stopping) after 30 epochs of non-improvement in the validation loss, restoring the weights to those that minimized the validation loss [16].

### C. *Statistical analysis*

The overall performance of the CNN was assessed by means of confusion matrices, which were used to calculate the 3-class accuracy (Acc) and Cohen's kappa index (kappa). The performance for each sleep stage was evaluated by means of precision (positive predictive value, proportion of epochs rightly classified among all the epochs assigned to the class), recall (sensitivity, percentage of epochs belonging to the class correctly classified), and F1-score (harmonic mean of the precision and recall). Furthermore, the total sleep time (TST) was calculated for each subject based on CNN scoring ($TST_{CNN}$) and compared with the TST obtained in PSG ($TST_{PSG}$). Bland-Altman plots and the intra-class correlation coefficient (ICC) were used to assess their agreement.

## III. RESULTS

### A. *CNN model performance*

Figure 2 shows the confusion matrices of the CNN models obtained using PR ($CNN_{PR}$), $SpO_2$ ($CNN_{SpO_2}$) and PR+$SpO_2$ ($CNN_{PR+SpO_2}$) data in the test set for the three-stage classification procedure (W/NREM/REM). Using $CNN_{PR}$and $CNN_{SpO_2}$, 80.4% and 75.2% of the 30-s epochs in the test set were correctly assigned to their actual sleep stage (sum of the main diagonal elements of the confusion matrix), respectively. Conversely, the combination of PR and $SpO_2$ ($CNN_{PR+SpO_2}$) reached higher accuracy (83.1%) than the single-channel models ($CNN_{PR}$ and $CNN_{SpO_2}$). Similarly, $CNN_{PR+SpO_2}$ reached a higher kappa (0.680) than $CNN_{PR}$ (0.626) and $CNN_{SpO_2}$ (0.478) in the test set.

Table II presents the diagnostic performance of $CNN_{PR}$, $CNN_{SpO_2}$, and $CNN_{PR+SpO_2}$ models for each individual sleep stage. The three CNN models obtained a higher performance for NREM than for REM and W stages. In addition, $CNN_{PR+SpO_2}$ outperformed $CNN_{PR}$ and $CNN_{SpO_2}$, as derived from the values of F1-score in W, NREM, and REM stages.

### B. *Estimation of TST*

Figure 3 shows the Bland-Altman plots of the TST derived from CNN scoring using PR ($TST_{PR}$), $SpO_2$ ($TST_{SpO_2}$) and

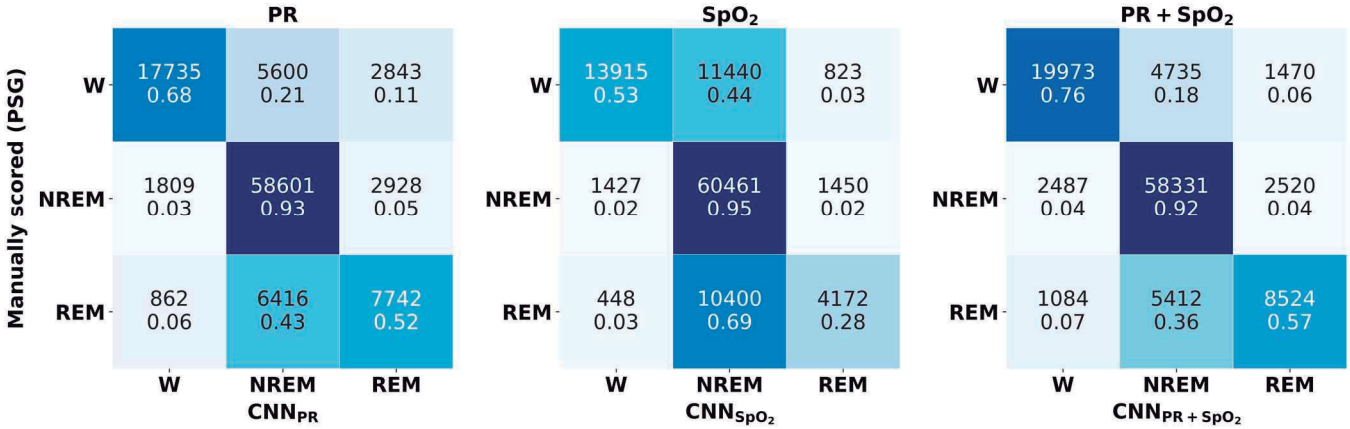



Fig. 2. Confusion matrices of $CNN_{PR}$, $CNN_{SpO_2}$, and $CNN_{PR+SpO_2}$ models in the test set. These matrices compare the sleep stages from standard PSG with those predicted by the CNN models.

TABLE II
DIAGNOSTIC PERFORMANCE OF $CNN_{PR}$, $CNN_{SpO_2}$, AND $CNN_{PR+SpO_2}$ MODELS FOR W, NREM, AND REM SLEEP STAGES

| | Test set | Precision | Recall | F1-score |
|---|---|---|---|---|
| W | $CNN_{PR}$ | 0.87 | 0.68 | 0.76 |
| | $CNN_{SpO_2}$ | 0.88 | 0.53 | 0.66 |
| | $CNN_{PR+SpO_2}$ | 0.85 | 0.76 | 0.80 |
| NREM | $CNN_{PR}$ | 0.83 | 0.93 | 0.87 |
| | $CNN_{SpO_2}$ | 0.73 | 0.95 | 0.83 |
| | $CNN_{PR+SpO_2}$ | 0.85 | 0.92 | 0.89 |
| REM | $CNN_{PR}$ | 0.57 | 0.52 | 0.54 |
| | $CNN_{SpO_2}$ | 0.65 | 0.28 | 0.39 |
| | $CNN_{PR+SpO_2}$ | 0.68 | 0.57 | 0.62 |

CNN = convolutional neural network, $CNN_{PR}$ = CNN model obtained using PR data, $CNN_{SpO_2}$ = CNN model obtained using $SpO_2$ data, $CNN_{PR+SpO_2}$ = CNN model obtained using PR and $SpO_2$ data, W: Wake; REM: Rapid Eye Movement; NREM: Non-REM.

PR+$SpO_2$ ($TST_{PR+SpO_2}$) compared to $TST_{PSG}$ in the test set. ICC is also shown. It can be seen that $TST_{PR}$, $TST_{SpO_2}$, and $TST_{PR+SpO_2}$ overestimated $TST_{PSG}$, as reported by their mean difference (bias). In this respect, $TST_{PR+SpO_2}$ reached lower bias and confidence interval than $TST_{PR}$ and $TST_{SpO_2}$. In addition, $TST_{PR+SpO_2}$ showed higher agreement with manual scoring (ICC=0.677) than $TST_{PR}$ (ICC=0.524) and $TST_{SpO_2}$ (ICC=0.276) in the test set.

## IV. DISCUSSION

In this preliminary work, we evaluated the diagnostic ability of a 2D CNN architecture fed with PR and $SpO_2$ signals from pulse oximetry to detect wake, NREM, and REM sleep stages in pediatric OSA patients. To our knowledge, this is the first study that proposes the joint analysis of PR and $SpO_2$ signals to detect sleep stages in children.

The proposed CNN models reached high performance for the classification of sleep stages, with 80.4% Acc and 0.626 kappa ($CNN_{PR}$), 75.2% Acc and 0.478 kappa ($CNN_{SpO_2}$), and 83.1% Acc and 0.680 kappa ($CNN_{PR+SpO_2}$). Specifically, the PR signal showed higher ability than $SpO_2$ to detect W/NREM/REM stages, which can be explained by wider shifts in heart rate associated with sleep stages. Conversely, the combination of PR and $SpO_2$ ($CNN_{PR+SpO_2}$) produced a slight improvement in performance comparing to using only PR. This indicates that $SpO_2$ provides additional information to PR in the detection of sleep stages, as also reported by Casal *et al*. in adult OSA patients [19]. According to the guidelines of McHugh [21], the kappa value obtained by $CNN_{PR+SpO_2}$ (in the range 0.61-0.79) indicates that there is a substantial agreement between our deep-learning models and manual PSG-based scoring. Hence, our CNN-based approach could be used to detect W, NREM, and REM stages and hence characterize the sleep architecture in at-home pulse oximetry tests for pediatric OSA diagnosis [9], [11].

Regarding the estimation of TST, it has been observed that TST derived from CNN-based automatic scoring ($TST_{PR}$, $TST_{SpO_2}$, and $TST_{PR+SpO_2}$) overestimates $TST_{PSG}$, which can be explained by the slight trend of the CNN models to assign W epochs to the NREM class. This also agrees with the higher values of F1-score obtained for NREM stage. Nonetheless, the TST obtained from the CNN models showed promising results, particularly $TST_{PR+SpO_2}$, which reached an ICC of 0.677. In this respect, an ICC value in the range 0.50-0.75 indicates a moderate agreement [22]. Accordingly, our approach could be used to estimate TST in nocturnal oximetry tests.

In the last years, multiple studies showed the usefulness of deep-learning techniques to automatically classify sleep stages in adult OSA patients from raw biomedical signals [6], including those from pulse oximetry [19], [23]. In this respect, Korkalainen *et al*. [23] applied a CNN combined with a recurrent neural network (RNN) to automatically classify sleep stages from raw PPG data, achieving 80.1% Acc and 0.65 kappa for W/NREM/REM detection. Similarly, Casal *et al*. [19] applied a RNN to raw PR and $SpO_2$ signals for sleep-wake (W/Sleep) classification, reaching 90.1% Acc. In contrast to these studies, our research addresses a slightly higher performance for the detection of W/NREM/REM (83.1% Acc and 0.68 kappa) and W/Sleep (90.7% Acc) in childhood OSA patients, whose sleep architecture and cardiac activity differ from adults [5].

In pediatric OSA patients, Dehkordi *et al*. [14] addressed the automated detection of W/Sleep and NREM/REM from the PPG-derived PR signal using a feature-based approach, reaching 77% Acc (W/Sleep) and 80% Acc (NREM/REM). Additionally, a preliminary study developed by our group applied a CNN to classify W/NREM/REM stages from raw PPG data, reaching 78.3% Acc and 0.57 kappa [15]. In contrast, our current results achieved higher performance with the use of a CNN fed with raw PR and $SpO_2$ data: 83.1% Acc

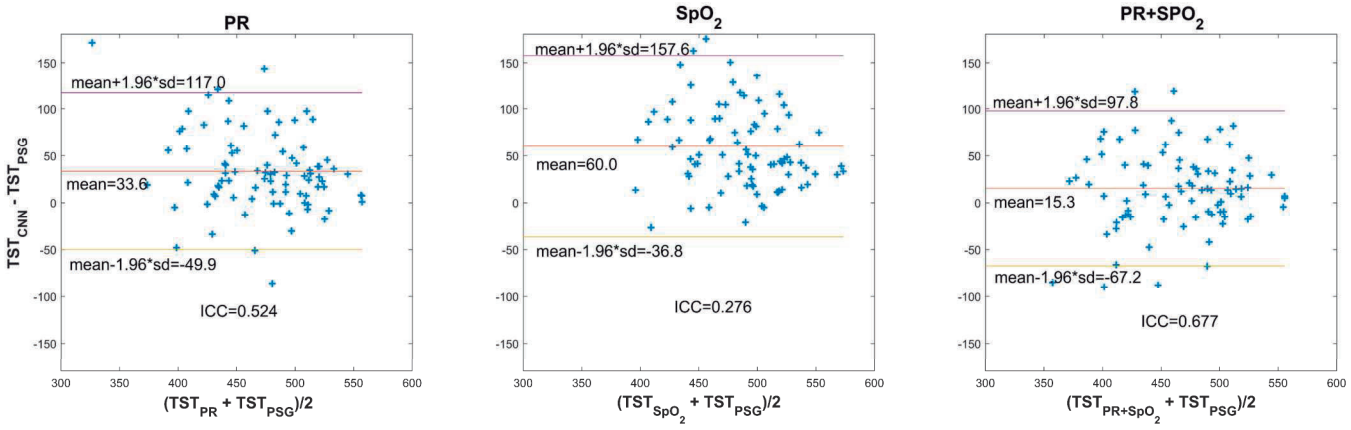



Fig. 3. Bland-Altman plots comparing $TST_{PR}$, $TST_{SpO_2}$, and $TST_{PR+SpO_2}$ with $TST_{PSG}$ in the test set.

for 3-class (W/NREM/REM) classification, 90.7% Acc for W/Sleep detection, and 89.4% for NREM/REM classification.

Despite the potential usefulness of our proposed approach, several limitations need to be considered. First, the database employed in this work did not contain healthy control subjects (AHI< 1 e/h). The inclusion of these subjects could help to analyze the performance of our CNN-based scoring in non-pathological sleep. Furthermore, the use of additional deep-learning techniques (e.g., RNN or time series transformers) may contribute to improve the automated classification of sleep stages at the cost of higher computational load. Finally, the application of eXplainable Artificial Intelligence algorithms could help to identify the patterns of PR and $SpO_2$ signals linked with wake, NREM, and REM stages.

## V. Conclusions

Our 2D CNN-based deep-learning model fed with PR and $SpO_2$ data exhibited high performance in the automated detection of sleep stages in symptomatic children at risk of OSA, outperforming 1D CNN models fed with PR and $SpO_2$ alone, as well as previous approaches based on pulse oximetry signals. The 2D CNN model also showed high agreement in the estimation of TST. Therefore, we conclude that deep-learning approaches can be used to reliably characterize sleep stage distribution in pulse oximetry tests for pediatric OSA diagnosis.

## Acknowledgment

This research has been developed under the grants PID2020-115468RB-I00, PDC2021-120775-I00 and RTC-2017-6516-1 funded by 'Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación/10.13039/501100011033/' and ERDF A way of making Europe; by Sociedad Española de Neumología y Cirugía Torácica (SEPAR) under project 649/2018, by Sociedad Española de Sueño (SES) under project "Beca de Investigación SES 2019" and by 'CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)' through 'Instituto de Salud Carlos III' co-funded with ERDF funds. The work of Daniel Álvarez was supported by a "Ramón y Cajal" grant (RYC2019-028566-I) from the 'Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación' co-funded by the European Social Fund. David Gozal and Leila Kheirandish-Gozal are supported by the Leda J. Sears Foundation.

## References

[1] C. L. Marcus *et al.*, "Diagnosis and management of childhood obstructive sleep apnea syndrome.," *Pediatrics*, vol. 130, no. 3, pp. 576–84, 2012.

[2] O. S. Capdevila, L. Kheirandish-Gozal, E. Dayyat, and D. Gozal, "Pediatric obstructive sleep apnea: Complications, management, and long-term outcomes," *Proc. Am. Thorac. Soc.*, vol. 5, no. 2, pp. 274–282, 2008.

[3] G. M. Loughlin *et al.*, "Standards and indications for cardiopulmonary sleep studies in children," *Am. J. Respir. Crit. Care Med.*, vol. 153, no. 2, pp. 866–878, 1996.

[4] H.-L. Tan, D. Gozal, H. M. Ramirez, H. P. R. Bandla, and L. Kheirandish-Gozal, "Overnight polysomnography versus respiratory polygraphy in the diagnosis of pediatric obstructive sleep apnea," *Sleep*, vol. 37, no. 2, pp. 255–260, 2014.

[5] C. Iber, S. Ancoli-Israel, A. Chesson, and S. F. Quan, "The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification," *J. Clin. Sleep Med.*, vol. 3, no. 7, p. 752, 2007.

[6] O. Faust, H. Razaghi, R. Barika, E. J. Ciaccio, and U. R. Acharya, "A review of automated sleep stage scoring based on physiological signals for the new millennia," *Comput. Methods Programs Biomed.*, vol. 176, pp. 81–91, 2019.

[7] E. S. Katz, R. B. Mitchell, and C. M. D. Ambrosio, "Obstructive Sleep Apnea in Infants," *Am. J. Respir. Crit. Care Med.*, vol. 185, no. 8, pp. 805–816, 2012.

[8] T. Penzel, X. Zhang, and I. Fietze, "Inter-scorer Reliability between Sleep Centers Can Teach Us What to Improve in the Scoring Rules," *J. Clin. Sleep Med.*, vol. 09, no. 01, pp. 89–91, 2013.

[9] F. del Campo, A. Crespo, A. Cerezo-Hernández, G. C. Gutiérrez-Tobal, R. Hornero, and D. Álvarez, "Oximetry use in obstructive sleep apnea," *Expert Rev. Respir. Med.*, vol. 12, no. 8, pp. 665–681, 2018.

[10] M. Shokoueinejad *et al.*, "Sleep apnea: a review of diagnostic sensors, algorithms, and therapies," *Physiol. Meas.*, vol. 38, no. 9, p. R204, 2017.

[11] G. C. Gutiérrez-Tobal, D. Álvarez, L. Kheirandish-Gozal, F. del Campo, D. Gozal, and R. Hornero, "Reliability of machine learning to diagnose pediatric obstructive sleep apnea: Systematic review and meta-analysis," *Pediatr. Pulmonol.*, no. April, pp. 1–13, 2021.

[12] T. Penzel, J. W. Kantelhardt, L. Grote, J. Peter, and A. Bunde, "Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 10, pp. 1143–1151, 2003.

[13] K. Spruyt and D. Gozal, "REM and NREM sleep-state distribution of respiratory events in habitually snoring school-aged community children," *Sleep Med.*, vol. 13, no. 2, pp. 178–184, 2012.

[14] P. Dehkordi, A. Garde, W. Karlen, D. Wensley, J. M. Ansermino, and G. A. Dumont, "Sleep stage classification in children using photo plethysmogram pulse rate variability," *Comput. Cardiol. (2010).*, vol. 41, no. January, pp. 297–300, 2014.

[15] F. Vaquerizo-villar *et al.*, "Automatic Sleep Staging in Children with Sleep Apnea using Photoplethysmography and Convolutional Neural Networks," in *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC 2021)*, 2021, pp. 216–219.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[17] 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, 2013.
[18] S. Redline *et al.*, "The Childhood Adenotonsillectomy Trial (CHAT): Rationale, Design, and Challenges of a Randomized Controlled Trial Evaluating a Standard Surgical Procedure in a Pediatric Population," *Sleep*, vol. 34, no. 11, pp. 1509–1517, 2011.
[19] R. Casal, L. E. Di Persia, and G. Schlotthauer, "Classifying sleep–wake stages through recurrent neural networks using pulse oximetry signals," *Biomed. Signal Process. Control*, vol. 63, no. April 2020, p. 102195, 2021.
[20] 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, 2021.
[21] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochem. Medica*, vol. 22, no. 3, pp. 276–282, 2012.
[22] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, 2016.
[23] H. Korkalainen *et al.*, "Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea," *Sleep*, vol. 43, no. 11, pp. 1–10, 2020.