Convolutional Neural Networks to Detect Pediatric Apnea-Hypopnea Events from Oximetry

Fernando Vaquerizo-Villar*, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, *Member, IEEE*, Verónica Barroso-García, Félix del Campo, David Gozal, Roberto Hornero, *Senior Member, IEEE*

Abstract— Pediatric sleep apnea-hypopnea syndrome (SAHS) is a highly prevalent breathing disorder that is related to many negative consequences for the children's health and quality of life when it remains untreated. The gold standard for pediatric SAHS diagnosis (overnight polysomnography) has several limitations, which has led to the search for alternative tests. In this sense, automated analysis of overnight oximetry has emerged as a simplified technique. Previous studies have focused on the extraction of ad-hoc features from the blood oxygen saturation (SpO_2) signal, which may miss useful information related to apnea and hypopnea (AH) events. In order to overcome this limitation of traditional approaches, we propose the use of convolutional neural networks (CNN), a deep learning technique, to automatically detect AH events from the SpO_2 raw data. CHAT-baseline dataset, composed of 453 SpO_2 recordings, was used for this purpose. A CNN model was trained using 60-s segments from the SpO_2 signal using a training set (50% of subjects). Optimum hyperparameters of the CNN architecture were obtained using a validation set (25% of subjects). This model was applied to a third test set (25% of subjects), reaching 93.6% accuracy to detect AH events. These results suggest that the application of CNN may be useful to detect changes produced in the oximetry signal by AH events in pediatric SAHS patients.

I. INTRODUCTION

Pediatric sleep apnea-hypopnea syndrome (SAHS) is defined as a breathing disorder characterized by the repetitive occurrence of episodes of complete cessation (apneas) and decreases (hypopneas) of breathing during sleep that lead to the disruption of normal oxygenation and normal sleep patterns [1]. SAHS is a highly prevalent condition in children (in the range of 1 to 5%). It is related to many adverse consequences on the overall children's health and quality of life when it remains untreated, including pulmonary hypertension, daytime sleepiness, neurocognitive deficits, and cardiac derangements [1].

Overnight polysomnography (PSG) is the gold standard for pediatric SAHS diagnosis [1]. It requires the presence of the children during the whole night in a specialized sleep laboratory while a wide range of biomedical signals are recorded [1], [2]. These signals are used to score apneas and hypopneas (AH) in order to calculate the apnea-hypopnea index (AHI), which is used to reach a diagnosis [1]. However, PSG is complex, costly, highly intrusive for children, and shows limited availability, thus delaying the treatment [3]. In addition, AH events must be manually scored by expert physicians, which is time-consuming and may lead to subjective diagnoses [2].

In order to overcome these drawbacks, simplified diagnostic techniques are needed [3]. In this regard, overnight oximetry has emerged as a simplified, reliable, and suitable technique for pediatric SAHS diagnosis [3], [4]. Overnight oximetry records the blood oxygen saturation signal (SpO_2), which measures the oxygen content in the hemoglobin [2]. SpO_2 value usually presents drops associated to AH events [2], thus being useful to detect these episodes. In this sense, previous works have employed a methodology based on the automated analysis of the SpO_2 signal in order to assist in the diagnosis of childhood SAHS [5]–[10].

These studies have applied conventional oximetric indices, as well as time, frequency, and nonlinear methods to obtain features that may characterize the properties of the SpO_2 signal [5]–[10]. In addition, feature selection algorithms were used to reduce the dimensionality and improve performance, whereas classification and regression algorithms were employed to detect pediatric SAHS [5]–[10]. However, these machine-learning based frameworks require to determine which features to extract from the physiological signal, as well as to decide how many features are needed, being a difficult task [11]. This could lead to miss useful information from these signals that may help to detect AH events.

In contrast to conventional machine-learning approaches, deep learning techniques provide the ability to automatically learn and extract relevant information from the raw physiological data needed for detection or classification tasks [11]. In this regard, convolutional neural networks (CNN) is the most popular deep learning technique. CNN are well suited

F. Vaquerizo-Villar, G. C. Gutiérrez-Tobal, V. Barroso-García and R. Hornero, are with the Biomedical Engineering Group, Universidad de Valladolid, Paseo de Belén 15, 47011, Valladolid, Spain (phone: +34 983 423000, ext. 3708) (e-mail: fernando.vaquerizo@gib.tel.uva.es).

D. Álvarez and F. del Campo are with the Hospital Universitario Río Hortega of Valladolid, Spain (e-mail: fsas@telefonica.net).

L. Kheirandish-Gozal and D. Gozal are with the Department of Child Health, The University of Missouri School of Medicine, Columbia, Missouri, USA (email: gozald@health.missouri.edu).

to process multidimensional arrays, such as 1D signals or 2D images, due to its multi-layer architecture with shared weights and pooling operations [11]. Previous studies have suggested the ability of CNNs to detect AH events in adult patients with SAHS using cardiorespiratory signals [12]–[16]. To the best of our knowledge, however, there is no study applying deep learning in the context of pediatric SAHS.

Based on the above-mentioned considerations, we hypothesized that CNNs could help to detect the changes in the SpO_2 signal elicited by AH events among pediatric SAHS patients. Thus, the objective of this study is to assess the utility of CNNs to detect AH events in children suffering from SAHS using the oximetry signal.

II. MATERIAL AND METHODS

A. Subjects and signals

The multicenter Childhood Adenotonsillectomy Trial (CHAT)-baseline dataset was employed in this study [17], [18]. Approval for the use of the CHAT database in this study was obtained from the National Sleep Research Resource in the website: www.sleepdata.org. This dataset is comprised of 453 pediatric patients ranging from 5 and 10 years of age. These patients met the criteria defined by Redline et al. [18] for being considered for adenotonsillectomy treatment. PSG was performed and AH events were scored according to the American Academy of Sleep Medicine 2007 guidelines [19]. The AHI was defined as the number of all apneas (central, obstructive, and mixed) and hypopneas associated with an arousal or a 3% desaturation. SpO_2 signals were acquired at sampling rates of 1, 2, 10, 12, 16, 200, 256, and 512 Hz.

The dataset was divided into three sets: a training set (50%), employed to train the CNNs, a validation set (25%), used to obtain the optimum values for the hyperparameters of the CNN, and a test set (25%), employed to assess the performance of our proposal. Table I shows clinical and demographic data of the population under study.

B. Signal preprocessing and segmentation

Firstly, SpO_2 recordings were resampled to a common sample rate of 1 Hz and were also rounded to the second decimal place in order to homogenize the frequency and resolution in all recordings [5]. The resampled signals were segmented into epochs of 60 non-overlapping seconds prior to train the CNN model. A segment was scored as apnea or hypopnea if at least 50% of an AH event was present in this segment. According to this rule, the distribution of total segments was 246655 normal segments and 30317 AH segments. However, an equal number of normal and AH segments was chosen from the training set, since CNN are sensitive to an imbalanced proportion between classes. Thus, the training set was composed of 20461 normal and 20461 AH segments. The validation set was composed of 64669 normal and 4570 AH segments, while the test set had 63271 normal and 5286 AH segments.

C. CNN-based deep learning architecture

CNN were originally inspired for image analysis. However, CNN have also proven to be useful to capture local connections in 1D time series [11]. Fig 1 shows the CNN-based deep learning architecture employed in this study. The whole CNN-based deep learning architecture has three main

TABLE I.   DEMOGRAPHIC AND CLINICAL DATA OF THE SUBJECTS UNDER STUDY

|	All	Training group	Validation group	Test group
Subjects (n)	453	227	113	113
Age (years)	6 [3]	6 [3]	6 [3]	6 [2]
Males (n)	219 (48.3%)	120 (52.9%)	45 (39.8%)	54 (47.8%)
BMI (kg/m2)	17.1 [6.6]	17.2 [6.3]	17.4 [6.2]	16.5 [7.3]
AHI (e/h)	5.7 [6.1]	5.9 [6.1]	4.6 [6.6]	6.3 [5.7]
AHI ≥ 5 (e/h)	256 (56.5%)	137 (60.3%)	52 (46.0%)	67 (59.3%)
AHI ≥ 10 (e/h)	106 (23.4%)	57 (25.1%)	24 (21.2%)	25 (22.1%)

Data are presented as median [interquartile range], n or %. BMI: Body Mass Index; AHI: Apnea-Hypopnea Index

blocks: input layer, CNN layers, and classification layers. Each component of the architecture is described as follows:
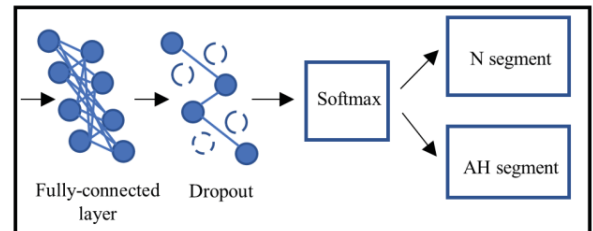
- Input layer. This layer receives the SpO_2 signals preprocessed and segmented. Each segment has 60 samples (60-s).

- CNN layers. Each CNN layer is composed of batch normalization, convolutional layers, rectified linear units (ReLU) activation layers, pooling layers, and dropout.
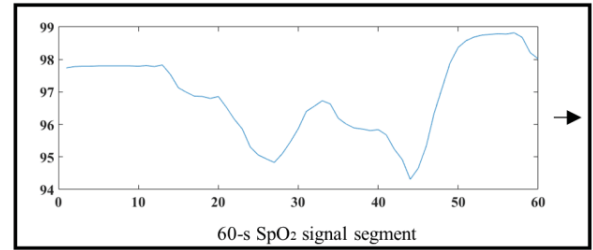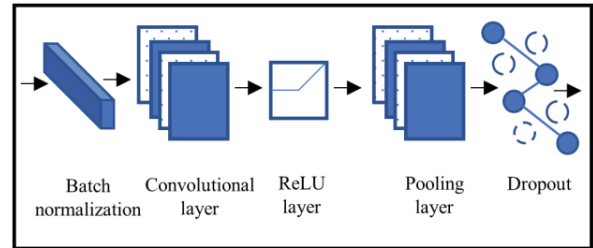
Figure 1. Schematic diagram of the proposed CNN-based deep learning architecture: (a) input layer contains the input SpO_2 segments (b) CNN layers with batch normalization, convolutional layers, ReLU layer, pooling layer, and dropout, and (c) classification layers, with the fully-connected layers, dropout, and the softmax layer. The ouput of this block is the label of a normal or AH segment.

First, batch normalization was employed to normalize the input data [20]. Then, in the convolutional layer, a convolutional filter (kernel) is applied to detect local patterns from the input data [20]:

$$x_k^l = b_k^l + \sum_{j=1}^N w_k^{l-1} * y_j^{l-1}, \qquad (1)$$

where x_k^l is the kth feature map in the layer l, b_k^l is the bias of the kth feature map in the layer l, w_k^{l-1} is the kth convolutional kernel in the layer l-1, y_j^{l-1} is the output of the jth feature map in the layer l-1, N is the total number of features in the layer l-1, and * is the convolution operator.

The ReLU layer is the activation layer that follows the convolutional layer. It has shown to produce a robust training performance [20]. This activation function performs a thresholding operation [20]:

$$f(x) = max(0, wx + b), \qquad (2)$$

where x is the feature map, w denotes the weight factor, and b is the bias. A pooling layer follows the convolutional and ReLU layers. It reduces the dimensionality by merging similar patterns into one. In this study, a max pooling layer was used [20].

Finally, dropout was used in order to keep the generalizability of the network [20]. Dropout randomly removes node connections with a probability p_{drop}, thus preventing overfitting [20].

- Classification layers: The classification block is composed of fully connected (FC) layers and a softmax layer. In these layers, all the units are connected to the units from the preceding layer, like in traditional neural networks [11]. A ReLU activation function was used in the FC layers, whereas a softmax activation function is used in the output layer to determine whether the input segment is a normal or AH segment.

D. Implementation and performance analysis

Keras library was used with a Tensorflow backend to implement and train the CNN model. A workstation with a GeForce GTX 1080 Ti GPU was used for this purpose. Diagnostic ability of the CNN model was assessed by means of sensitivity (Se, percentage of AH segments correctly classified), specificity (Sp, percentage of normal segments correctly classified), accuracy (Acc, percentage of segments correctly classified), and Cohen's kappa index (kappa).

III. RESULTS

A. Optimization of the CNN model

Several experiments were conducted to obtain the optimum values of the hyperparameters of the CNN model. The following hyperparameters were varied: number of CNN layers ($n_{CNNlayers}$), number of filters in each convolutional layer ($nfilters$), kernel size ($kern_{size}$), pooling size ($pool_{size}$), stride in the convolutional and pooling layers ($conv_{stride}$ and $pool_{stride}$), number of FC layers ($n_{FClayers}$), number of units in the FC layers ($nunits_{FClayers}$), dropout probability (p_{drop}) and batch size ($batch_{size}$). CNN models were trained using the training set to search for the optimal combination of hyperparameters. He-normal initializer was used for the convolutional filters in the

convolutional layers and for the units in the FC layers, whereas the adam algorithm with the default parameters was used to train the CNN models in 50 epochs [20]. Finally, the following values of the hyperparameters were obtained as those for which the kappa value was the highest in the validation set: $n_{CNNlayers}$=6, $nfilters$=16, $kern_{size}$=5, $conv_{stride}$=1, $pool_{size}$=5, $pool_{stride}$=1, $n_{FClayers}$=1 $nunits_{FClayers}$=1000, p_{drop}=0.1, and $batch_{size}$ =256.

B. Performance of the CNN model

The obtained CNN model was further assessed in an independent test set. Table 2 shows the confusion matrix of the CNN model for the detection of AH events in pediatric SAHS patients, as well as the diagnostic performance parameters (Se, Sp, Acc, and kappa) in the test set. Notice that the proposed CNN model reached high accuracy (93.6%), as well as high specificity (96.7%).

IV. DISCUSSION AND CONCLUSIONS

This study assessed the usefulness of CNN to detect AH events in pediatric SAHS events when only using the oximetry signal. To our knowledge, the application of deep learning techniques to detect AH events is novel in the context of pediatric SAHS. A CNN model trained with 60-s segments from the SpO_2 signal reached high performance to detect AH events (93.6% Acc). A high specificity (96.7%) was also achieved. In this sense, only a low proportion of normal segments (3.3%) are considered as AH events. These misclassified segments might be due to desaturations in the SpO_2 signal that are not related to apneas or hypopneas. On the other hand, a low sensitivity (56.5%) was obtained. The misclassified AH events could occur in some AH events that may not produce any perturbation in the SpO_2 signal.

Previous studies applied CNN to detect AH events in adult patients with SAHS [12]–[16]. Dey et al. [12] applied CNN to detect apneic events using 35 electrocardiogram (ECG) recordings from the Apnea-ECG database divided in 1-minute segments. The CNN model achieved 98.2% Acc (97.8% Se and 99.2% Sp). Wang et al. [13] analyzed the same database. In their study, they applied CNN to the RR intervals extracted from ECG, reaching 97.8% Acc (93% Se and 100% Sp). However, the database used in these studies does not contain annotations for hypopnea events [12], [13]. Urtnasan et al. [14], [15] also applied CNN models to ECG recordings in order to detect AH events. A CNN model was applied by Urtansan et al. [14] to 10-s ECG segments from a database of 82 adult SAHS patients. An accuracy of 96% (96% Se) was reached to detect apneic events. A multiclass CNN architecture was also designed by Urtnasan et al. [15] to detect apnea and hypopnea events in 10-s ECG segments from 86 adult SAHS patients. They reached 90.8% Acc (87.0% Se) to

TABLE II.   DIAGNOSTIC PERFORMANCE OF THE CNN MODEL IN THE TEST SET

| | | Reference | | Se | Sp | Acc | kappa |
		N	AH				
Estimated	N	61157	2298	56.5	96.7	93.6	0.54
	AH	2114	2988				

differentiate between normal segments, apneas, and hypopneas. Nonetheless, patients with central and mixed apneas were not included in these studies [14], [15]. Choi et al. [16] used a CNN-based approach to detect AH events from the nasal-pressure (NP) signal. They analyzed 10-s NP segments, reaching 96.6% Acc (81.1% Se, and 98.5% Sp). Despite these studies reported high accuracies, they applied CNN to detect AH events in adult patients, while our study uses CNN in the context of pediatric SAHS. In this regard, scoring rules for AH events are more restrictive in children than in adults [19]. In addition, these studies used the ECG and NP signals [12]–[16], whereas our study uses only the SpO_2 signal. The use of the oximetry signal to assist in the diagnosis of pediatric SAHS has been frequently advocated, being especially suitable for children [3]. The development of portable monitoring devices has allowed to record the SpO_2 signals with a pulse oximeter at the patient's home in children suffering from SAHS [10].

Recent works have focused on the application of signal processing techniques to the SpO_2 signal in order to assist in the diagnosis of pediatric SAHS [5]–[10]. These studies used conventional machine-learning approaches: feature extraction, feature selection, and classification and regression algorithms. These studies reported accuracies in the range 75-85%, 81-84%, and 85-91% for the detection of mild (AHI≥1), moderate (AHI≥5), and severe SAHS (AHI≥10), respectively [5]–[10]. However, these studies focused on detecting the presence and severity of pediatric SAHS [5]–[10], whereas our study aims at detecting AH events. In this regard, an accurate detection of AH events could be useful to estimate AHI with a higher accuracy than these studies. In addition, our methodology avoids the need to determine which features to extract from the SpO_2 recordings.

This study presents some limitations that should be considered. First, the database used in this study does not contain no-SAHS subjects (AHI<1 e/h). The inclusion of these subjects could help to better characterize normal segments. Another limitation concerns the size of 60-s SpO_2 segments to detect AH events. In this sense, there may be more than one AH event in a 60-s segment. However, it is necessary to have a long time window, since desaturations may occur more than 30 seconds after the onset of an AH event [21]. The only use of the SpO_2 signal also limits our results, since the oximetry signal may not contain information of some AH events [19]. In this regard, the use of SpO_2 together with other physiological signals may enhance the detection of AH events. Finally, the use of recurrent neural networks (RNN) might help to improve the performance of our proposal, since RNN are well suited to model time-dependences in 1D data [20].

In summary, we applied CNN, a deep learning technique, to detect the changes produced in the oximetry signal by AH events among pediatric SAHS patients. A CNN model trained with 60-s SpO_2 segments showed promising results, reaching 93.6% Acc to differentiate between normal segments and AH events in the test set. These results suggest that deep learning approaches could be useful to detect AH events in pediatric SAHS patients.

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] S. M. *et al.*, "Sleep apnea: a review of diagnostic sensors, algorithms, and therapies," *Physiol. Meas.*, vol. 38, no. 9, pp. R204–R252, 2017.

[3] G. M. Nixon, A. S. Kermack, G. M. Davis, J. J. Manoukian, A. Brown, and R. T. Brouillette, "Planning adenotonsillectomy in children with obstructive sleep apnea: the role of overnight oximetry," *Pediatrics*, vol. 113, no. 1, pp. e19–e25, 2004.

[4] N. Netzar, A. H. Eliasson, C. Netzar, and D. A. Kristo, "Overnight Pulse Oximetry for Sleep- Disordered Breathing in Adults," *CHEST J.*, vol. 120, no. 2, pp. 625–633, 2001.

[5] R. Hornero *et al.*, "Nocturnal Oximetry-based Evaluation of Habitually Snoring Children.," *Am. J. Respir. Crit. Care Med.*, vol. 196, no. 12, pp. 1591–1598, 2017.

[6] D. Álvarez *et al.*, "Automated Screening of Children With Obstructive Sleep Apnea Using Nocturnal Oximetry: An Alternative to Respiratory Polygraphy in Unattended Settings," *J Clin Sleep Med*, vol. 13, no. 5, pp. 7–11, 2017.

[7] F. Vaquerizo-Villar *et al.*, "Utility of bispectrum in the screening of pediatric sleep apnea-hypopnea syndrome using oximetry recordings," *Comput. Methods Programs Biomed.*, vol. 156, pp. 141–149, 2018.

[8] F. Vaquerizo-Villar *et al.*, "Wavelet analysis of oximetry recordings to assist in the automated detection of moderate-to-severe pediatric sleep apnea-hypopnea syndrome," *PLoS One*, vol. 13, no. 12, p. e0208502, 2018.

[9] F. Vaquerizo-Villar *et al.*, "Detrended fluctuation analysis of the oximetry signal to assist in paediatric sleep apnoea-hypopnoea syndrome diagnosis," *Physiol. Meas.*, vol. 39, no. 11, p. 114006, 2018.

[10] D. Álvarez *et al.*, "Symbolic dynamics to enhance diagnostic ability of portable oximetry from the phone oximeter in the detection of paediatric sleep apnoea," *Physiol. Meas.*, 2018.

[11] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[12] D. Dey, S. Chaudhuri, and S. Munshi, "Obstructive sleep apnoea detection using convolutional neural network based deep learning framework," *Biomed. Eng. Lett.*, vol. 8, no. 1, pp. 95–100, 2018.

[13] X. Wang *et al.*, "Obstructive sleep apnea detection using ecg-sensor with convolutional neural networks," *Multimed. Tools Appl.*, pp. 1–15, 2018.

[14] E. Urtnasan, J. U. Park, E. Y. Joo, and K. J. Lee, "Automated Detection of Obstructive Sleep Apnea Events from a Single-Lead Electrocardiogram Using a Convolutional Neural Network," *J. Med. Syst.*, vol. 42, no. 6, 2018.

[15] E. Urtnasan, J. U. Park, and K. J. Lee, "Multiclass classification of obstructive sleep apnea/hypopnea based on a convolutional neural network from a single-lead electrocardiogram," *Physiol. Meas.*, vol. 39, no. 6, 2018.

[16] S. H. Choi *et al.*, "Real-time apnea-hypopnea event detection during sleep by convolutional neural networks," *Comput. Biol. Med.*, vol. 100, no. February, pp. 123–131, 2018.

[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] 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.

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

[21] A. Kulkas, P. Tiihonen, P. Julkunen, E. Mervaala, and J. Töyräs, "Desaturation delay, parameter for evaluating severity of sleep disordered breathing," *IFMBE Proc.*, vol. 39 IFMBE, pp. 336–339, 2013.