Thomas Penzel • Roberto Hornero
Editors

Advances in the Diagnosis and Treatment of Sleep Apnea

Filling the Gap Between Physicians and Engineers

Springer

Editors
Thomas Penzel
Sleep Medicine Center
Charite Universitätsmedizin Berlin
Berlin, Berlin, Germany

Roberto Hornero
Biomedical Engineering Group
University of Valladolid
Valladolid, Spain

Centro de Investigación Biomédica en
Red Bioingeniería
Biomateriales y Nanomedicina
(CIBER-BBN)
Valladolid, Spain

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-031-06412-8 ISBN 978-3-031-06413-5 (eBook)
https://doi.org/10.1007/978-3-031-06413-5

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

Preface

Sleep apnea is a sleep disorder with a very high prevalence and many health consequences. As such it is a major health burden (Benjafield et al., 2019). Sleep apnea has been systematically explored only a little more than 40 years now (Guilleminault & Dement, 1978). Major impacts of sleep apnea are sleepiness and associated risks for accidents (Bonsignore et al., 2021). Major health impacts are cardiovascular risk and pathophysiological traits, even if this is currently much debated when focusing on the apnea-hypopnea index as the measure for sleep apnea severity (Arnaud et al., 2020). Sleep apnea is a disorder which is a chronic condition and can be treated successfully.

The disorders of sleep-disordered breathing have largely supported the growth of sleep medicine in general from a small specialty field to a major spectrum of disorders in the arena of medical specialties. This activity helped to convert the niche field of sleep research into sleep medicine, a clinical discipline with its own departments, its own center certification, physician certification, dedicated conferences, journals, and research activities. The recognition and importance have grown so much that the new International Classification of Disorders by WHO in its 11th version, being launched in 2022, has added a new section on sleep and wake disorders with its own range of codes. This worldwide recognition will enable the growth of medical education on sleep physiology, sleep pathology, and specific sleep disorders.

The diagnostic field for sleep disorders, and for sleep apnea specifically, is strongly linked to the development of new and recent methods, which allow long-term recording and analysis of physiological functions during sleep. Sleep and sleep apnea are not just identified by taking a single blood sample or by a single measurement by a physician at a visit, but sleep recording requires the continuous recording of biosignals. This is comparable to monitoring of vital functions during anesthesia or intensive care. Because of this methodological challenge, biomedical engineering as well as new sensor and analysis technologies are closely linked to the development of sleep apnea diagnosis. New technologies helped to a large extent develop new diagnostic and treatment modalities for sleep-disordered breathing. Sleep apnea diagnostic research is now linked to the development of new wearables, nearables, and smartphone apps, and profits much from the ubiquitous development of photoplethysmography recording everywhere.

Artificial intelligence is playing a very important role in analyzing sleep recordings and, particularly, in automatizing several of the stages of sleep apnea diagnosis. Since the generalization of computerized analysis in the 1990s, automated processing of cardiorespiratory and neuromuscular signals from polysomnographic studies provided a number of indices able to assist sleep experts in the characterization of the disease (Shokoueinejad et al., 2017). Parameterization of the influence of apneic events on biological system dynamics has relied on widely known techniques from the engineering field, such as spectral and nonlinear analysis. Currently, there is a demand for novel alternative metrics able to overcome the limitations of the standard apnea-hypopnea index concerning its low association with patient symptoms and outcomes (Malhotra et al., 2021). In this regard, signal processing and pattern recognition are going to play a key role. In addition, machine learning has also shown its usefulness in the last decades (Uddin et al., 2018) and, like many other areas in our society, sleep apnea diagnosis is rapidly entering the deep learning era (Mostafa et al., 2019) and big data. These new analytical techniques, along with the advances in health device development, are the main hope for reaching a reliable diagnostic paradigm shift. One that finally could cope with the disease prevalence, personalized interventions, and runaway spending.

Beyond the widespread application of machine learning methods to automate polysomnography scoring and to provide sleep experts with tools for automated diagnosis, artificial intelligence has also the potential to significantly improve the management of sleep apnea treatment. Recent advances in the framework of big data together with remote monitoring capability of novel treatment devices are able to promote conventional sleep medicine towards a real personalized medicine. Identification of refined clinical phenotypes of patients will allow the development of precision interventions, enabling the quick identification of the treatment option that best fits the particular characteristics of a patient (Watson & Fernández., 2021). Similarly, machine learning is able to accurately model patient's adherence from usage data (pressure setting, residual respiratory events, mask leaks) derived from portable treatment devices, improving the efficacy of available therapies (Goldstein et al., 2020). Thus, artificial intelligence is going to significantly change the management of sleep apnea treatment in the short term.

This volume gives a basis of current knowledge on sleep research, sleep medicine, and sleep apnea, with a strong focus on new challenges and new research directions in the diagnosis of sleep apnea and its treatment. The volume contains three sections: the first one is on physiology and pathophysiology, the second one is on diagnostic advances, and the third one is on treatment advances. Each chapter author was asked to not only describe the state of the art but also develop visions for future research as seen from their special angle and viewpoint.

As editors, we think that the volume can serve as an introduction to the field of sleep-disordered breathing, can serve as a basis for educating in sleep-disordered breathing, and can immediately stimulate and trigger new research in physiology, clinical trials, and biomedical engineering for sensors and analysis methodologies.

Berlin, Berlin, Germany — Thomas Penzel
Valladolid, Spain — Roberto Hornero

References

Arnaud, C., Bochaton, T., Pépin, J. L., & Belaidi, E. (2020). Obstructive sleep apnoea and cardiovascular consequences: Pathophysiological mechanisms. *Archives of Cardiovascular Disease, 113*:350—358

Benjafield, A. V., Ayas, N. T., Eastwood, P. R., Heinzer, R., Ip, M. S. M., Morrell, M. J., Nunez, C. M., Patel, S. R., Penzel, T., Pepin, J. L. D., Peppard, P. E., Sinha, S., Tufik, S., Valentine, K., & Malhotra, A. (2019). Estimating the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respiratory Medicine, 7*:687–698. 10.1016/S2213-2600(19)30198-5

Bonsignore, M.R., Randerath, W., Schiza, S., Verbraecken, J., Elliott, M. W., Riha, R., Barbe, F., Bouloukaki, I., Castrogiovanni, A., Deleanu, O., Goncalves, M., Leger, D., Marrone, O., Penzel, T., Ryan, S., Smyth, D., Teran-Santos, J., Turino, C., McNicholas, W. T. (2021). European Respiratory Society statement on sleep apnoea, sleepiness and driving risk. *European Respiratory Journal, 57*: 2001272 doi: 10.1183/13993003.01272-2020

Goldstein, C. A., Berry, R. B., Kent, D. T., Kristo, D. A., Seixas, A. A., Redline, S., Westover, M. B., Abbasi-Feinberg, F., Aurora, R. N., Carden, K. A., Kirsch, D. B., Malhotra, R. K., Martin, J. L., Olson, E. J., Ramar, K., Rosen, C. L., Rowley, J. A., Shelgikar, A. V. (2020). Artificial intelligence in sleep medicine: An American Academy of Sleep Medicine position statement. *Journal of Clinical Sleep Medicine, 16*(4):605-607. 10.5664/jcsm.8288

Guilleminault, C., & Dement, W. C. (eds) (1978). *Sleep apnea syndromes*. New York: Alan R. Liss Inc.

Malhotra, A., Ayappa, I., Ayas, N., Collop, N., Kirsch, D., Mcardle, N., Mehra, R, Pack, A. I., Punjabi, N., White, D. P., & Gottlieb, D. J. (2021). Metrics of sleep apnea severity: beyond the apnea-hypopnea index. *Sleep, 44*(7):1-16. 10.1093/sleep/zsab030

Mostafa, S. S., et al. (2019). A systematic review of detecting sleep apnea using deep learning. *Sensors, 19.22*: 4934. 10.3390/s19224934

Shokoueinejad, M., Fernandez, C., Carroll, E., Wang, F., Levin, J., Rusk, S., Glattard, N., Mulchrone, A., Zhang, X., Xie, A., Teodorescu, M., Dempsey, J., & Webster, J. (2017). Sleep apnea: a review of diagnostic sensors, algorithms, and therapies. *Physiol Meas, 38*:R204–R252 doi: 10.1088/1361-6579/aa6ec6

Uddin, M. B., Chow, C. M., & S. W. Su. (2018). Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: A systematic review. *Physiological Measurement, 39.3*: 03TR01. 10.1088/1361-6579/aaafb8

Watson, N. F., & Fernandez, C. R. (2021). Artificial intelligence and sleep: Advancing sleep medicine. *Sleep Med Rev, 59*:101512. 10.1016/j.smrv.2021.101512

Contents

15 Deep-Learning Model Based on Convolutional Neural Networks to Classify Apnea–Hypopnea Events from the Oximetry Signal

Fernando Vaquerizo-Villar, Daniel Álvarez, Gonzalo C. Gutiérrez-Tobal, C. A. Arroyo-Domingo, F. del Campo, and Roberto Hornero

Abstract

Automated analysis of the blood oxygen saturation (SpO_2) signal from nocturnal oximetry has shown usefulness to simplify the diagnosis of obstructive sleep apnea (OSA), including the detection of respiratory events. However, the few preceding studies using SpO_2 recordings have focused on the automated detection of respiratory events versus normal respiration, without making any distinction between apneas and hypopneas. In this sense, the characteristics of oxygen desaturations differ between obstructive apnea and hypopnea episodes. In this chapter, we use the SpO_2 signal along with a convolutional neural network (CNN)-based deep-learning architecture for the automatic identification of apnea and hypopnea events. A total of 398 SpO_2 signals from adult OSA patients were used for this purpose. A CNN architecture was trained using 30-s epochs from the SpO_2 signal for the automatic classification of three classes: normal respiration, apnea, and hypopnea. Then, the apnea index (AI), the hypopnea index (HI), and the apnea–hypopnea index (AHI) were obtained by aggregating the outputs of the CNN for each subject (AI_{CNN}, HI_{CNN}, and AHI_{CNN}). This model showed a promising diagnostic performance in an independent test set, with 80.3% 3-class accuracy and 0.539 3-class Cohen's kappa for the classification of respiratory events. Furthermore, AI_{CNN}, HI_{CNN}, and AHI_{CNN} showed a high agreement with the values obtained from the standard PSG: 0.8023, 0.6774, and 0.8466 intra-class corre-

F. Vaquerizo-Villar (✉) · G. C. Gutiérrez-Tobal
R. Hornero
Biomedical Engineering Group,
University of Valladolid, Valladolid, Spain

Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Instituto de Salud Carlos III, Valladolid, Spain
e-mail: fernando.vaquerizo@gib.tel.uva.es

D. Álvarez
Biomedical Engineering Group, University of Valladolid, Valladolid, Spain

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

Pneumology Service, Río Hortega University Hospital, Valladolid, Spain

C. A. Arroyo-Domingo · F. del Campo
Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Instituto de Salud Carlos III, Valladolid, Spain

Pneumology Service, Río Hortega University Hospital, Valladolid, Spain

T. Penzel, R. Hornero (eds.), *Advances in the Diagnosis and Treatment of Sleep Apnea*, Advances in Experimental Medicine and Biology 1384, https://doi.org/10.1007/978-3-031-06413-5_15

lation coefficients (ICCs), respectively. This suggests that CNN can be used to analyze SpO_2 recordings for the automated diagnosis of OSA in at-home oximetry tests.

Keywords

Apnea · Apnea index (AI) · Apnea–hypopnea index (AHI) · Blood oxygen saturation (SpO_2) · Convolutional neural networks (CNN) · Deep learning · Hypopnea · Hypopnea index (HI) · Obstructive sleep apnea (OSA) · Oximetry

15.1 Introduction

Obstructive sleep apnea (OSA) has become a major issue in recent years (Senaratna et al., 2017). OSA is marked by recurrent episodes of apneas (complete absences of airflow) and hypopneas (considerable reductions of airflow), which leads to fragmented and restless sleep (Berry et al., 2012). Despite its high prevalence in the adult population (9–38%), OSA is an underdiagnosed condition (Benjafield et al., 2020; Senaratna et al., 2017). This contributes to an increased risk of cardiovascular, metabolic, and psychiatric alterations, such as hypertension, cerebrovascular diseases, diabetes, and depression (Eastwood et al., 2010; Park et al., 2011).

Despite serving as the gold standard for OSA diagnosis, overnight polysomnography (PSG) presents important limitations. PSG is a costly test, highly intrusive for the patients, and technically complex and lacks availability (del Campo et al., 2018; Redline, 2017). In addition, apneas and hypopneas must be manually annotated by trained specialists, which is labor intensive and may lead to errors and inconsistencies in the diagnosis (Shokoueinejad et al., 2017). In order to overcome these PSG limitations, multiple investigations have focused on the use of simplified approaches aimed at the automated detection of OSA from a reduced subset of cardiorespiratory signals. Among these approaches, the automated analysis of the single-channel blood oxygen saturation (SpO_2) signal from nocturnal oximetry has been frequently proposed due to its easy acquisition and interpretation (del Campo et al., 2018). SpO_2 signal provides a continuous measure of the oxygen content in the hemoglobin (McClatchey, 2002), which allows to detect oxygen desaturations induced by OSA-related respiratory events, i.e., apneas and hypopneas (Berry et al., 2012).

Different studies have examined the SpO_2 signal as a simplified alternative to PSG in the automated detection of respiratory events and in the automated diagnosis of OSA (del Campo et al., 2018). A majority of these studies have followed conventional feature-engineering methodologies, which are based on feature extraction and selection stages (del Campo et al., 2018). Nonetheless, these methodologies require a substantial human-based knowledge to identify, a priori, a set of relevant features to extract from the signal under study (Goodfellow et al., 2016), which limits its ability to obtain all the ad-hoc information from the SpO_2 recordings related to respiratory events. This limitation can be overcome by deep-learning methods, which can directly analyze raw data and automatically make decisions based on non-human-driven knowledge (Faust et al., 2018; Goodfellow et al., 2016).

In the last few years, deep-learning algorithms have outperformed conventional approaches in many fields (Goodfellow et al., 2016), such as image recognition, autonomous driving, natural language processing, and time series analysis (Faust et al., 2018; Goodfellow et al., 2016). In the OSA context, recent studies have demonstrated the usefulness of deep-learning approaches to analyze cardiorespiratory signals in the automated detection of apneic events (Mostafa et al., 2019). Particularly, Mostafa et al. (2020a, b) and Vaquerizo-Villar et al. (2019) applied a deep-learning architecture based on convolutional neural networks (CNNs) to the oximetry signal to detect respiratory events in adult and pediatric OSA patients, respectively. However, these studies have only addressed the automated detection of respiratory events versus normal respiration, without making any distinction between apneas and hypopneas (Mostafa, Baptista, et al., 2020a; Mostafa, Mendonca, et al., 2020b; Vaquerizo-

Villar et al., 2019). Conversely, Kulkas et al. (2017) stated that the severity of oxygen desaturations differs between obstructive apnea and hypopnea events.

In the present chapter, a CNN architecture is proposed to automatically identify apnea and hypopnea events. Despite being originally designed for image analysis (Goodfellow et al., 2016), CNNs have become one of the most relevant deep-learning methods for time series classification (Ismail Fawaz et al., 2019) in many fields, including biomedical signal processing (Ebrahimi et al., 2020; Faust et al., 2019; Murat et al., 2020; Roy et al., 2019). Accordingly, we hypothesized that a deep-learning architecture based on CNNs could help to automatically learn the most relevant information from the oximetry signal in the detection and classification of apnea and hypopnea events. Consequently, the main objective of this chapter is to design and evaluate a deep-learning model based on CNNs to automatically classify respiratory events from the SpO_2 signal in OSA patients. In addition, the secondary goal of this research is to assess the usefulness of the CNN model to estimate the apnea–hypopnea index (AHI: the number of apneas and hypopneas per sleep hour), which is the clinical parameter used to establish OSA diagnosis.

15.2 Materials and Methods

15.2.1 Subjects and Signals

This chapter involved a database composed of 398 adult patients diagnosed with OSA (AHI ≥5 events per hour). All of them were referred to the sleep laboratory of the Hospital Universitario Río Hortega (Valladolid, Spain), where they underwent overnight PSG. The Ethics and Drugs Research Committee of the hospital approved the protocol (CEIm 47/16).

All subjects were diagnosed by medical specialists following the standards of the American Academy of Sleep Medicine (AASM) (Berry et al., 2012). Accordingly, an episode of apnea was annotated when there was a drop in the amplitude of the oronasal thermal airflow signal higher than 90% during at least 10 seconds (Berry et al., 2012). Similarly, a hypopnea was scored when there was a minimum of 30% reduction in the amplitude of the nasal pressure airflow signal, lasting at least 10 seconds and accompanied by an oxygen desaturation of at least 3% or/and an electroencephalographic arousal (Berry et al., 2012). Subsequently, the apnea index (AI: the number of apneas per hour), hypopnea index (HI: the number of hypopneas per hour), and AHI from each subject were computed as the total number of each type of event divided by the total sleep time.

SpO_2 signals were acquired during PSG at a sampling rate of 16 Hz. In order to reduce the computational requirements, all the SpO_2 recordings were downsampled to a sample rate of 1 Hz. SpO_2 recordings from each subject were then divided into 30-second non-overlapping epochs, being each epoch labelled as normal respiration (N), apnea (A), or hypopnea (H) using the annotations provided by the clinicians. The dataset was divided into three groups: training set (first 199 subjects, 50%), employed to train the CNN architecture; validation set (the following 79 subjects, 20%), used to monitor the convergence of the CNN; and test set (the last 120 subjects, 30%), employed to evaluate the proposed CNN-based methodology. Table 15.1 summarizes polysomnographic and clinical data from the population under study. No statistically significant differences (p-value < 0.05) were found in age, sex, body mass index (BMI), AI, HI, or AHI between the three groups.

15.2.2 Proposed CNN Architecture

Figure 15.1 shows the main components of the proposed CNN architecture. The input section of the CNN consists of the SpO_2 samples for the 30-s epoch (i.e., 30 samples) to be classified, concatenated with the four preceding and the five following epochs, thus having a 10-epoch length (300 samples) 1D input vector. The reason for using preceding and following epochs is twofold: (i) it enhances the identification of oxygen

Table 15.1 Clinical and polysomnographic data of the population under study

	All	Training	Validation	Test
Clinical characteristics in each data subset				
Subjects (*n*)	398	199	79	120
Age (years)	56 [47–65]	56 [47–64]	56 [47–67]	55 [44–65]
Males (*n*)	278 (69.9%)	144 (72.4%)	59 (74.6%)	75 (62.5%)
BMI (kg/m^2)	29.1 [26.1–33.1]	29.4 [26.0–33.5]	28.9 [26.8–31.7]	29.2 [25.6–34.0]
AI (e/h)	8.2 [1.8–24.8]	9.1 [1.6–25.3]	6.1 [2.1–24.1]	7.7 [2.0–24.8]
HI (e/h)	19.4 [10.4–32.5]	19.0 [10.1–29.7]	19.8 [10.9–32.9]	19.9 [10.6–36.5]
AHI (e/h)	35.0 [17.3–59.4]	33.2 [15.9–59.7]	35.6 [18.1–59.2]	36.2 [18.3–61.4]
Number and type of events in each data subset				
Normal (*n*)	250,669 (72.5%)	127,292 (73.2%)	49,932 (72.5%)	73,445 (71.3%)
Apnea (*n*)	40,838 (11.8%)	20,174 (11.6%)	7972 (11.6%)	12,692 (12.3%)
Hypopnea (*n*)	54,165 (15.7%)	26,350 (15.2%)	10,989 (15.9%)	16,826 (16.4%)

Data are presented as median [interquartile range], *n*, or %
BMI: body mass index, *AHI*: apnea index, *HI*: hypopnea index, *AHI*: apnea–hypopnea index, *e/h* events per hour

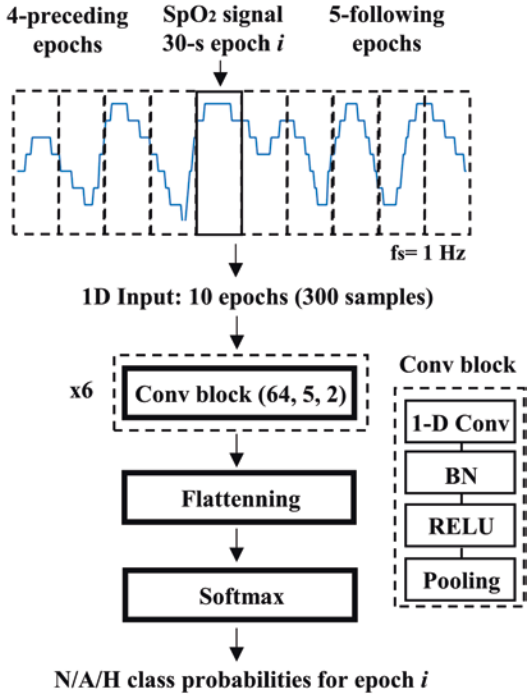

Fig. 15.1 Overview of the proposed CNN architecture. Each convolutional block (conv block) includes a 1D convolution (1D Conv), BN, a RELU activation function, and pooling

desaturations associated to respiratory events, since the onset of oxygen desaturations may occur more than 30 seconds after the start of the respiratory events (Kulkas et al., 2013); and (ii) it allows for a better modeling of the temporal distribution of respiratory events, which are typically grouped in clusters.

The proposed CNN architecture processes this input using six convolutional blocks (conv block), and each one composed of the following:

- 1D convolution (1D conv). This layer extracts feature maps using the 1D convolution operation (Goodfellow et al., 2016):

$$x_i^j[n] = \sum_{k=1}^{k_{\text{size}}} w_k^j * a_i[n-k+1] + b_k^j, \quad (15.1)$$

- where x^j_i is the feature map generated by the jth convolutional filter (j = 1, …, 64) in the ith convolutional block (i = 1, …, 6); k_{size} = 5 is the filter (kernel) size; w^j_k and b^j_k are the filter weights and biases, respectively; and a_i is the input of the ith convolutional block. The number of convolutional blocks, the number of filters, and the kernel size were chosen according to the optimum values obtained in Vaquerizo-Villar et al. (2021).
- Batch normalization (BN). BN is applied to normalize the feature maps obtained in the 1D convolution layer (Goodfellow et al., 2016).
- Rectified linear unit (ReLU). ReLU is the standard activation function in CNNs. It is applied to introduce nonlinearity to the normalized feature maps, which provides univer-

sal approximation to any function (Goodfellow et al., 2016):

$$f\left(x_i^j\right) = \max\left(0, x_i^j\right), \tag{15.2}$$

- Pooling. After the ReLU function, a max-pooling layer was applied to the activations with a pooling factor of 2 to reduce dimensionality, while the most relevant features are kept (Goodfellow et al., 2016).

Following the last convolutional block, the 2D feature maps are converted into 1D feature vectors using a flattening operation. Finally, a softmax activation function is used to obtain the output of the CNN architecture, i.e., the probability of belonging to each class (N/A/H) for the input 30-s SpO_2 epoch.

15.2.3 CNN Training Process

The CNN architecture was implemented using the Keras framework with TensorFlow backend. A workstation with a NVIDIA GeForce RTX 2080 GPU running on a Windows 10 environment was used for this purpose. The training data were fed into the CNN using minibatches of size 100 during 200 epochs. The weights of each layer of the network were initialized using He-normal initialization (He et al., 2014). Then, the adaptive moment estimation (Adam) algorithm was used with an initial learning rate of 0.0001 (Kingma & Ba, 2015), and a categorical cross entropy loss function was applied to update the weights and biases at each minibatch. As the whole training data does not fit on the memory of the workstation, training data were fed at each epoch in random order from different patients to the network using 50 reading queues (Sors et al., 2018), which also improves the convergence of the Adam algorithm (Goodfellow et al., 2016; Sors et al., 2018). The validation data was used during the training process to monitor the convergence of the CNN by means of the validation loss. In this respect, the learning rate was reduced by a factor of 2 when the validation loss did not improve for ten consecutive epochs, and early stopping was applied to finish the learning process after 30 epochs of non-improvement in the validation loss, restoring the network weights to those that minimized the validation loss (Goodfellow et al., 2016).

15.2.4 Statistical Analysis

The Kruskal–Wallis test was used to assess statistical differences (p-value < 0.05) between groups. The overall performance of the CNN architecture to automatically classify respiratory events was assessed by means of confusion matrices (3-class), which were used to compute the Cohen's kappa index (kappa) and the 3-class accuracy. The performance for each individual class was measured by means of sensitivity (percentage of epochs belonging to the class rightly classified), specificity (percentage of epochs not belonging to the class rightly classified), positive predictive value (proportion of epochs assigned to the class that are true positives), negative predictive value (proportion of epochs not assigned to the class that are true negatives), and accuracy (proportion of epochs rightly classified). In addition, AI, HI, and AHI were obtained for each subject based on the CNN scoring (AI_{CNN}, HI_{CNN}, and AHI_{CNN}) and compared with those from the standard PSG (AI_{PSG}, HI_{PSG}, and AHI_{PSG}) using Bland–Altman plots and the intra-class correlation coefficient (ICC).

15.3 Results

15.3.1 CNN Model Performance

Figure 15.2 shows the confusion matrix of the CNN model in the test set for the 3-class classification procedure (N/A/H). This model rightly classified 80.3% (82,628/102963) of the 30-s SpO_2 epochs in the test set, with a 3-class kappa value of 0.539. Table 15.2 presents the diagnostic ability for each individual class. Notice that higher performance metrics were obtained for the

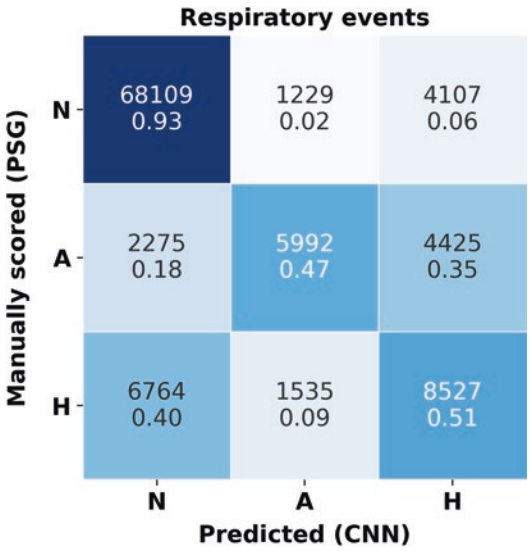

Fig. 15.2 Confusion matrix of the CNN architecture in the test set. This matrix compares the type of respiratory event from standard PSG with the corresponding assignation using the CNN model

Table 15.2 Diagnostic ability of the CNN model in the test set for the detection of normal respiration, apnea, and hypopnea events

Epoch type	Se (%)	Sp (%)	PPV (%)	NPV (%)	Acc (%)
Normal respiration	92.7	69.4	88.3	79.3	86.0
Apnea	47.2	96.9	68.4	92.9	90.8
Hypopnea	50.7	90.1	50.0	90.3	83.7

Se: sensitivity, *Sp:* specificity, *PPV:* positive predictive value, *NPV:* negative predictive value, *Acc:* accuracy

detection of normal respiration than for apnea and hypopnea events.

15.3.2 Estimation of Respiratory Indices

Figure 15.3 shows the Bland–Altman plots comparing AI_{CNN}, HI_{CNN}, and AHI_{CNN} with AI_{PSG}, HI_{PSG}, and AHI_{PSG} in the test set, respectively. ICC is also shown. It can be seen that the respiratory indices predicted by the CNN (AI_{CNN}, HI_{CNN}, and AHI_{CNN}) are underestimating those from standard PSG (AI_{PSG}, HI_{PSG}, and AHI_{PSG}), as reported by their mean difference (bias). HI_{CNN} reached a lower bias (−4.22) than AI_{CNN} (−7.87) and AHI_{CNN} (−12.09), whereas AHI_{CNN} achieved a slightly lower confidence interval (40.82) than AI_{CNN} (45.49) and HI_{CNN} (45.66). In addition, AHI_{CNN} showed a higher agreement with manual scoring (ICC = 0.8466) than AI_{CNN} (ICC = 0.8023) and HI_{CNN} (ICC = 0.6774).

15.4 Discussion

In this chapter, we evaluated the potential usefulness of a CNN architecture to automatically classify respiratory events (apnea, hypopnea, and normal respiration) from the SpO_2 signal in adult OSA patients. To our knowledge, this is the first study applying a deep-learning model to automatically identify apnea and hypopnea events from the oximetry signal.

The proposed CNN-based deep-learning model reached a high performance, with 80.3% 3-class Acc and 0.539 kappa for the classification of respiratory events. According to the guidelines of McHugh (2012), a kappa value between 0.41 and 0.60 indicates that there is a moderate agreement between our CNN architecture and manual PSG-based scoring (McHugh, 2012). Hence, our approach could be potentially used to detect respiratory events in at-home pulse oximetry tests for OSA diagnosis (del Campo et al., 2018).

Looking at the confusion matrix of Fig. 15.2., it can be seen that 93% of normal respiration epochs are rightly classified by the CNN model, which may indicate that oxygen desaturations infrequently occur without being associated to a respiratory event. Furthermore, 47% of apnea and 51% of hypopnea epochs are rightly detected by the CNN architecture, which indicates that the characteristics of SpO_2 desaturations caused by apneas that differ from those related to hypopneas. This agrees with Kulkas et al. (2017), who reported that oxygen desaturations associated to obstructive apneas have significantly larger duration and depth than SpO_2 desaturations related to hypopneas. In this sense, 35% of apnea epochs are misclassified as hypopneas. This can be explained by the fact that oxygen desaturations related to

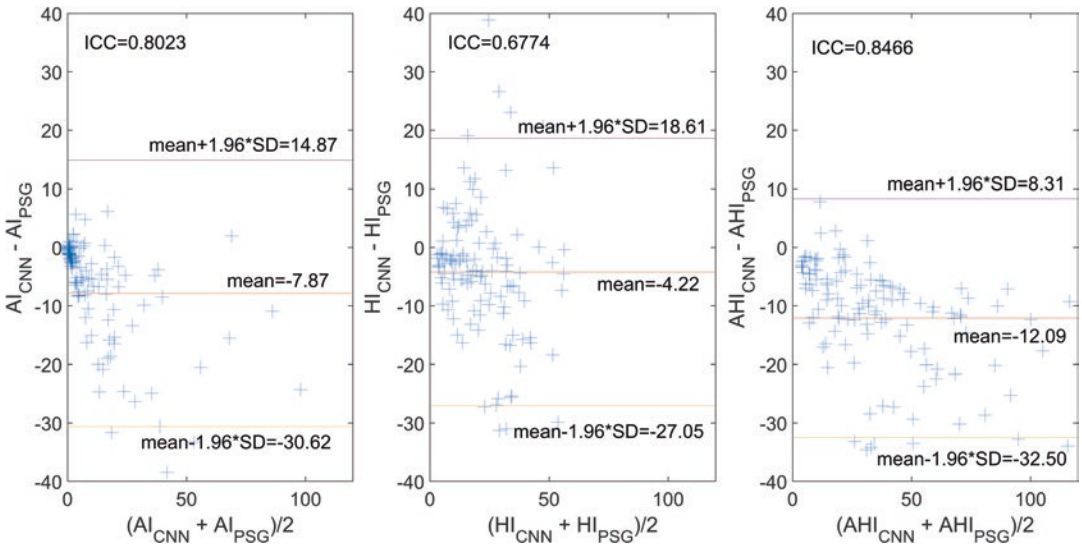

Fig. 15.3 Bland–Altman plots comparing (**a**) AI_{CNN} with AI_{PSG}, (**b**) HI_{CNN} with HI_{PSG}, and (**c**) AHI_{CNN} with AHI_{PSG}

obstructive apnea events of short duration may have similar characteristics to those related to long-duration hypopneas, as the duration and area of SpO_2 desaturations are significantly correlated to the duration of obstructive apnea and hypopnea events (Kulkas et al., 2017). Conversely, 40% of hypopnea epochs are predicted as normal respiration by the CNN. These misclassified hypopneas may be associated to electroencephalographic arousals that do not produce any physiological perturbation in the oximetry signal (Berry et al., 2012).

Regarding the respiratory indices, the CNN model shows a trend to underestimate them, especially AI_{CNN} and AHI_{CNN}. Nonetheless, the CNN model showed promising results, reaching ICCs of 0.8023 (AI_{CNN}), 0.6774 (HI_{CNN}), and 0.8466 (AHI_{CNN}). The higher ICC obtained by AI_{CNN} and AHI_{CNN} can be explained by the fact that their Bland–Altman plots show a linear underestimation trend, whereas HI_{CNN} has outliers in both directions. In this respect, an ICC value in the range 0.50–0.75 indicates a moderate agreement, whereas an ICC value in the range 0.75–0.90 indicates a good reliability (Koo & Li, 2016). Accordingly, our CNN-based deep-learning approach could be used to calculate these respiratory indices in oximetry tests.

Recent studies showed the usefulness of deep-learning techniques to automatically score respiratory events from raw cardiorespiratory signals in OSA patients, outperforming conventional feature-based methodologies (Mostafa et al., 2019). Particularly, some studies faced the automated detection of normal respiration, apneas and hypopneas from airflow; thoracic, abdominal, and chest respiratory signals; and the electrocardiogram (Haidar et al., 2020; McCloskey et al., 2018; Nikkonen et al., 2021; Urtnasan et al., 2018; Van Steenkiste et al., 2020; Yue et al., 2021), reaching a 3-class accuracy (normal, apnea, and hypopnea) in the range 73–91%. In contrast to these studies, our work achieved a 3-class accuracy of 80% using only the SpO_2 signal. In this regard, the oximetry signal has been frequently advocated for OSA screening due to its accessibility, simplicity, and reliability (del Campo et al., 2018).

Vaquerizo-Villar et al. (2019) and Mostafa et al. (2020a, b) have also focused on the automated classification of respiratory events using oximetry-based deep-learning approaches. These studies employed CNNs to differentiate respiratory events from normal respiration episodes using 60-s SpO_2 segments, reaching accuracies in the range 85–95% (Mostafa, Baptista, et al., 2020a; Mostafa, Mendonca, et al., 2020b;

Vaquerizo-Villar et al., 2019). In the present chapter, the 2-class accuracy (normal versus apnea/hypopnea) was included in this range (86%) with a 30-s segment size, which is more appropriate for the detection of clusters of respiratory events that contain more than one respiratory event in a 60-s segment. Furthermore, our CNN-based model addresses for the first time the distinction between apneas and hypopneas from raw oximetry data and the estimation of respiratory indices (AI, HI, and AHI).

Despite the potential usefulness of our proposed approach, some limitations need to be considered. First, the database employed in this work did not contain healthy control subjects (AHI < 5 e/h). The inclusion of these subjects could help to improve the characterization of normal respiration. Another limitation concerns the use of 30-s SpO_2 segments to automatically detect respiratory events, which does not allow to identify the onset and end of apneas and hypopneas. Nonetheless, SpO_2 does not contain this information, as the delay of oxygen desaturations occurring after respiratory events is variable (Kulkas et al., 2013). Similarly, the proposed CNN does not differentiate between obstructive and central respiratory events. However, this would require information about breathing effort (Berry et al., 2012), which is not included in the oximetry signal. In this respect, the acquisition of the photoplethysmography (PPG) signal with the pulse oximetry sensor may contribute to enhance the diagnostic ability of our proposal, as it contains information related to respiratory events (Karmakar et al., 2014; Papini et al., 2020). Furthermore, the use of novel deep-learning techniques (e.g., transformer or generative adversarial networks) may help improve the automatic classification of respiratory events at the cost of higher computational complexity. Finally, the application of eXplainable artificial intelligence techniques could help to further understand the perturbations in the oximetry signal linked with apnea and hypopnea events and the differences between them.

15.5 Conclusions

Our CNN-based deep-learning model exhibited a high performance in the automatic identification of apnea and hypopnea events from the SpO_2 signal. The CNN model also showed a high agreement in the estimation of OSA-related respiratory indices (AI, HI, and AHI). According to our findings, we can conclude that CNN-based SpO_2 approaches could be potentially used to provide an automated diagnosis of OSA in at-home oximetry studies.

Acknowledgments This work was supported by 'Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación/10.13039/501100011033/', ERDF A way of making Europe, and NextGenerationEU/PRTR under projects PID2020-115468RB-I00, PDC2021-120775-I00, and RTC-2017-6516-1, 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 -Consorcio Centro de Investigación Biomédica en Red-' (CB19/01/00012) through "Instituto de Salud Carlos III". The work of Fernando Vaquerizo-Villar was supported by a "Ayuda para contratos predoctorales para la Formación de Profesorado Universitario (FPU)" grant from the Ministerio de Educación, Cultura y Deporte (FPU16/02938). 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.

References

Benjafield, A. V., Eastwood, P. R., Heinzer, R., Morrell, M. J., Federal, U., Paulo, D. S., … Valentine, K. (2020). Estimation of the global prevalence and burden of obstructive sleep apnoea: A literature-based analysis. *The Lancet Respiratory Medicine, 7*(8), 687–698. https://doi.org/10.1016/S2213-2600(19)30198-5. Estimation

Berry, R. B., Budhiraja, R., Gottlieb, D. J., Gozal, D., Iber, C., Kapur, V. K., … others. (2012). Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events: Deliberations of the sleep apnea definitions task force of the American Academy of Sleep Medicine. *Journal of Clinical Sleep Medicine, 8*(5), 597.

del Campo, F., Crespo, A., Cerezo-Hernández, A., Gutiérrez-Tobal, G. C., Hornero, R., & Álvarez, D. (2018). Oximetry use in obstructive sleep apnea.

Expert Review of Respiratory Medicine, 12(8), 665–681. https://doi.org/10.1080/17476348.2018.1495563

Eastwood, P. R., Malhotra, A., Palmer, L. J., Kezirian, E. J., Horner, R. L., Ip, M. S., … Hillman, D. R. (2010). Obstructive sleep apnoea: From pathogenesis to treatment: Current controversies and future directions. *Respirology, 15*(4), 587–595. https://doi.org/10.1111/j.1440-1843.2009.01699.x

Ebrahimi, Z., Loni, M., Daneshtalab, M., & Gharehbaghi, A. (2020). A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications: X, 7*, 100033. https://doi.org/10.1016/j.eswax.2020.100033

Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine, 161*, 1–13. https://doi.org/10.1016/j.cmpb.2018.04.005

Faust, O., Razaghi, H., Barika, R., Ciaccio, E. J., & Acharya, U. R. (2019). A review of automated sleep stage scoring based on physiological signals for the new millennia. *Computer Methods and Programs in Biomedicine, 176*, 81–91.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Haidar, R., Koprinska, I., & Jeffries, B. (2020). Sleep apnea event prediction using convolutional neural networks and Markov chains. *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN48605.2020.9207345

He, K., Zhang, X., Ren, S., & Sun, J. (2014). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026–1034). https://doi.org/10.1109/ICCV.2015.123

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery, 33*(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1

Karmakar, C., Khandoker, A., Penzel, T., Schobel, C., & Palaniswami, M. (2014). Detection of respiratory arousals using photoplethysmography (PPG) signal in sleep apnea patients. *IEEE Journal of Biomedical and Health Informatics, 18*(3), 1065–1073. https://doi.org/10.1109/JBHI.2013.2282338

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (pp. 1–15).

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kulkas, A., Tiihonen, P., Julkunen, P., Mervaala, E., & Töyräs, J. (2013). Desaturation delay, parameter for evaluating severity of sleep disordered breathing. *IFMBE Proceedings, 39 IFMBE*, 336–339. https://doi.org/10.1007/978-3-642-29305-4_90

Kulkas, A., Duce, B., Leppänen, T., Hukins, C., & Töyräs, J. (2017). Severity of desaturation events differs between hypopnea and obstructive apnea events and is modulated by their duration in obstructive sleep apnea. *Sleep & Breathing, 21*(4), 829–835. https://doi.org/10.1007/s11325-017-1513-6

McClatchey, K. D. (2002). *Clinical laboratory medicine*. Lippincott Williams & Wilkins.

McCloskey, S., Haidar, R., Koprinska, I., & Jeffries, B. (2018). Detecting hypopnea and obstructive apnea events using convolutional neural networks on wavelet spectrograms of nasal airflow. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* https://doi.org/10.1007/978-3-319-93034-3_29

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemica Medica, 22*(3), 276–282. Retrieved from https://hrcak.srce.hr/89395

Mostafa, S. S., Mendonça, F., Ravelo-García, A. G., & Morgado-Dias, F. (2019). A systematic review of detecting sleep apnea using deep learning. *Sensors (Switzerland), 19*(22), 1–26. https://doi.org/10.3390/s19224934

Mostafa, S. S., Baptista, D., Ravelo-García, A. G., Juliá-Serdá, G., & Morgado-Dias, F. (2020a). Greedy based convolutional neural network optimization for detecting apnea. *Computer Methods and Programs in Biomedicine, 197*, 105640. https://doi.org/10.1016/j.cmpb.2020.105640

Mostafa, S. S., Mendonca, F., Ravelo-Garcia, A. G., Julia-Serda, G., & Morgado-Dias, F. (2020b). Multi-objective hyperparameter optimization of convolutional neural network for obstructive sleep apnea detection. *IEEE Access, 8*, 129586–129599. https://doi.org/10.1109/ACCESS.2020.3009149

Murat, F., Yildirim, O., Talo, M., Baloglu, U. B., Demir, Y., & Acharya, U. R. (2020). Application of deep learning techniques for heartbeats detection using ECG signals-analysis and review. *Computers in Biology and Medicine, 120*(April), 103726. https://doi.org/10.1016/j.compbiomed.2020.103726

Nikkonen, S., Korkalainen, H., Leino, A., Myllymaa, S., Duce, B., Leppanen, T., & Toyras, J. (2021). Automatic respiratory event scoring in obstructive sleep apnea using a long short-term memory neural network. *IEEE Journal of Biomedical and Health Informatics, 25*(8), 2917–2927. https://doi.org/10.1109/JBHI.2021.3064694

Papini, G. B., Fonseca, P., Gilst, M. M. V., Bergmans, J. W. M., Vullings, R., & Overeem, S. (2020). Respiratory activity extracted from wrist-worn reflective photoplethysmography in a sleep-disordered population. *Physiological Measurement, 41*(6). https://doi.org/10.1088/1361-6579/ab9481

Park, J. G., Ramar, K., & Olson, E. J. (2011). Updates on definition, consequences, and management of obstructive sleep apnea concise review for clinicians.

Mayo Clinic Proceedings, 86(6), 549–555. https://doi.org/10.4065/mcp.2010.0810

Redline, S. (2017). Screening for obstructive sleep apnea: Implications for the sleep health of the population. *JAMA, 317*(4), 368–370.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering, 16*(5). https://doi.org/10.1088/1741-2552/ab260c

Senaratna, C. V., Perret, J. L., Lodge, C. J., Lowe, A. J., Campbell, B. E., Matheson, M. C., … Dharmage, S. C. (2017). Prevalence of obstructive sleep apnea in the general population: A systematic review. *Sleep Medicine Reviews, 34*, 70–81. https://doi.org/10.1016/j.smrv.2016.07.002

Shokoueinejad, M., Fernandez, C., Carroll, E., Wang, F., Levin, J., Rusk, S., … others. (2017). Sleep apnea: A review of diagnostic sensors, algorithms, and therapies. *Physiological Measurement, 38*(9), R204.

Sors, A., Bonnet, S., Mirek, S., Vercueil, L., & Payen, J. F. (2018). A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomedical Signal Processing and Control, 42*, 107–114. https://doi.org/10.1016/j.bspc.2017.12.001

Urtnasan, E., Park, J. U., & Lee, K. J. (2018). Multiclass classification of obstructive sleep apnea/hypopnea based on a convolutional neural network from a single-lead electrocardiogram. *Physiological Measurement, 39*(6). https://doi.org/10.1088/1361-6579/aac7b7

Van Steenkiste, T., Groenendaal, W., Dreesen, P., Lee, S., Klerkx, S., De Francisco, R., … Dhaene, T. (2020). Portable detection of apnea and hypopnea events using bio-impedance of the chest and deep learning. *IEEE Journal of Biomedical and Health Informatics, 24*(9), 2589–2598. https://doi.org/10.1109/JBHI.2020.2967872

Vaquerizo-Villar, F., Álvarez, D., Kheirandish-gozal, L., Gutiérrez-tobal, G. C., Barroso-garcía, V., Campo, F., … Hornero, R. (2019). Convolutional neural networks to detect pediatric apnea- hypopnea events from oximetry. In *41st annual international conference of the IEEE engineering in medicine and biology society* (pp. 3555–3558).

Vaquerizo-Villar, F., Alvarez, D., Kheirandish-Gozal, L., Gutierrez-Tobal, G. C., Barroso-Garcia, V., Santamaria-Vazquez, E., … Hornero, R. (2021). A convolutional neural network architecture to enhance oximetry ability to diagnose pediatric obstructive sleep apnea. *IEEE Journal of Biomedical and Health Informatics, 25*(8). https://doi.org/10.1109/JBHI.2020.3048901

Yue, H., Lin, Y., Wu, Y., Wang, Y., Li, Y., Guo, X., … Lei, W. (2021). Deep learning for diagnosis and classification of obstructive sleep apnea: A nasal airflow-based multi-resolution residual network. *Nature and Science of Sleep, 13*, 361–373. https://doi.org/10.2147/NSS.S297856