

RESEARCH ARTICLE

On the Generalization of Sleep Apnea Detection Methods Based on Heart Rate Variability and Machine Learning

DANIELE PADOVANO¹, ARTURO MARTINEZ-RODRIGO¹, JOSE M. PASTOR¹,
JOSE J. RIETA², (Member, IEEE), AND RAUL ALCARAZ¹

¹Research Group in Electronic, Biomedical, and Telecommunication Engineering, Instituto de Tecnologías Audiovisuales, University of Castilla-La Mancha, 16071 Cuenca, Spain

²BioMIT.org, Electronic Engineering Department, Universitat Politècnica de Valencia, 46022 Valencia, Spain

Corresponding author: Daniele Padovano (daniele.padovano@uclm.es)

This research has received financial support from public grants PID2021-00X128525-IV0 and PID2021-123804OB-I00 of the Spanish Government 10.13039/501100011033 jointly with the European Regional Development Fund, SBPLY/17/180501/000411 and SBPLY/21/180501/000186 from Junta de Comunidades de Castilla-La Mancha, and AICO/2021/286 from Generalitat Valenciana. Moreover, Daniele Padovano holds a predoctoral scholarship 2022-PRED-20642, which is cofinanced by the operating program of European Social Fund (ESF) 2014-2020 of Castilla-La Mancha.

ABSTRACT Obstructive sleep apnea (OSA) is a respiratory disorder highly correlated with severe cardiovascular diseases that has unleashed the interest of hundreds of experts aiming to overcome the elevated requirements of polysomnography, the gold standard for its detection. In this regard, a variety of algorithms based on heart rate variability (HRV) features and machine learning (ML) classifiers have been recently proposed for epoch-wise OSA detection from the surface electrocardiogram signal. Many researchers have employed freely available databases to assess their methods in a reproducible way, but most were purely tested with cross-validation approaches and even some using solely a single database for training and testing procedures. Hence, although promising values of diagnostic accuracy have been reported by some of these methods, they are suspected to be overestimated and the present work aims to analyze the actual generalization ability of several epoch-wise OSA detectors obtained through a common ML pipeline and typical HRV features. Precisely, the performance of the generated OSA detectors has been compared on two validation approaches, i.e., the widely used epoch-wise, k -fold cross-validation and the highly recommended external validation, both considering different combinations of well-known public databases. Regardless of the used ML classifiers and the selected HRV-based features, the external validation results have been 20 to 40% lower than those obtained with cross-validation in terms of accuracy, sensitivity, and specificity. Consequently, these results suggest that ML-based OSA detectors trained with public databases are still not sufficiently general to be employed in clinical practice, as well as that larger, more representative public datasets and the use of external validation are mandatory to improve the generalization ability and to obtain reliable assessment of the true predictive power of these algorithms, respectively.

INDEX TERMS Electrocardiography, heart rate variability, machine learning, sleep apnea.

I. INTRODUCTION

The obstructive sleep apnea (OSA) syndrome is a condition characterized by repetitive episodes of respiratory arrests during sleep [1]. According to recent global screenings, OSA

The associate editor coordinating the review of this manuscript and approving it for publication was Venkata Rajesh Pamula¹.

prevalence is considered high, ranging from 9 to 38% in the general population [2]. Patients suffering from this disorder generally describe excessive sleepiness and a characteristic loud snoring during sleep. These symptoms often lead to bad school performance, job loss, family or marital problems, and road accidents in the most severe cases [3]. Besides, there exists substantial evidence on OSA correlation with major

cardiovascular diseases, such as atrial fibrillation, stroke, and heart failure [4]. Such comorbidities today represent the main form of non-communicable diseases and, since they are responsible of millions of deaths every year, the early detection of OSA has become a major public health concern [5]. However, most OSA cases are still undiagnosed [6], which may be in part because polysomnography (PSG) remains as the gold standard for its detection [7]. This diagnostic procedure is a resource-intensive and time-consuming method, which limits its global coverage to a large extent and leads to delayed diagnosis and increased waiting lists [8]. Hence, many researchers have committed to find a convenient alternative method to PSG [9], [10].

For that purpose, OSA detection from single or a limited set of physiological signals among those involved in PSG have been widely analyzed in the last years [9]. The problem has been addressed from two different perspectives, such that whilst some works have specifically focused on OSA severity diagnosis, others have pursued the development of epoch-wise OSA detectors [9]. Generally, OSA severity is quantified by means of the apnea-hypopnea index (AHI), a widely used parameter in clinical practice measuring the number of apnea or hypopnea events per hour of sleep [11]. A broad variety of studies have searched direct correlation between OSA severity (i.e., mild, moderate, and severe) and parameters derived from diverse physiological signals to globally diagnose subjects suffering from the disease. Although they have reported promising results, the proposed methods are unable to accurately identify and locate OSA episodes, thus hindering real-time and continuous monitoring of subjects outside of a sleep laboratory and without supervision of specialized technicians, such as at home [10], [12]. Thus, epoch-wise detection with minimum hardware requirements and computational burden currently plays a critical role in the OSA realm and will be the focus of the present work.

To date, epoch-wise OSA detectors based on single-lead ECG recordings have reported the greatest diagnostic accuracy when compared to other physiological signals, also providing a more convenient opportunity for the design and development of single-sensor wearable technology [8], [10]. Furthermore, beyond ECG morphology analyses from statistical [13], [14] and transformed domains [15], [16], the ECG signal also allows to easily obtain heart rate variability (HRV) and then study activation of the sympathetic nervous system, which is the main drive of breathing control and a major contributor to the metabolic dysfunction and elevated blood pressure noticed in OSA presence [8]. Actually, most of the recently proposed epoch-wise OSA detectors are based on the combination of HRV features through common machine learning (ML) classifiers. A complete list of these algorithms can be found in Table 1. As an illustrative example, Rajesh *et al.* [17] presented an ensemble random forest-based classifier to detect OSA episodes from traditional HRV time-frequency features. In the same line, Liang *et al.* [18] proposed a single-feature discriminant analysis for the identification of OSA epochs by means of a non-parametric form

of sample entropy, and Tang and Liu [19] presented a similar approach based on the temporal dependency complexity analysis of the 2-Dimensional form of sample entropy. Eventually, Afrakhteh and Soltani [20] appealed to several ML algorithms, such as support vectors machine (SVM), the k -nearest neighbors (KNN), linear discriminant analysis (LDA), and diverse boosted ensemble variants, to detect OSA episodes from the empirical mode decomposition of the HRV information and the ECG-derived respiration.

The methods proposed in these works, as well as in most of those listed in Table 1, have reported very promising classification results when discerning between normal and apnea epochs, exhibiting accuracy rates higher than 85% with well-balanced values of sensitivity and specificity. However, these performance measures have been mostly obtained by training and validating the OSA detectors on the same set of ECG intervals through an epoch-wise, k -fold cross-validation approach and, therefore, they are suspected to be overestimated [21], [22], [23]. Compared with resubstitution validation, where the classifier is learned from all the available data and then tested on the same set of data, k -fold cross-validation offers a more general and unbiased view of the algorithm's performance [24]. Nonetheless, the inclusion of ECG segments from the same patients in the subsets of learning and testing could facilitate classification, thus involving a bias [23], [25]. To avoid this problem, after developing a classification model, it is strongly recommended to evaluate its performance in a different set of patients from those used for training [24], [26]. Unfortunately, among the recently proposed OSA detectors, only a few ones, e.g., those presented by Varon *et al.* [27], Martín-Gonzalez *et al.* [13], and Lin *et al.* [28], have been validated using this external approach (such as Table 1 shows).

Given this context and bearing in mind that external validation in clinical applications has been vigorously advocated by the Transparent Reporting of a multivariate prediction model for Individual Prognosis Or Diagnosis (TRIPOD) initiative [26], the main goal of the present work is to analyze the actual generalization ability of several epoch-wise OSA detectors obtained through a common ML pipeline and typical HRV features. Precisely, the performance of the generated OSA detectors will be compared on two validation approaches, i.e., the widely used epoch-wise, k -fold cross-validation and the recommended external validation. For both cases, diverse combinations of the most widely used databases in the development of epoch-wise OSA detectors will be considered, which are freely available on PhysioNet [29]. Indeed, the Apnea-ECG database [30] has been the favorite one among most authors, being the only one used for training and validation of many epoch-wise OSA detectors (see Table 1). Nonetheless, the MIT-BIH Polysomnographic Database [31] and the St. Vincent's University Hospital/University College Dublin Sleep Apnea Database [32] have also been used by some authors [7], [14], [33] and will be employed in the present study.

TABLE 1. Most recent epoch-wise OSA detectors based on HRV and ML approaches presented in the scientific literature. Results are orientate, and note that maximum values for the best classifier of each work were collected only for epoch-wise OSA detection. QDA: Quadratic Discriminant Analysis, LR: Linear Regression, DERE: Differential Evolution Rule Estimation, KELM: Kernel Extreme Learning Machine, ENS: Ensemble classifier.

REF	Year	Authors	Databases	Classifier	Validation method	Ac (%)	Se (%)	Sp (%)
[34]	2010	Mendez <i>et al.</i>	Apnea-ECG (50 recordings)	QDA	Training set=25 recordings Testing set=25 recordings	89.07	90.37	86.73
[35]	2012	Al-Angari and Sahakian.	Sleep Heart Health Study (100 Subjects)	SVM	Training set=50 subjects. Testing set=50 subjects.	82.4	89.8	94.10
[36]	2014	Sannino <i>et al.</i>	Apnea-ECG (released set)	DERE	10-fold cross-val	97.85	97.84	97.85
[37]	2014	Nguyen <i>et al.</i>	Apnea-ECG (released set)	SVM	3-fold cross-val	85.26	86.37	83.47
[38]	2014	Rachim <i>et al.</i>	Apnea-ECG (released set)	SVM	10-fold cross-val	93.91	95.20	92.65
[39]	2015	Atri and Mohebbi	Apnea-ECG	SVM	10-fold cross-val	95.57	98.64	92.48
[40]	2015	Ravelo-García <i>et al.</i>	Apnea-ECG	QDA	Training set=released Testing set=withheld	-	67.20	86.80
[27]	2015	Varon <i>et al.</i>	Apnea-ECG, Leuven Hospital	SVM	External validation	84.74	84.71	84.69
[25]	2016	Cheng <i>et al.</i>	Apnea-ECG (released set)	LR	Average of 100 random sets	85.00	83.00	82.00
[15]	2016	Song <i>et al.</i>	Apnea-ECG	SVM	Training set=released Testing set=withheld	86.20	82.60	88.40
[41]	2016	Sharma and Sharma	Apnea-ECG	SVM	Training set=released Testing set=withheld	83.80	79.50	88.40
[42]	2016	Rashik and Anyal	Apnea-ECG (released set)	ENS	Out-of-bag error method	85.97	84.14	86.83
[43]	2017	Martín-Gonzalez <i>et al.</i>	Apnea-ECG, HuGCDN2014	QDA	Training set= Apnea-ECG, Testing set= HuGCDN2014	84.76	81.45	86.82
[44]	2018	Tripathy	Apnea-ECG (released set)	KELM	10-fold cross-val	-	78.02	74.64
[13]	2018	Martín-Gonzalez <i>et al.</i>	Apnea-ECG, HuGCDN2014	LDA	Train with one, test with remaining	86.33	-	-
[45]	2018	Kumar and Kanhangad	Apnea-ECG (released set)	SVM	10-fold cross-val	93.31	93.05	93.46
[46]	2019	Shao <i>et al.</i>	Apnea-ECG (60 recordings)	SVM	10-fold cross-val	91.89	88.01	93.98
[47]	2019	Haoyu <i>et al.</i>	UCD-DB, custom DB	SVM	10-fold cross-val	98.54	97.05	98.95
[48]	2019	Pinho <i>et al.</i>	Apnea-ECG (released set)	SVM	10-fold cross-val	82.10	88.40	72.30
[16]	2020	Zarei and Asl	Apnea-ECG	ENS	10-fold cross-val	93.26	91.52	94.36
[49]	2020	Kalaivani	Apnea-ECG	SVM	V-fold cross-val (V=0.1, 1, 4)	95.00	95.00	95.00
[50]	2020	Pombo <i>et al.</i>	Apnea-ECG (released set)	SVM	10-fold cross-val	82.12	88.41	72.29
[51]	2020	Zarei and Asl	Apnea-ECG	ENS	10-fold cross-val	93.90	92.26	94.92
[52]	2020	Singh <i>et al.</i>	Apnea-ECG (released set), UCD-DB, You Snooze You Win	SVM	10-fold cross-val	81.06	82.45	79.72
[53]	2020	Tripathy <i>et al.</i>	Apnea-ECG (released set, 31 rec.)	SVM	10-fold cross-val	80.54	82.27	78.67
[14]	2020	Sharma and Sharma	Apnea-ECG	KNN	Training set=released Testing set=withheld	87.50	84.90	88.20
[54]	2020	Baty <i>et al.</i>	Custom database (241 night recordings)	SVM	ROC-based	72.00	88.00	61.00
[55]	2020	Bozkurt <i>et al.</i>	Custom database (10 subjects)	ENS	Leave-one-out cross-val	85.12	85.00	86.00
[28]	2021	Lin <i>et al.</i>	Apnea-ECG, National Cheng Kung Uni. Hospital Sleep Center	SVM	5-fold cross-val	90.43	88.72	91.55
[20]	2021	Afrakhteh <i>et al.</i>	Apnea-ECG (released set)	SVM	ROC-based	91.43	91.67	91.30
[19]	2021	Tang and Liu	Apnea-ECG (60 recordings)	SVM	5-fold cross-validation	93.30	92.50	95.00
[17]	2021	Rajesh <i>et al.</i>	Apnea-ECG (released set)	ENS	10-fold cross-val	90.30	-	-
[56]	2022	Karimi	Apnea-ECG	SVM	ROC-based	-	95.46	97.57

II. MATERIALS AND METHODS

A. STUDY POPULATION

The three aforementioned databases were annotated by clinical experts. However, in knowledge that the labeling system

for each one was inherently different, the original annotations were retrieved to build the compatibility framework represented in Figure 1. This is explained in further detail in the following subsections.

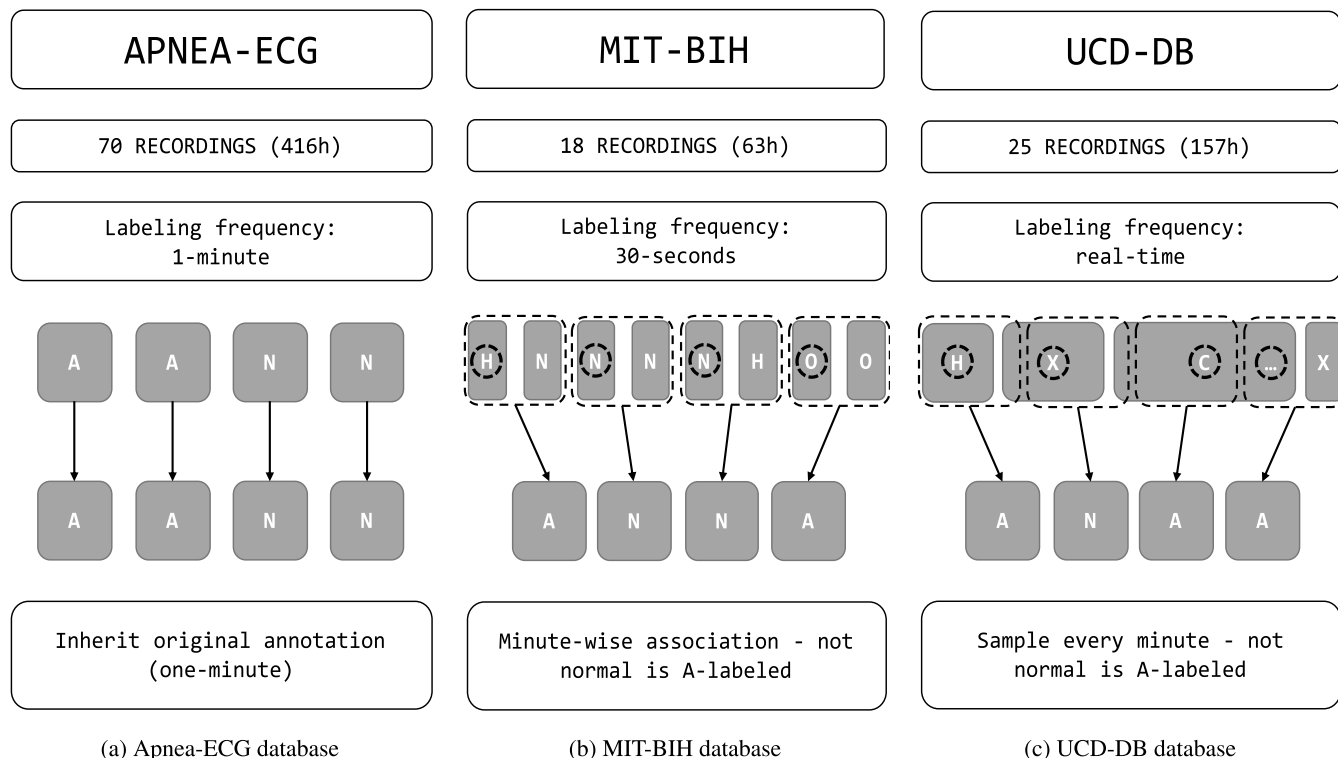


FIGURE 1. Labeling system adaptation for each database. Possible respiratory episodes were: apnea (A), normal (N), obstructive apnea (O), central apnea (C), hypopnea (H), and others (X).

1) APNEA-ECG DATABASE

The Apnea-ECG Database was first published in the 27th International Conference of Computers in Cardiology (CinC) 2000 Challenge [30] and consists of 30 male and 5 female subjects from 27 to 63 years of age. These subjects were in turn classified into three groups according to their AHI level. Notwithstanding, since OSA severity was out of the competence of the present work, the AHI-related information was ignored. From the patients, 70 single-lead ECG recordings from 7 to 9 hours of duration, sampled at 100 Hz and 200 ADC units per mV, were obtained. These recordings were grouped into a released set and a withheld set of 35 recordings each (accordingly used as the training and testing subsets for the CinC Challenge). The released-as-training and withheld-as-testing paradigm has been widely used by many authors for global subject diagnosis based on OSA severity, even after the end of the Challenge [9]. Conversely, many epoch-wise OSA detectors have only been trained and validated on the released subset (see Table 1). In the present work, all 70 ECG recordings were considered as a complete database. Last but not least, the original annotation system aimed to assess whether right at the beginning of each 1 minute-length ECG interval there was an apnea-related epoch (A-labeled), or not (N-labeled). This labeling system was adopted as a standard reference for the entire study since, compared to the rest of databases, this one presents the lowest annotation time resolution (see Figure 1a).

2) MIT-BIH POLYSOMNOGRAPHIC DATABASE

The MIT-BIH Polysomnographic Database [31], henceforward the MIT-BIH, consists of 18 PSG recordings between 2 and 7 hours of duration acquired from 16 male subjects within 32 and 56 years old. The single-lead ECG recording was sampled at 250 Hz and 12-bit per sample. Besides, the ECG recordings were annotated every 30 seconds by clinical experts, identifying OSA, central apnea, and hypopnea episodes with and without arousals. Considering these annotations, the presence of an apnea or hypopnea episode was evaluated right at the beginning of each one minute-length ECG interval to identify apnea (A-labeled) and normal (N-labeled) epochs, and then maintain a consistent labeling system with the ECG-Apnea database (see Figure 1b).

3) ST. VINCENT’S UNIVERSITY HOSPITAL/UNIVERSITY COLLEGE DUBLIN SLEEP APNEA DATABASE

The St. Vincent’s University Hospital / University College Dublin Sleep Apnea Database, henceforward the UCD-DB, consists of 25 full overnight PSG recordings acquired from 25 subjects within the range of 28 and 68 years old [32]. It comprises 21 male and 4 female subjects. Each entry of the dataset contains a single-lead ECG recording of 6 to nearly 8 hours of duration, sampled at 125 Hz and ADC units not specified. Annotations were made by a sleep technologist following the standard Rechtschaffen and Kales rules, indicating the exact hour, date, and duration of

each obstructive/central apnea, hypopnea, and other kinds of non-sleep-related episodes. Again, on the basis of these annotations every minute in the ECG recordings was accordingly re-labeled as apnea (A-labeled) or normal (N-labeled), following the criterion employed in the ECG-Apnea database to maintain coherency between datasets (see Figure 1c).

B. DATA PREPROCESSING

To improve further analysis of the ECG recordings, baseline wander removal, signal filtering, and R-peak detection were initially developed. Firstly, every single-lead ECG was re-sampled at 500 Hz to enhance R-peak detection quality. Secondly, the baseline wander and high-frequency noise were subtracted from the original signal through a band-pass second-order Chebyshev filter with cut-off frequencies of 0.5 Hz and 100 Hz. In order to preserve phase and amplitude characteristics, a zero-phase digital filter was implemented [57]. Secondly, R-peak detection was performed with the Pan-Tompkins algorithm, which is well-known due to its effectiveness and low computational cost for real-time purposes [58]. However, since this algorithm may produce slight delays in the output, these were corrected by analyzing the range of samples around the corresponding R-peak and updating the location to the maximum value. Thirdly, the ECG recording was broken down into one minute-length segments to then compute the RR intervals as the consecutive differences of R-peaks in time. Such differences, in general, represent what is usually known as the HRV series [59].

Thereafter, a manual signal screening was conducted to discard segments that contained high presence of noise or artifacts, such as those located at the boundaries of the ECG signal, which are often caused by non-physiological reasons (e.g., lead detaching, signal saturation by motion, sneezes, cough, etc.). After this data clearance, 81.83% of Apnea-ECG, 87.2% of MIT-BIH, and 91.1% of UCD-DB 1-minute length ECG segments remained as the actual *payload* of each database for the present study. Should any of the employed databases be named in the present work, it will be referred to as this corresponding payload. Segments altogether from the three selected databases added up to 38,150 ECG intervals of 1-minute length, i.e., about 635 hours of recording.

Eventually, given that both groups of ECG intervals (i.e., A-labeled and N-labeled) were notably unbalanced in the three databases, a balanced subset for each one was also built. The balancing process was conducted by randomly discarding segments from the bigger group. Thus, the group with more segments was downsized until it equaled the number of segments of the other group. Precisely, the balanced subset for the Apnea-ECG was 83.5% of the corresponding payload, whilst that subset was 79.6% for the MIT-BIH, and 21.4% for the UCD-DB. These balanced subgroups were used to train the generated ML-based OSA detectors, whereas the complete payloads were reserved for their external validation.

C. FEATURE EXTRACTION

Most relevant markers and indices of HRV present in the literature of epoch-wise OSA detection were computed following the guidelines of the Electrophysiology Task Force of the European Society of Cardiology and the North American Society of Pacing, hereafter, the Task Force [59]. In this regard, the Task Force contemplated three different domains of study: time domain, frequency domain, and complexity domain. In the time domain, multiple statistical parameters have been extracted from the HRV series. It is worth mentioning that all 1-minute length ECG segments selected in the present work contained only normal RRi (NNi). In this particular case, a determined RRi was considered *normal* if it was free of artifacts and a high presence of noise [59]. Be that as it may, the following features were computed from the ECG segments:

- Maximum (MAX): defined as the maximum of all RRi.
- Minimum (MIN): defined as the minimum of all RRi.
- Mean (MEAN): defined as the statistical mathematical hope measured from all RRi.
- Median (MED): statistical middle value of all RRi.
- Standard deviation of normal RRi (SDNN): represents the standard deviation (SD) according to all NNi.
- Standard deviation of the differences between adjacent RRi (SDSD): defines the SD of the differences between a determined RRi and its successive value.
- Root mean square of differences between adjacent RRi (RMSSD): defined as the square root of the mean squared differences of successive RRi.
- NN50: defined as the number of adjacent NNi pairs whose time difference is greater than 50 ms.
- pNN50: is the quotient between the NN50 and the number of all available NNi.
- Interquartile range (IQR): equal to the difference between the 75th and the 25th percentiles.

In the frequency domain, the Task Force contemplates three different spectral bands or spectral components of interest of study [59], i.e.:

- Very low frequency (VLF, 0.003 – 0.04 Hz): carries information on the recording baseline and may be affected by breathing modulation.
- Low frequency (LF, 0.04 – 0.15 Hz): it is considered to vary by means of changes in autonomic modulations of the heart period.
- High frequency (HF, 0.15 – 0.4 Hz): carries further information on autonomic modulation of RRi available in short-term recordings (less than 5 minutes).

In the present work, these parameters were computed using two spectral analysis, i.e., the fast Fourier transform (FFT) and the Lomb-Scargle periodogram (LSP) [60]. Although FFT-based spectrum bands were widely used in the corresponding literature [9], the LSP is deemed to provide further information on power spectral density (PSD), particularly in unevenly sampled signals, such as the HRV data [61].

In the complexity domain, the Task Force initially included some non-linear methods with the aim of characterizing HRV sequences influenced by complex physiological behaviors. Such methods included the estimation of Lyapunov exponents, Kolgomorov entropy, and spectral scaling indices. Afterward, Richman & Moorman introduced the concept of sample entropy (SampEn) [62], which proved to outperform the former non-linear markers in several works [9], [63].

In brief, the computation of SampEn of N points, with tolerance r and maximum vicinity of m points, can be defined as follows [62]. Let $u(n)$ be a set of N points,

$$u(n) = \{u(1), u(2), \dots, u(N)\},$$

two sequences can be conformed, the *target* sequence,

$$x_m(i) = \{u(i + k)\}, \quad \forall i \in [1, N - m + 1], \quad \forall k \in [0, m - 1],$$

and the *posterior* sequence,

$$x_{m+1}(i) = \{u(i + k)\}, \quad \forall i \in [1, N - m], \quad \forall k \in [0, m].$$

Now, let the distance between two vectors be defined as:

$$d[x_m(i), x_m(j)] = \max\{|x_m(i + k) - x_m(j + k)|\}, \\ \forall k \in [0, m - 1], \quad \forall j \in [1, N - m + 1], \quad j \neq i.$$

From the *target* sequence, the A pairs of vectors such that the distance $d[x_m(i), x_m(j)] \leq r$ are counted. Similarly, from the *posterior* sequence, the B pairs of vectors such that the distance $d[x_{m+1}(i), x_{m+1}(j)] \leq r$ are counted as well. Thus, SampEn can be obtained following the expression in Equation (1). Also, the effect of m and r may change significantly the final SampEn value. For that reason, the selected parameters were $m = 2$ and $r = 0.2$ times the SD of the evaluated series (HRV), as suggested in Richman & Moorman's works on empirical grounds [62], [64].

$$SampEn(N, m, r) = -\ln\left(\frac{A}{B}\right). \quad (1)$$

In recent years, optimized versions of SampEn, such as the quadratic SampEn (QSampEn) [65] and the non-parametric SampEn (NPSampEn) [18], as well as modified alternatives, like the dispersion entropy (DispEn) [66], the distribution entropy (DistEn) [67], the fuzzy entropy (FuzzEn) [68], and the measure of fuzzy entropy (MFuzzEn) [69], have been proposed. In the present work, beside the common SampEn, all these alternatives have also been computed following similar parameters of SampEn, i.e., $m = 2$, and $r = 0.2$ times the SDNN, from the ECG segments and inputted to the developed ML-based OSA detectors.

Additionally, recurrence plot-based features have also revealed interesting results in the scope of epoch-wise OSA detection [13], [16]. A recurrence plot (RP) is an advanced technique of non-linear data analysis, usually applied to the study of dynamical chaotic systems [70]. Specifically, a RP determines the relationship between instants of time for each state in a phase space (which is comparable to each sequence $x_m(i)$ defined for SampEn), which results in a squared matrix

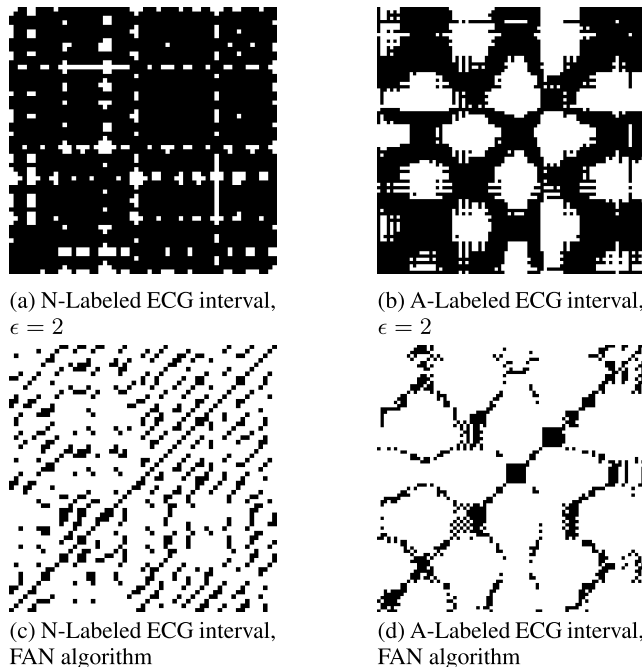


FIGURE 2. Examples of RPs obtained for two common ECG intervals (one N-labeled and another A-labeled) using a fix threshold and the FAN algorithm.

that represents the remoteness between states according to a threshold distance, expressed as [71]:

$$RP_{i,j} = \Theta(\epsilon - d[x_m(i), x_m(j)]), \quad \forall i, j = 1, \dots, N - m + 1. \quad (2)$$

The value of $RP_{i,j}$ denotes the recurrence of a state at time i at a different time j , ϵ the threshold distance, and $\Theta(\cdot)$ the Heaviside function. Note that this function will only bring a binary result for each $RP_{i,j}$ depending on ϵ , for which a black dot is represented if the norm is lower than the threshold, whilst a white dot is represented otherwise. This would output a symmetric graph, very similar to those shown in Figure 2a and Figure 2b, which may vary depending on the selected threshold distance [72].

As for computation of aforementioned entropy-based indices, an embedded dimension of $m = 2$ was used. Moreover, to minimize the dependence of the resulting RP on ϵ and establish a common and reproducible criterion, the neighborhood distance ϵ was determined using the fixed amount of nearest neighbors (FAN) algorithm, in which for each state \vec{x}_i all columns of the RP have the same recurrence density. This neighborhood criterion leads to an asymmetric RP, as illustrated in Figure 2c and Figure 2d [73]. From this map, the most common RP-based features were extracted as [73]:

- Recurrence rate (REC): retrieves the percentage of recurrence points in an RP.
- Determinism (DET): retrieves the percentage of recurrence points which form diagonal lines.

- Shannon entropy (ENTR): is the Shannon entropy of the probability distribution of the diagonal line lengths.
- Average diagonal line length (L): is defined as the average length of all diagonal lines within an RP.
- Divergence (DIV): is the inverse of the length of the longest diagonal line, which also corresponds to the sum of the positive Lyapunov exponents.

For the sake of completeness, a summary of all extracted features is provided in Table 2. Additionally, it should be noted that the computational software tool for feature extraction and data analysis was MATLAB R2020b, alongside with PhysioNet Cardiovascular Signal Toolbox [74], Marcus Vollmer's HRVTool [75], and Gaoxiang Ouyang's Recurrence Quantification Analysis (RQA) Toolbox [76]. Eventually, for database downloading and preparation, PhysioNet's WaveForms DataBase Toolbox for MATLAB was employed [77].

D. FEATURE SELECTION AND DEVELOPMENT OF THE ML-BASED OSA DETECTORS

Before the development of diverse ML models, two sequential feature selection (SFS) algorithms were employed, i.e., the sequential forward feature selection (SFFS) and the sequential backward feature selection (SBFS). In SFFS, features are sequentially added to an empty candidate set until the addition of further features does not decrease the criterion function [78]. Conversely, in SBFS, features are sequentially removed from a full candidate set until the removal of further features increases the criterion function [78]. It is worth noting that both procedures use an internal cross-validation approach for the evaluation of each feature set [79].

For both cases, the features selected by minimizing the classification error between A-labeled and N-labeled ECG intervals were employed to build several OSA detectors using five common ML classifiers, such as support vector machine (SVM), k-nearest neighbors (KNN), decision tree (TREE), and two random forest-based algorithms, i.e., adaptive boosting (ADA) and bootstrap aggregation (BAG). Moreover, to serve as a reference, additional five OSA detectors were also obtained by combining all the computed features (see Table 2) through the same five ML classifiers. As a summary, three predictive models were built using each ML classifier and a total of 15 OSA detectors were then obtained.

In this regard, the SVM models were trained with Gaussian radial basis function kernel, kernel scale = 1, and polynomial order = 3. The KNN models were trained with distance method = euclidean, and equal distance weights. The TREE models were trained with maximum number of splits = 1, minimum leaf size = 1, minimum parent size = 10, and following the standard CART (Classification and Regression Trees) predictor. Eventually, the ensemble models were both trained with the same number of learning cycles (100 cycles) under the tree decision learner. The only difference in ensemble models consisted in the ADA's learning rate, which value was established at 1.

TABLE 2. Summary of the 28 obtained HRV-based features.

Domain	Feature	Description
Time	MAX	Maximum of RRi
	MIN	Minimum of RRi
	MEAN	Mean of RRi
	MED	Median of RRi
	SDNN	Standard deviation of normal RRi
	SDSD	Standard deviation of the differences between adjacent RRi
	RMSSD	Root mean square of differences between adjacent RRi
	NN50	Pairs of adjacent NNi differing by more than 50 ms
	pNN50	Determined by dividing NN50 by the total number of all NNi
Frequency	IQR	Interquartile range
	VLF	FFT very low frequency component
	LF	FFT low frequency component
	HF	FFT high frequency component
	LS-VLF	LSP very low frequency component
	LS-LF	LSP low frequency component
Complexity	LS-HF	LSP high frequency component
	SampEn	Sample entropy
	QSampEn	Quadratic SampEn
	NPSampEn	Non-parametric SampEn
	DispEn	Dispersion Entropy
	DistEn	Distribution Entropy
	FuzzEn	Fuzzy Entropy
	MFuzzEn	Measure of Fuzzy Entropy
	REC	RP Recurrence rate
	DET	RP Determinism
ENTR	RP Shannon entropy	
L	RP average diagonal line length	
DIV	RP divergence	

E. TRAINING AND VALIDATION OF THE ML MODELS

To thoroughly analyze the generalization ability achieved by the obtained ML-based OSA detectors, they were trained and validated by making use of two different approaches. The first one consisted in a common epoch-wise, k -fold cross-validation strategy, where the training set is divided into k stratified subsets of randomly selected observations. The most common value for k was chosen based on the reviewed literature, that is, $k = 10$ [9]. Thereafter, the corresponding model is trained with $k - 1$ subsets and tested with the remaining one [80]. This procedure is subsequently carried out until all combinations of training and testing subsets are evaluated.

TABLE 3. Summary of the six experiments conducted by combining the three analyzed databases for the two considered validation approaches.

Experiment	Approach 1: 10-fold cross-validation (all balanced)	Approach 2: external validation	
		Training (balanced)	Test (unbalanced)
1	Apnea-ECG	Apnea-ECG	MIT-BIH & UCD-DB
2	MIT-BIH	MIT-BIH	Apnea-ECG & UCD-DB
3	UCD-DB	UCD-DB	Apnea-ECG & MIT-BIH
4	Apnea-ECG & MIT-BIH	Apnea-ECG & MIT-BIH	UCD-DB
5	Apnea-ECG & UCD-DB	Apnea-ECG & UCD-DB	MIT-BIH
6	MIT-BIH & UCD-DB	MIT-BIH & UCD-DB	Apnea-ECG

The approach was repeated for the six combinations of the balanced subsets of the three analyzed databases, such as described in Table 3. Thus, each obtained OSA detector was trained and validated six times using the balanced datasets of Apnea-ECG, MIT-BIH, UCD-DB, Apnea-ECG along with MIT-BIH, Apnea-ECG along with UCD-DB, and MIT-BIH along with UCD-DB, respectively.

For direct and fair comparison, the second validation approach consisted in training the ML-based OSA detectors with the same aforementioned six combinations of the balanced subsets of the analyzed databases and then testing them with the non-balanced sets of the remaining databases, such as Table 3 describes. Hereafter, this latter approach will be referred to as *external validation*. More exactly, the followed combinations of training and testing sets were collected as experiments, these were: Apnea-ECG along with MIT-BIH & UCD-DB, MIT-BIH along with Apnea-ECG & UCD-DB, UCD-DB along with Apnea-ECG & MIT-BIH, Apnea-ECG & MIT-BIH along with UCD-DB, Apnea-ECG & UCD-DB along with MIT-BIH, and MIT-BIH & UCD-DB along with Apnea-ECG. For the sake of exemplification, Figure 3 shows the workflow of the first experiment, in which Apnea-ECG balanced payload was used as the training set and the full datasets of UCD-DB and MIT-BIH constituted the testing set.

For both validation approaches, performance classification of the OSA detectors was assessed in terms of the common values of accuracy (Ac), sensitivity (Se), and specificity (Sp). These measures were computed as:

$$\begin{aligned}
 Ac(\%) &= \frac{TP + TN}{TP + TN + FP + FN} \cdot 100, \\
 Se(\%) &= \frac{TP}{TP + FN} \cdot 100, \\
 Sp(\%) &= \frac{TN}{TN + FP} \cdot 100,
 \end{aligned} \tag{3}$$

where true positives (TP) represent the number of 1 minute-length ECG intervals correctly classified as apneic,

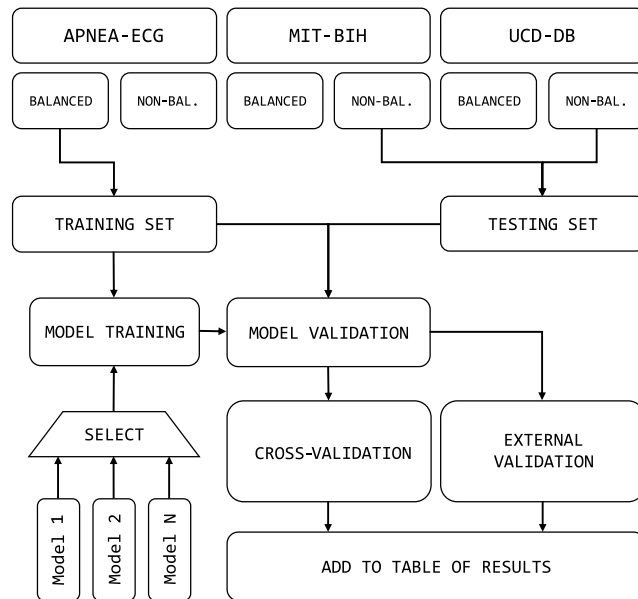


FIGURE 3. Experiment #1 exemplification. For a 10-fold cross-validation approach, the generated ML-based OSA detectors were trained and tested on the balanced subset of the Apnea-ECG. For the external validation approach, the OSA detectors were trained on the same balanced subset, but then tested on the whole payload of the MIT-BIH and UCD-DB databases.

true negatives (TN) represent the number of ECG segments correctly classified as normal, false positives (FP) correspond to the number of normal ECG intervals incorrectly classified as apneic, and false negatives (FN) correspond to the number of apneic epochs incorrectly classified as normal.

III. RESULTS

On the grounds of the large amount of experiments conducted and the high similarity between the results provided by all the generated ML-based OSA detectors, only those from the algorithms based on the BAG classifier are presented below. Nonetheless, the outcomes for the remaining OSA detectors, based on the SVM, KNN, TREE and ADA classifiers, can be found in Appendix.

Table 4 shows the classification results obtained for the six conducted experiments, the three analyzed feature sets, and the two considered validation approaches. As can be seen, SBFS usually brought more features than SFFS for all the experiments. An overview of how often each feature was selected by any of the two selection methods, i.e., SFFS or SBFS, is displayed in Figure 4. In general terms, the measures computed from the time domain of the HRV data were more frequently selected than those from the frequency domain, and these latter ones more frequently than those from the complexity domain. Within each domain, the variables MEAN, LS-LF and FuzzEn were respectively the most prevalent in the obtained OSA detectors. However, regardless the set of features, the classification performance noticed in all the experiments maintained the same trends for both validation approaches. Of note, SFFS only selected a feature in the experiment #2 (i.e., NN50), and classification obtained

TABLE 4. Results for BAG.

SFS	# Features	Experiment (training dataset)	10-fold Cross-Validation			External Validation		
			Ac (%)	Se (%)	Sp (%)	Ac (%)	Se (%)	Sp (%)
SFBS	14	1 (Apnea-ECG)	85.46	85.47	85.45	67.48	69.21	60.16
	1	2 (MIT-BIH)	73.78	74.00	73.57	55.00	63.07	36.70
	7	3 (UCD-DB)	81.89	82.65	81.13	50.11	29.07	83.18
	8	4 (Apnea-ECG & MIT-BIH)	82.02	81.35	82.69	69.01	71.18	50.89
	11	5 (Apnea-ECG & UCD-DB)	81.84	80.86	82.81	64.19	59.98	70.54
	8	6 (MIT-BIH & UCD-DB)	77.11	77.10	77.12	63.39	62.47	64.86
SBFS	24	1 (Apnea-ECG)	85.46	84.96	85.96	66.63	68.54	58.48
	27	2 (MIT-BIH)	80.89	78.84	82.95	64.39	75.45	39.31
	26	3 (UCD-DB)	81.65	81.73	81.57	50.85	28.21	86.44
	27	4 (Apnea-ECG & MIT-BIH)	82.34	81.43	83.25	68.60	70.63	51.69
	27	5 (Apnea-ECG & UCD-DB)	81.99	81.42	82.57	64.69	60.87	70.47
	27	6 (MIT-BIH & UCD-DB)	77.13	77.81	76.44	64.19	64.75	63.31
ALL	28	1 (Apnea-ECG)	83.42	82.16	84.67	67.20	69.05	59.32
	28	2 (MIT-BIH)	81.16	80.59	81.74	63.90	74.62	39.58
	28	3 (UCD-DB)	82.24	82.16	82.32	50.80	28.73	85.51
	28	4 (Apnea-ECG & MIT-BIH)	82.40	81.19	83.61	68.76	70.65	53.08
	28	5 (Apnea-ECG & UCD-DB)	81.96	81.54	82.39	64.34	60.47	70.21
	28	6 (MIT-BIH & UCD-DB)	77.31	78.08	76.53	63.33	64.26	61.86

by both 10-fold cross-validation and external validation was notably lower than when a larger number of variables was considered (i.e., in the cases of SBFS and ALL). In fact, about 7–10% lower values of Ac were seen regarding the other experiments for both validation approaches.

Anyway, for the six experiments the OSA detectors presented drastically lower performance values with external validation than with 10-fold cross-validation, globally around 10 to 30% lower in terms of Ac, Se, and Sp. Moreover, this tendency was more intense when the ML models were trained with the balanced form of the UCD-DB (e.g., highest performance drop in experiment #3 for all cases, see Table 4 and tables in Appendix), which corresponds to the smallest training subset. As previously mentioned, similar results were observed when the OSA detectors based on other classifiers than BAG were analyzed (see Appendix). However, the SVM-based models presented slightly higher resistance to performance dropping from one validation approach to another, because differences about 5 to 15% between both validation approaches were observed. Conversely, the TREE-based OSA detectors presented the lowest resistance with drops in performance about 15-25%, as well as the lowest classification rates for all the experiments in general terms, with values of Ac about 50% (see Table 7 in Appendix).

IV. DISCUSSION

Validation is a very important step in the development of every ML-based prediction model. To the best of our

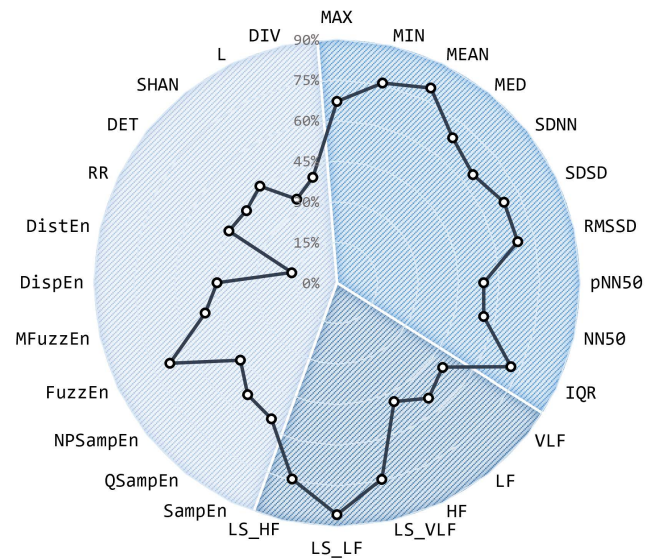


FIGURE 4. Overview about how often each feature was selected by both SFBS and SBFS approaches to be included in OSA detectors.

knowledge, the present work is the first one assessing the generalization ability of diverse epoch-wise HRV-based OSA detectors by comparing their classification performance with two different validation approaches. The results obtained by all the generated ML models for an epoch-wise, 10-fold cross validation approach were similar to those reported in most of the previous works, which also used the same validation

methodology (although some variety in the number of folds was noticed). In fact, in line with the values of A_c , S_e , and S_p about 85% reported by the ML-based OSA detectors in the present work (see Section III and Appendix), Table 1 shows that many previous works obtained A_c values above 80%, with well-balanced classification rates both for normal and apnea ECG segments.

These outcomes are initially promising, suggesting that the obtained OSA detectors were able to properly generalize the apnea detection problem in a minute-by-minute basis, such as some previous authors have claimed [25], [28], [47]. However, the same OSA detectors performed significantly worse when externally validated with alien datasets to those employed for their training, thus demonstrating that epoch-wise cross-validated results overestimated the prediction accuracy of the ML models. As can be seen in Table 4, drops in diagnostic accuracy between 20 and 40% were noticed in most of the conducted analyses, along with notably unbalanced values of sensitivity and specificity. This outcome was previously suspected by other authors [21] and partially proven by only training two ML-based OSA detectors with ECG-based features on the Apnea-ECG [22]. Nonetheless, in the present work a wider study has been conducted by considering a longer set of freely available databases, combinations of them, HRV features, and ML classifiers.

The main reason behind that overestimated classification performance of the ML models reported by the epoch-wise, 10-fold cross validation approach could be the bias introduced by involving ECG intervals from the same patients into training and testing subsets [23]. In this setting, complex classifiers can pick up a confounding relationship between a subject-specific feature (e.g., the heart rate) and diagnostic status, and so produce unrealistically high prediction accuracy [23], [81]. Such association is still easier to be established when more limited and unrepresentative databases are analyzed, because poor diversity of subjects could be found [23]. However, even when the balanced subsets of the three databases analyzed in this work were considered together, an epoch-wise, 10-fold cross validation approach was unable to offer a realistic view of the classification performance of the ML models. In this case, BAG-based OSA detectors still obtained values of A_c , S_e , and S_p of 80.77%, 80.94%, and 80.61% for the SFFS-based feature set, 80.72%, 80.67%, and 80.77% for the SBFS-based feature set, and 78.74%, 78.25%, and 79.22% for the set of all HRV features, respectively. These overestimated results, along with the drastic drops in classification observed when external validation was used, suggest that the three most famous public databases are not sufficiently large and representative to train general ML-based OSA detectors.

To this last respect, although more than 38,000 ECG intervals were analyzed, they were obtained from 70 patients collected in the Apnea ECG database, 16 in the MIT-BIH, and 25 in the UCD-DB. Moreover, whereas Apnea-ECG and UCD-DB involved both males and female subjects (notably unbalanced), in MIT-BIH all subjects were male. It is known

that OSA has been considered a risk factor for cardiovascular death in men, but whether it was also in women was unknown until recent years [82]. Namely, it has been demonstrated that severe OSA is also associated with cardiovascular death in women [83], especially in women with polycystic ovary syndrome [84]. Another aspect is the age range, which is wider in Apnea-ECG and UCD-DB than in MIT-BIH. Knowing that age represents one of the highest risk factors for OSA right after weight and sex [1], this aspect should be considered in future proposed OSA detectors. To a lesser extent than age and sex, ethnicity may also influence the OSA behavior in HRV patterns, as Quin *et al.* recently stated in their latest work [85]. As a consequence, the collection of larger, more representative, and multi-center datasets is mandatory to achieve general and accuracy ML-based OSA detectors, which can be transferred into clinical practice.

On the other hand, it is worth noting that in the context of OSA detection, two main dimensional reduction algorithms are frequently mentioned, i.e., principal component analysis (PCA) and SFS [10]. Whilst PCA represents an approach focused on the variance of relative components computed from an altered space of features, SFS adds or removes features from a candidate subset while evaluating a criterion function, which usually depends on the classifier internal belief [38]. In the present work, SFS was chosen instead of PCA because of its ability to quantify the exact contribution of each combination of features for every experiment [16]. Anyway, the main goal of both algorithms is to lessen the computational burden and improve the model generalization, but no significant differences in the classification performance of the generated ML-based OSA detectors were noticed for the three feature sets (SFFS, SBFS and ALL). Indeed, although SFFS-based and SBFS-based models were slightly lower in A_c than those models trained with all HRV-based features when a epoch-wise, 10-fold cross-validation approach was used, a considerable drop in performance was still reported for the three cases in the setting of an external validation (see Table 4).

Despite this last outcome, decreasing the feature set size is always convenient to minimize computational burden and some interesting observations were noticed to this respect. Given the usefulness of HRV non-linear characterization in the OSA realm [8], some of the most recent and innovative non-linear features were implemented beyond the well-known SampEn, such as DispEn, DistEn, QSampEn, RP-based features, etc., to verify their global contribution with SFS. However, they did not perform as well as expected, since they were apparently unable to properly separate A-labeled and N-labeled ECG segments. Only the original form of FuzzEn appeared in almost all the feature sets selected by SFS algorithms, demonstrating certain improvement in the ML models' performance, providing that it was used in combination with other time and frequency measures. Moreover, both the frequency domain and time domain features were the most selected ones by the SFS algorithms (near the 90% of the cases), therefore they are

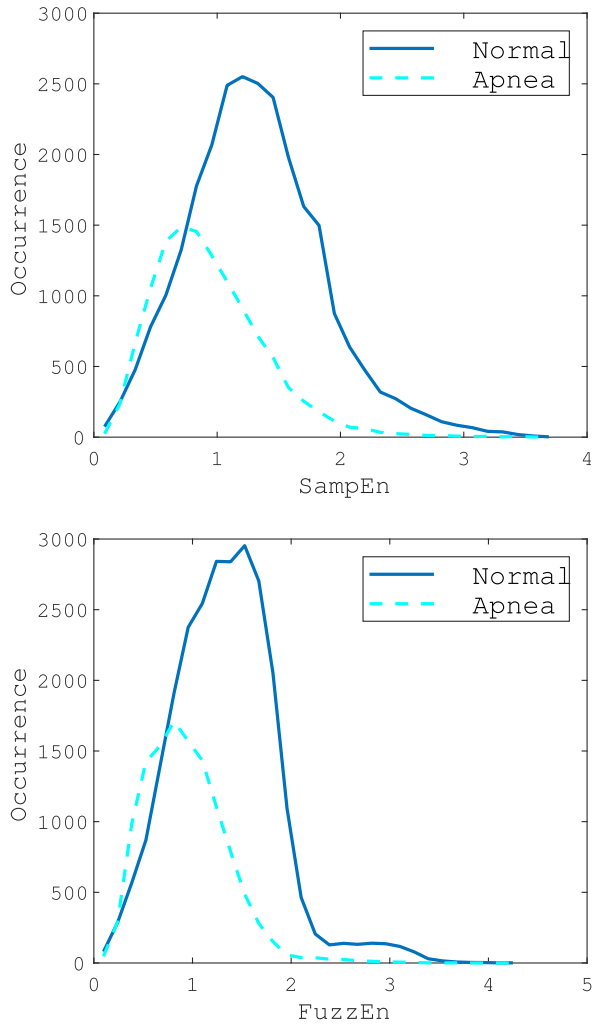


FIGURE 5. Distribution of SampEn for Apnea-ECG groups.

also useful in combination with other time features to detect OSA [10], [51]. It is worth saying that the proportion of feature appearance is not totally indicative of the individual feature efficiency. For instance, the individual separation of groups using solely SampEn (Figure 5a) is very similar to FuzzEn (Figure 5b), but this latter one is more prone to be used in combination with further features (according to SFS) to contribute to the ML model's performance.

Regarding the performance provided by the OSA detectors based on the analyzed ML classifiers others than BAG, no relevant differences were noticed (see Section III and Appendix). Nonetheless, it is remarkable that more advanced ML classifiers, such as ensemble classifiers (ADA and BAG), did not bring significantly better results than traditional ML classifiers, like SVM or KNN, especially in the case of external validation. Moreover, the SVM-based models presented the highest resistance to performance dropping from cross validation to external validation, e.g., experiments #1 and #2 in SVM with all features shown a performance drop of approximately 5%, whereas for KNN, TREE, ADA, and

BAG, have shown over 10% drop in the same conditions (see Appendix). These results makes SVM an attractive choice for OSA detection, such as several authors like Pombo *et al.* [50] and Zarei *et al.* [16] have also addressed previously.

Finally, in the interest of understanding and reproducibility of the results, solely well-known and publicly available databases were analyzed in the present work. The generalization ability achieved by the ML models for OSA prediction using other wider private or limited databases is out of scope of this work, but it will be conducted in the future. Additionally, the generalization ability of the growing deep learning algorithms for OSA detection will also be tackled in further works. These algorithms have not been included in the present work because, unlike the case of ML models, there is no a generic pipeline for their development and, even worse, a broad variety of factors can impact the resulting model. In this respect, many kinds of networks can be found, such as recurrent neural networks and convolutional neural networks, presenting different forms of layer architectures and several hyperparameters. Besides, diverse initial training conditions, learning rates, and techniques of early stopping are available. Also, many alternatives to transform the HRV signal into a 1-D sequence (e.g., the signal itself, instantaneous frequency, entropy-based indices, etc.), and into a 2-D image (spectrograms, wavelet scalograms, Poincare plots, etc.) are possible. This convoluted pool of options may hamper appropriate comparison among deep learning models, as well as their impact on OSA detection, and their thorough analysis will require a totally different work.

V. CONCLUSION

The present study has shown that epoch-wise OSA detectors based on common ML models and HRV features, and trained on the three most famous freely available databases, such as most of those published in the last decade, are not sufficiently general to be employed in clinical practice. To improve the generalization ability of these OSA detectors, larger, more representative, and multi-center public databases are still required. Also, external validation, where separate subjects must be considered for training and testing, has resulted to be essential for reliable assessment of the true predictive power of these algorithms. This finding is in line with the recommendations found in recent guidelines to progress in the transition into ML-driven, data-rich medicine, such as the TRIPOD guidelines.

APPENDIX CLASSIFICATION RESULTS FOR THE GENERATED OSA DETECTORS BASED ON SVM, KNN, TREE, AND ADA

In previous Section III, only classification results obtained by the OSA detectors based on the BAG classifier were presented. Thus, the corresponding results for the OSA detectors based on SVM, KNN, TREE and ADA are presented below in Tables 5, 6, 7 and 8, respectively.

TABLE 5. Results for SVM.

SFS	# Features	Experiment (training dataset)	10-fold Cross-Validation			External Validation		
			Ac (%)	Se (%)	Sp (%)	Ac (%)	Se (%)	Sp (%)
SFFS	20	1 (Apnea-ECG)	82.49	82.50	82.48	71.13	76.84	46.83
	8	2 (MIT-BIH)	79.76	79.56	79.96	66.75	78.60	39.87
	6	3 (UCD-DB)	77.88	79.96	75.79	47.45	24.20	83.99
	4	4 (Apnea-ECG & MIT-BIH)	80.94	80.95	80.93	75.09	79.71	36.61
	13	5 (Apnea-ECG & UCD-DB)	80.74	80.80	80.68	62.03	65.77	56.38
	4	6 (MIT-BIH & UCD-DB)	72.73	72.19	73.28	63.82	59.65	70.42
SBFS	22	1 (Apnea-ECG)	81.91	80.98	82.85	71.13	78.29	40.68
	22	2 (MIT-BIH)	76.75	76.21	77.29	69.96	92.33	19.22
	21	3 (UCD-DB)	79.51	79.28	79.74	43.42	16.52	85.70
	19	4 (Apnea-ECG & MIT-BIH)	78.37	80.10	76.63	76.90	82.37	31.35
	18	5 (Apnea-ECG & UCD-DB)	79.54	78.76	80.33	61.45	68.86	50.23
	19	6 (MIT-BIH & UCD-DB)	75.85	76.28	75.42	64.88	72.16	53.35
ALL	28	1 (Apnea-ECG)	78.56	81.23	75.90	72.64	81.67	34.21
	28	2 (MIT-BIH)	73.48	75.75	71.20	69.85	94.38	14.21
	28	3 (UCD-DB)	78.12	75.82	80.43	41.41	12.32	87.13
	28	4 (Apnea-ECG & MIT-BIH)	77.72	77.66	77.77	77.52	83.20	30.16
	28	5 (Apnea-ECG & UCD-DB)	77.47	79.56	75.37	59.69	74.16	37.81
	28	6 (MIT-BIH & UCD-DB)	72.85	72.78	72.92	66.55	84.10	38.79

TABLE 6. Results for KNN.

SFS	# Features	Experiment (training dataset)	10-fold Cross-Validation			External Validation		
			Ac (%)	Se (%)	Sp (%)	Ac (%)	Se (%)	Sp (%)
SFFS	11	1 (Apnea-ECG)	84.61	84.78	84.44	65.00	66.03	60.64
	8	2 (MIT-BIH)	79.32	78.91	79.75	64.80	74.21	43.46
	10	3 (UCD-DB)	79.22	79.48	78.95	50.60	34.35	76.16
	11	4 (Apnea-ECG & MIT-BIH)	80.44	79.24	81.64	66.37	67.76	54.76
	12	5 (Apnea-ECG & UCD-DB)	80.40	79.79	81.00	61.98	59.14	66.27
	11	6 (MIT-BIH & UCD-DB)	75.15	75.01	75.29	61.81	63.23	59.58
SBFS	23	1 (Apnea-ECG)	83.67	83.42	83.93	64.27	64.48	63.39
	17	2 (MIT-BIH)	79.42	79.76	79.09	63.74	73.28	42.08
	19	3 (UCD-DB)	79.91	80.28	79.53	49.62	31.53	78.07
	22	4 (Apnea-ECG & MIT-BIH)	80.41	80.53	80.29	65.35	66.75	53.67
	22	5 (Apnea-ECG & UCD-DB)	80.46	80.33	80.60	60.83	56.10	68.00
	22	6 (MIT-BIH & UCD-DB)	75.11	75.30	74.92	62.03	64.24	58.53
ALL	28	1 (Apnea-ECG)	81.73	79.68	83.78	64.19	64.26	63.87
	28	2 (MIT-BIH)	79.43	79.42	79.42	64.31	73.87	42.63
	28	3 (UCD-DB)	79.12	78.57	79.66	49.36	31.08	78.11
	28	4 (Apnea-ECG & MIT-BIH)	80.38	79.67	81.09	65.15	66.51	53.77
	28	5 (Apnea-ECG & UCD-DB)	80.23	80.80	79.66	62.40	57.77	69.41
	28	6 (MIT-BIH & UCD-DB)	74.97	75.00	74.94	61.83	62.94	60.06

TABLE 7. Results for TREE.

SFS	# Features	Experiment (training dataset)	10-fold Cross-Validation			External Validation		
			Ac (%)	Se (%)	Sp (%)	Ac (%)	Se (%)	Sp (%)
SFFS	12	1 (Apnea-ECG)	78.81	79.70	77.91	60.65	61.52	56.97
	2	2 (MIT-BIH)	73.98	73.39	74.58	54.99	63.07	36.64
	7	3 (UCD-DB)	75.50	74.45	76.53	51.84	36.79	75.49
	5	4 (Apnea-ECG & MIT-BIH)	74.84	74.78	74.91	63.08	64.40	52.08
	1	5 (Apnea-ECG & UCD-DB)	65.29	66.37	64.22	57.70	65.28	46.23
	5	6 (MIT-BIH & UCD-DB)	68.60	67.87	69.34	59.79	63.65	53.69
SBFS	18	1 (Apnea-ECG)	79.28	79.25	79.32	62.71	64.73	54.09
	21	2 (MIT-BIH)	72.48	71.94	73.01	59.45	65.52	45.68
	19	3 (UCD-DB)	74.70	74.50	74.90	50.23	39.59	66.95
	21	4 (Apnea-ECG & MIT-BIH)	75.16	74.95	75.36	62.41	63.87	50.30
	22	5 (Apnea-ECG & UCD-DB)	74.53	74.37	74.69	58.44	57.33	60.12
	21	6 (MIT-BIH & UCD-DB)	70.32	70.60	70.03	54.91	56.84	51.86
ALL	28	1 (Apnea-ECG)	75.95	75.42	76.48	62.46	63.72	57.09
	28	2 (MIT-BIH)	73.51	72.29	74.73	59.41	66.19	44.02
	28	3 (UCD-DB)	74.75	76.09	73.41	52.11	42.87	66.64
	28	4 (Apnea-ECG & MIT-BIH)	74.66	74.28	75.04	62.38	63.58	52.38
	28	5 (Apnea-ECG & UCD-DB)	74.36	74.31	74.42	56.45	54.20	59.85
	28	6 (MIT-BIH & UCD-DB)	69.72	69.54	69.90	57.15	57.88	55.98

TABLE 8. Results for ADA.

SFS	# Features	Experiment (training dataset)	10-fold Cross-Validation			External Validation		
			Ac (%)	Se (%)	Sp (%)	Ac (%)	Se (%)	Sp (%)
SFFS	7	1 (Apnea-ECG)	83.21	83.05	83.38	67.26	69.79	56.53
	4	2 (MIT-BIH)	77.22	77.07	77.37	64.92	73.89	44.58
	4	3 (UCD-DB)	77.88	80.18	75.58	52.44	32.78	83.35
	9	4 (Apnea-ECG & MIT-BIH)	80.39	80.38	80.41	69.41	71.78	49.70
	12	5 (Apnea-ECG & UCD-DB)	79.79	80.15	79.43	63.84	57.99	72.68
	9	6 (MIT-BIH & UCD-DB)	74.41	74.24	74.58	65.89	68.42	61.91
SBFS	22	1 (Apnea-ECG)	84.00	83.36	84.64	67.33	68.37	62.87
	23	2 (MIT-BIH)	79.32	79.91	78.74	63.93	73.74	41.68
	22	3 (UCD-DB)	79.71	81.87	77.56	50.00	31.67	78.81
	25	4 (Apnea-ECG & MIT-BIH)	80.69	79.28	82.11	72.18	75.04	48.31
	25	5 (Apnea-ECG & UCD-DB)	79.77	79.68	79.86	65.75	60.78	73.28
	25	6 (MIT-BIH & UCD-DB)	75.15	76.80	73.49	65.70	70.01	58.87
ALL	28	1 (Apnea-ECG)	81.37	81.42	81.32	68.11	69.96	60.24
	28	2 (MIT-BIH)	78.46	78.43	78.49	64.22	75.09	39.57
	28	3 (UCD-DB)	80.16	80.54	79.79	50.45	28.80	84.49
	28	4 (Apnea-ECG & MIT-BIH)	80.66	80.72	80.59	71.21	73.89	48.91
	28	5 (Apnea-ECG & UCD-DB)	79.74	79.83	79.64	64.66	59.23	72.88
	28	6 (MIT-BIH & UCD-DB)	75.21	74.66	75.76	65.42	66.74	63.33

REFERENCES

- [1] G. A. Gould, K. F. Whyte, G. B. Rhind, M. A. A. Airlie, J. R. Catterall, C. M. Shapiro, and N. J. Douglas, "The sleep hypopnea syndrome," *Amer. Rev. Respiratory Disease*, vol. 137, no. 4, pp. 895–898, Apr. 1988, doi: [10.1164/ajrccm/137.4.895](https://doi.org/10.1164/ajrccm/137.4.895).
- [2] C. V. Senaratna, J. L. Perret, C. J. Lodge, A. J. Lowe, B. E. Campbell, M. C. Matheson, G. S. Hamilton, and S. C. Dharmage, "Prevalence of obstructive sleep apnea in the general population: A systematic review," *Sleep Med. Rev.*, vol. 34, pp. 70–81, Aug. 2017, doi: [10.1016/j.smrv.2016.07.002](https://doi.org/10.1016/j.smrv.2016.07.002).
- [3] T. Young, P. E. Peppard, and D. J. Gottlieb, "Epidemiology of obstructive sleep apnea: A population health perspective," *Amer. J. Respiratory Crit. Care Med.*, vol. 165, no. 9, pp. 1217–1239, May 2002, doi: [10.1164/rccm.2109080](https://doi.org/10.1164/rccm.2109080).
- [4] A. Zapater et al., "The effect of sleep apnea on cardiovascular events in different acute coronary syndrome phenotypes," *Amer. J. Respiratory Crit. Care Med.*, vol. 202, no. 12, pp. 1698–1706, Jul. 2020, doi: [10.1164/rccm.202004-1127OC](https://doi.org/10.1164/rccm.202004-1127OC).
- [5] World Health Organization, "Noncommunicable diseases country profiles 2018," Publications NCDs, 2018. Accessed: Jul. 10, 2021. [Online]. Available: <https://apps.who.int/iris/bitstream/handle/10665/274512/9789241514620-eng.pdf>
- [6] A. V. Benjafield et al., "Estimation of the global prevalence and burden of obstructive sleep apnoea: A literature-based analysis," *Lancet Respiratory Med.*, vol. 7, no. 8, pp. 687–698, Aug. 2019, doi: [10.1016/S2213-2600\(19\)30198-5](https://doi.org/10.1016/S2213-2600(19)30198-5).
- [7] M. Bsoul, H. Minn, and L. Tamil, "Apnea MedAssist: Real-time sleep apnea monitor using single-lead ECG," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 416–427, Oct. 2011, doi: [10.1109/TITB.2010.2087386](https://doi.org/10.1109/TITB.2010.2087386).
- [8] E. S. Jeyajothi, J. Anitha, S. Rani, and B. Tiwari, "A comprehensive review: Computational models for obstructive sleep apnea detection in biomedical applications," *BioMed Res. Int.*, vol. 2022, p. e7242667, Feb. 2022, doi: [10.1155/2022/7242667](https://doi.org/10.1155/2022/7242667).
- [9] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, "A review of obstructive sleep apnea detection approaches," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 825–837, Apr. 2019.
- [10] M. Bahrami and M. Forouzanfar, "Sleep apnea detection from single-lead ECG: A comprehensive analysis of machine learning and deep learning algorithms," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, Feb. 2022, doi: [10.1109/TIM.2022.3151947](https://doi.org/10.1109/TIM.2022.3151947).
- [11] M. R. Mannarino, F. Di Filippo, and M. Pirro, "Obstructive sleep apnea syndrome," *Eur. J. Int. Med.*, vol. 23, no. 7, pp. 586–593, Oct. 2012, doi: [10.1016/j.ejim.2012.05.013](https://doi.org/10.1016/j.ejim.2012.05.013).
- [12] B. Xie and H. Minn, "Real-time sleep apnea detection by classifier combination," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 77–469, May 2012, doi: [10.1109/TITB.2012.2188299](https://doi.org/10.1109/TITB.2012.2188299).
- [13] S. Martín-González, J. L. Navarro-Mesa, G. Juliá-Serdá, G. M. Ramírez-Ávila, and A. G. Ravelo-García, "Improving the understanding of sleep apnea characterization using recurrence quantification analysis by defining overall acceptable values for the dimensionality of the system, the delay, and the distance threshold," *PLoS ONE*, vol. 13, no. 4, p. e0194462, Apr. 2018, doi: [10.1371/journal.pone.0194462](https://doi.org/10.1371/journal.pone.0194462).
- [14] H. Sharma and K. K. Sharma, "Sleep apnea detection from ECG using variational mode decomposition," *Biomed. Phys. Eng. Exp.*, vol. 6, no. 1, p. 015026, Jan. 2020, doi: [10.1088/2057-1976/ab68e9](https://doi.org/10.1088/2057-1976/ab68e9).
- [15] C. Song, K. Liu, X. Zhang, L. Chen, and X. Xian, "An obstructive sleep apnea detection approach using a discriminative hidden Markov model from ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1532–1542, Jul. 2016, doi: [10.1109/TBME.2015.2498199](https://doi.org/10.1109/TBME.2015.2498199).
- [16] A. Zarei and B. M. Asl, "Automatic detection of obstructive sleep apnea using wavelet transform and entropy-based features from single-lead ECG signal," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 3, pp. 1011–1021, May 2019, doi: [10.1109/JBHI.2018.2842919](https://doi.org/10.1109/JBHI.2018.2842919).
- [17] K. N. V. P. S. Rajesh, R. Dhuli, and T. S. Kumar, "Obstructive sleep apnea detection using discrete wavelet transform-based statistical features," *Comput. Biol. Med.*, vol. 130, p. 104199, Mar. 2021, doi: [10.1016/j.compbiomed.2020.104199](https://doi.org/10.1016/j.compbiomed.2020.104199).
- [18] D. Liang, S. Wu, L. Tang, K. Feng, and G. Liu, "Short-term HRV analysis using nonparametric sample entropy for obstructive sleep apnea," *Entropy*, vol. 23, no. 3, p. 267, Feb. 2021, doi: [10.3390/e23030267](https://doi.org/10.3390/e23030267).
- [19] L. Tang and G. Liu, "The novel approach of temporal dependency complexity analysis of heart rate variability in obstructive sleep apnea," *Comput. Biol. Med.*, vol. 135, p. 104632, Aug. 2021, doi: [10.1016/j.compbiomed.2021.104632](https://doi.org/10.1016/j.compbiomed.2021.104632).
- [20] S. Afrakhteh, A. Ayatollahi, and F. Soltani, "Classification of sleep apnea using EMD-based features and PSO-trained neural networks," *Biomed. Eng./Biomedizinische Technik*, vol. 66, no. 5, pp. 459–472, Oct. 2021, doi: [10.1515/bmt-2021-0025](https://doi.org/10.1515/bmt-2021-0025).
- [21] A. R. Hassan and M. A. Haque, "An expert system for automated identification of obstructive sleep apnea from single-lead ECG using random under sampling boosting," *Neurocomputing*, vol. 235, pp. 122–130, Apr. 2017, doi: [10.1016/j.neucom.2016.12.062](https://doi.org/10.1016/j.neucom.2016.12.062).
- [22] G. B. Papini et al., "On the generalizability of ECG-based obstructive sleep apnea monitoring: Merits and limitations of the apnea-ECG database," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 6022–6025, doi: [10.1109/EMBC.2018.8513660](https://doi.org/10.1109/EMBC.2018.8513660).
- [23] M. A. Little et al., "Using and understanding cross-validation strategies. Perspectives on Saeb et al.," *GigaScience*, vol. 6, no. 5, pp. 1–6, May 2017, doi: [10.1093/gigascience/gix020](https://doi.org/10.1093/gigascience/gix020).
- [24] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, Eds. Boston, MA, USA: Springer, 2009, pp. 532–538, doi: [10.1007/978-0-387-39940-9_565](https://doi.org/10.1007/978-0-387-39940-9_565).
- [25] C. Cheng, C. Kan, and H. Yang, "Heterogeneous recurrence analysis of heartbeat dynamics for the identification of sleep apnea events," *Comput. Biol. Med.*, vol. 75, pp. 10–18, Aug. 2016, doi: [10.1016/j.compbiomed.2016.05.006](https://doi.org/10.1016/j.compbiomed.2016.05.006).
- [26] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement," *Brit. J. Surg.*, vol. 102, no. 3, pp. 148–158, Jan. 2015, doi: [10.1002/bjs.9736](https://doi.org/10.1002/bjs.9736).
- [27] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, and S. Van Huffel, "A novel algorithm for the automatic detection of sleep apnea from single-lead ECG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 9, pp. 2269–2278, Sep. 2015, doi: [10.1109/TBME.2015.2422378](https://doi.org/10.1109/TBME.2015.2422378).
- [28] C.-Y. Lin, Y.-W. Wang, F. Setiawan, N. T. H. Trang, and C.-W. Lin, "Sleep apnea classification algorithm development using a machine-learning framework and bag-of-features derived from electrocardiogram spectrograms," *J. Clin. Med.*, vol. 11, no. 1, p. 192, Dec. 2021, doi: [10.3390/jcm11010192](https://doi.org/10.3390/jcm11010192).
- [29] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–E220, Jun. 2000, doi: [10.1161/01.cir.101.23.e215](https://doi.org/10.1161/01.cir.101.23.e215).
- [30] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter. (2000). *Apnea-ECG Database*. [Online]. Available: <https://physionet.org/content/apnea-ecg/>, doi: [10.13026/C23W2R](https://doi.org/10.13026/C23W2R).
- [31] Y. Ichimaru and G. B. Moody. (1992). *MIT-BIH Polysomnographic Database*. [Online]. Available: <https://physionet.org/content/slpdb/>, doi: [10.13026/C23K5S](https://doi.org/10.13026/C23K5S).
- [32] W. McNicholas, L. Doherty, S. Ryan, and J. Garvey. (2004). *St. Vincent's University Hospital/University College Dublin Sleep Apnea Database*. [Online]. Available: <https://physionet.org/content/ucddb/>, doi: [10.13026/C26C7D](https://doi.org/10.13026/C26C7D).
- [33] L. Almazaydeh, K. Elleithy, and M. Faezipour, "Obstructive sleep apnea detection using SVM-based classification of ECG signal features," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 4938–4941, doi: [10.1109/EMBC.2012.6347100](https://doi.org/10.1109/EMBC.2012.6347100).
- [34] M. O. Mendez, J. Corthout, S. Van Huffel, M. Matteucci, T. Penzel, S. Cerutti, and A. M. Bianchi, "Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis," *Physiol. Meas.*, vol. 31, no. 3, pp. 273–289, Mar. 2010, doi: [10.1088/0967-3334/31/3/001](https://doi.org/10.1088/0967-3334/31/3/001).
- [35] M. Al-Angari and V. Sahakian, "Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 463–468, May 2012, doi: [10.1109/TITB.2012.2185809](https://doi.org/10.1109/TITB.2012.2185809).
- [36] G. Sannino, I. De Falco, and G. De Pietro, "Monitoring obstructive sleep apnea by means of a real-time mobile system based on the automatic extraction of sets of rules through differential evolution," *J. Biomed. Informat.*, vol. 49, pp. 84–100, Jun. 2014, doi: [10.1016/j.jbi.2014.02.015](https://doi.org/10.1016/j.jbi.2014.02.015).
- [37] H. D. Nguyen, B. A. Wilkins, Q. Cheng, and B. A. Benjamin, "An online sleep apnea detection method based on recurrence quantification analysis," *IEEE J. Biomed. Health. Inf.*, vol. 18, no. 4, pp. 1285–1293, Jul. 2014, doi: [10.1109/JBHI.2013.2292928](https://doi.org/10.1109/JBHI.2013.2292928).
- [38] V. P. Rachim, G. Li, and W.-Y. Chung, "Sleep apnea classification using ECG-signal wavelet-PCA features," *Biomed. Mater. Eng.*, vol. 24, no. 6, pp. 2875–2882, Sep. 2014, doi: [10.3233/BME-141106](https://doi.org/10.3233/BME-141106).

- [39] R. Atri and M. Mohebbi, "Obstructive sleep apnea detection using spectrum and bispectrum analysis of single-lead ECG signal," *Physiol. Meas.*, vol. 36, no. 9, pp. 1963–1980, Sep. 2015, doi: [10.1088/0967-3334/36/9/1963](https://doi.org/10.1088/0967-3334/36/9/1963).
- [40] A. Ravelo-García et al., "Application of the permutation entropy over the heart rate variability for the improvement of electrocardiogram-based sleep breathing pause detection," *Entropy*, vol. 17, no. 3, pp. 914–927, Mar. 2015, doi: [10.3390/e17030914](https://doi.org/10.3390/e17030914).
- [41] H. Sharma and K. K. Sharma, "An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions," *Comput. Biol. Med.*, vol. 77, pp. 116–124, Oct. 2016, doi: [10.1016/j.compbiomed.2016.08.012](https://doi.org/10.1016/j.compbiomed.2016.08.012).
- [42] A. R. Hassan and M. A. Haque, "Computer-aided obstructive sleep apnea screening from single-lead electrocardiogram using statistical and spectral features and bootstrap aggregating," *Biocybern. Biomed. Eng.*, vol. 36, no. 1, pp. 256–266, Jan. 2016, doi: [10.1016/j.bbe.2015.11.003](https://doi.org/10.1016/j.bbe.2015.11.003).
- [43] S. Martín-González, J. L. Navarro-Mesa, G. Juliá-Serdá, J. F. Kraemer, N. Wessel, and A. G. Ravelo-García, "Heart rate variability feature selection in the presence of sleep apnea: An expert system for the characterization and detection of the disorder," *Comput. Biol. Med.*, vol. 91, pp. 47–58, Dec. 2017, doi: [10.1016/j.compbiomed.2017.10.004](https://doi.org/10.1016/j.compbiomed.2017.10.004).
- [44] R. K. Tripathy, "Application of intrinsic band function technique for automated detection of sleep apnea using HRV and EDR signals," *Biocybern. Biomed. Eng.*, vol. 38, no. 1, pp. 136–144, Jan. 2018, doi: [10.1016/j.bbe.2017.11.003](https://doi.org/10.1016/j.bbe.2017.11.003).
- [45] T. S. Kumar and V. Kanhangad, "Gabor filter-based one-dimensional local phase descriptors for obstructive sleep apnea detection using single-lead ECG," *IEEE Sensors Lett.*, vol. 2, no. 1, pp. 1–4, Mar. 2018, doi: [10.1109/LSENS.2018.2807584](https://doi.org/10.1109/LSENS.2018.2807584).
- [46] S. Shao, T. Wang, C. Song, X. Chen, E. Cui, and H. Zhao, "Obstructive sleep apnea recognition based on multi-bands spectral entropy analysis of short-time heart rate variability," *Entropy*, vol. 21, no. 8, p. 812, Aug. 2019, doi: [10.3390/e21080812](https://doi.org/10.3390/e21080812).
- [47] L. Haoyu, L. Jianxing, N. Arunkumar, A. F. Hussein, and M. M. Jaber, "An IoMT cloud-based real time sleep apnea detection scheme by using the SpO2 estimation supported by heart rate variability," *Future Gener. Comput. Syst.*, vol. 98, pp. 69–77, Sep. 2019, doi: [10.1016/j.future.2018.12.001](https://doi.org/10.1016/j.future.2018.12.001).
- [48] A. Pinho, N. Pombo, B. M. C. Silva, K. Bousson, and N. Garcia, "Towards an accurate sleep apnea detection based on ECG signal: The quintessential of a wise feature selection," *Appl. Soft Comput.*, vol. 83, p. 105568, Oct. 2019, doi: [10.1016/j.asoc.2019.105568](https://doi.org/10.1016/j.asoc.2019.105568).
- [49] V. Kalaivani, "Automated recognition of obstructive sleep apnea using ensemble support vector machine classifier," *Int. J. Biomed. Eng. Technol.*, vol. 33, no. 3, pp. 274–289, Jan. 2020, doi: [10.1504/IJBET.2020.107756](https://doi.org/10.1504/IJBET.2020.107756).
- [50] N. Pombo, B. M. C. Silva, A. M. Pinho, and N. Garcia, "Classifier precision analysis for sleep apnea detection using ECG signals," *IEEE Access*, vol. 8, pp. 200477–200485, 2020, doi: [10.1109/ACCESS.2020.3036024](https://doi.org/10.1109/ACCESS.2020.3036024).
- [51] A. Zarei and B. M. Asl, "Automatic classification of apnea and normal subjects using new features extracted from HRV and ECG-derived respiration signals," *Biomed. Signal Process. Control*, vol. 59, p. 101927, May 2020, doi: [10.1016/j.bspc.2020.101927](https://doi.org/10.1016/j.bspc.2020.101927).
- [52] H. Singh, R. K. Tripathy, and R. B. Pachori, "Detection of sleep apnea from heart beat interval and ECG derived respiration signals using sliding mode singular spectrum analysis," *Digit. Signal Process.*, vol. 104, p. 102796, Sep. 2020, doi: [10.1016/j.dsp.2020.102796](https://doi.org/10.1016/j.dsp.2020.102796).
- [53] R. K. Tripathy, P. Gajbhiye, and U. R. Acharya, "Automated sleep apnea detection from cardio-pulmonary signal using bivariate fast and adaptive EMD coupled with cross time-frequency analysis," *Comput. Biol. Med.*, vol. 120, p. 103769, May 2020, doi: [10.1016/j.compbiomed.2020.103769](https://doi.org/10.1016/j.compbiomed.2020.103769).
- [54] F. Baty et al., "Classification of sleep apnea severity by electrocardiogram monitoring using a novel wearable device," *Sensors*, vol. 20, no. 1, p. 286, Jan. 2020, doi: [10.3390/s20010286](https://doi.org/10.3390/s20010286).
- [55] F. Bozkurt, M. K. Uçar, M. R. Bozkurt, and C. Bilgin, "Detection of abnormal respiratory events with single channel ECG and hybrid machine learning model in patients with obstructive sleep apnea," *IRBM*, vol. 41, no. 5, pp. 241–251, Oct. 2020, doi: [10.1016/j.irbm.2020.05.006](https://doi.org/10.1016/j.irbm.2020.05.006).
- [56] M. K. Moridani, "A novel clinical method for detecting obstructive sleep apnea using of nonlinear mapping," *J. Biomed. Phys. Eng.*, vol. 12, no. 1, pp. 31–34, Feb. 2022, doi: [10.31661/jbpe.v0i0.1211](https://doi.org/10.31661/jbpe.v0i0.1211).
- [57] M. Kaur, B. Singh, and Seema, "Comparison of different approaches for removal of baseline wander from ECG signal," in *Proc. Int. Conf. Workshop Emerg. Trends Technol. (ICWET)*, New York, NY, USA, Feb. 2011, pp. 1290–1294, doi: [10.1145/1980022.1980307](https://doi.org/10.1145/1980022.1980307).
- [58] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: [10.1109/TBME.1985.325532](https://doi.org/10.1109/TBME.1985.325532).
- [59] J. Camm, "Task force of the European society of cardiology and the North American society of pacing and electrophysiology. Heart rate variability: Standards of measurement, physiological interpretation and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996.
- [60] J. D. Scargle, "Studies in astronomical time series analysis. II—statistical aspects of spectral analysis of unevenly spaced data," *Astrophys. J.*, vol. 263, pp. 835–853, Dec. 1982, doi: [10.1086/160554](https://doi.org/10.1086/160554).
- [61] Z. Dong, X. Li, and W. Chen, "Frequency network analysis of heart rate variability for obstructive apnea patient detection," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1895–1905, Nov. 2018, doi: [10.1109/JBHI.2017.2784415](https://doi.org/10.1109/JBHI.2017.2784415).
- [62] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000, doi: [10.1152/ajpheart.2000.278.6.H2039](https://doi.org/10.1152/ajpheart.2000.278.6.H2039).
- [63] H. M. Al-Angari and A. V. Sahakian, "Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 10, pp. 1900–1904, Oct. 2007, doi: [10.1109/TBME.2006.889772](https://doi.org/10.1109/TBME.2006.889772).
- [64] J. S. Richman, D. E. Lake, and J. R. Moorman, "Sample entropy," in *Methods in Enzymology* (Numerical Computer Methods, Part E), vol. 384. New York, NY, USA: Academic, 2004, pp. 172–184. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0076687904840114>, doi: [10.1016/S0076-6879\(04\)84011-4](https://doi.org/10.1016/S0076-6879(04)84011-4).
- [65] D. E. Lake and J. R. Moorman, "Accurate estimation of entropy in very short physiological time series: The problem of atrial fibrillation detection in implanted ventricular devices," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 300, no. 1, pp. H319–H325, Jan. 2011, doi: [10.1152/ajp-heart.00561.2010](https://doi.org/10.1152/ajp-heart.00561.2010).
- [66] M. Rostaghi and H. Azami, "Dispersion entropy: A measure for time-series analysis," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 610–614, May 2016, doi: [10.1109/LSP.2016.2542881](https://doi.org/10.1109/LSP.2016.2542881).
- [67] P. Li, C. Liu, K. Li, D. Zheng, C. Liu, and Y. Hou, "Assessing the complexity of short-term heartbeat interval series by distribution entropy," *Med. Biol. Eng. Comput.*, vol. 53, no. 1, pp. 77–87, Jan. 2015, doi: [10.1007/s11517-014-1216-0](https://doi.org/10.1007/s11517-014-1216-0).
- [68] W. Chen, J. Zhuang, W. Yu, and Z. Wang, "Measuring complexity using FuzzyEn, ApEn, and SampEn," *Med. Eng. Phys.*, vol. 31, no. 1, pp. 61–68, Jan. 2009, doi: [10.1016/j.medengphy.2008.04.005](https://doi.org/10.1016/j.medengphy.2008.04.005).
- [69] C. Liu, K. Li, L. Zhao, F. Liu, D. Zheng, C. Liu, and S. Liu, "Analysis of heart rate variability using fuzzy measure entropy," *Comput. Biol. Med.*, vol. 43, no. 2, pp. 100–108, Feb. 2013, doi: [10.1016/j.compbiomed.2012.11.005](https://doi.org/10.1016/j.compbiomed.2012.11.005).
- [70] I. Ozken, D. Eroglu, S. F. M. Breitenbach, N. Marwan, L. Tan, U. Tirnakli, and J. Kurths, "Recurrence plot analysis of irregularly sampled data," *Amer. Phys. Soc.*, vol. 98, no. 5, p. 052215, Nov. 2018, doi: [10.1103/PhysRevE.98.052215](https://doi.org/10.1103/PhysRevE.98.052215).
- [71] J.-P. Eckmann, K. S. Oliffson, and D. Ruelle, "Recurrence plots of dynamical systems," *World Sci. Ser. Nonlinear Sci. A*, vol. 16, pp. 441–446, Sep. 1995.
- [72] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, "Recurrence-plot-based measures of complexity and their application to heart-rate-variability data," *Amer. Phys. Soc.*, vol. 66, no. 2, p. 026702, Aug. 2002, doi: [10.1103/PhysRevE.66.026702](https://doi.org/10.1103/PhysRevE.66.026702).
- [73] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Phys. Rep.*, vol. 438, nos. 5–7, pp. 237–329, Jan. 2007, doi: [10.1016/j.physrep.2006.11.001](https://doi.org/10.1016/j.physrep.2006.11.001).
- [74] A. N. Vest et al., "An open source benchmarked toolbox for cardiovascular waveform and interval analysis," *Physiol. Meas.*, vol. 39, no. 10, p. 105004, Oct. 2018.
- [75] M. Vollmer, "HRVTool—An open-source MATLAB toolbox for analyzing heart rate variability," in *Proc. CINC. IEEE*, 2019, p. 1.
- [76] O. Gaoxiang, (2020). *Recurrence Quantification Analysis (RQA)*. [Online]. Available: <https://es.mathworks.com/matlabcentral/fileexchange/46765-recurrence-quantification-analysis-rqa>
- [77] I. Silva and G. B. Moody, "An open-source toolbox for analysing and processing PhysioNet databases in MATLAB and octave," *J. Open Res. Softw.*, vol. 2, no. 1, p. e27, Sep. 2014, doi: [10.5334/jors.bi](https://doi.org/10.5334/jors.bi).
- [78] T. Rückstieß, C. Osendorfer, and P. van der Smagt, "Sequential feature selection for classification," in *Proc. AAIR*, in Lecture Notes in Computer Science, D. Wang and M. Reynolds, Eds., Berlin, Germany, 2011, pp. 132–141, doi: [10.1007/978-3-642-25832-9_14](https://doi.org/10.1007/978-3-642-25832-9_14).
- [79] D. W. Aha and R. L. Bankert, "A comparative evaluation of sequential feature selection algorithms," in *Learning From Data: Artificial Intelligence and Statistics V* (Lecture Notes in Statistics), D. Fisher and H.-J. Lenz, Eds. New York, NY, USA: Springer, 1996, pp. 199–206, doi: [10.1007/978-1-4612-2404-4_19](https://doi.org/10.1007/978-1-4612-2404-4_19).

- [80] M. W. Browne, "Cross-validation methods," *J. Math. Psychol.*, vol. 44, no. 1, pp. 108–132, Mar. 2000, doi: [10.1006/jmps.1999.1279](https://doi.org/10.1006/jmps.1999.1279).
- [81] S. Saeb, L. Lonini, A. Jayaraman, D. C. Mohr, and K. P. Kording, "The need to approximate the use-case in clinical machine learning," *Giga-Science*, vol. 6, no. 5, pp. 1–9, May 2017, doi: [10.1093/gigascience/gix019](https://doi.org/10.1093/gigascience/gix019).
- [82] G. Gutiérrez-Tobal, D. Álvarez, J. Gomez-Pilar, F. del Campo, and R. Hornero, "Assessment of time and frequency domain entropies to detect sleep apnoea in heart rate variability recordings from men and women," *Entropy*, vol. 17, no. 1, pp. 123–141, Jan. 2015, doi: [10.3390/e17010123](https://doi.org/10.3390/e17010123).
- [83] F. Campos-Rodriguez, M. A. Martinez-Garcia, I. de la Cruz-Moron, C. Almeida-Gonzalez, P. Catalan-Serra, and J. M. Montserrat, "Cardiovascular mortality in women with obstructive sleep apnea with or without continuous positive airway pressure treatment," *Ann. Internal Med.*, vol. 156, no. 2, pp. 115–122, Jan. 2012, doi: [10.7326/0003-4819-156-2-201201170-00006](https://doi.org/10.7326/0003-4819-156-2-201201170-00006).
- [84] H. Kahal et al., "The prevalence of obstructive sleep apnoea in women with polycystic ovary syndrome: A systematic review and meta-analysis," *Sleep Breathing*, vol. 24, no. 1, pp. 339–350, Mar. 2020, doi: [10.1007/s11325-019-01835-1](https://doi.org/10.1007/s11325-019-01835-1).
- [85] H. Qin et al., "The different facets of heart rate variability in obstructive sleep apnea," *Frontiers Psychiatry*, vol. 12, p. 1128, Jul. 2021.



JOSE M. PASTOR received the M.Sc. degree in electronics and automatic control engineering and the Ph.D. degree in robotics and artificial intelligence from the Polytechnic University of Madrid (UPM), in 1991 and 1997, respectively. Since 2005, he has been an Associate Professor with the Department of Computer Science, University of Castilla-La Mancha (UCLM). He is currently the Director of Research Institute of Audiovisual Technologies and the Head of the Biomedical, Electronics and Telecommunication Research Group at UCLM. His research interests include biomedical signal processing, biomedical sensors, wearables, the IoT, and A.I.



JOSE J. RIETA (Member, IEEE) received the Ph.D. degree in biomedical signal processing from the Universitat Politècnica de València (UPV), Valencia, Spain, in 2003. He is a Full Professor with the Electronic Engineering Department, UPV, where he currently coordinates the Biomedical Synergy Research Group. His research interests include biomedical signal processing, blind signal separation, engineering education, arrhythmias, and educational data mining.



DANIELE PADOVANO received the B.Sc. degree in telecommunications technologies engineering from the University of Castilla-La Mancha (UCLM), Cuenca, Spain, in 2020, and the M.Sc. degree in telecommunications engineering from UCLM, where he is pursuing the Ph.D. degree in Advanced Computing Technologies. He joined the Biomedical, Electronics and Telecommunication Research Group, in 2020. His research interests include biomedical signal processing and A.I.



ARTURO MARTINEZ-RODRIGO received the M.Sc. degree in telecommunications engineering and the Ph.D. degree in medical care research from the University of Castilla-La Mancha (UCLM), Cuenca, Spain, in 2010 and 2013, respectively. He has been an Assistant Professor with the Department of Mathematics, UCLM, since 2010, where he is currently with the Department of Computer Science. His research interests include signal processing, sensor networks, and applied A.I.



RAUL ALCARAZ received the Ph.D. degree in electronic engineering from the Universitat Politècnica de València, Valencia, Spain, in 2008. He is an Associate Professor with the Department of Electrical, Electronic, Automatic Control and Communications, University of Castilla-La Mancha, Cuenca, Spain. His research interests include biomedical signal processing, engineering education, and educational data mining.

...