

## Article

# Gaussian Mixture Models for Detecting Sleep Apnea Events Using Single Oronasal Airflow Record

Hisham ElMoaqet <sup>1,\*</sup>, Jungyoon Kim <sup>2,\*</sup>, Dawn Tilbury <sup>3</sup>, Satya Krishna Ramachandran <sup>4</sup>,  
Mutaz Ryalat <sup>1</sup> and Chao-Hsien Chu <sup>5</sup>

<sup>1</sup> Department of Mechatronics Engineering, German Jordanian University, Amman 11180, Jordan; mutaz.ryalat@gju.edu.jo

<sup>2</sup> Department of Computer Science, Kent State University, Kent, OH 23185, USA

<sup>3</sup> Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA; tilbury@umich.edu

<sup>4</sup> Department of Anesthesiology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA; skrama@bidmc.harvard.edu

<sup>5</sup> College of Information Sciences and Technology, Penn State University, College Park, PA 16802, USA; chc4@psu.edu

\* Correspondence: hisham.elmoaqet@gju.edu.jo (H.E.); jkim78@kent.edu (J.K.)

Received: 9 October 2020; Accepted: 2 November 2020; Published: 6 November 2020



**Abstract:** Sleep apnea is a common sleep-related disorder that significantly affects the population. It is characterized by repeated breathing interruption during sleep. Such events can induce hypoxia, which is a risk factor for multiple cardiovascular and cerebrovascular diseases. Polysomnography, the gold standard, is expensive, inaccessible, uncomfortable and an expert technician is needed to score sleep-related events. To address these limitations, many previous studies have proposed and implemented automatic scoring processes based on fewer sensors and machine learning classification algorithms. However, alternative device technologies developed for both home and hospital still have limited diagnostic accuracy for detecting apnea events even though many of the previous investigational algorithms are based on multiple physiological channel inputs. In this paper, we propose a new probabilistic algorithm based on (only) oronasal respiration signal for automated detection of apnea events during sleep. The proposed model leverages AASM recommendations for characterizing apnea events with respect to dynamic changes in the local respiratory airflow baseline. Unlike classical threshold-based classification models, we use a Gaussian mixture probability model for detecting sleep apnea based on the posterior probabilities of the respective events. Our results show significant improvement in the ability to detect sleep apnea events compared to a rule-based classifier that uses the same classification features and also compared to two previously published studies for automated apnea detection using the same respiratory flow signal. We use 96 sleep patients with different apnea severity levels as reflected by their Apnea-Hypopnea Index (AHI) levels. The performance was not only analyzed over obstructive sleep apnea (OSA) but also over other types of sleep apnea events including central and mixed sleep apnea (CSA, MSA). Also the performance was comprehensively analyzed and evaluated over patients with varying disease severity conditions, where it achieved an overall performance of  $TPR = 88.5\%$ ,  $TNR = 82.5\%$ , and  $AUC = 86.7\%$ . The proposed approach contributes a new probabilistic framework for detecting sleep apnea events using a single airflow record with an improved capability to generalize over different apnea severity conditions

**Keywords:** sleep apnea; airflow signal; Gaussian Mixture Models (GMM)

## 1. Introduction

Sleep apnea is a highly prevalent sleep disorder that can cause significant daytime sleepiness and result in many cardiovascular comorbidities [1–4]. It is characterized by repetitive significant airflow reductions during sleep causing recurrent hypoxia and sleep fragmentation [5–7]. When breathing doesn't completely stop but the volume of air going into the lungs is significantly reduced, then the respiratory event is called a hypopnea. More than 200 million patients worldwide are affected with sleep apnea [8].

Sleep apnea events are three types: Obstructive, central, and mixed [9]. Obstructive sleep apnea (OSA) is characterized by repetitive upper airway obstruction that limit airflow from going in to the lungs with the presence of continued respiratory effort. Central sleep apnea (CSA) is characterized by the loss of all respiratory effort during sleep due to a neurological disorder. Mixed sleep apnea (MSA) is combination of both obstructive and central sleep apnea symptoms.

Polysomnography (PSG), often called a sleep study, is the gold standard for detecting sleep apnea. Polysomnography records basic human body activities during sleep in an attended setting (sleep laboratory). This includes electrocardiogram (ECG) for heart, oronasal thermal airflow signal (FlowTh) and nasal pressure signal (NPPE) for respiration, electroencephalogram (EEG) for brain, electromyogram (EMG) for muscles, and oxygen level in the blood (SpO<sub>2</sub>) [10,11]. Connecting a large number of sensors and wires to a subject in a dedicated sleep lab makes PSG uncomfortable, expensive, and unavailable to a large number of sleep patients in many parts of the world [9]. Moreover, clinicians need an offline inspection of the recordings to score apnea and derive the apnea-hypopnea index (AHI), which is the parameter used to establish sleep apnea and its severity [12]. Thus, the analysis process is labor-intensive and time-consuming, leading to a delayed diagnostic process and increased patient waiting lists [13–16] as well as being highly susceptible to human errors [17].

To overcome limitations of PSG, several studies have been proposed for automated detection of sleep apnea using a limited subset of signals among those involved in PSG [18]. This includes respiratory signals [19–29], ECG [16,30–32], SpO<sub>2</sub> [33–35], tracheal sound signals [36], or some combinations of signals listed above [37]. A number of portable devices for sleep apnea monitoring and diagnosis have been developed and are available. LifeShirt, SleepStrip, and ApneaLink are among the most popular products [38].

Respiratory airflow signal is a straightforward choice to look for simpler alternatives to PSG, since apneas are primarily defined on the basis of its amplitude oscillations [12]. According to the American Academy of Sleep Medicine (AASM), the primary sensor for identifying apnea in sleep diagnostic studies is the oronasal thermal airflow sensor [11]. Thus, several studies focused on automated detection of sleep apnea events based exclusively on the analysis of this signal. In these studies, the airflow signal is analyzed in different analytic domains (linear, nonlinear, time and frequency domains) to extract features which are then used in a rule-based threshold classifier or in a “black box” machine learning model. Rule-based threshold classification has been used in [24,39–42]. Support Vector Machines (SVM) [43,44], Artificial Neural Networks (ANN) [20,45–47], linear discriminant analysis (LDA), and regression trees (CART) with the AdaBoost (AB) [22] are among the most popular machine learning models that used respiratory flow signal.

Despite their popularity in sleep apnea problems, a major limitation in classical rule-based threshold detectors is that they provide classification based on simple comparison for the features against experimentally derived thresholds while overlooking the statistical distributions for the input features as well as the output classes. Even more complex discriminative (black box) methods are based on learning a function that maps the features directly into decisions. There hasn't been much research considering probabilistic view of classification for sleep apnea detection.

Gaussian Mixture Model (GMM) is a probabilistic machine learning framework that aims at providing a richer class of density models than single Gaussian using a finite weighted mixture of Gaussian densities. It is well known as a rich framework capable of characterizing any continuous

density. This framework has also shown promising results in classification problems including noisy features [48]. Nevertheless, it hasn't been well evaluated for sleep apnea detection problems.

The contribution of this paper is two fold. First, we develop a probability based classification approach for automated detection of sleep events using single oronasal airflow record. Second, we study the performance of the proposed approach over a large data set of 96 patients of different sleep apnea severity levels. Finally, we conduct a comprehensive evaluation and comparison of the proposed probabilistic framework against a rule-based classifier for the same input features as well as two previously published algorithms for apnea detection using airflow signal.

This paper is organized as follows. Section 2 describes the data set, the proposed algorithm, classification methods, and evaluation metrics. Section 3 presents results for the proposed algorithm along with a detailed comparison with related works. Section 4 discusses the results and lessons learned and 5 summarizes conclusion of the paper.

## 2. Materials and Methods

### 2.1. Data Set

The Institutional Review Board (IRB) at the University of Michigan approved this study (IRB#HUM00069035). Full polysomnography (PSG) data was collected for 96 patients at the University of Michigan Sleep Disorders Center. For each patient, polysomnography consisted of electroencephalography (EEG), electrooculography (EOG), submental and tibial electromyography (EMG), electrocardiography (ECG), two piezoelectric belts for recording plethysmography (PPG), oronasal airflow sensor (FlowTH), air pressure transducer (NPPE), digital micro-phone and pulse oximeter.

The oronasal airflow sensor used in this study is a thermocouple-based one from Respirationics Model: Pro-Tech-P1273 (Philips Healthcare, Eindhoven, The Netherlands). Clinical annotations for respiratory events were carried out by expert clinicians from the Sleep Disorders Center at the University of Michigan (Ann Arbor, MI) and according to recommendations of the AASM [11]. Apneic events in the data set are either obstructive (OSA), central (CSA), or mixed (MSA). The data set spans different apnea severity levels as reflected by the apnea-hypopnea index (AHI) computed over night for patients in the study. 10 are non/minimal sleep apnea patients ( $AHI < 5$ ), 36 are mild sleep apnea patients ( $5 \leq AHI < 15$ ), 27 are moderate sleep apnea patients and 23 are severe sleep apnea patients ( $AHI \geq 30$ ). Table 1 provides distribution for the numbers and types of the apneic events per each class of the patient severity levels.

The oronasal airflow signals of 66 patients (with different apnea severity levels) were used for training the proposed modeling framework. The developed framework was then tested on 30 patients (distinct from the training data) composed of 5 none/minimal apnea, 5 mild, 5 moderate, and 15 severe sleep apnea patients.

**Table 1.** Distribution of Number and Types of Apneic Events per each Class of the Severity Levels.

Severity Level	OSA	CSA	MSA	Total
None/Minimal	20	312	0	332
Mild	1302	1887	78	3267
Moderate	3056	4240	400	7696
Severe	10,922	23,960	2143	37,025
Total	15,300	30,399	2621	48,320

### 2.2. A Data-Driven Approach for Characterizing Changes in Respiratory Baseline

According to the AASM, an apnea event is scored if there is a drop in peak thermal airflow signal excursions by  $\geq 90\%$  of the corresponding baseline for a duration  $\geq 10$  s [11]. Nevertheless, the airflow baseline is not precisely defined neither in the AASM Scoring Manual nor in sleep literature.

To overcome this limitation, a data driven approach will be used to derive the respiratory flow baseline from the airflow signal (FlowTH). The derived baseline will then be used to characterize dynamic changes in respiration with respect to this (dynamic) baseline in order to detect the occurrence of apneic events. To establish respiratory baseline, we will consider two important respiratory features: Inter-breath intervals and breath amplitudes.

A sliding window method will be used for detecting apnea events in the oronasal airflow signal (FlowTH). At time step  $t$ , two windows will be established. The first window (baseline window- $W_b$ ) will be used to derive the local respiratory baseline. The second window (detection window- $W_m$ ) will be used to detect apneic events based on relative changes in inter-breath intervals and breath amplitudes in  $W_m$  with respect to those in the  $W_b$ . In this study, we considered a  $W_b$  of length  $L_b = 600$  s that contains the airflow measurements up to time  $t$  and a  $W_m$  of length  $L_m = 100$  s that contains airflow measurements starting from time step  $t + 1$ .

After constructing  $W_b$  and  $W_m$ , peaks and valleys of the respiratory airflow signal are detected in both windows. An example of  $W_b$  and  $W_m$  that both include an apneic event along with peak and valley detections is illustrated in Figure 1a. The inter-breath intervals and the breath amplitudes can now be extracted from  $W_b$  and  $W_m$  as follows:

$$PP_i = t_{i+1} - t_i \quad (1)$$

$$PV_i = P_i - V_i \quad (2)$$

where the airflow breath  $i$  has a peak  $P_i$  that occurs at time instance  $t_i$ , a valley  $V_i$ , an inter-breath interval  $PP_i$ , and a breath amplitude  $PV_i$ . These Equations generate sequences of inter-breath intervals and breath amplitudes in  $W_b$  and  $W_m$  as illustrated in Figure 1b,c,f,g.

After getting these sequences, it is required to extract the (inter-breath) intervals and (breath) amplitudes in  $W_b$  that contribute most to the respiratory baseline estimate of this window. Similarly, it is required to extract the intervals and amplitudes in  $W_m$  that belong to the apneic event to be detected. This can be effectively done by sorting sequences of intervals and amplitudes in both  $W_b$  and  $W_m$  based on corresponding values. For convenience,  $PP_i$  and  $PV_i$  in  $W_b$  are sorted in a descending order while those in  $W_m$  are sorted in an ascending order as illustrated in Figure 1d,e,h,i. This process will generate ordered sequences  $PP_b = \{PP_i^d\}$ ,  $PV_b = \{PV_i^d\}$ ,  $PP_m = \{PP_i^a\}$ , and  $PV_m = \{PV_i^a\}$  where subscripts  $b$  and  $m$  specify  $W_b$  and  $W_m$  respectively, and superscripts  $a$  and  $d$  specify ascending and descending orders respectively.

Although the length of  $W_b$  and  $W_m$  are fixed, the number of airflow breaths in these windows typically vary during different sleep stages and across different patients. Thus, the ordered sequences ( $PP_b$ ,  $PV_b$ ,  $PP_m$ , and  $PV_m$ ) will be filtered to keep only the intervals and amplitudes that contribute most to the baseline estimate in  $W_b$  and the apneic events in  $W_m$  respectively. This can be mathematically expressed as follows:

$$L_{PP_b} = \lfloor F_{PP_b} N_{PP_b} \rfloor \quad (3)$$

$$L_{PV_b} = \lfloor F_{PV_b} N_{PV_b} \rfloor \quad (4)$$

$$L_{PP_m} = \lfloor F_{PP_m} N_{PP_m} \rfloor \quad (5)$$

$$L_{PV_m} = \lfloor F_{PV_m} N_{PV_m} \rfloor \quad (6)$$

where  $F_s$  is the cut-off filter applied to the ordered sequence  $s$  of length  $N_s$  in order to generate a filtered sequence of length  $L_s$  where  $s \in \{PP_b, PP_m, PV_b, PV_m\}$ . Accordingly, the filtered sequences include the highest  $L_{PP_b}$  inter-breath intervals and  $L_{PV_b}$  breath amplitudes in  $W_b$  and the lowest  $L_{PP_m}$  intervals and  $L_{PV_m}$  amplitudes in  $W_m$ . The filter values were defined individually for each of the sequences to allow them to be tuned separately to maximize the ability to detect apneic windows. The mathematical means of the filtered sequences can now expressed as follows:

$$B_{PP_b} = \frac{1}{L_{PP_b}} \sum_{i=1}^{L_{PP_b}} PP_i^d \quad (7)$$

$$B_{PV_b} = \frac{1}{L_{PV_b}} \sum_{i=1}^{L_{PV_b}} PV_i^d \quad (8)$$

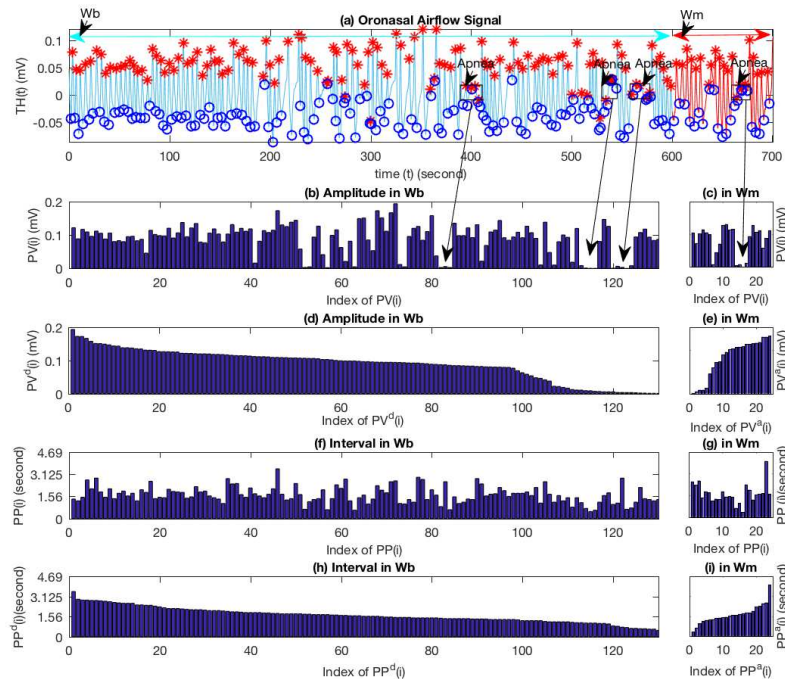
$$B_{PP_m} = \frac{1}{L_{PP_m}} \sum_{i=1}^{L_{PP_m}} PP_i^a \quad (9)$$

$$B_{PV_m} = \frac{1}{L_{PV_m}} \sum_{i=1}^{L_{PV_m}} PV_i^a, \quad (10)$$

where  $B_s$  is the mathematical mean of the filtered sequence  $s$ . The relative changes in the inter-breath intervals ( $I_c$ ) and the amplitude of the breaths ( $A_c$ ), with respect to the respiratory baseline, can now be computed as follows:

$$I_c = \frac{B_{PP_b} - B_{PP_m}}{B_{PP_b}} \quad (11)$$

$$A_c = \frac{B_{PV_b} - B_{PV_m}}{B_{PV_b}} \quad (12)$$



**Figure 1.** (a) Oronasal respiration signal with peak/ valley detections illustrated by (red ‘\*’)/ (blue ‘o’) respectively. (b) Sequence of extracted breath amplitudes of  $W_b$ . (c) Sequence of extracted breath amplitudes of  $W_m$ . (d)  $PV_b$  Sequence of descending ordered amplitudes in  $W_b$ . (e)  $PV_m$  Sequence of ascending ordered amplitudes of  $W_m$ . (f) Extracted inter-breath intervals of  $W_b$ . (g) Extracted inter-breath intervals of  $W_m$ . (h)  $PP_b$  sequence of descending ordered inter-breath intervals of  $W_b$ . (i)  $PP_m$  sequence of ascending ordered intervals of  $W_m$ . Annotations in subplot (a) illustrate example of apnea events present in  $W_b$ ,  $W_m$ . Arrows point to the breath amplitudes from apnea events shown in  $W_b$ ,  $W_m$ .



### 2.3. Detection of Apnea Events based on Relative Changes in Respiratory Baseline

For the classification part, we propose a probabilistic view of classification for automated detection of apnea events. We leverage a Gaussian Mixture Model (GMM) to derive a decision boundary based on probabilistic assumptions about the underlying distribution of the respiratory input features. We denote this modeling scheme as a Gaussian Mixture Model (GMM) classifier. In order to demonstrate the improvement achieved by considering GMM as a generative machine learning model, we compare results with a classical threshold-based detector that uses the same input features for automated detection of apnea events.

To prepare data for classification, a sliding window that is being successively updated each 20 s (step-size = 20 s) was applied over the oronasal airflow signal. At each of the steps,  $W_b$  and  $W_m$  are constructed. Then, Equations (1)–(12) are applied to compute  $I_c$  and  $A_c$  (input features to the classification model) while  $W_m$  provides classification label based on whether or not an apnea event was clinically scored in this window. Considering our data set with 66 patients for training and 30 for testing, Table 2 shows the distribution of the data segments and corresponding labels for both training and test sets.

**Table 2.** Segments for training and testing classification models.

Labels/ Segments	Training Set	Testing Set	Total
Normal	208,456	20,789	229,245
Apnea	81,055	8048	89,103
Total	289,511	28,837	318,348

#### 2.3.1. Rule-Based Threshold Based Classification

The rule-based classifier detects apnea in detection windows ( $W_m$ ) when the input features  $I_c$  and  $A_c$  both activate the classification rules  $\hat{I}_{lb} \leq I \leq \hat{I}_{ub}$  and  $A_c > \hat{A}$  where  $\hat{I}_{lb}$ ,  $\hat{I}_{ub}$  are the classification thresholds for  $I_c$  and  $\hat{A}$  is the classification threshold for  $A_c$ . An exhaustive search approach is applied for each of these thresholds in order to learn their optimal values. A novel approach was used to fit the rule-based classifier and learn the classification rules using a two step optimization method. The classification thresholds for  $I_c$  are optimized first to obtain the receiver operating characteristics curve (ROC) with the maximum area under ROC (AUC) over our training data. Once the  $I_c$  classification rule is learned, the optimal classification threshold for  $A_c$  is learned by searching along the selected receiver operating curve (with maximum AUC) for the threshold that provides the maximum sensitivity (TPR) that constrains (FPR) not to exceed the maximum acceptable limit of 20% ( $FPR \leq 20\%$ ). More details about the derivation and tuning of the rule-based classifier can be found in our recent study [49].

#### 2.3.2. Classification with Gaussian Mixture Models (GMM)

A Gaussian mixture model (GMM) is a probabilistic modeling framework. In this model, the probability density function (PDF) of  $\mathbf{x} \in R^d$  is defined as a finite weighted sum of  $k$  Gaussian distributions:

$$p(\mathbf{x}|\Theta) = \sum_{i=1}^k \gamma_i p(\mathbf{x}|\theta_m) \quad (13)$$

such that  $\mathbf{x}$  is the 2-dimensional feature vector  $[I_c, A_c]^T$  computed every time  $W_b$  and  $W_m$  are constructed,  $\sum_{i=1}^k \gamma_i = 1$ ,  $\Theta$  is the mixture model,  $\gamma_i$  corresponds to the weight of component  $i$ , and the density of each component is given by the normal probability distribution:

$$p(\mathbf{x}|\theta_m) = \frac{|\Sigma_m|^{-\frac{1}{2}}}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_m)^T \Sigma_m^{-1}(\mathbf{x} - \mu_m)\right\} \quad (14)$$

The parameters  $\gamma$ , the mean  $\mu$ , and the covariance  $\Sigma$  are optimized during the training process using the expectation maximization algorithm [50] such that the log-likelihood of the model is maximized. During testing, a likelihood estimate is obtained for the apnea class, defined by the model  $\Theta_A$ , and for the non-apneic (normal respiration) class, defined by the model  $\Theta_N$ . Using the Bayesian classification formula, the likelihood estimates are combined to compute the posterior probability of apnea for the sample  $\mathbf{x}$ :

$$P(A|\mathbf{x}) = \frac{p(\mathbf{x}|\Theta_A)P(A)}{p(\mathbf{x}|\Theta_A)P(A) + p(\mathbf{x}|\Theta_N)P(N)} \quad (15)$$

where  $P(A)$  and  $P(N)$  are the prior probabilities of the apnea and non-apnea (normal) classes respectively. These probabilities were set by symmetry to be equal  $P(A) = P(N) = 0.5$  assuming we have no prior knowledge about them. The combination of the two GMMs and the Bayesian classification formula in Equation (15) form the GMM classifier [51].

## 2.4. Evaluation of Apnea Detection Results

### 2.4.1. Classification Performance over Detection Windows

In this paper, five statistical metrics of accuracy ( $ACC$ ), true positive rate ( $TPR$ ), true negative rate ( $TNR$ ), positive predictive value ( $PPV$ ), and  $F_1$  score are applied to assess the performance of the proposed modeling framework over all detection windows ( $W_m$ ):

$$ACC = \frac{\sum TP + \sum TN}{\sum TP + \sum FP + \sum FN + \sum TN} \times 100\% \quad (16)$$

$$TPR = \frac{\sum TP}{\sum TP + \sum FN} \times 100\% \quad (17)$$

$$TNR = \frac{\sum TN}{\sum FP + \sum TN} \times 100\% \quad (18)$$

$$PPV = \frac{\sum TP}{\sum TP + \sum FP} \times 100\% \quad (19)$$

$$F_1 = 2 \frac{TPR \cdot PPV}{TPR + PPV} \times 100\% \quad (20)$$

where  $TP$  (true positive) is the number of apneic windows that were correctly classified as such,  $TN$  (true negative) is the number of normal windows that were correctly classified as such,  $FP$  (false positive) is the number of normal windows that were falsely classified as apneic,  $FN$  (false negative) is the number of apneic windows that were missed by the classifier.  $ACC$  is a classical measure for binary classification but is not enough in this problem due to class imbalance between the apnea and normal classes [52–54]. Thus,  $TPR$ ,  $TNR$ , and  $PPV$  are used to report a more detailed performance in detecting apneic and normal windows. The  $F_1$  score considers  $TP$  and  $FP$  detections simultaneously and thus accounts to the  $TPR/PPV$  tradeoff reporting a more comprehensive idea on the overall performance of the proposed model.

### 2.4.2. Receiver Operating Characteristics (ROC) Curve

The Receiver Operating Characteristics (ROC) curve is an effective tool used to graphically illustrate the diagnostic ability of a binary classification system as its classification threshold is varied [55]. This curve simply plots the  $TPR$  against the false positive rate ( $FPR = 100\% - TNR$ ) at various discrimination threshold settings. The Area Under receiver operating characteristics Curve ( $AUC$ ) is used as a measure of the overall ability of the classification model to automatically detect sleep apnea events. A greater  $AUC$  indicates a more useful and effective classification model. Additionally, the ROC curve can be used for optimizing classification models by finding the operating threshold

that provides the highest  $TPR$  for the allowable  $FPR$  level. This approach was used in learning the classification rules of the rule-based classifier.

### 3. Results

For the proposed GMM classifier (AICPV with GMM) and the rule-based threshold classifier (AICPV with Threshold), we used the PSGs of 66 patients for training, tuning, and optimizing the classifiers. The trained classifiers were then tested over the PSGs of the other distinct 30 patients.

#### 3.1. Rule-Based Threshold Classifier

The optimal filter values were set to  $F_{PP_b} = 0.3$ ,  $F_{PV_b} = 0.4$ ,  $F_{PP_m} = 0.1$ , and  $F_{PV_m} = 0.3$ . The classification rules for detecting apneic windows were identified as follows

$$0.05 \leq I_c \leq 0.95 \quad (21)$$

$$0.957 \leq A_c \quad (22)$$

such that an apneic window is detected by the rule-based classifier whenever both rules are active.

#### 3.2. GMM Classifier

Cross validation over the training data set was used for investigating and selecting the choices of parameters for the GMM models. These parameters include the number of Gaussian distributions needed to model each class, and the type of covariance matrix used (diagonal or full symmetrical). Our results show that 12 Gaussian components are needed to model the GMM of the apneic class and 11 Gaussian components are needed to model the GMM of the normal class along with diagonal covariance matrices for both GMMs.

#### 3.3. Classification Performance Comparison over the Testing Data Set

We performed an overall evaluation for the proposed model (AICPV with GMM—AICPVwGMM) over the 30 patient test data and we compared the performance results with the rule-based classification model that uses the same input features (AICPV with Threshold - AICPVwTH). Also, we considered performance comparison with two well known published algorithms for automated apnea detection using the oronasal thermal airflow signal and similar time-domain based features [39,45]. The first algorithm implements a classical threshold based classification model [39] while the other one uses an artificial neural network classification model [45]. Note that [39] includes an additional module for classifying the type of detected apnea using a neural network classifier trained on the thoracic effort signal. Nevertheless, we just included the apnea detection module from this study since the proposed algorithm uses only a single channel of oronasal airflow.

In order to do a fair comparison, the four algorithms AICPVwGMM, AICPVwTH, Refs. [39,45] were all trained, evaluated, and tested on identical data within our data set. Table 3 comprehensively compares classification performance over the test data between these four algorithms. First, it can be noticed that the AICPVwGMM and the AICPVwTH outperform the two previously published algorithms in all performance measures of this paper. The AICPV algorithms demonstrate a higher ability to detect apnea events (reflected by their  $TPR$ ) as well a higher ability to detect normal respiration patterns (reflected by their  $TNR$ ) as opposed to [39,45]. An overall better classification performance of the AICPV algorithms can be demonstrated with the  $F_1$  and  $AUC$  values compared to [39,45]. The improvement achieved with AICPV algorithms is mainly caused by the dynamic approach considered in these algorithms such that apneic events are characterized based on relative changes in airflow breath amplitudes and inter-breath intervals with respect to local respiration baseline.



**Table 3.** Comparison of performance over a 30-patient test data set between the rule-based threshold classifier (AICPVwTH), the proposed Gaussian Mixture Models (GMM) classifier (AICPVwGMM), and two previous related studies.

Algorithm	TPR (%)	TNR (%)	PPV (%)	ACC (%)	F <sub>1</sub> (%)	AUC (%)
AICPVwTH	88.3	79.4	28.2	80.2	42.7	83.9
AICPVwGMM	88.5	82.5	46.6	83.4	61.1	86.7
[45]	62.0	76.9	19.3	77.6	29.4	75.7
[39]	60.3	79.5	9.2	78.9	16.0	69.9

Comparing the performance obtained with the proposed GMM based classifier (AICPVwGMM) and the rule-based classifier (AICPVwTH), we can notice a 65.2% increase in the *PPV* of the apnea detections obtained with the GMM based classifier as opposed to the rule-based threshold classifier. Recognizing *TPR/PPV* tradeoff, and that *TPR* and *PPV* are performance metrics of competing natures, it can be noticed that using a GMM-based classifier, we can achieve a high *TPR* with a significantly improved *PPV* compared to the rule-based classifier. This can be also noticed by observing the significant increase in the *F<sub>1</sub>* score for the detections with the proposed GMM model as opposed to the rule-based one. Also, higher *TNR* and *ACC* are obtained with the proposed algorithm reflecting a higher ability to detect normal respiratory patterns. The overall classification performance indicated by *AUC* also shows an improved detection with the proposed GMM modeling framework.

To provide an in-depth analysis and understand the sources of improvement with the proposed model as opposed to the rule-based classification model, we did a comprehensive analysis for the test performance of the proposed algorithms over different apnea types and different apnea severity conditions. Tables 4 and 5 provide a detailed comparison between the GMM-based classifier and the rule based classifier over different apnea types and different apnea severity levels. A clear increase in the ability to detect OSA and CSA events can be noticed in the *TPR* of these detections using the proposed algorithm along with a significant increase in the *PPV* of all types of apnea detections. It can be also noticed that the *TPR* of the MSA detections is superior with both algorithms and that it didn't change between them which is mainly due to the fact the MSA events are minority compared to OSA and CSA events.

Looking at the detailed performance of the proposed GMM model and the rule-based classifier over different apnea severity conditions, we can notice that the best performance for the rule-based threshold classifier is in severe apnea patients and that the performance significantly degrades over less severe cases. On other hand, the GMM based classifier maintains a high ability to detect apnea events in severe patients, but more importantly, it has a significantly higher ability to detect apneic events in less severe apnea patients which can be clearly seen through the *TPR* of these detections. Moreover, looking at the *AUC* values, we can also notice an excellent overall ability to detect apnea events in severe patients for the GMM modeling framework as well as a significantly improved overall ability to detect apnea in less severe patients compared to the rule-based classification model.

**Table 4.** Test performance of AICPV with GMM classifier over different types of apnea and different apnea severities.

AICPVwGMM		TPR (%)	TNR (%)	PPV (%)	ACC (%)	F <sub>1</sub> (%)	AUC (%)
Type of Apnea	MSA	97.2	80.6	9.9	80.9	18.0	88.9
	CSA	88.7	80.6	41.8	81.7	56.8	84.6
	OSA	92.7	80.6	25.3	81.4	39.8	86.7
AHI Severity	Severe	96.4	77.3	68.7	83.8	80.3	86.7
	Moderate	98.1	77.5	32.6	79.5	48.9	87.8
	Mild	100.0	84.8	12.2	85.2	21.7	92.4
	None/Minimal	27.3	71.8	0.0	83.4	0.0	86.7
All		88.5	82.5	46.6	83.4	42.7	86.7

**Table 5.** Test performance of AICPV with threshold classifier over different types of apnea and different apnea severities.

AICPVwTH		TPR (%)	TNR (%)	PPV (%)	ACC (%)	F <sub>1</sub> (%)	AUC (%)
Type of Apnea	MSA	97.5	81.1	7.8	81.4	14.4	89.3
	CSA	86.8	81.1	17.1	81.4	28.6	83.4
	OSA	88.4	81.1	20.6	81.6	33.4	85.6
AHI Severity	Severe	92.4	90.3	82.5	91.0	87.2	91.4
	Moderate	79.5	80.7	32.8	80.6	46.4	80.1
	Mild	76.6	76.1	3.7	76.1	7.1	76.3
	None/Minimal	6.3	97.0	2.0	96.1	3.0	51.7
All		88.4	79.4	28.2	80.2	61.1	83.4

Finally, we performed a comprehensive analysis for the performance per class of apnea types among different patient severity levels. Table 6 evaluates how the proposed model performs on detecting different apnea types in each of apnea severity classes. As it can be seen in the table, the proposed model AICPVwGMM maintains a high ability to detect different apnea types regardless disease severity in the test patients. MSA events are very rare in mild patients which caused low and skewed detection rates for these patients. Also, the final row in Table 6 was left empty since there are no MSA events in the class of none/minimal apnea patients. In general, the class of none/minimal apnea reflects patients that are healthy or with few significant apnea events and so it is a class of less interest compared to other disease states. Nevertheless, we kept performance results on all disease severity classes to report comprehensive assessment of the modeling framework.

It is worthy to be mentioned that the implemented algorithms were trained and tested using full overnight PSG records. Our goal is to test the proposed framework and to compare it with existing works in a more practical setting as opposed to many previous studies that considered shorter records avoiding the class imbalance problem and excluding segments with low signal to noise ratio (SNR). False positives were affected by the class imbalance problem, segments of low airflow signal quality, and irregular breathing patterns and artifacts. High airflow peak amplitudes resulting from increased respiratory effort after the end of apnea events may affect false positives by contributing falsely to the respiratory baseline. Future work may consider adding more advanced signal filtration algorithms that can allow more accurate detection in case of artifacts as well as to reject airflow segments with very low SNR.

**Table 6.** Detailed test performance of AICPV with GMM classifier (AICPVwGMM) over different Apnea–Hypopnea Index (AHI) severities per apnea class.

Type	AHI	TPR (%)	TNR (%)	PPV (%)	ACC (%)	F <sub>1</sub> (%)	AUC (%)
OSA	All	92.7	80.6	25.3	81.4	39.8	86.7
	Severe	97.2	77.5	59.8	82.5	74.0	87.4
	Moderate	92.8	84.8	31.3	85.4	46.8	88.8
	Mild	87.6	71.8	1.9	71.9	3.7	79.7
	Normal	0.0	93.3	0.0	93.2	NA	46.6
CSA	All	88.6	80.6	41.8	81.7	56.8	84.6
	Severe	95.6	77.5	52.8	81.3	68.0	86.6
	Moderate	85.9	84.8	23.9	84.9	37.4	85.4
	Mild	90.1	71.8	1.7	71.9	3.3	80.9
	Normal	7.6	93.3	1.1	92.4	1.9	50.4
MSA	All	97.2	80.6	9.9	80.9	18.0	88.8
	Severe	98.1	77.5	32.6	79.5	48.9	87.8
	Moderate	100	84.8	12.2	85.2	21.7	92.4
	Mild	27.3	71.8	0.1	71.8	0.0	49.5
	Normal	-	-	-	-	-	-

#### 4. Discussion of Results

A probabilistic-based framework was developed in this study for automated apnea detection using single channel data from oronasal airflow record (FlowTH). The proposed framework leverages AASM recommendations to define apnea based on relative changes with respect to respiratory airflow baseline. To overcome the absence of a precise mathematical definition for airflow baseline, a data-driven method is developed to represent it based on two features: The breath amplitudes and inter-breath intervals. The apnea is then characterized based on relative changes in these features between two sliding windows: The baseline window which represents the current respiratory baseline and the detection window where an apneic event is to be detected.

For automatic detection of apneic events, we considered classification based on a probabilistic view using a GMM-based modeling framework. The proposed framework showed a significantly improved performance in detecting apnea compared to a rule-based classifier that uses the same input features as well as two previously published algorithms that respectively use threshold-based classification and neural networks, applied on time domain features from the same respiratory signal. Using relative changes in respiratory features to define apnea enabled a dynamic approach that accounts for continuous changes in the respiratory baseline making AICPVwTH and AICPVwGMM algorithms significantly more capable to detect apneic events than previous studies that considered absolute changes in classification features overlooking relative ones. Comparing the proposed model AICPVwGMM with AICPVwTH that uses a rule-based classifier showed a significantly improved performance for the GMM model characterized by achieving high *TPR* with a significantly improved overall *PPV* for all types of apnea detections. The proposed model also allows much better performance over different apnea severity levels compared to the rule-based classifier which showed best performance over severe apnea patients only with significantly degraded performance over less severe disease classes.

In recent years, many studies considered oronasal thermal airflow signal for automated apnea detection [19–21,39,42–45,56]. Nevertheless, patients with severe and moderate OSA conditions have been a major focus of many of these studies while not giving sufficient attention to less severe patient populations or other types of apnea events. The present results highlight the importance of evaluating apnea models on patients of varying severity conditions as well as on different apnea types. This also agrees with previous literature which demonstrated that high performance accuracies achieved with patients with high severity levels may not be generalizable to other groups of patients [57,58]. The detailed analysis presented in this paper using patients of varying apnea severity conditions and different apnea types provides the basis for a more comprehensive understanding of the performance of apnea detection systems.

Comparative analysis between the performance of proposed modeling framework and previously published research highlights some of the previously reported limitations for single-respiration channel based apnea detection methods. In particular, previous studies have reported significant fall in the diagnostic accuracy of automated sleep systems that measure two or fewer physiological parameters as opposed to those that measure three or more physiologic variables [38,57]. Although many automated sleep systems have proved effectiveness assisting lab PSGs and at home, they still cannot completely replace dedicated centers for sleep studies [59].

The present study leverages AASM recommendations for apnea detection in many aspects. We considered the AASM recommended sensor for scoring apneic events in PSG diagnostic studies which is the oronasal thermal airflow sensor. The criteria in the AASM manual were also employed for scoring an event after a drop in peak thermal airflow signal excursions by  $\geq 90\%$  of the corresponding baseline for a duration  $\geq 10$  s [11]. Nevertheless, there are many sources of uncertainty in the criteria defined in literature. First, the AASM doesn't provide a precise definition for the respiratory baseline. Second, the published criteria for mathematical scoring apnea are not consistent and vary with different standards [60]. Moreover, the high intraobserver and interobserver variability due to human scoring and human errors [17] significantly affect the robustness and performance of automated sleep systems.

Adopting a probabilistic framework in the proposed study provides an efficient approach to propagate different sources of uncertainty using a data-driven modeling framework optimized with respect to the ability to detect apnea events. Interestingly, using the proposed GMM-based classification system showed an overall improved performance in apnea detection and a more consistent and generalizable performance across patients with different severity levels.

Finally, the presented approach focuses on automated detection for apneic events using oronasal airflow respiration signal. Future work may extend the algorithm by adding an additional input through the nasal pressure signal in order to study dynamical changes in this signal during hypopnea as recommended by the AASM. Using the oronasal airflow and the nasal pressure signals will enable detecting both apnea and hypopnea events needed to estimate the AHI in order to provide a diagnosis for sleep apnea severity. Additionally, this will improve the ability to estimate respiratory baseline by incorporating two signals instead of one leading potentially to a decreased number of false positive detections. Moreover, future work may consider adding input from the respiratory effort signal to provide diagnosis for the type of detected events (OSA, OSA, or MSA).

## 5. Conclusions

In this study, a new algorithm is developed for automated detection of sleep apnea events using single channel data from oronasal thermal airflow sensor (FlowTH). The algorithm leverages AASM recommendations by defining apnea using relative changes in the oronasal airflow signal with respect to the evolving respiratory airflow baseline. A novel approach is developed to represent the respiratory airflow baseline based on two features: The inter-breath intervals and breath amplitudes. In order to detect apneic events, we considered a probabilistic view for classification using a GMM modeling framework. The proposed framework showed a significantly improved apnea detection performance for different apnea types and severity conditions as opposed to a rule-based classifier that uses the same input features as well as two previous methods for automated detection using the same respiratory source. The proposed modeling framework achieved an overall summary performance of  $TPR = 88.5\%$ ,  $TNR = 82.5\%$ , and  $AUC = 86.7\%$ .

**Author Contributions:** Conceptualization, H.E. and J.K.; methodology, H.E. and J.K.; software, J.K.; validation, H.E. and J.K.; formal analysis, J.K.; investigation, H.E., J.K., and M.R.; resources, H.E., J.K., D.T., and S.K.R.; data curation, H.E. and J.K.; writing—original draft preparation, H.E.; writing—review and editing, H.E., J.K., and M.R.; visualization, H.E. and J.K.; supervision, D.T., S.K.R., and C.-H.C.; project administration, D.T. and S.K.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Somers, V.K.; White, D.P.; Amin, R.; Abraham, W.T.; Costa, F.; Culebras, A.; Daniels, S.; Floras, J.S.; Hunt, C.E.; Olson, L.J.; et al. Sleep apnea and cardiovascular disease: An American heart association/American college of cardiology foundation scientific statement from the American heart association council for high blood pressure research professional education committee, council on clinical cardiology, stroke council, and council on cardiovascular nursing in collaboration with the national heart, lung, and blood institute national center on sleep disorders research (national institutes of health). *J. Am. Coll. Cardiol.* **2008**, *52*, 686–717. [[PubMed](#)]
2. Botros, N.; Concato, J.; Mohsenin, V.; Selim, B.; Doctor, K.; Yaggi, H.K. Obstructive sleep apnea as a risk factor for type 2 diabetes. *Am. J. Med.* **2009**, *122*, 1122–1127. [[CrossRef](#)] [[PubMed](#)]
3. Mandal, S.; Kent, B.D. Obstructive sleep apnoea and coronary artery disease. *J. Thorac. Dis.* **2018**, *10*, S4212. [[CrossRef](#)] [[PubMed](#)]
4. Sharma, S.; Culebras, A. Sleep apnoea and stroke. *Stroke Vasc. Neurol.* **2016**, *1*, 185–191. [[CrossRef](#)] [[PubMed](#)]

5. Quan, S.; Gillin, J.C.; Littner, M.; Shepard, J. Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. editorials. *Sleep* **1999**, *22*, 662–689. [[CrossRef](#)]
6. American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*; American Academy of Sleep Medicine: Westchester, IL, USA, 2007.
7. Dempsey, J.A.; Veasey, S.C.; Morgan, B.J.; O'Donnell, C.P. Pathophysiology of sleep apnea. *Physiol. Rev.* **2010**, *90*, 47–112. [[CrossRef](#)] [[PubMed](#)]
8. Zhang, J.; Zhang, Q.; Wang, Y.; Qiu, C. A real-time auto-adjustable smart pillow system for sleep apnea detection and treatment. In Proceedings of the 2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Philadelphia, PA, USA, 9–11 April 2013; pp. 179–190.
9. De Chazal, P.; Penzel, T.; Heneghan, C. Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram. *Physiol. Meas.* **2004**, *25*, 967. [[CrossRef](#)]
10. Patil, S.P.; Schneider, H.; Schwartz, A.R.; Smith, P.L. Adult obstructive sleep apnea: Pathophysiology and diagnosis. *Chest J.* **2007**, *132*, 325–337. [[CrossRef](#)]
11. Berry, R.B.; Brooks, R.; Gamaldo, C.E.; Harding, M.; Ioy, R.M.; Marcos, C.; Vaughn, B.V. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*; Version 2.6.0; American Academy of Sleep Medicine: Darien, IL, USA, 2020.
12. Berry, R.B.; Budhiraja, R.; Gottlieb, D.J.; Gozal, D.; Iber, C.; Kapur, V.K.; Marcus, C.L.; Mehra, R.; Parthasarathy, S.; Quan, S.F.; et al. Rules for scoring respiratory events in sleep: Update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. *J. Clin. Sleep Med.* **2012**, *8*, 597–619. [[CrossRef](#)]
13. Agarwal, R.; Gotman, J. Computer-assisted sleep staging. *IEEE Trans. Biomed. Eng.* **2001**, *48*, 1412–1423. [[CrossRef](#)]
14. Flemons, W.W.; Littner, M.R.; Rowley, J.A.; Gay, P.; Anderson, W.M.; Hudgel, D.W.; McEvoy, R.D.; Loube, D.I. Home diagnosis of sleep apnea: A systematic review of the literature: an evidence review cosponsored by the American Academy of Sleep Medicine, the American College of Chest Physicians, and the American Thoracic Society. *CHEST J.* **2003**, *124*, 1543–1579. [[CrossRef](#)]
15. de Almeida, F.R.; Ayas, N.T.; Otsuka, R.; Ueda, H.; Hamilton, P.; Ryan, F.C.; Lowe, A.A. Nasal pressure recordings to detect obstructive sleep apnea. *Sleep Breath.* **2006**, *10*, 62–69. [[CrossRef](#)]
16. Khandoker, A.H.; Gubbi, J.; Palaniswami, M. Automated scoring of obstructive sleep apnea and hypopnea events using short-term electrocardiogram recordings. *Inf. Technol. Biomed. IEEE Trans.* **2009**, *13*, 1057–1067. [[CrossRef](#)] [[PubMed](#)]
17. Whitney, C.W.; Gottlieb, D.J.; Redline, S.; Norman, R.G.; Dodge, R.R.; Shahar, E.; Surovec, S.; Nieto, F.J. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep* **1998**, *21*, 749–757. [[CrossRef](#)]
18. Bennett, J.; Kinnear, W. *Sleep on the Cheap: The Role of Overnight Oximetry in the Diagnosis of Sleep Apnoea Hypopnoea Syndrome*; BMJ Publishing Group Ltd.: London, UK, 1999.
19. Koley, B.; Dey, D. Automated detection of apnea and hypopnea events. In Proceedings of the 2012 Third International Conference on Emerging Applications of Information Technology, Kolkata, India, 30 November–1 December 2012; pp. 85–88.
20. Gutiérrez-Tobal, G.C.; Álvarez, D.; Marcos, J.V.; Del Campo, F.; Hornero, R. Pattern recognition in airflow recordings to assist in the sleep apnoea–hypopnoea syndrome diagnosis. *Med Biol. Eng. Comput.* **2013**, *51*, 1367–1380. [[CrossRef](#)]
21. Gutiérrez-Tobal, G.; Hornero, R.; Álvarez, D.; Marcos, J.; Del Campo, F. Linear and nonlinear analysis of airflow recordings to help in sleep apnoea–hypopnoea syndrome diagnosis. *Physiol. Meas.* **2012**, *33*, 1261. [[CrossRef](#)]
22. Gutiérrez-Tobal, G.C.; Álvarez, D.; del Campo, F.; Hornero, R. Utility of adaboost to detect sleep apnea-hypopnea syndrome from single-channel airflow. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 636–646. [[CrossRef](#)] [[PubMed](#)]
23. Nigro, C.A.; Dibur, E.; Aimaretti, S.; González, S.; Rhodius, E. Comparison of the automatic analysis versus the manual scoring from ApneaLink™ device for the diagnosis of obstructive sleep apnoea syndrome. *Sleep Breath.* **2011**, *15*, 679–686. [[CrossRef](#)] [[PubMed](#)]
24. Nakano, H.; Tanigawa, T.; Furukawa, T.; Nishima, S. Automatic detection of sleep-disordered breathing from a single-channel airflow record. *Eur. Respir. J.* **2007**, *29*, 728–736. [[CrossRef](#)]



25. BaHammam, A.; Sharif, M.; Gacuan, D.E.; George, S. Evaluation of the accuracy of manual and automatic scoring of a single airflow channel in patients with a high probability of obstructive sleep apnea. *Med Sci. Monit. Int. Med J. Exp. Clin. Res.* **2011**, *17*, MT13. [\[CrossRef\]](#)
26. Lin, Y.Y.; Wu, H.T.; Hsu, C.A.; Huang, P.C.; Huang, Y.H.; Lo, Y.L. Sleep apnea detection based on thoracic and abdominal movement signals of wearable piezoelectric bands. *IEEE J. Biomed. Health Inf.* **2016**, *21*, 1533–1545. [\[CrossRef\]](#)
27. Vaughn, C.M.; Clemmons, P. Piezoelectric belts as a method for measuring chest and abdominal movement for obstructive sleep apnea diagnosis. *Neurodiagn. J.* **2012**, *52*, 275–280.
28. Avcı, C.; Akbaş, A. Sleep apnea classification based on respiration signals by using ensemble methods. *Bio-Med Mater. Eng.* **2015**, *26*, S1703–S1710. [\[CrossRef\]](#)
29. Azimi, H.; Gilakjani, S.S.; Bouchard, M.; Goubran, R.A.; Knoefel, F. Automatic apnea-hypopnea events detection using an alternative sensor. In Proceedings of the 2018 IEEE Sensors Applications Symposium (SAS), Seoul, Korea, 12–14 March 2018; pp. 1–5.
30. Penzel, T.; McNames, J.; De Chazal, P.; Raymond, B.; Murray, A.; Moody, G. Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. *Med Biol. Eng. Comput.* **2002**, *40*, 402–407. [\[CrossRef\]](#) [\[PubMed\]](#)
31. De Chazal, P.; Heneghan, C.; Sheridan, E.; Reilly, R.; Nolan, P.; O'Malley, M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 686–696. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Bsoul, M.; Minn, H.; Tamil, L. Apnea MedAssist: Real-time sleep apnea monitor using single-lead ECG. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *15*, 416–427. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Burgos, A.; Goñi, A.; Illarramendi, A.; Bermúdez, J. Real-time detection of apneas on a PDA. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *14*, 995–1002. [\[CrossRef\]](#)
34. Magalang, U.J.; Dmochowski, J.; Veeramachaneni, S.; Draw, A.; Mador, M.J.; El-Solh, A.; Grant, B.J. Prediction of the apnea-hypopnea index from overnight pulse oximetry. *Chest J.* **2003**, *124*, 1694–1701. [\[CrossRef\]](#)
35. Alvarez, D.; Hornero, R.; Marcos, J.V.; del Campo, F. Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2816–2824. [\[CrossRef\]](#)
36. Yadollahi, A.; Giannouli, E.; Moussavi, Z. Sleep apnea monitoring and diagnosis based on pulse oximetry and tracheal sound signals. *Med Biol. Eng. Comput.* **2010**, *48*, 1087–1097. [\[CrossRef\]](#)
37. Xie, B.; Minn, H. Real-time sleep apnea detection by classifier combination. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 469–477. [\[CrossRef\]](#)
38. El Shayeb, M.; Topfer, L.A.; Stafinski, T.; Pawluk, L.; Menon, D. Diagnostic accuracy of level 3 portable sleep tests versus level 1 polysomnography for sleep-disordered breathing: A systematic review and meta-analysis. *Cmaj* **2014**, *186*, E25–E51. [\[CrossRef\]](#)
39. Fontenla-Romero, O.; Guijarro-Berdiñas, B.; Alonso-Betanzos, A.; Moret-Bonillo, V. A new method for sleep apnea classification using wavelets and feedforward neural networks. *Artif. Intell. Med.* **2005**, *34*, 65–76. [\[CrossRef\]](#)
40. Han, J.; Shin, H.B.; Jeong, D.U.; Park, K.S. Detection of apneic events from single channel nasal airflow using 2nd derivative method. *Comput. Methods Programs Biomed.* **2008**, *91*, 199–207. [\[CrossRef\]](#)
41. Selvaraj, N.; Narasimhan, R. Detection of sleep apnea on a per-second basis using respiratory signals. In Proceedings of the Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, Osaka, Japan, 3–7 July 2013; pp. 2124–2127.
42. Ciolek, M.; Niedźwiecki, M.; Sieklicki, S.; Drozdowski, J.; Siebert, J. Automated detection of sleep apnea and hypopnea events based on robust airflow envelope tracking in the presence of breathing artifacts. *IEEE J. Biomed. Health Inf.* **2015**, *19*, 418–429. [\[CrossRef\]](#)
43. Koley, B.L.; Dey, D. Automatic detection of sleep apnea and hypopnea events from single channel measurement of respiration signal employing ensemble binary SVM classifiers. *Measurement* **2013**, *46*, 2082–2092. [\[CrossRef\]](#)
44. Koley, B.L.; Dey, D. Real-time adaptive apnea and hypopnea event detection methodology for portable sleep apnea monitoring devices. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 3354–3363. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Várady, P.; Micsik, T.; Benedek, S.; Benyó, Z. A novel method for the detection of apnea and hypopnea events in respiration signals. *Biomed. Eng. IEEE Trans.* **2002**, *49*, 936–942. [\[CrossRef\]](#)



46. Tian, J.; Liu, J. Apnea detection based on time delay neural network. In Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Shanghai, China, 17–18 January 2005; pp. 2571–2574.
47. Norman, R.G.; Rapoport, D.M.; Ayappa, I. Detection of flow limitation in obstructive sleep apnea with an artificial neural network. *Physiol. Meas.* **2007**, *28*, 1089. [\[CrossRef\]](#)
48. Ozerov, A.; Lagrange, M.; Vincent, E. GMM-based classification from noisy features. In Proceedings of the International Workshop on Machine Listening in Multisource Environments (CHiME), Florence, Italy, 1 September 2011; pp. 30–35.
49. Kim, J.; ElMoaqet, H.; Tilbury, D.M.; Ramachandran, S.K.; Penzel, T. Time domain characterization for sleep apnea in oronasal airflow signal: A dynamic threshold classification approach. *Physiol. Meas.* **2019**, *40*, 054007. [\[CrossRef\]](#)
50. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.
51. Thomas, E.; Temko, A.; Lightbody, G.; Marnane, W.; Boylan, G. Gaussian mixture models for classification of neonatal seizures using EEG. *Physiol. Meas.* **2010**, *31*, 1047. [\[CrossRef\]](#) [\[PubMed\]](#)
52. ElMoaqet, H.; Tilbury, D.M.; Ramachandran, S.K. Predicting oxygen saturation levels in blood using autoregressive models: A threshold metric for evaluating predictive models. In Proceedings of the 2013 American Control Conference, Washington, DC, USA, 17–19 June 2013; pp. 734–739.
53. ElMoaqet, H.; Tilbury, D.M.; Ramachandran, S.K. Evaluating predictions of critical oxygen desaturation events. *Physiol. Meas.* **2014**, *35*, 639. [\[CrossRef\]](#)
54. ElMoaqet, H.; Tilbury, D.M.; Ramachandran, S.K. Multi-step ahead predictions for critical levels in physiological time series. *IEEE Trans. Cybern.* **2016**, *46*, 1704–1714. [\[CrossRef\]](#)
55. Zweig, M.H.; Campbell, G. Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clin. Chem.* **1993**, *39*, 561–577. [\[CrossRef\]](#)
56. Koley, B.; Dey, D. Adaptive classification system for real-time detection of apnea and hypopnea events. In Proceedings of the 2013 IEEE Point-of-Care Healthcare Technologies (PHT), Bangalore, India, 16–18 January 2013; pp. 42–45.
57. Canadian Agency for Drugs and Technologies in Health (CADTH). Portable monitoring devices for diagnosis of obstructive sleep apnea at home: Review of Accuracy, cost-effectiveness, guidelines, and coverage in Canada. *CADTH Technol. Overviews* **2010**, *1*.
58. Collop, N.A.; Anderson, W.M.; Boehlecke, B.; Claman, D.; Goldberg, R.; Gottlieb, D.J.; Hudgel, D.; Sateia, M.; Schwab, R. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. *J. Clin. Sleep. Med.* **2007**, *3*, 737–747.
59. Baron, K.G.; Duffecy, J.; Berendsen, M.A.; Mason, I.C.; Lattie, E.G.; Manalo, N.C. Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. *Sleep Med. Rev.* **2018**, *40*, 151–159. [\[CrossRef\]](#)
60. Ruehland, W.R.; Rochford, P.D.; O'Donoghue, F.J.; Pierce, R.J.; Singh, P.; Thornton, A.T. The new AASM criteria for scoring hypopneas: Impact on the apnea hypopnea index. *Sleep* **2009**, *32*, 150. [\[CrossRef\]](#)

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).