Received: 27 February 2023 | Revised: 18 June 2023 | Accepted: 23 June 2023

DOI: 10.1111/jsr.13991

RESEARCH ARTICLE

Detection of preceding sleep apnea using ECG spectrogram during CPAP titration night: A novel machine-learning and bag-of-features framework

Tran Thanh Duy Linh[1,2] | Nguyen Thi Hoang Trang[3] | Shang-Yang Lin[4] | Dean Wu[5,6,7] | Wen-Te Liu[7,8,9,10] | Chaur-Jong Hu[5,6]

[1]International Ph.D. Program of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[2]Family Medicine Training Center, University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City, Vietnam

[3]Department of Biomedical Engineering, College of Engineering, National Cheng Kung University, Tainan, Taiwan

[4]Research Center of Sleep Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[5]Department of Neurology, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan

[6]Department of Neurology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[7]Sleep Center, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan

[8]Division of Pulmonary Medicine, Department of Internal Medicine, Shuang Ho Hospital, Taipei Medical University, New Taipei City, Taiwan

[9]School of Respiratory Therapy, College of Medicine, Taipei Medical University, Taipei, Taiwan

[10]Research Center of Artificial Intelligence in Medicine, Taipei Medical University, Taipei, Taiwan

Correspondence
Wen-Te Liu, Sleep Center, Shuang Ho Hospital, Taipei Medical University, New Taipei City 235, Taiwan.
Email: lion5835@gmail.com

Chaur-Jong Hu, Department of Neurology, Shuang Ho Hospital, Taipei Medical University, New Taipei City 235, Taiwan.
Email: chaurjongh@tmu.edu.tw

Summary

Obstructive sleep apnea (OSA) has a heavy health-related burden on patients and the healthcare system. Continuous positive airway pressure (CPAP) is effective in treating OSA, but adherence to it is often inadequate. A promising solution is to detect sleep apnea events in advance, and to adjust the pressure accordingly, which could improve the long-term use of CPAP treatment. The use of CPAP titration data may reflect a similar response of patients to therapy at home. Our study aimed to develop a machine-learning algorithm using retrospective electrocardiogram (ECG) data and CPAP titration to forecast sleep apnea events before they happen. We employed a support vector machine (SVM), k-nearest neighbour (KNN), decision tree (DT), and linear discriminative analysis (LDA) to detect sleep apnea events 30–90 s in advance. Preprocessed 30 s segments were time–frequency transformed to spectrograms using continuous wavelet transform, followed by feature generation using the bag-of-features technique. Specific frequency bands of 0.5–50 Hz, 0.8–10 Hz, and 8–50 Hz were also extracted to detect the most detected band. Our results indicated that SVM outperformed KNN, LDA, and DT across frequency bands and leading time segments. The 8–50 Hz frequency band gave the best accuracy of 98.2%, and a F1-score of 0.93. Segments 60 s before sleep events seemed to exhibit better performance than other pre-OSA segments. Our findings demonstrate the feasibility of detecting sleep apnea events in advance using only a single-lead ECG signal at CPAP titration, making our proposed framework a novel and promising approach to managing obstructive sleep apnea at home.

KEYWORDS
adherence, bag-of-features, machine learning, sleep apnea, spectrogram, time-frequency transforms

1 | INTRODUCTION

Obstructive sleep apnea (OSA) is a highly prevalent respiratory sleep disorder (Benjafield et al., 2019), characterised by recurrent upper airway collapse during sleep (Malhotra & White, 2002), leading to poor sleep quality, excessive daytime sleepiness, increased risk of traffic or work-related accidents, and long-term health-related complications (Huang et al., 2021; Legault et al., 2021; X. Wang et al., 2013). Continuous positive airway pressure (CPAP) is a standard treatment for moderate to severe OSA, which applies a fixed positive pressure to keep the upper airway open. The pressure adjustment is managed by a sleep technician during the CPAP titration night, based on patient sleep data, including respiratory patterns, and oxygen levels. However, adherence to CPAP is low (Contal et al., 2021; Pack et al., 2021), partly due to suboptimal pressure management, which can result in patients experiencing residual apneic or hypopneic episodes. These episodes cause patient discomfort, and negatively impact sleep quality, contributing to a vicious cycle of poor treatment adherence.

To address this issue, the current study aims to develop a novel machine-learning model for the early detection of sleep apnea events using single-lead electrocardiogram (ECG) signals. Inspired by the work of De Falco (Falco et al., 2015) and Huseyin Nasifoglu and Osman Erogul (Nasifoglu & Eroğul, 2021), we hypothesised that the changes in respiratory patterns could impact the autonomic nervous activity even before actual respiratory events occur, leading to alterations in ECG signals. ECG signals present a more convenient, less intrusive, and widely available alternative to respiratory flow trace across various clinical or home-based settings. By capitalising on these advantages, ECG signals offer a practical and easily accessible method to capture these pre-event alterations, making them an ideal choice for our study.

In this study, we used ECG-derived spectrograms as inputs to detect apneas and hypopneas 30–90 s in advance. Spectrograms allow for a comprehensive understanding of heart signals, capturing both time and frequency domain information, which can potentially offer early warning signs for sleep apnea events. Furthermore, we decided to use data collected during CPAP titration, as it reflects patients' responses to PAP treatment at home and changes in automatic activity under PAP pressure (W. Guo et al., 2018). By employing a machine-learning approach and a bag-of-features technique, we aimed to improve the accuracy of detecting preceding respiratory events and increase adherence to PAP treatment in patients with OSA. This innovative approach expands on the success of previous studies, that used various methods and signals to forecast sleep apnea events in advance, namely the models of Dagum and Galper (Dagum & Galper, 1995), Bock and Gough (Bock & Gough, 1998), and Waxman and his colleagues (Waxman et al., 2010), Maali and Al-Jumailey (Maali & Al-Jumaily, 2013).

2 | MATERIALS AND METHODS

2.1 | Database

This retrospective cross-sectional study utilised the dataset of PSG recording during CPAP titration nights obtained from Sleep Center, Shuangho hospital. With the approval of the Institutional Review Board of Taipei Medical University (TMU-JIRB No: N201911007), we collected a subset of the database from February to November 2022, including subjects with a diagnosis of obstructive sleep apnea and an indication for CPAP treatment. We excluded subjects having drop-out ECG data (<80% of signal available to analyse), atrial fibration, and pacemakers (Eiseman et al., 2012). In total, 71 sets of PSG were collected, and 71 continuous single-lead ECGs were extracted with annotations for apnea, hypopnea, and normal breathing labelled by sleep specialists. Obstructive apnea was defined as a reduction of airflow by equal to or more than 90% over 10 s with respiratory–effort chest wall movement, while hypopnea was defined as a decrease in baseline airflow of 30% or more, combined with either arousal in electromyogram signal for ≥3 s or oxygen desaturation ≥3% (Berry et al., 2015). The dataset included 71 subjects with a mean age of 52.9 ± 13.8 years old (range 29–90); a mean BMI of 27.3 ± 4.1 kg/cm^2; a mean total recording time of 6.2 ± 0.3 h; a mean total sleep time of 4.9 ± 0.9 h; and a mean AHI of 4.8 ± 4.0 events/hour with a mean CPAP pressure of 8.6 ± 2.9 mmHg.

The dataset was split into 80% for training and 20% for testing, and 5-fold cross-validation was performed on the training dataset to evaluate the fitting of the models.

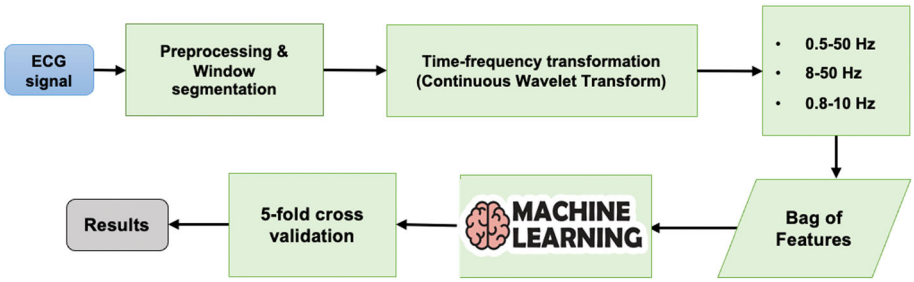
2.2 | Proposed framework

In this study, we developed a framework to forecast sleep apnea in advance using machine learning and the bag-of-features (BoF) technique. The method contains several stages, explained in detail below and in Figure 1.

2.3 | Data preprocessing

First, we preprocessed the single-lead ECG signals by filtering out noises and artefacts, followed by z-score normalisation. We then divided the standardised data into non-overlapping 30 s segments. Based on nasal pressure annotation of obstructive sleep apnea or

FIGURE 1 Proposed framework.



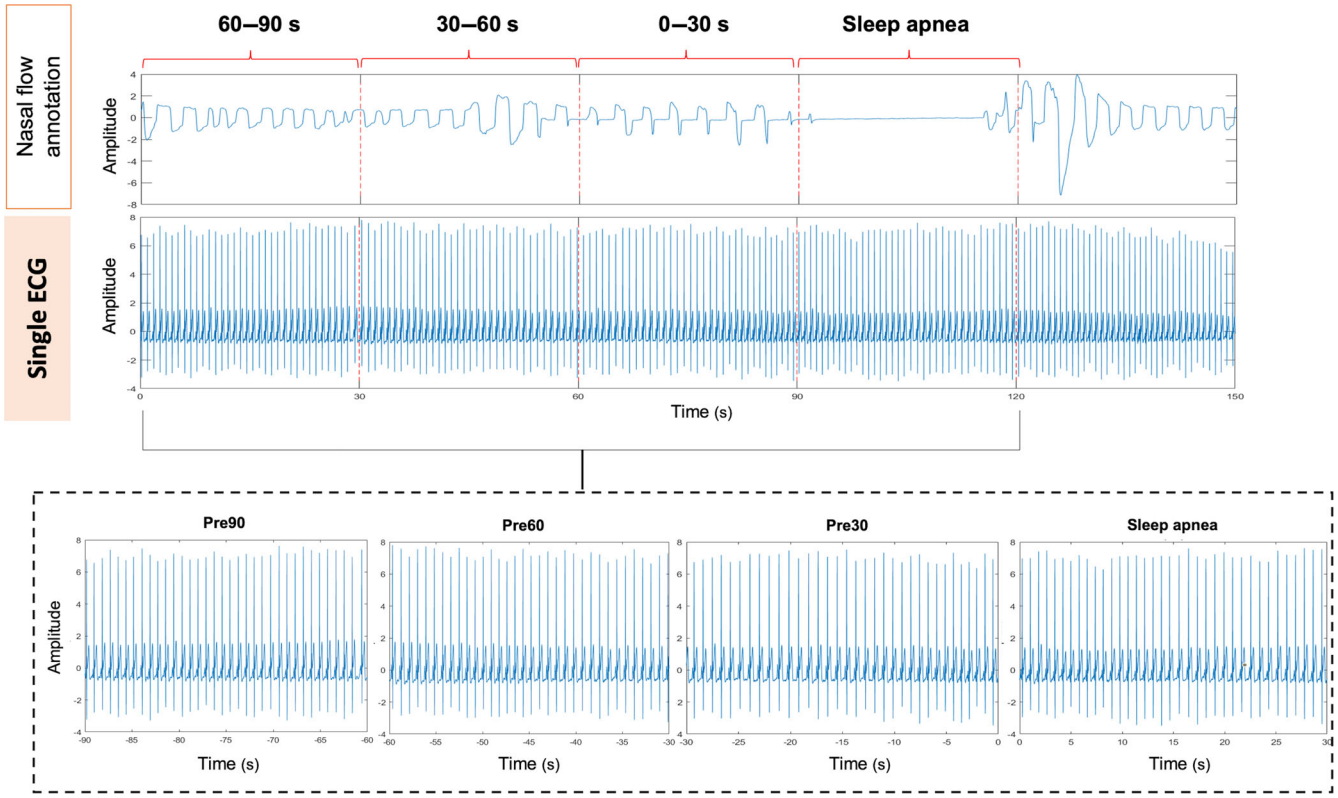


FIGURE 2 Feature preprocessing and segmentation.

hypopnea, we defined pre-30, pre-60, pre-90, and normal breathing segments. Pre-OSA segments, including pre-30, pre-60, and pre-90, are defined as an interval occurring 30, 60 and 90 s immediately before an obstructive sleep apnea or hypopnea event, respectively. These pre-OSA segments did not include other types of sleep apnea such as central or mixed apnea, and arousals. The duration of a 30 s segment was guided by Waxman's study, which indicated that a 30 s segment provided the best performance in detecting pre-OSA events, and a more extended duration resulted in decreased performance (Waxman et al., 2010) (Figure 2).

After segmentation, we collected normal breathing segments accounting for 94.3% of the database, while pre-30, pre-60, and pre-90 segments accounted for only 2.3%, 1.9%, and 1.5%, alternatively. Due to the imbalance in the numbers of these pre-OSA segments compared with normal breathing segments (1:40 ratio), we addressed this issue by using the synthetic minority oversampling technique (SMOTE) to create additional samples for minority class (pre-OSA), making the dataset more balanced (Figure 3). Technically, SMOTE randomly selects k-nearest neighbours for each minority sample, draws a line segment between the sample and its selected neighbours, and generates a synthetic sample along the line (Chawla et al., 2002). This step is crucial in ensuring that our machine learning models are not biased toward the majority class (normal breathing). We opted for SMOTE over other techniques such as geometric transformation since SMOTE is designed to work with time-series data and to maintain temporal patterns while generating artificial samples.

2.4 | Time-frequency transformation

We transformed 30 s time-domain ECG segments into time–frequency domain spectrograms using continuous wavelet transform (CWT). This step allowed us to capture both time and frequency information from ECG signals which are essential in identifying early

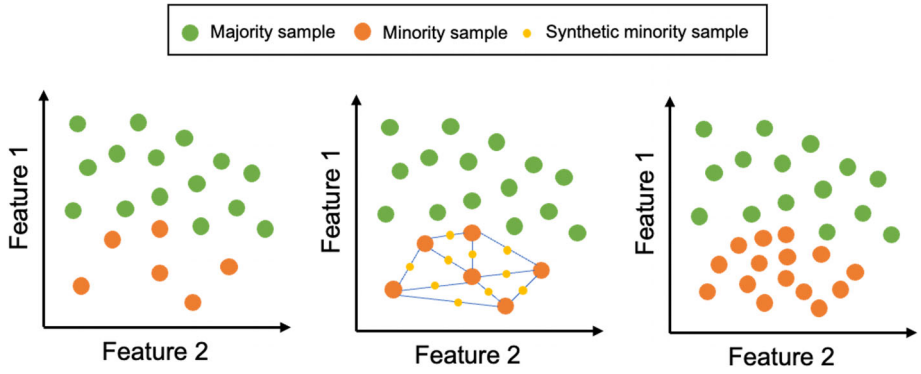


FIGURE 3 Synthetic minority oversampling technique.

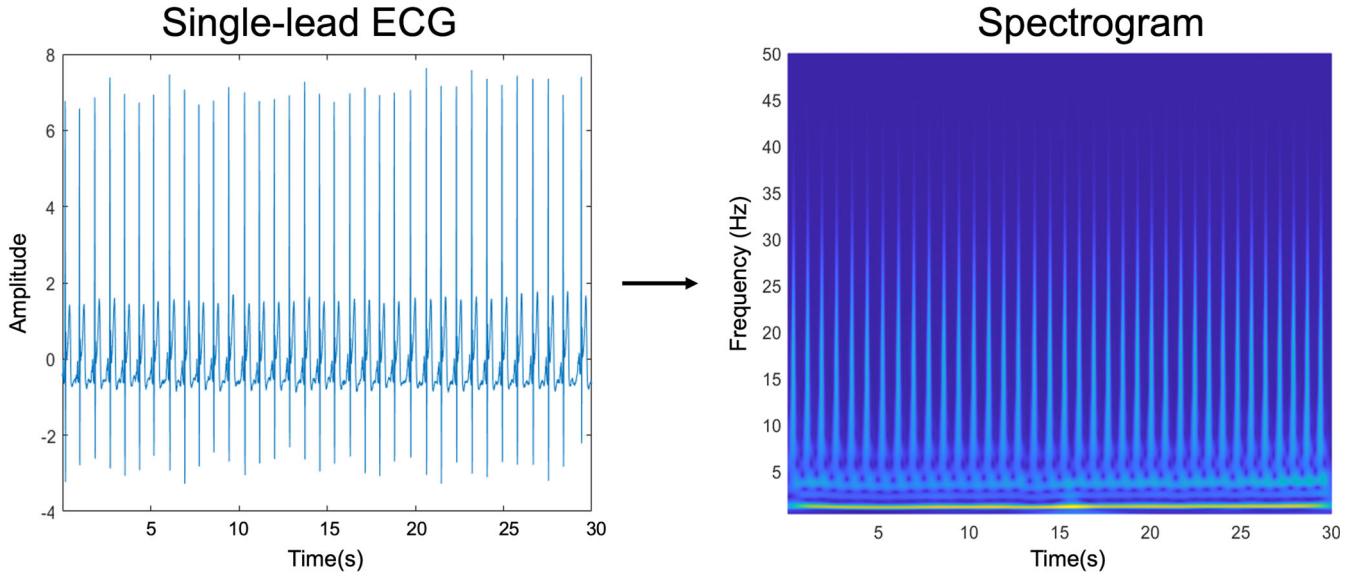


FIGURE 4 Spectrogram generation via time–frequency transform.

warning of respiratory events. The CWT employs a family of wavelet functions that can be scaled and shifted in time length to fit the resolution of signals in different frequency ranges. CWT can be presented by Equation (1):

$$\mathrm{CWT}(a,\tau)\frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(t)\times\psi\left(\frac{t-\tau}{a}\right) \quad (1)$$

In which, a is a scale factor, τ is a shift factor, $f(t)$ represents ECG segment, t is time variable, and ψ denotes the mother wavelet (Addison, 2005). Morlet wavelet was used in this study.

To achieve high temporal resolution visualisation, spectrograms were extracted in three frequency bands: 0.5–50 Hz, 8–50 Hz, and 0.8–10 Hz, corresponding to the frequency ranges of the overall ECG waveform, QRS complex, and T wave, respectively (Tereshchenko & Josephson, 2015). Some studies suggested the frequency band 0.01–0.1 Hz for periodic respiration during sleep disorder breathing and 0.1–0.4 Hz for physiology sinus arrhythmia (D. Guo et al., 2011; Thomas et al., 2005). Still, we ignored these bands because they did not provide good resolution for ECG spectrograms in pre-OSA detection (Lin et al., 2022) (Figure 4).

2.5 | Feature extraction

After obtaining the spectrogram, we utilised the bag-of-features (BoF) to extract features. The BoF is a robust technique identifying key visual patterns that were shared across different spectrograms, even in the presence of noises or variability. This process involved treating ECG spectrograms as images, and extracting key points and descriptors using the speeded up robust feature (SURF) tool. These features were then clustered by a k-means clustering algorithm and quantised to form a visual vocabulary, with each centroid cluster representing a visual word. By clustering the features into a visual word, we could identify common patterns of structures in the spectrogram that corresponds to specific pre-OSA or normal segments.

During testing, features from a new spectrogram were assigned to their nearest neighbour word in the visual vocabulary, and the resulting image was represented as "a bag of features". This bag of features described the spectrogram by quantifying the occurrence of visual words. The frequency of the visual words describing the image was encoded to a frequency histogram which was used as input for machine learning classifiers or sleep apnea detection algorithms (O'Hara & Draper, 2011) (Figure 5).

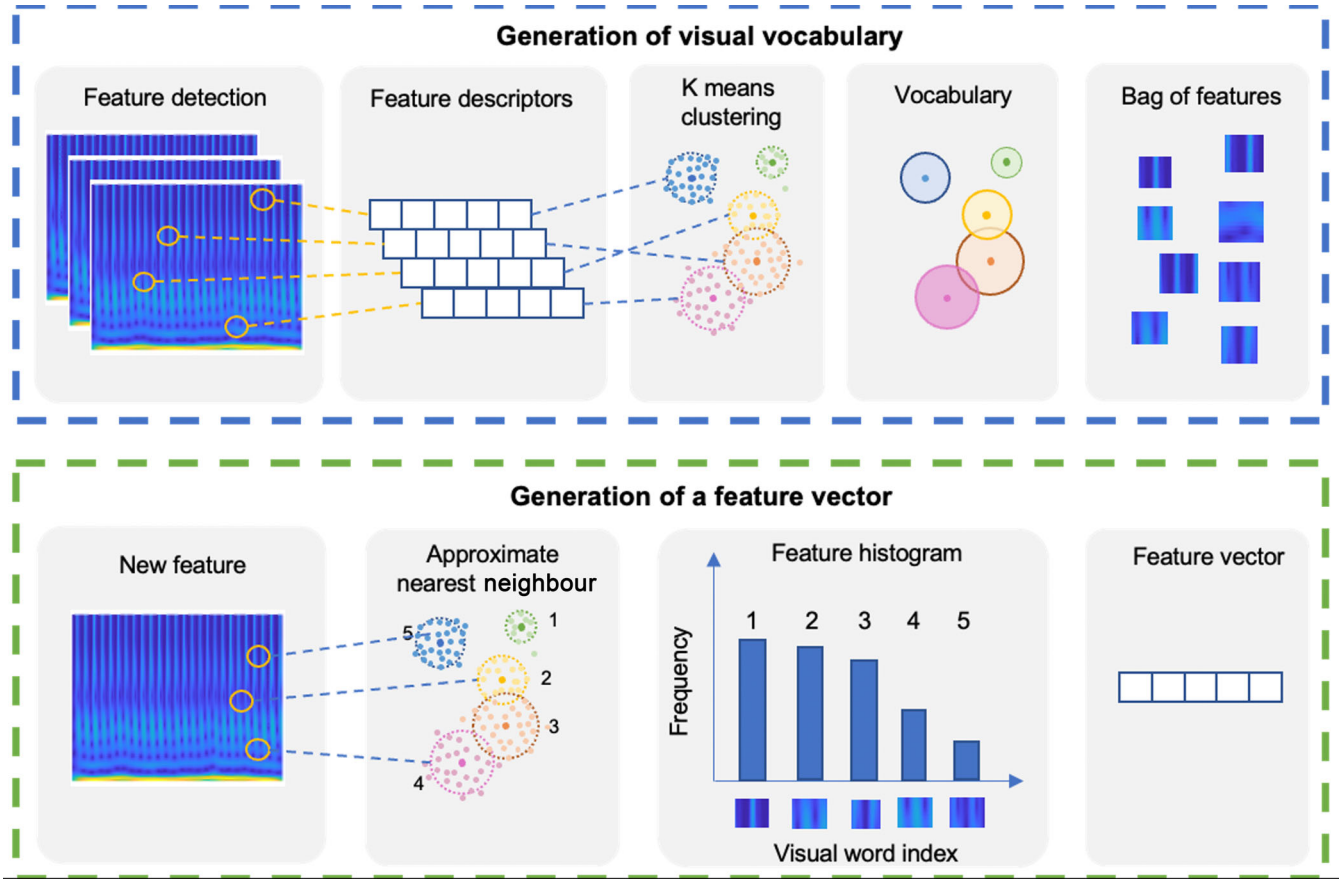


FIGURE 5 Feature extraction via bag-of-features method.

2.6 | Machine learning algorithms

For detecting sleep apnea events in advance, we chose four commonly used machine learning algorithms: support vector machine (SVM); k-nearest neighbour (K-NN); decision tree (DT); and linear discriminant analysis (LDA). These models were chosen due to their varying performance and complexity, which enables us to compare their effectiveness in the detection of sleep apnea.

The support vector machine is a popular classification algorithm that aims to find the optimal hyperplane, maximally separating data into two classes (Hearst et al., 1998). We used a polynomial kernel and a C value of 1 to control the trade-off between the correct classification of training features and the smoothness of the decision boundary.

The k-nearest neighbour is a simple supervised learning algorithm that identifies the k-nearest neighbours of a new given feature based on the distance measure. KNN then assigns that feature to a class by voting on which neighbours are closest to it. In this study, we chose k of 10.

The decision tree is a decision-making algorithm that uses a tree-like structure to make decisions based on the known conditions of features. The process starts from the tree's root, moving through branches until reaching a leaf node, providing an answer for the detection. For DT, we set the max. depth to 100, which allowed the model to capture more complex relationships in the data.

TABLE 1 Confusion matrix and evaluation metrics.

Confusion matrix		Actual class A	Actual class B
Detected class	A	TP	FP
	B	FN	TN
Total		TP + FN	FP + TN

Abbreviations: FN, false-negative value; FP, false-positive value; TN, true-negative value; TP, true-positive value.

Finally, linear discriminant analysis is a robust tool that classifies and reduces the dimensions of data points for machine learning. It creates a low-dimensional space and draws a hyperplane to separate classes while minimising the variance within each class. We set the gamma value to 0 to use the default for this hypermeter.

2.7 | Performance evaluation

We validated the performance of selected algorithms using k-fold cross-validation. The dataset was divided into k-folds of equal size,

where one-fold data were randomly selected as a testing dataset, while the remaining (k−1) folds were used for training. The process was repeated k times, and k different accuracies were collected. The absolute accuracy was calculated as a mean of those k accuracy values where k was set to 5 in this study.

We computed accuracy (Acc), sensitivity (Sen), precision (Pre), and F1-score to compare the performance of the models. F1-score is a weighted average of precision and sensitivity, where the best value is 1 and the worst is 0. Receiver operating characteristics (ROC) curves were also plotted to compare the performance among classifiers and frequency bands (Table 1).

$$\mathrm{Acc} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FN}+\mathrm{TN}+\mathrm{FP}}.$$

$$\mathrm{Sen} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

$$\mathrm{F1\text{-}score} = \frac{2ñ\mathrm{Precision}ñ\mathrm{Sensitivity}}{\mathrm{Precision}+\mathrm{Sensitivity}} = \frac{2ñ\mathrm{TP}}{2ñ\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$$

$$\text{with } \mathrm{Pre} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

For this study, the preprocessing and machine learning implementation were conducted using Matlab (R2020a) environment on a computer with 12th Gen Intel® Core™ i7-12700H 2.30 Hz, 40GB RAM, NVIDIA GeForce GTX 3050 8GB, and Windows 10 64-bit operation system.

TABLE 2 The numbers of pre-OSA and normal breathing segments before and after SMOTE resampling.

Frequency	Type	Primary[a]	SMOTE	Broken[b]	Total
0.5–50 Hz	Pre-30	1306	10,000	57	11,249
	Pre-60	1004	10,000	50	10,954
	Pre-90	846	10,000	56	10,790
	Normal	49,968	0	4132	45,836
8.0–50 Hz	Pre-30	1306	10,000	186	11,120
	Pre-60	1004	10,000	86	10,918
	Pre-90	846	10,000	86	10,760
	Normal	49,968	0	14,967	35,001
0.8–10 Hz	Pre-30	1306	10,000	54	11,252
	Pre-60	1004	10,000	67	10,937
	Pre-90	846	10,000	40	10,806
	Normal	49,968	0	4053	45,915

Abbreviations: Pre-30, Pre-60, and Pre-90 = 30 s segments prior to sleep apnea events of 30, 60, and 90 s, alternatively.
[a]The number of segments after filtering and segmentation.
[b]The number of spectrograms broken after time–frequency transformation via CWT.

3 | RESULTS

In this study, we generated ECG spectrograms from ECG recordings obtained during the CPAP titration night. We then applied machine learning algorithms to features obtained from ECG spectrograms to detect pre-OSA events and evaluated the performance of four models: SVM, KNN, DT, and LDA.

TABLE 3 Five-fold cross-validation detection performance.

	Accuracy (%)			Sensitivity (%)			Precision (%)			F1-score		
Frequency band	Pre-30	Pre-60	Pre-90	Pre-30	Pre-60	Pre-90	Pre-30	Pre-60	Pre-90	Pre-30	Pre-60	Pre-90
KNN												
0.5–50 Hz	95.8	94.7	94.4	80.9	86.9	71.8	87.46	77.72	86.11	0.85	0.83	0.79
8–50 Hz	95.0	95.0	**95.9**	80.1	88.3	80.2	86.16	77.70	79.37	0.82	0.84	**0.85**
0.8–10 Hz	95.6	95.3	92.5	79.6	**89.9**	66.5	**89.11**	80.89	76.58	0.84	0.85	0.68
SVM												
0.5–50 Hz	96.9	96.7	97.8	81.5	80.9	87.6	96.11	96.77	97.71	0.88	0.88	0.92
8–50 Hz	97.5	**98.2**	97.6	83.8	87.7	86.8	98.59	**99.34**	99.23	0.91	**0.93**	0.91
0.8–10 Hz	96.4	98.1	93.0	79.0	**90.9**	57.7	95.40	96.99	90.61	0.87	0.93	0.71
DT												
0.5–50 Hz	94.5	94.8	95.6	75.5	67.2	81.1	84.0	94.99	96.37	0.80	0.79	0.84
8–50 Hz	97.3	**97.8**	97.3	84.5	**87.9**	84.7	95.81	**97.46**	96.81	0.90	**0.92**	0.90
0.8–10 Hz	91.0	90.3	87.0	47.5	40.9	20.1	84.36	81.40	69.12	0.61	0.55	0.31
LDA												
0.5–50 Hz	94.7	93.6	94.8	69.7	61.6	71.5	90.47	92.27	94.18	0.79	0.74	0.80
8–50 Hz	96.9	**97.9**	95.1	79.7	**85.3**	77.4	98.93	**98.98**	98.98	0.88	**0.92**	0.82
0.8–10 Hz	93.0	93.1	87.7	59.6	66.6	30.8	88.20	83.46	68.47	0.71	0.74	0.42

Note: The bold values highlight the best performance of the machine learning classifiers, rather than statistical significance.

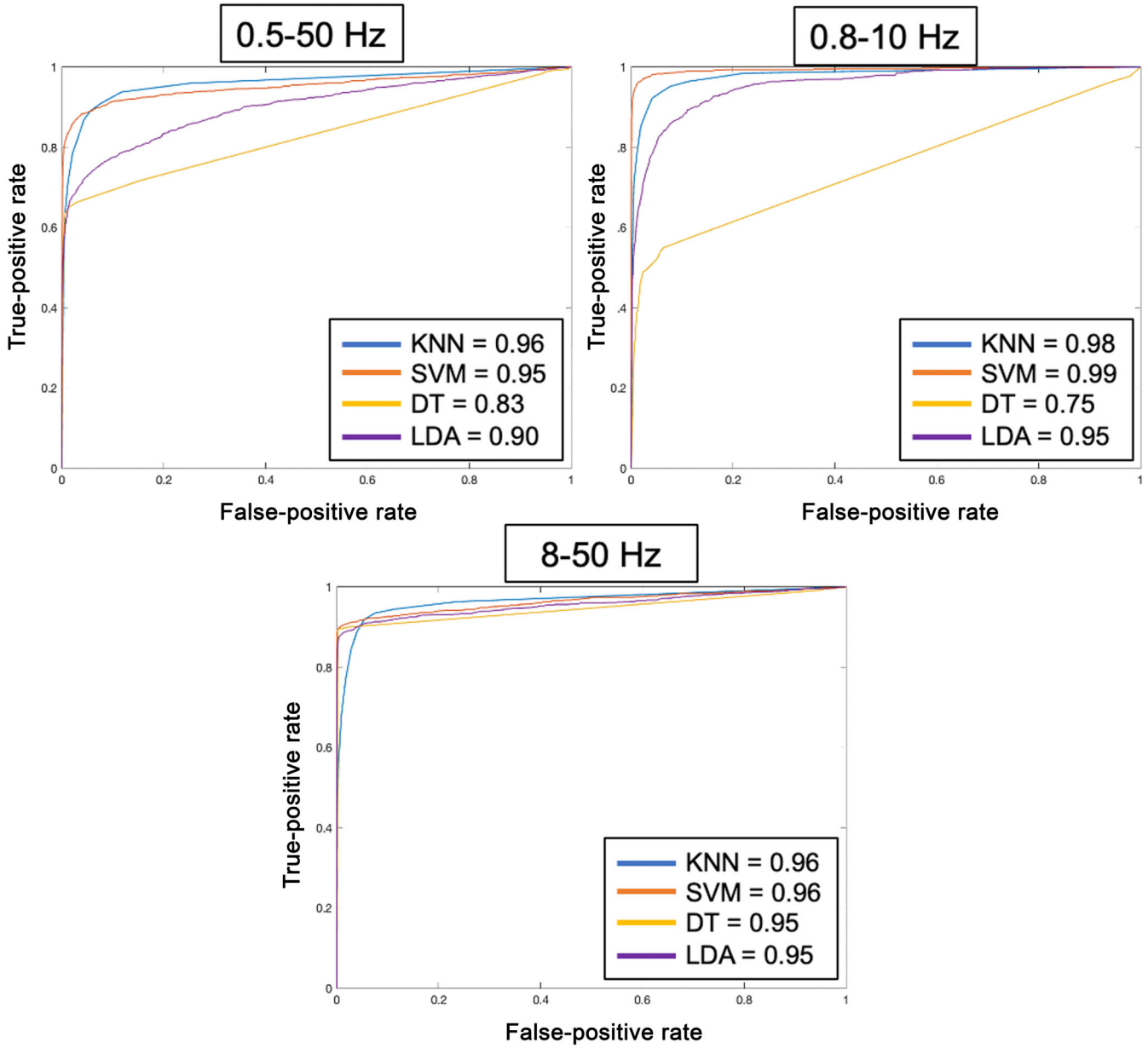


FIGURE 6 ROC curve of detection performance using the segments 30–60 s before sleep apnea at three frequency bands.

While addressing the dataset imbalance using SMOTE technique, we encountered considerable noises in the ECG spectrograms of artificially imputed pre-OSA segments. As a solution, we added 10,000 segments to the numbers of each pre-OSA group, resulting in an acceptable level of noises. Following the time–frequency transformation of ECG signals, we removed some broken spectrograms and presented the final numbers of each pre-OSA segment at different frequency bands in Table 2.

In terms of classifier performance, the SVM model outperformed the other models with the highest accuracy and F1-score across all three frequency bands and three pre-OSA segments. For instance, at the 8–50 Hz frequency band, SVM obtained its best accuracy of 98.2% and F1-score of 0.93, which is higher than the other models. On the other hand, the KNN model achieved the lowest performance across all frequency bands and pre-OSA segments. The performance of the DT and LDA models varied across all pre-OSA segments or frequency bands. At the 0.8–50 Hz band, DT achieved an accuracy of 94.5%, while LDA achieved 94.7%. However, at the 8–50 Hz band, LDA achieved a higher accuracy (97.9%) than DT (97.8%). Additionally, at the 0.8–10 Hz band, LDA attained the highest sensitivity (90.9%) among all the models, while DT had the lowest F1-score (0.61) (Table 3 and Figure 6)

Among the frequency bands, 8–50 Hz frequency provided the best performance, while the 0.8–10 Hz band had the worst. At the 8–50 Hz band, all the models achieved an accuracy of at least 95%, while at the 0.8–10 Hz band, the accuracy of all models was mostly below 95%.

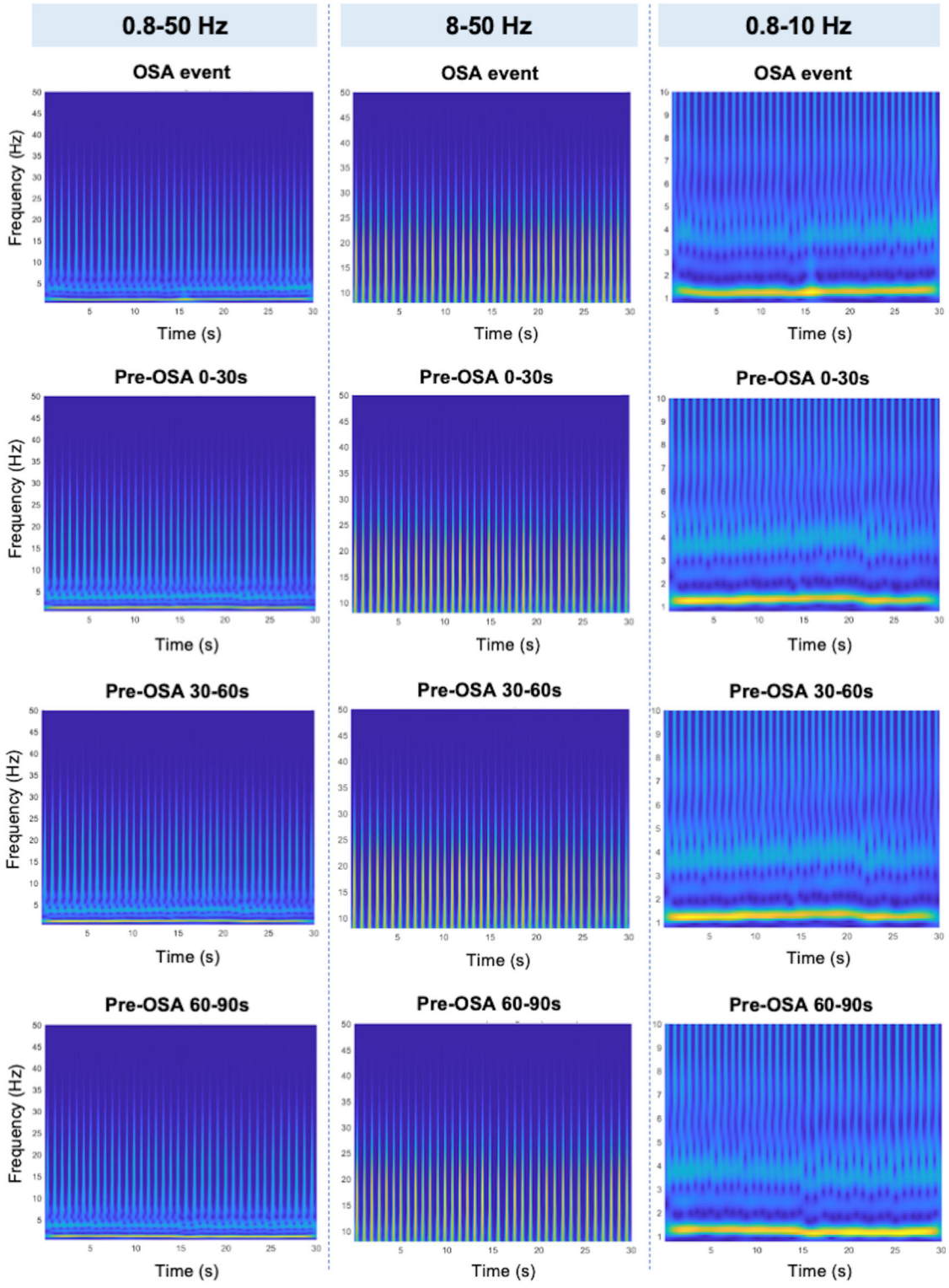


FIGURE 7 The spectrograms for sleep apnea event and 30 s proceeding segments.

Moreover, the pre-60 segments performed better than pre-30 and pre-90, particularly at the 8–50 Hz frequency band. For instance, at this frequency band, the SVM model achieved an accuracy of 98.2% and an F1-score of 0.93 for the pre-60 segment, overperforming the pre-30 and pre-90 segments were lower.

4 | DISCUSSION

Our study is the first to examine the performance of the pre-OSA detection algorithms under the effect of CPAP, using PSG data from the CPAP titration night to generate ECG spectrograms. Despite the

smaller number of respiratory events detected in our study, which resulted in a significant imbalance between normal breathing and pre-OSA segments, our proposed framework achieved acceptable performance. Specifically, the SVM model outperformed with the highest accuracy and F1-score across all three frequency bands and three pre-OSA segments. The KNN model achieved the lowest performance of the other two models, LDA and DT, which varied across different frequency bands and pre-OSA segments.

An intriguing observation in our study pertains to the performance differences among segments at different time points before respiratory events. As changes in cardiac activity are expected to be more profound during respiratory events than in lead time duration, it is reasonable that pre-90 segments performed worst among the pre-OSA segments, as they were furthest from the respiratory events. Interestingly, we found that pre-60 segments seemed to perform better than pre-30 segments, contrary to our expectation that the performance would improve as segments got closer to a respiratory event as in other studies (Falco et al., 2015; Nasifoglu & Eroğul, 2021; Waxman et al., 2010). A possible explanation for this observation is that pre-30 segments lay under CPAP effects in normalising the patient's respiration, as CPAP pressure opens the upper airway and normalises the cardiac activity (Karasulu et al., 2010; Kufoy et al., 2012). Moreover, it is possible that ECG-derived features used in our model capture discriminative patterns of the autonomic nervous system that occurs 60 s prior to sleep apnea events, which allowed for better classification performance than pre-30 segments. However, it is important to note that the differences in performance among pre-OSA segments were not statistically significant. The underlying reasons for this better performance are still unclear and require more investigation to confirm the finding.

Additionally, our findings suggest that the 8–50 Hz frequency band showed the best performance, while the 0.8–10 Hz frequency band had the worst performance, likely due to the noises in the lower frequency range (Figure 7). The 8–50 Hz band provided a clearer visualisation with a brighter region representing higher energy. This region corresponds to the R wave, providing crucial ECG variation information to detect specific events in cardiovascular diseases or sleep-related respiratory conditions. Meanwhile, the 0.5–50 Hz band, including two other frequency bands, may be affected by noises from the low frequency and consume more computational cost for drawing spectrograms due to covering a larger range. Therefore, the 8–50 Hz band is a potential selection for developing a simple classifier using minimum features.

It is also important to consider that sleep stages can impact ECG signals due to changes in the autonomic nervous system during sleep, which may, in turn, affect detection models. In OSA, autonomic activity is disrupted by chronic hypoxia, leading to increased parasympathetic modulation during non-rapid eye movement sleep (NREM), and reverse to less parasympathetic modulation during rapid eye movement sleep (REM) (Kesek et al., 2009; Liao et al., 2010). CPAP may help to correct this modulation (Karasulu et al., 2010). To assess the

TABLE 4 NREM and REM detection performance based on SVM classifier and 8–50 Hz frequency band

Sleep stage	Accuracy %	Sensitivity %	Precision %	F1-score
NREM	97.07	75.67	91.54	0.83
REM	86.13	57.32	89.32	0.70

Abbreviations: NREM, non-rapid eye movement; REM, rapid eye movement.

TABLE 5 Proposed algorithm comparison with state-of-art algorithms.

Authors	Segment windows	Lead time	Signal	Signal transform	Classifiers	Performance results
(Waxman et al., 2010)	30, 60, 90, & 120 s	30, 60, 90, or 120 s	EEG, ECG, EMG, EOG, nasal press, oronasal temperature	DWT	ANN	Sn = 80.6 Sp = 72.8
(Maali & Al-Jumaily, 2013)	30, 60, 90, & 120 s	30, 60, 90, or 120 s	Airflow, abdominal & thoracic movement	Wavelet transform	ANN	AUC = 0.866
(Falco et al., 2015)	1 min	1, 2, or 3 min	ECG	HRV	IF-THEN rules	Sn = 84.2 Sp = 84.7
(Ozdemir et al., 2016)	40 s	40 s	Nasal flow	Time series	SVM, KNN, LR	Acc = 86.9 Sn = 95.5
(Nasifoglu & Eroğul, 2021)	30 s	30, 60, or 90 s	ECG	CWT & STFT	CNN	Acc = 82.3 Sn = 83.2 Sp = 82.2
(Wang et al., 2021)	30 s	1 min	Snore sound	Acoustic analysis	XGBoost	Acc = 92 Sn = 91 Sp = 90

Abbreviations: Acc, accuracy; ANN, artificial neural network; AUC, area under the curve; CNN, convolutional neural network; DWT, discrete wavelet transform; ECG, electrocardiogram; EEG, electroencephalogram; EMG, electromyogram; EOG, electrooculogram; KNN, k-nearest neighbour; LR, linear regression; Sn, sensitivity; Sp, specificity; STFT, short-time Fourier transform; SVM, support vector machine.

influence of sleep stages on sleep apnea detection performance, we additionally partitioned ECG signals into NREM and REM sleep stages, generated spectrograms, and compared the performance of detection models for each group. The results revealed subtle differences in performance between NREM and REM sleep, particularly a reduced sensitivity during REM sleep compared with NREM sleep, as shown in Table 4. Regardless of this, acceptable performance was achieved across both sleep stages, with F1-scores of 0.83 for NREM and 0.70 for REM sleep. Despite the physiological differences between NREM and REM sleep affecting ECG signals, our analysis indicated that these differences do not significantly impact sleep apnea detection performance. This suggests that our proposed model is robust enough to handle variations in ECG signals across sleep stages and provide reliable OSA detection.

There are many resampling techniques to solve the imbalance, such as geometric transformation, but we favoured SMOTE, a commonly used tool for time series data. However, SMOTE also has its drawback, as in our case, inducing additional noises in the resampling dataset. It is due to SMOTE ignoring the possibility that neighbour samples could come from other classes. Choosing a different resampling ratio may overcome this SMOTE problem and improve the classification performance (Kraiem et al., 2021).

The comparison between our proposed pre-OSA detection algorithm and those in other studies is presented in Table 5. Our study used ECG signals from CPAP titration, while others elicited signals from baseline PSG. Using CPAP titration data presents a challenge due to fewer apnea and hypopnea events, leading to a significant imbalance between pre-OSA and normal breathing segments (1:40 ratio in our case). We addressed this issue with SMOTE, similar to Huseyin Nasifoglu's study, which had a 1:5 imbalance ratio in their data. We found two studies also used single-lead ECG for detection models, while some studies used a set of signals (Maali & Al-Jumaily, 2013; Waxman et al., 2010) or single respiratory signals such as nasal flow (Ozdemir et al., 2016), or snore sound (B. Wang et al., 2021). Despite reflecting indirect respiratory pattern changes, single-lead ECG signals are widely available and less intrusive in nature. Given the acceptable performance in previous studies and ours, the models using ECG provide an accessible and feasible option for sleep apnea detection in a real-world setting.

Our study has some limitations. Firstly, we acknowledge the challenges that may occur with non-airway OSA endotypes, such as high loop gain, low arousal threshold, and those with the co-morbidities of insomnia and OSA (COMISA). Those sleep and wake states are more freely admixed. These conditions can lead to biological oscillations in autonomic nervous activity and, consequently, ECG signals, which can potentially confound the algorithm's ability to detect sleep apnea events in advance. As our primary focus in this study was using single ECG signals to detect respiratory events during CPAP titration, we have not yet detected specific OSA endotypes or COMISA. However, we recognise the potential value of extending our approach to address this issue in future studies. Using a more robust and diverse dataset for training the model, including the data from patients with specific endotypes, could potentially enhance the generalisability and applicability of the proposed approach to a broader range of OSA populations. Secondly, we only consider obstructive sleep apnea and exclude the other types, such as central or mixed sleep apnea. This may raise false detections if those sleep-related disorders are experienced, particularly in real-life. Future research should be conducted to address these limitations.

Despite these limitations, our findings imply promising applications of a proceeding sleep apnea detection model in the follow-up of OSA treatment. Early detection of sleep apnea events allows sleep technologists proactively to adjust pressure settings. Developing a computer-aided design (CAD) device is another potential application of our research in practice, particularly with the remarkable advancement in wearable technology. Such a device could help to monitor the OSA patients' respiratory condition during use of the auto-titrated CPAP (APAP) machine, commonly used in home-based settings, and alert the machine to adjust the pressure. This could lead to better sleep quality, enhanced daytime function, reduced risks of complications associated with untreated OSA, and improved therapy adherence, ultimately breaking the vicious cycle of poor adherence leading to suboptimal treatment outcomes.

5 | CONCLUSION

This paper proposed a new approach to detect sleep apnea in advance using ECG at CPAP titration. Our results showed that SVM outperformed KNN, LDA, and DT models across all frequency bands and pre-OSA segments. The 8–50 Hz band provided much information about cardiac variation and could be used as a potential selection for developing a simple classifier using minimum features. Moreover, we found that increasing the lead time duration does not influence the detection performance, which may be attributable to CPAP normalisation on respiratory obstruction. Overall, our findings suggested that ECG signals at CPAP titration can be used to input into pre-OSA detection models. Detection models under the impact of CPAP may be closer to the real-life condition when patients use the PAP machine at home. This could be a promising approach for people interested in developing CAD to care for patient adherence. Nevertheless, future research is needed to address the limitations of this study, such as the generalised applicability of the proposed approach to a wide range of OSA endotypes or co-morbidities.

AUTHOR CONTRIBUTIONS

Wen-Te Liu and Tran Thanh Duy Linh conceived the presented idea. Tran Thanh Duy Linh and Nguyen Thi Hoang Trang designed the models and computational frameworks and analysed the data, with support from Shang-Yang Lin and Dean Wu in collecting data. Tran Thanh Duy Linh wrote the manuscript. Wen-Te Liu and Chaur-Jong Hu supervised the project and had an equal contribution. All authors have discussed and agreed to the final version of the manuscript.

CONFLICT OF INTEREST STATEMENT

No potential conflict of interest is reported by the authors.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Tran Thanh Duy Linh https://orcid.org/0000-0001-8762-0832
Wen-Te Liu https://orcid.org/0000-0003-1281-8718
Chaur-Jong Hu https://orcid.org/0000-0002-4900-5967

REFERENCES

Addison, P. S. (2005). Wavelet transforms and the ECG: A review. *Physiological Measurement*, *26*, R155–R199. https://doi.org/10.1088/0967-3334/26/5/R01

Benjafield, A. V., Ayas, N. T., Eastwood, P. R., Heinzer, R., Ip, M. S. M., Morrell, M. J., … Malhotra, A. (2019). Estimation of the global prevalence and burden of obstructive sleep apnea: A literature-based analysis. *The Lancet Respiratory Medicine*, *7*(8), 687–698. https://doi.org/10.1016/s2213-2600(19)30198-5

Berry, R. B., Gamaldo, C. E., Harding, S. M., Lloyd, R. M., Marcus, C. L., & Vaughn, B. V. (2015). *The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications, version 2.2*. American Academy of Sleep Medicine.

Bock, J., & Gough, D. A. (1998). Toward prediction of physiological state signals in sleep apnea. *IEEE Transactions on Biomedical Engineering*, *45*(11), 1332–1341. https://doi.org/10.1109/10.725330

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Contal, O., Poncin, W., Vaudan, S., De Lys, A., Takahashi, H., Bochet, S., … Charbonnier, F. (2021). One-year adherence to continuous positive airway pressure with telemonitoring in sleep apnea hypopnea syndrome: A randomized controlled trial. *Front Med (Lausanne)*, *8*, 626361. https://doi.org/10.3389/fmed.2021.626361

Dagum, P., & Galper, A. (1995). Time series prediction using belief network models. *International Journal of Human-Computer Studies*, *42*(6), 617–632. https://doi.org/10.1006/ijhc.1995.1027

Eiseman, N. A., Westover, M. B., Mietus, J. E., Thomas, R. J., & Bianchi, M. T. (2012). Classification algorithms for predicting sleepiness and sleep apnea severity. *Journal of Sleep Research*, *21*(1), 101–112. https://doi.org/10.1111/j.1365-2869.2011.00935.x

Falco, I. D., Pietro, G. D., & Sannino, G. (2015). On finding explicit rules for personalized forecasting of obstructive sleep apnea episodes. Paper presented at the 2015 IEEE international conference on information reuse and integration.

Guo, D., Peng, C.-K., Wu, H.-L., Mietus, J. E., Liu, Y., Sun, R.-S., & Thomas, R. J. (2011). ECG-derived cardiopulmonary analysis of pediatric sleep-disordered breathing. *Sleep Medicine*, *12*(4), 384–389. https://doi.org/10.1016/j.sleep.2010.09.011

Guo, W., Lv, T., She, F., Miao, G., Liu, Y., He, R., … Zhang, P. (2018). The impact of continuous positive airway pressure on heart rate variability in obstructive sleep apnea patients during sleep: A meta-analysis. *Heart & Lung*, *47*(5), 516–524. https://doi.org/10.1016/j.hrtlng.2018.05.019

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, *13*(4), 18–28. https://doi.org/10.1109/5254.708428

Huang, K., Zhou, Y., Huang, Z., Wang, R., Liu, Y., Chen, P., … Wang, L. (2021). Associations between nocturnal continuous blood pressure fluctuations and the characteristics of oxygen desaturation in patients with obstructive sleep apnea: A pilot study. *Sleep Medicine*, *84*, 1–7. https://doi.org/10.1016/j.sleep.2021.05.005

Karasulu, L., Epöztürk, P. O., Sökücü, S. N., Dalar, L., & Altin, S. (2010). Improving heart rate variability in sleep apnea patients: Differences in treatment with auto-titrating positive airway pressure (APAP) versus conventional CPAP. *Lung*, *188*(4), 315–320. https://doi.org/10.1007/s00408-010-9237-4

Kesek, M., Franklin, K. A., Sahlin, C., & Lindberg, E. (2009). Heart rate variability during sleep and sleep apnea in a population based study of 387 women. *Clinical Physiology and Functional Imaging*, *29*(4), 309–315. https://doi.org/10.1111/j.1475-097X.2009.00873.x

Kraiem, M. S., Sánchez-Hernández, F., & Moreno-García, M. N. (2021). Selecting the suitable resampling strategy for imbalanced data classification regarding dataset properties. An approach based on association models. *Applied Sciences*, *11*(18), 8546 Retrieved from https://www.mdpi.com/2076-3417/11/18/8546

Kufoy, E., Palma, J.-A., Lopez, J., Alegre, M., Urrestarazu, E., Artieda, J., & Iriarte, J. (2012). Changes in the heart rate variability in patients with obstructive sleep apnea and its response to acute CPAP treatment. *PLoS One*, *7*(3), e33769. https://doi.org/10.1371/journal.pone.0033769

Legault, J., Thompson, C., Martineau-Dussault, M., André, C., Baril, A. A., Martinez Villar, G., … Gosselin, N. (2021). Obstructive sleep apnea and cognitive decline: A review of potential vulnerability and protective factors. *Brain Sciences*, *11*(6), 706. https://doi.org/10.3390/brainsci11060706

Liao, D., Li, X., Vgontzas, A. N., Liu, J., Rodriguez-Colon, S., Calhoun, S., & Bixler, E. O. (2010). Sleep-disordered breathing in children is associated with impairment of sleep stage-specific shift of cardiac autonomic modulation. *Journal of Sleep Research*, *19*(2), 358–365. https://doi.org/10.1111/j.1365-2869.2009.00807.x

Lin, C.-Y., Wang, Y.-W., Setiawan, F., Trang, N. T. H., & Lin, C.-W. (2022). Sleep apnea classification algorithm development using a machine-learning framework and bag-of-features derived from electrocardiogram spectrograms. *Journal of Clinical Medicine*, *11*(1), 192 Retrieved from https://www.mdpi.com/2077-0383/11/1/192

Maali, Y., & Al-Jumaily, A. (2013). Multi neural networks investigation based sleep apnea prediction. *Procedia Computer Science*, *24*, 97–102. https://doi.org/10.1016/j.procs.2013.10.031

Malhotra, A., & White, D. P. (2002). Obstructive sleep apnea. *The Lancet*, *360*(9328), 237–245. https://doi.org/10.1016/S0140-6736(02)09464-3

Nasifoglu, H., & Eroğul, O. (2021). Convolutional neural networks based OSA event prediction from ECG Scalograms and spectrograms.

O'Hara, S., & Draper, B. (2011). Introduction to the bag of features paradigm for image classification and retrieval.

Ozdemir, G., Nasifoglu, H., & Eroğul, O. (2016). An early warning algorithm to predict obstructive sleep apnea (OSA) episodes. *Journal of Biomedical Engineering and Biosciences*, *3*, 34–42. https://doi.org/10.11159/jbeb.2016.007

Pack, A. I., Magalang, U. J., Singh, B., Kuna, S. T., Keenan, B. T., & Maislin, G. (2021). Randomized clinical trials of cardiovascular disease in obstructive sleep apnea: Understanding and overcoming bias. *Sleep*, *44*(2). https://doi.org/10.1093/sleep/zsaa229

Tereshchenko, L. G., & Josephson, M. E. (2015). Frequency content and characteristics of ventricular conduction. *Journal of Electrocardiology*, *48*(6), 933–937. https://doi.org/10.1016/j.jelectrocard.2015.08.034

Thomas, R. J., Mietus, J. E., Peng, C.-K., & Goldberger, A. L. (2005). An electrocardiogram-based technique to assess cardiopulmonary coupling during sleep. *Sleep*, *28*(9), 1151–1161. https://doi.org/10.1093/sleep/28.9.1151

Wang, B., Yi, X., Gao, J., Li, Y., Xu, W., Wu, J., & Han, D. (2021). Real-time prediction of upcoming respiratory events via machine learning using

snoring sound signal. *Journal of Clinical Sleep Medicine*, *17*(9), 1777–1784. https://doi.org/10.5664/jcsm.9292

Wang, X., Ouyang, Y., Wang, Z., Zhao, G., Liu, L., & Bi, Y. (2013). Obstructive sleep apnea and risk of cardiovascular disease and all-cause mortality: A meta-analysis of prospective cohort studies. *International Journal of Cardiology*, *169*(3), 207–214. https://doi.org/10.1016/j.ijcard.2013.08.088

Waxman, J. A., Graupe, D., & Carley, D. W. (2010). Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network. *American Journal of Respiratory and Critical Care Medicine*, *181*(7), 727–733. https://doi.org/10.1164/rccm.200907-1146OC

How to cite this article: Linh, T. T. D., Trang, N. T. H., Lin, S.-Y., Wu, D., Liu, W.-T., & Hu, C.-J. (2023). Detection of preceding sleep apnea using ECG spectrogram during CPAP titration night: A novel machine-learning and bag-of-features framework. *Journal of Sleep Research*, 1–12. https://doi.org/10.1111/jsr.13991