Hindawi
Journal of Sensors
Volume 2021, Article ID 2998202, 9 pages
https://doi.org/10.1155/2021/2998202

Research Article

A Pilot Study of Detecting Individual Sleep Apnea Events Using Noncontact Radar Technology, Pulse Oximetry, and Machine Learning

Ståle Toften[1], **Jonas T. Kjellstadli**[1], **Stig S. Tyvold**[2] and **Mads H. S. Moxness**[2,3]

[1]*Department of Data Science and Research, VitalThings AS, Trondheim, Norway*
[2]*Aleris Hospital Solsiden, Trondheim, Norway*
[3]*Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Norway*

Correspondence should be addressed to Ståle Toften; staletoften@gmail.com

Received 7 April 2021; Revised 8 June 2021; Accepted 24 June 2021; Published 16 July 2021

Academic Editor: Qiang Wu

The gold standard for assessing sleep apnea, polysomnography, is resource intensive and inconvenient. Thus, several simpler alternatives have been proposed. However, validations of these alternatives have focused primarily on estimating the apnea-hypopnea index (apnea events per hour of sleep), which means information, clearly important from a physiological point of view such as apnea type, apnea duration, and temporal distribution of events, is lost. The purpose of the present study was to investigate if this information could also be provided with the combination of radar technology and pulse oximetry by classifying sleep apnea events on a second-by-second basis. Fourteen patients referred to home sleep apnea testing by their medical doctor were enrolled in the study (6 controls and 8 patients with sleep apnea; 4 mild, 2 moderate, and 2 severe) and monitored by Somnofy (radar-based sleep monitor) in parallel with respiratory polygraphy. A neural network was trained on data from Somnofy and pulse oximetry against the polygraphy scorings using leave-one-subject-out cross-validation. Cohen's kappa for second-by-second classifications of no event/event was 0.81, or almost perfect agreement. For classifying no event/hypopnea/apnea and no event/hypopnea/obstructive apnea/central apnea/mixed apnea, Cohen's kappa was 0.43 (moderate agreement) and 0.36 (fair agreement), respectively. The Bland-Altman 95% limits of agreement for the respiratory event index (apnea events per hour of recording) were -8.25 and 7.47, and all participants were correctly classified in terms of sleep apnea severity. Furthermore, the results showed that the combination of radar and pulse oximetry could be more accurate than the two technologies separately. Overall, the results indicate that radar technology and pulse oximetry could reliably provide information on a second-by-second basis for no event/event which could be valuable for management of sleep apnea. To be clinically useful, a larger study is necessary to validate the algorithm on a general population.

1. Introduction

Sleep apnea is characterized by repetitive reduction or cessation of airflow during sleep resulting in microarousals and is associated with increased risk of daytime sleepiness, coronary artery disease, stroke, and early death [1]. Despite being a serious disease, sleep apnea is underrecognized and underdiagnosed [2]. The gold standard for diagnosing sleep apnea is inlaboratory polysomnography (PSG) [3]. PSG uses a comprehensive set of sensors to measure brain, muscular, respiratory, and cardiovascular activity, and collected data is manually analyzed by a sleep specialist. While PSG is accurate, it is also resource intensive and can be inconvenient for the patient, who must sleep with several sensors attached to their body. Overnight respiratory polygraphy (RP), which in contrast to PSG does not measure brain activity, is often used as a simpler alternative when diagnosing sleep apnea. However, RP is still resource intensive and inconvenient for the patient. To reduce the amount of manual work required, PSG and RP software have been enhanced with algorithms for automatically scoring apnea events at the cost of slightly reduced precision [4].

More convenient alternatives for assessing sleep apnea have been investigated [5, 6], such as using only pulse oximetry [7, 8]. Furthermore, recent papers have shown that radar technology could accurately assess sleep apnea without any sensors attached to the patient [9–11]. However, the combination of pulse oximetry and radar technology has not been investigated. Though, the combination of pulse oximetry and respiratory inductance plethysmography (RIP), which also measure respiratory effort, has been studied [12]. Moreover, previous research has focused primarily on measuring the apnea-hypopnea index (AHI = overall number of hypopneas and apneas per hour of sleep). The AHI is today used as the most important metric for categorizing sleep apnea severity. However, this practice is debatable as the AHI does not take individual apnea type (hypopnea, obstructive apnea, central apnea, or mixed apnea), degree of desaturation, apnea duration, temporal distribution of events, or sleep disruption due to respiratory effort related arousals into account, which are clearly important from a physiological point of view [13, 14]. An ideal tool for assessing sleep apnea would be less resource intensive and more convenient than PSG/RP, while still providing the same information.

The aim of the present study was to analyze how accurately the sleep assistant Somnofy can classify individual sleep apnea events using a combination of radar technology, pulse oximetry, and machine learning. For this purpose, sleep apnea events from Somnofy were compared to scorings from a RP-based home sleep apnea test (HSAT) on a second-by-second basis. Though, the agreement between Somnofy and HSAT was also analyzed for the respiratory event index (REI = number of apnea and hypopnea per hour recording).

Figure 1: A visualization of the Somnofy set up. One Somnofy unit was placed on the nightstand, and one unit mounted in a wall position. Both units were aiming towards the participants' chest from approximately one meter. Both units were connected to a Somnofy Recorder placed under the nightstand.

2. Methods

2.1. Participants. Fourteen patients (9 males, 5 females) referred to HSAT by their medical doctor were enrolled in this study. The average age was 50.1 years, and average body mass index (BMI) was 30.3. The inclusion criteria were age between 18 and 70, a medical history that indicated possible sleep apnea, and no history of upper airway surgery or use of nasal decongestants or anti-inflammatory medication the last three months prior to the study. No participants were excluded from the study. The study was approved by the Norwegian Ethical Committee (REK, id number 10445). Written informed consent was obtained from all participants. All methods were performed in accordance with relevant guidelines and regulations.

2.2. Procedure. All participants underwent HSAT for one night while also being monitored by Somnofy. The participants slept in a preorganized bedroom at a university hospital hotel where one Somnofy unit was placed in a nightstand position (by the head) and one in a wall position (above the head), both aiming at the participant's chests from approximately 1 meter distance. The setup is visualized in Figure 1. Since both Somnofy units recorded properly for all the nights, one unit was randomly picked per patient to reduce bias to sensor location in the analysis (9 nightstand, 5 wall).

2.3. Home Sleep Apnea Test. Nox T3 (Nox Medical, Iceland), a type 3 HSAT monitor, was used in this study [15]. Nox T3 measure respiration using a nasal cannula, a thermistor, thoracic and abdominal respiratory inductance plethysmography, and a pulse oximeter. Sleep apneas were manually scored by a trained specialist in accordance with The AASM Manual for scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications [3] in the Noxturnal software (version 5.1.3.20388, Nox Medical, Iceland). Hypopneas were scored based on the recommended rules of ≥30% reduction in flow for ≥10 seconds and ≥3% oxygen desaturation from preevent baseline. Apneas, on the other hand, were scored if there was ≥90% reduction in flow from preevent baseline for ≥10 seconds. Apneas were classified as obstructive if there was inspiratory effort throughout the event, as central apnea if there was no inspiratory effort, and as mixed apneas if there was no inspiratory effort at the beginning of the event followed by inspiratory effort at the final part of the event.

To simplify comparison of events with Somnofy, the scorings from Nox T3 were transformed into second-by-second classifications where each second could take the values no event, hypopnea, obstructive apnea, central apnea, or mixed apnea.

2.4. Somnofy. Somnofy (SM-100, version 0.9.3, VitalThings, Norway) with the Somnofy Recorder (SM-REC2, version 3.0.1, VitalThings, Norway) was used in this study. Somnofy utilizes an impulse radio ultra-wideband radar with an average sampling rate of 23.8 GHz to measure body movements. Through configuration, the samples are sampled into a 3-meter-long frame of 5 cm bins updated at a frequency of approximately 17 Hz. Periodic movements that have frequencies that could be generated by respiration are further used to extract an effort-based respiratory waveform. The Somnofy Recorder enables storage of the 17 Hz data instead of the standard 1 Hz resolution. The respiratory waveform was synchronized with the RP scorings by maximizing the crosscorrelation between movements from Somnofy and RP. Somnofy supports connection with other devices through Bluetooth, which enables collection of oxygen desaturation (SpO_2) data from a pulse oximeter. WristOx$_2$ 3150 (Nonin Medical, USA) was used in this study.

An envelope was created around the respiratory waveform to extract an instantaneous amplitude, as previously done by Coronel et al. for sleep apnea detection with RIP [12]. The amplitude, the respiratory rate, and the SpO_2 were resampled to a 1 Hz resolution and fed into a long short-term memory neural network (dense–>LSTM–>dense). The network was trained against the manual HSAT scored events with supervised learning. This way, the algorithm could learn the relationship between the radar-measured respiratory effort, the SpO_2, and the different apnea types. Leave-one-subject-out crossvalidation was utilized to both train and validate the algorithm on the same dataset. In other words, sleep apnea scorings from Somnofy for one patient were based on an algorithm which had been trained only on the other patients.

Somnofy is harmless to human beings and certified according to the Federal Communication Commission (FCC) and "Conformité Européene" (CE). The frequency of the radar pulses enables them to travel through softer materials like bed sheets while being reflected at denser objects like the human body. The sensor will measure respiration on only the closest person if there are multiple persons in the room. For more details on Somnofy, the reader is referred to the validation study of Somnofy for sleep stage classification in healthy adults [16]. Somnofy is currently not an FDA-approved medical device.

2.5. Statistics. In order to analyze the performance of Somnofy for classifying individual sleep apnea events (providing information on event type, event duration, and temporal distribution of events), the agreement between Somnofy and HSAT was investigated for second-by-second classifications for each participant using Cohen's kappa. Cohen's kappa values were interpreted in the following way: values higher than .80 were considered as almost perfect agreement, .80 to .61 as substantial agreement, .60 to .41 as moderate agreement, .40 to .21 as fair agreement, .20 to.11 as slight agreement, and values less than.10 as no agreement [17, 18].

Table 1: Overview of participant demography and HSAT data.

	Average (SD)	Range (min, max)
N (count)	14	
N female (count)	5	
Age (years)	50.1 (14.1)	(20, 75)
BMI (kg/m^2)	30.3 (6.1)	(21.2, 39.6)
Recording duration (hours)	7.7 (1.0)	(5.9, 9.5)
REI	14.3 (14.0)	(0.5, 51.2)
HI	5.9 (6.2)	(0.1, 24.5)
OAI	7.3 (11.5)	(0.0, 44.0)
CAI	0.7 (1.0)	(0.0, 3.0)
MAI	0.5 (1.1)	(0.0, 3.5)
ODI	12.4 (11.2)	(0.3, 39.2)

N: number of participants; BMI: body mass index; REI: respiratory event index: HI/OAI/CAI/MAI: hypopnea/obstructive apnea/central apnea/mixed apnea index (number of hypopneas/obstructive apneas/central apneas/mixed apneas per hour of recording); ODI: oxygen desaturation index.

The agreement on the REI between Somnofy and manual HSAT was analyzed using Bland-Altman analysis [19]. Bland-Altman plots were also generated for the hypopnea index (HI = number of hypopneas per hour recording), the obstructive apnea index (OAI = number of obstructive apneas per hour recording), the central apnea index (CAI = number of central apneas per hour recording), and the mixed apnea index (MAI = number of mixed apneas per hour recording). A confusion matrix was generated to compare the agreement on sleep apnea severity (control: REI < 5, mild: 5 ≤ REI < 15, moderate: 15 ≤ REI < 30, and severe: REI ≥ 30).

Calculations were performed in Python (v. 3.8.6), and Cohen's kappa was calculated with the scikit-learn package (v 0.23.2).

3. Results

3.1. Data. Table 1 shows demographic data and sleep parameters for the study participants. In total, the dataset consisted of 1 584 (34 010 seconds) apnea events for which 628 (14 383 seconds), 823 (16 370 seconds), 76 (1 452 seconds), and 56 (1 805 seconds) were manually scored by HSAT as hypopnea, obstructive apnea, central apnea, and mixed apnea, respectively.

Figure 2 displays the development of the respiratory waveform from Somnofy during one example of each event type. In these examples, the amplitude is clearly reduced during apnea events.

3.2. Second-by-Second Analysis. Cohen's kappa for the binary classification no event/event between Somnofy with pulse oximetry (Somnofy+SpO_2) and manually scored HSAT on a second-by-second basis for the whole dataset was 0.81, or almost perfect agreement. For classifying the three classes, no event/hypopnea/apnea Somnofy+SpO_2 showed moderate agreement (Cohen's kappa = 0.43), while for classifying all five classes no event/hypopnea/obstructive apnea/central apnea/mixed apnea, the agreement was fair (Cohen's kappa = 0.36).

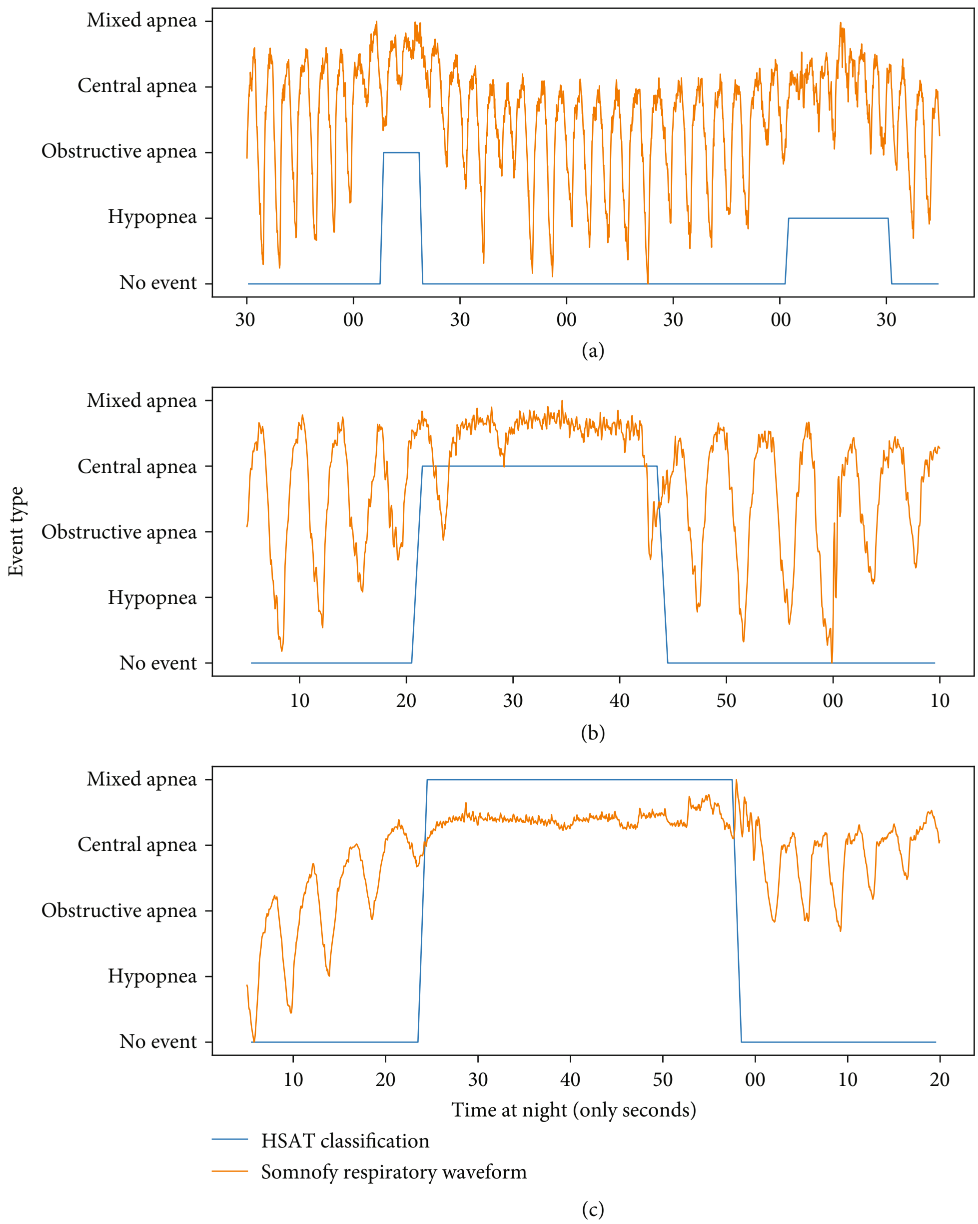

Figure 2: Illustration of the respiratory waveform from Somnofy during hypopnea, obstructive apnea, central apnea, and mixed apnea. The respiratory waveform measured by Somnofy is clearly reduced during HSAT classified hypopnea and obstructive apnea (a), central apnea (b), and mixed apnea (c).

The temporal distribution of events is further visualized in Figure 3, which shows classifications by both Somnofy +SpO_2 and HSAT for two random nights in each severity group.

3.3. Analysis of Sleep Apnea Indexes and Severity. Bland-Altman plots for the different sleep apnea indexes, calculated from individual nights, are provided in Figure 4. Somnofy +SpO_2 tended to overestimate REI for low REI values and underestimate REI for high REI values. This trend was driven mostly by OAI. Furthermore, Somnofy+SpO_2 underestimated both CAI and MAI. Consequently, the differences with HSAT were not normally distributed, and the Bland-Altman limits of agreement and bias are not statistically valid for these plots.

The differences in sleep apnea severity between Somnofy +SpO_2 and manual scored HSAT are shown in the classification matrix in Table 2. HSAT and Somnofy agreed on the severity of all participants.

3.4. Using Only Radar Technology or Only Pulse Oximetry. Table 3 shows the results for the same algorithm but trained with only radar data or only pulse oximetry. Both technologies alone showed promising results on a second-by-second basis and for overall indexes but were inferior to the results for Somnofy+SpO_2.

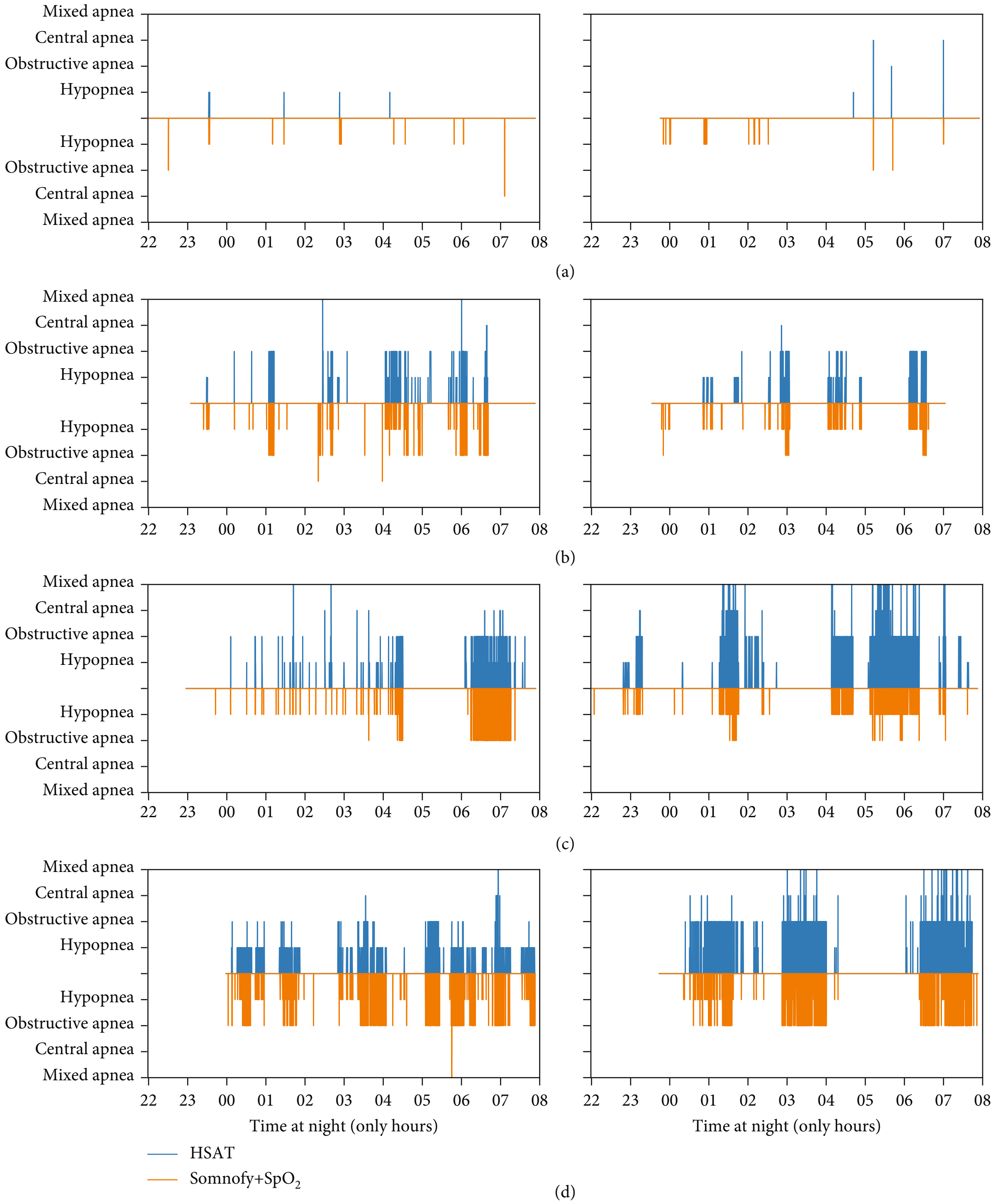

FIGURE 3: Comparison of HSAT and Somnofy+SpO_2 for temporal distribution of events for patients with no, mild, moderate, and severe sleep apnea. Classifications second by second for HSAT (blue) and Somnofy+SpO_2 (orange) plotted on the positive and negative y-axis, respectively, against time (only denoted by hours) on the x-axis. Parts (a), (b), (c), and (d) show two randomly picked nights with no, mild, moderate, and severe sleep apnea, respectively.

4. Discussion

The results in the present paper indicate that the combination of radar technology and pulse oximetry can classify sleep apnea more accurately than the two technologies separately. Furthermore, the no event/event classification on a second-by-second basis showed almost perfect agreement with HSAT providing information on temporal distribution of events and event duration. Though, Cohen's kappa was lower than between manual PSG scorers on epoch (30 seconds)

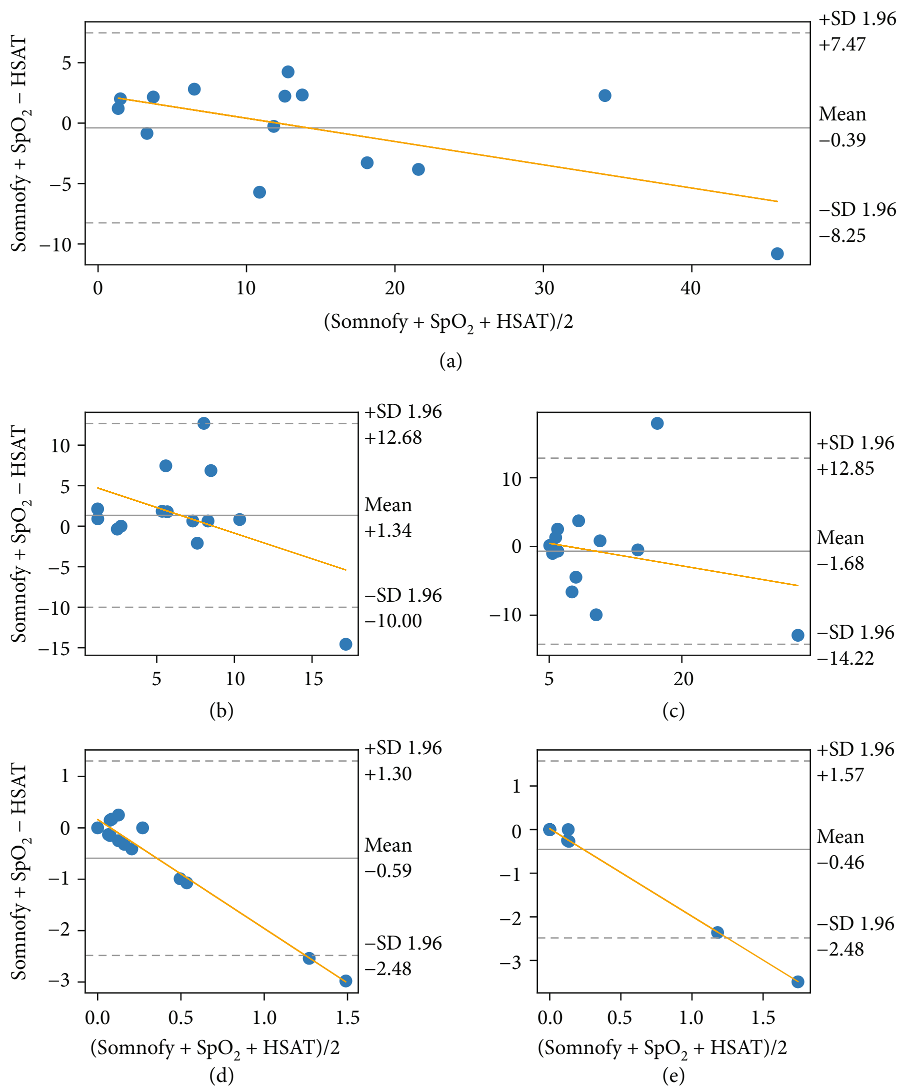

FIGURE 4: Bland-Altman plots of agreement between Somnofy+SpO_2 and HSAT on different sleep apnea indexes. The disagreement between Somnofy+SpO_2 and HSAT on the y-axis is plotted against the average of Somnofy+SpO_2 and HSAT on the x-axis for different sleep apnea indexes. The blue dots represent values for individual nights, the solid grey line represents the bias, the dotted lines represent the Bland-Altman 95% limits of agreement, and the orange line represents a linear regression to visualize any trend. REI: respiratory event index (number of hypopneas and apneas per hour of recording); HI/OAI/CAI/MAI: hypopnea/obstructive apnea/central apnea/mixed apnea index (number of hypopneas/obstructive apneas/central apneas/mixed apneas per hour of recording).

resolution [20]. REI per night for only radar technology and only pulse oximetry was comparable but less accurate than in previous research [7–10]. Nikkonen et al. [8] also utilized a neural network on pulse oximetry, but their training dataset contained 1 692 nights compared to our 14. It is likely that the accuracy of the algorithm in the present study would improve with a larger dataset. Nevertheless, the results for Somnofy with pulse oximetry were as accurate as previous research [7–10], and the agreement was higher than shown between RP and PSG [15, 21]. Furthermore, HSAT and Somnofy with pulse oximetry agreed on the sleep apnea severity of all patients.

The algorithm showed only moderate agreement for distinguishing no event/hypopnea/apnea and only fair agreement for classifying all the different apnea types.

TABLE 2: Concordance of sleep apnea severity between HSAT and Somnofy+SpO_2.

Severity according to HSAT	Severity according to Somnofy+SpO_2			
	Control	Mild	Moderate	Severe
Control	4	0	0	0
Mild	0	6	0	0
Moderate	0	0	2	0
Severe	0	0	0	2

Control = REI < 5, mild = 5 ≤ REI < 15, moderate 15 ≤ REI < 30, and severe REI ≥ 30.

Table 3: Agreement for only pulse oximetry, only radar, and the combination of radar and pulse oximetry (Somnofy+SpO_2) against HSAT.

	5 class κ	3 class κ	2 class κ	REI 95% LoA
Only pulse oximetry	0.28	0.34	0.71	-13.96, 11.18
Only radar	0.26	0.34	0.78	-14.48, 9.98
Somnofy+SpO_2	0.36	0.43	0.81	-8.25, 7.47

κ = Cohen's kappa for the whole dataset on seconds resolution, 5 class = no event/hypopnea/obstructive apnea/central apnea/mixed apnea, 3 class = no event/hypopnea/apnea, 2 class = no event/event, and REI 95% LoA = Bland-Altman 95% limits of agreement for REI.

Thus, it could not reliably detect the event type. As shown in Figure 2(a) it is not always straightforward to distinguish the apnea types. Here, the radar-measured respiratory effort behaves quite similarly for obstructive apnea and hypopnea, but also the HSAT scorings are not necessary always correct [15, 21], or the sensor data could be noisy. This motivates a machine learning approach. It is likely that the machine learning algorithm would improve with more data, especially for central and mixed apnea where the dataset contained few events. However, this warrants further investigations.

Recently, radar technology has been shown to accurately classify sleep in healthy adults [16] and to possibly detect body position during sleep (supine, prone, side) [22, 23]. If radar technology also could reliably classify sleep in persons with sleep apnea, the proposed solution could calculate AHI, possibly detect RERAs, and investigate sleep disruption. With classifications on a second-by-second basis, sleep apneas could also be analyzed across the different sleep stages (light, deep, and REM sleep) and body positions. This would probably not be possible using only pulse oximetry. Using only radar technology, on the other hand, would not provide exact information on desaturation levels, unless this could be accurately estimated from the characteristics of the events.

Other sensors could be used to measure respiration instead of radars. Wearables, such as RIP [12], nasal cannula, and thermistor, are generally less convenient for the patient who has to sleep with sensors attached to his/her body. This could disrupt sleep and make it more difficult to assess patient groups that do not accept wearing wires, or the data quality could be affected if the sensor is attached suboptimal or disrupted from movements during sleep. Nasal cannula and thermistor cannot measure inspiratory effort, and nasal cannula cannot measure mouth breathing. Though, there are also other nonwearable alternatives like using sound [24] (subject to audio noise from surroundings), vision-based solutions (affected by bed sheets), infrared solutions [25], or under-the-mattress solutions [26, 27]. To the authors' knowledge, neither of these alternatives has been shown to reliably classify individual sleep apnea type, apnea duration, or temporal distribution of events.

The ability to classify sleep apnea second by second as well as using exact measurements of oxygen desaturation levels may contribute to a more detailed and profound understanding of sleep apnea. Individual sleep apnea event severity, event duration, oxygen desaturation, temporal distribution of events, and sleep disruption are all clearly important from a physiological point of view [13, 14]. Sleep apnea has a multifactorial pathogenesis [28] which has led to a multitude of options in both diagnostic and therapeutic measures. More detailed data could thus enable a more patient-specific tailored sleep apnea management. Furthermore, a more thorough understanding of the underlying pathophysiology will likely be instrumental in understanding comorbidities [29, 30]. In the present study, we have shown that this information might not be limited to only PSG/RP.

As the proposed solution does not require manual scoring and the equipment does not need expertise to install (set radar on nightstand and put oximeter on finger), it should be more scalable and cost efficient than PSG/RP. If this solution could also reliably measure and diagnose sleep apnea, more people could receive sleep apnea assessments, diagnosis could be based on several nights of measurements to counteract the night-to-night variability in severity of sleep apnea [31], assessments could be performed in the patient's own bed, patients could be continuously monitored during treatment, and more data could be gathered for research purposes. Optimally, such an alternative should be as accurate and detailed as possible.

4.1. Limitations and Future Work. The dataset in the present study includes 1 584 events which should be more than enough to validate the algorithm. However, the patient population is relatively small, and a larger study is needed to assess the clinical usefulness across age, sex, AHI, respiratory disturbance index (RDI), BMI, and on people with selected comorbidities. Accuracy should also be analyzed across sensor location and sleeping position. In contrast to the pilot study, a larger study should use PSG. HSAT does not measure sleep and is therefore unable to detect RERAs and AHI.

5. Conclusion

The present study indicates that radar technology and pulse oximetry could assess sleep apnea more accurately than the two technologies separately. Furthermore, the results show that classifications of no event/event could be performed reliably on a second-by-second basis, providing information on apnea duration and temporal distribution of events. This information is clearly important form a physiological point of view but has not been validated for radar technology or pulse oximetry as the focus has been primarily on the AHI. AHI is the most important clinical parameter today, but do not give the complete picture of the disease. To increase the understanding and improve the management of sleep apnea more information is needed. PSG/RP provides this information but is not scalable due to high cost and inconvenience. A scalable solution could collect data from a larger population and measure patients for longer periods of time. A larger study is needed to validate the clinical usefulness of the present algorithm across age, sex, AHI, RDI, BMI, and on people with selected comorbidities.

Abbreviations

AHI: Apnea-hypopnea index
BMI: Body mass index
CAI: Central apnea index (number of central apneas per hour recording)
HI: Hypopnea index (number of hypopneas per hour recording)
HSAT: Home sleep apnea test
MAI: Mixed apnea index (number of mixed apneas per hour recording)
OAI: Obstructive apnea index (number of obstructive apneas per hour recording)
PSG: Polysomnography
RDI: Respiratory disturbance index
REI: Respiratory event index (number of apneas and hypopneas per hour recording)
RERA: Respiratory effort-related arousals
RP: Respiratory polygraphy.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

(i) Ståle Toften: Ståle Toften works for VitalThings, the company that has developed Somnofy. (ii) Jonas T. Kjellstadli: Jonas T. Kjellstadli works for VitalThings, the company that has developed Somnofy. (iii) Stig S. Tyvold declares that there is no conflict of interest regarding the publication of this article. (iv) Mads H. S. Moxness declares that there is no conflict of interest regarding the publication of this article

Acknowledgments

We would like to thank Aleris and VitalThings for organizing data collection and especially Guttorm Bruskeland for manually scoring the HSAT dataset.

References

[1] K. A. Franklin and E. Lindberg, “Obstructive sleep apnea is a common disorder in the population-a review on the epidemiology of sleep apnea,” *Journal of thoracic disease*, vol. 7, no. 8, pp. 1311–1322, 2015.

[2] J. Arnold, M. Sunilkumar, V. Krishna, S. P. Yoganand, M. S. Kumar, and D. Shanmugapriyan, “Obstructive sleep apnea,” *Journal of Pharmacy & Bioallied Sciences*, vol. 9, Suppl 1, pp. S26–S28, 2017.

[3] R. B. Berry, “The AASM manual for the scoring of sleep and associated events,” in *Rules, Terminology and Technical Specifications*, American Academy of Sleep Medicine, 2018.

[4] S. Kristiansen, G. M. Traaen, B. Øverland et al., “Comparing manual and automatic scoring of sleep monitoring data from portable polygraphy,” *Journal of Sleep Research*, vol. 30, no. 2, 2021.

[5] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, “A review of obstructive sleep apnea detection approaches,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 825–837, 2019.

[6] P. Leelaarporn, P. Wachiraphan, T. Kaewlee et al., “Sensor-driven achieving of smart living: a review,” *IEEE Sensors Journal*, vol. 21, no. 9, pp. 10369–10391, 2021.

[7] F. Mendonça, S. S. Mostafa, F. Morgado-Dias, and A. G. Ravelo-García, “An oximetry based wireless device for sleep apnea detection,” *Sensors*, vol. 20, no. 3, p. 888, 2020.

[8] S. Nikkonen, I. O. Afara, T. Leppänen, and J. Töyräs, “Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea,” *Scientific Reports*, vol. 9, no. 1, 2019.

[9] Y. Zhou, D. Shu, H. Xu et al., “Validation of novel automatic ultra-wideband radar for sleep apnea detection,” *Journal of Thoracic Disease*, vol. 12, no. 4, pp. 1286–1295, 2020.

[10] S. Kang, D.-K. Kim, Y. Lee et al., “Non-contact diagnosis of obstructive sleep apnea using impulse-radio ultra- wideband radar,” *Scientific Reports*, vol. 10, no. 1, p. 5261, 2020.

[11] M. Kagawa, H. Tojima, and T. Matsui, “Non-contact diagnostic system for sleep apnea–hypopnea syndrome based on amplitude and phase analysis of thoracic and abdominal Doppler radars,” *Medical & Biological Engineering & Computing*, vol. 54, no. 5, pp. 789–798, 2016.

[12] C. Coronel, C. Wiesmeyr, H. Garn et al., “Detection of respiratory events by respiratory effort and oxygen desaturation,” *Journal of Medical and Biological Engineering*, vol. 40, no. 4, pp. 517–525, 2020.

[13] N. M. Punjabi, “Counterpoint: is the apnea-hypopnea undex the best way to quantify the severity of sleep-disordered breathing? No,” *Chest*, vol. 149, no. 1, pp. 16–19, 2016.

[14] C. M. Cielo and I. E. Tapia, “Diving deeper: rethinking AHI as the primary measure of OSA severity,” *Journal of Clinical Sleep Medicine*, vol. 15, no. 8, pp. 1075-1076, 2019.

[15] L. Xu, F. Han, B. T. Keenan et al., “Validation of the Nox-T3 portable monitor for diagnosis of obstructive sleep apnea in Chinese adults,” *Journal of Clinical Sleep Medicine*, vol. 13, no. 5, pp. 675–683, 2017.

[16] S. Toften, S. Pallesen, M. Hrozanova, F. Moen, and J. Grønli, “Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®),” *Sleep Medicine*, vol. 75, pp. 54–61, 2020.

[17] P. Ranganathan, C. S. Pramesh, and R. Aggarwal, “Common pitfalls in statistical analysis: measures of agreement,” *Perspectives in Clinical Research*, vol. 8, no. 4, pp. 187–191, 2017.

[18] J. R. Landis and G. G. Koch, “The measurement of observer agreement for categorical data,” *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[19] J. M. Bland and D. G. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement,” *The Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[20] R. S. Rosenberg and S. van Hout, “The American academy of sleep medicine inter-scorer reliability program: respiratory events,” *Journal of Clinical Sleep Medicine*, vol. 10, no. 4, pp. 447–454, 2014.

[21] A. Cairns, E. Wickwire, E. Schaefer, and D. Nyanjom, “A pilot validation study for the NOX T3TM portable monitor for the

detection of OSA," *Sleep and Breathing*, vol. 18, no. 3, pp. 609–614, 2014.

[22] J. E. Kiriazi, S. M. M. Islam, O. Borić-Lubecke, and V. M. Lubecke, "Sleep posture recognition with a dual-frequency cardiopulmonary Doppler radar," *IEEE Access*, vol. 9, pp. 36181–36194, 2021.

[23] M. Piriyajitakonkij, P. Warin, P. Lakhan et al., "SleepPoseNet: multi-view learning for sleep postural transition recognition using UWB," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 1305–1314, 2021.

[24] N. Ben-Israel, A. Tarasiuk, and Y. Zigel, "Nocturnal sound analysis for the diagnosis of obstructive sleep apnea," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 6146–6149, Buenos Aires, Argentina, 2010.

[25] J. N. Murthy, J. van Jaarsveld, J. Fei et al., "Thermal infrared imaging: a novel method to monitor airflow during polysomnography," *Sleep*, vol. 32, no. 11, pp. 1521–1527, 2009.

[26] D. Huysmans, P. Borzée, D. Testelmans et al., "Evaluation of a commercial ballistocardiography sensor for sleep apnea screening and sleep monitoring," *Sensors*, vol. 19, no. 9, p. 2133, 2019.

[27] P. Edouard, D. Campo, P. Bartet et al., "Validation of the Withings Sleep Analyzer, an under-the-mattress device for the detection of moderate-severe sleep apnea syndrome," *Journal of Clinical Sleep Medicine*, vol. 17, no. 6, pp. 1217–1227, 2021.

[28] D. J. Eckert and A. Malhotra, "Pathophysiology of adult obstructive sleep apnea," *Proceedings of the American Thoracic Society*, vol. 5, no. 2, pp. 144–153, 2008.

[29] I. H. Stevenson, K. C. Roberts-Thomson, P. M. Kistler et al., "Atrial electrophysiology is altered by acute hypercapnia but not hypoxemia: implications for promotion of atrial fibrillation in pulmonary disease and sleep apnea," *Heart Rhythm*, vol. 7, no. 9, pp. 1263–1270, 2010.

[30] M. R. Bonsignore, P. Baiamonte, E. Mazzuca, A. Castrogiovanni, and O. Marrone, "Obstructive sleep apnea and comorbidities: a dangerous liaison," *Multidisciplinary Respiratory Medicine*, vol. 14, 2019.

[31] E. Anitua, J. Duran-Cantolla, G. Z. Almeida, and M. H. Alkhraisat, "Predicting the night-to-night variability in the severity of obstructive sleep apnea: the case of the standard error of measurement," *Sleep Science*, vol. 12, no. 2, pp. 72–78, 2019.