



## OPEN ACCESS

EDITED BY  
Ding Zou,  
University of Gothenburg, Sweden

REVIEWED BY  
Johan A. Verbraecken,  
Antwerp University Hospital, Belgium  
Martin O. Mendez,  
Universidad Autónoma de San Luis Potosí,  
Mexico

\*CORRESPONDENCE  
Jacopo Piccini  
✉ jacopop@ru.is

SPECIALTY SECTION  
This article was submitted to  
Precision Sleep Medicine,  
a section of the journal  
Frontiers in Sleep

RECEIVED 19 December 2022  
ACCEPTED 14 February 2023  
PUBLISHED 02 March 2023

CITATION  
Piccini J, August E, Óskarsdóttir M and  
Arnardóttir ES (2023) Using the electrodermal  
activity signal and machine learning for  
diagnosing sleep. *Front. Sleep* 2:1127697.  
doi: 10.3389/frsle.2023.1127697

COPYRIGHT  
© 2023 Piccini, August, Óskarsdóttir and  
Arnardóttir. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums is  
permitted, provided the original author(s) and  
the copyright owner(s) are credited and that  
the original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Using the electrodermal activity signal and machine learning for diagnosing sleep

Jacopo Piccini<sup>1,2\*</sup>, Elias August<sup>1,2</sup>, María Óskarsdóttir<sup>1,3</sup> and  
Erna Sif Arnardóttir<sup>1,2,3,4</sup>

<sup>1</sup>Reykjavik University Sleep Institute, School of Technology, Reykjavik University, Reykjavik, Iceland, <sup>2</sup>Department of Engineering, Reykjavik University, Reykjavik, Iceland, <sup>3</sup>Department of Computer Science, Reykjavik University, Reykjavik, Iceland, <sup>4</sup>Landspítali—The National University Hospital of Iceland, Reykjavik, Iceland

**Introduction:** The use of the electrodermal activity (EDA) signal for health diagnostics is becoming increasingly popular. The increase is due to advances in computational methods such as machine learning (ML) and the availability of wearable devices capable of better measuring EDA signals. One field where work on EDA has significantly increased is sleep research, as changes in EDA are related to different aspects of sleep and sleep health such as sleep stages and sleep-disordered breathing; for example, obstructive sleep apnoea (OSA).

**Methods:** In this work, we used supervised machine learning, particularly the extreme gradient boosting (XGBoost) algorithm, to develop models for detecting sleep stages and OSA. We considered clinical knowledge of EDA during particular sleep stages and OSA occurrences, complementing a standard statistical feature set with EDA-specific variables.

**Results:** We obtained an average macro F1-score of 57.5% and 66.6%, depending on whether we considered five or four sleep stages, respectively. When detecting OSA, regardless of the severity, the model reached an accuracy of 83.7% or 78.4%, depending on the measure used to classify the participant's sleep health status.

**Conclusion:** The research work presented here provides further evidence that, in the future, most sleep health diagnostics might well do without complete polysomnography (PSG) studies, as wearables can detect well the EDA signal.

## KEYWORDS

sleep, electrodermal activity, sleep stages, obstructive sleep apnea, machine learning

## 1. Introduction

Electrodermal activity (EDA) is one of the longest-known and most accessible physiological signals (Boucsein, 2012). Electrodermal activity reflects changes in skin potential due to sweating, which, during sleep, has a thermoregulatory function. Eccrine sweat glands, the sweat glands that are activated during sleep (Boucsein, 2012), are innervated by the sympathetic nervous system (SNS) only, with no parasympathetic input (Baker, 2019). Despite this direct connection between EDA and the SNS during the night, the signal has been so far mostly used in studies of diurnal phenomena. For instance, it has been used for detecting stress (Zontone et al., 2019), epileptic seizures (Poh et al., 2010), and students' emotional engagement in classrooms (Di Lascio et al., 2018).

One of the main reasons for neglecting EDA in sleep studies is the complexity of the recorded signals. Long-term EDA recordings are susceptible to noise from various sources

that cause artifacts in the signals, that is, sudden out-of-scale spikes; the most prominent sources of noise are body movements and poor skin-to-electrode connection. While in laboratory-controlled settings it is possible to log the patient's movements and to discard those signal segments when analyzing data, in free-living conditions, it is more difficult to do so. Because removing artifacts is important, much of the research on EDA signals has focused on automating their detection. Various methods have been proposed, often using supervised or unsupervised machine learning (ML) algorithms (Taylor et al., 2015; Hossain et al., 2022; Subramanian et al., 2022). Electrodermal activity has been only scarcely and only recently used for sleep staging or to infer sleep quality (Anusha et al., 2022; Gashi et al., 2022).

Abnormal sweating patterns may indicate the presence of various sleep disorders (Broman and Hetta, 1994; Idiaquez et al., 2022). In this work, we focused on sleep-breathing disorders, particularly obstructive sleep apnoea (OSA) (Jordan et al., 2014). Obstructive sleep apnoea causes unexpected SNS activity, resulting in frequent nocturnal sweating (Arnardottir et al., 2013). Despite the relationships between EDA and OSA has been studied (Lajos, 2004; Arnardottir et al., 2010), there is still a need for a quantitative model relating EDA and OSA.

In this paper, we applied supervised ML to EDA data to predict sleep stages and the presence of OSA. Currently, diagnosing it requires performing a full polysomnography (PSG) study in a laboratory setting, followed by manual scoring of the recordings. This procedure is time-consuming and can lead to atypical sleep patterns because of the differences between sleeping in a controlled environment, such as a sleep lab, and sleeping at home (Arnardottir et al., 2021). We present an ML-based approach that uses features extracted from the EDA signal, recorded in a home-setting, to automatically detect sleep stages and OSA.

## 2. Materials and methods

We used a set of 60 full-night PSG recordings from participants in the Sleep Revolution Project (Arnardottir et al., 2022). We describe the cohort in detail in Table 1. The consent of the National Bioethics Committee and the Data Protection Authority of Iceland was granted for this study (VSN-21-070). All participants received and signed an informed consent for study participation.

### 2.1. Instrumentation

Polysomnography (PSG) studies were recorded using A1 devices from Nox Medical (Reykjavik, Iceland). As the traditional PSG setup does not include EDA recordings, we added a channel for the EDA signal. A1 devices measured EDA at a sampling frequency of 200 Hz. For the measurement of the EDA signal, we used the same technique as in Arnardottir et al. (2010).

### 2.2. Sleep stage labeling

Sleep experts manually scored the electroencephalogram (EEG) and determined the sleep stage: wake (W), rapid eye movement

TABLE 1 Dataset content according to the apnoea-hypopnoea index (AHI) or the oxygen desaturation index (ODI).

	Non-OSA	Mild OSA	Moderate to severe OSA
Number of participants (AHI)	19	24	17
Female participants (AHI)	47.4%	67.0%	29.4%
AHI	2.8 ± 1.3	10.0 ± 2.8	24.9 ± 10.5
BMI	25.8 ± 3.6	26.0 ± 3.6	25.8 ± 3.8
Age	36.2 ± 10.4	49.6 ± 14.7	52.0 ± 14.4
Percentage of epochs (AHI)	32.1%	39.0%	28.9%
Number of participants (ODI)	21	26	13
Female participants (ODI)	42.9%	61.5%	38.5%
ODI	1.5 ± 2.5	9.0 ± 2.5	24.1 ± 7.7
BMI	25.8 ± 3.8	27.7 ± 4.5	29.2 ± 2.8
Age	38.4 ± 12.4	48.2 ± 14.8	53.9 ± 14.0
Percentage of epochs (ODI)	33.9%	44.7%	21.4%

TABLE 2 Distribution of sleep stages for 4 and 5 stages architectures.

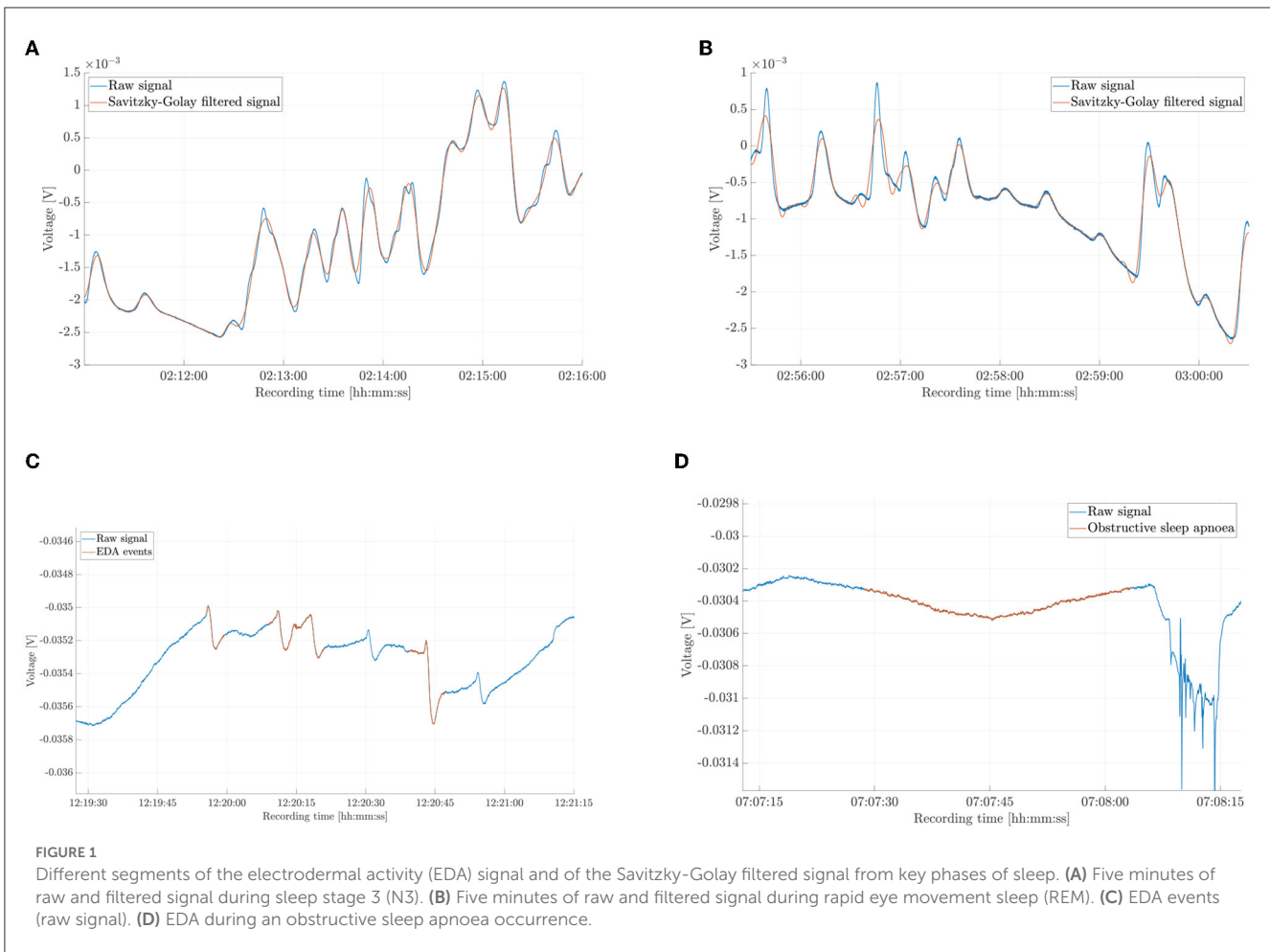
Wake	N1	N2	N3	REM
12.2% ± 0.1	16.5% ± 0.1	32.5% ± 0.1	18.2% ± 0.1	20.6% ± 0.1
Wake	Light		Deep	REM
12.2% ± 0.1	49.0% ± 0.1		18.2% ± 0.1	20.6% ± 0.1

We report mean values and standard deviations.

(REM) sleep, sleep stage 1 (N1), sleep stage 2 (N2), and sleep stage 3 (N3). The scoring procedure was performed according to the American Academy of Sleep Medicine guidelines (Berry et al., 2020), using the Noxturnal software (Nox Medical, Reykjavik, Iceland). In this work, for detection, we considered both the above mentioned five stages or only four stages, by merging the N1 and N2 stages and relabeling them as light sleep. Additionally, we relabeled the N3 stage as deep sleep. The stages that we considered are then W, light sleep, deep sleep, and REM sleep, as is often done in the literature (Genzel et al., 2014). We report the distribution of sleep stages in Table 2.

### 2.3. Obstructive sleep apnoea labeling

Currently, OSA detection requires either manual scoring of a full PSG study or a home sleep apnoea testing, and the evaluation of two parameters: the apnoea-hypopnoea index (AHI) and the oxygen desaturation index (ODI) per hour of sleep (Berry et al., 2020). A shortcoming of the AHI is that it does not quantify one of the main consequences of OSA, which is oxygen desaturation. For this reason, sleep experts have defined the ODI value as the number of oxygen desaturation events  $\geq 3\%$



or  $\geq 4\%$  divided by the total sleep time (Chung et al., 2012; Berry et al., 2020). In this work, the sleep experts used 3% as threshold value.

We obtained a participant’s OSA status from the manual scoring of PSG. We merged the moderate and severe OSA conditions to obtain three classes. To define them, we used the following modified version of the standard guidelines (AASM, 1999):

- Non-OSA: AHI  $< 5$ ,
- Mild OSA:  $5 \leq$  AHI  $< 15$ ,
- Moderate to severe OSA: AHI  $\geq 15$ .

We also classified the samples based on the ODI and computed the correlation between the two indexes and the EDA signal. Note that the ranges used for the ODI-based classification are the same as the ones for the AHI classification (Chung et al., 2012). Each epoch in an individual’s data sample was labeled as either belonging to a non-OSA participant, one with mild OSA, or one with moderate to severe OSA. By epoch, we refer to a 30 s signal window. We adopted this time length to be consistent with the epochs’ length used by sleep experts during manual scoring. Note that only seven samples were classified differently depending on whether we used the AHI or the ODI. Finally, we present the distribution of non-OSA, mild OSA, and moderate to severe OSA epochs in Table 1.

## 2.4. Signal pre-processing

From the Noxturnal software environment, we exported EDA signals using the EDF file format and imported them in MATLAB® (MATLAB, 2022) for pre-processing and feature extraction. We down-sampled the original signal from 200 to 35 Hz to reduce the computational burden, following the guidelines presented in Braithwaite et al. (2013). We then pre-processed the original signal to obtain different kinds of data required by our detection algorithm.

First, because individual sweating patterns lead to significantly different-looking EDA signals (Boucsein, 2012), we computed the second-order polynomial best approximating the raw signal and subtracted it from the raw signal. Second, we applied a seventh-order Savitzky-Golay filter (Schafer, 2011) to the original signal to eliminate high-frequency contributions. We also applied a discrete wavelet transform (DWT) to the original signal. We computed the approximate and detailed discrete wavelet coefficients and soft thresholded the detail coefficients to remove possible recording noise (Coifman and Donoho, 1995). We then subtracted the Savitzky-Golay filtered signal from the discrete wavelet filtered signal; we referred to it as diffEDA.

Third, we computed the first and second-order derivatives of previously described three signals using a differentiator finite impulse response (FIR) filter. We used this method rather

**TABLE 3** Set of variables extracted from the electrodermal activity (EDA) signal.

Index	Signal	Computed features
1–18	EDA detEDA	Mode, median, maximum of absolute value, line length, 10 <sup>th</sup> quantile, 75 <sup>th</sup> quantile, singular value decomposition (SVD) entropy, non-linear energy, Shannon entropy
19–34	$\partial_t$ EDA, $\partial_t^2$ EDA $\partial_t$ detEDA, $\partial_t^2$ detEDA	Mean value, variance, median value, numbers above zero
35–40	EDA detEDA	Maximum power spectral density (PSD) estimate, frequency of the maximum PSD estimate, Fisher's g (Posada-Quintero et al., 2016)
41–64	EDA detail coefficients decomposition levels (DL) 1–4	Maximum, mean, standard deviation, median, Euclidian norm, normalized numbers above zero
65–70	EDA detEDA	Lyapunov exponent, maximum value of the upper envelope, minimum value of the lower envelope
71–72	diffEDA	Sum of cross-correlation, maximum convolution value
73–76	EDA	Normalized number of event samples, normalized event energy, normalized number of storm samples, normalized storm energy
77	Individual	Sex

than a finite-differences scheme to prevent noise propagation. Particularly, we used a 50th-order filter with a passband frequency of 10 Hz and a stop-band frequency of 12.5 Hz. We disregarded the transient to avoid including artificial oscillations caused by applying the filter by discarding  $N = 50$  samples. Note that, we denoted time derivatives by placing  $\partial_t$  or  $\partial_t^2$  before the signal of interest; for example, we referred to the second time derivative of the de-trended signal as  $\partial_t^2$ detEDA.

Figure 1 shows the complexity of the EDA signal. We show 5-min time windows of continuous N3 and REM sleep in Figures 1A, B, respectively. We then highlight EDA events in Figure 1C. Finally, we show the EDA signal during an OSA occurrence in Figure 1D.

## 2.5. Feature extraction and selection

We defined a feature set in the time-domain, frequency-domain, as well as time-frequency domain (these are wavelet-related variables) in a process called feature engineering (Verdonck et al., 2021). In addition to standard statistical features, we used number and energy content of EDA events and storms, as they are known to differ for different sleep stages (Sano et al., 2014) and OSA severity (Arnardottir et al., 2010). Electrodermal activity events are oscillations of the skin voltage of defined amplitudes and frequencies. We are particularly interested in the following three types of oscillations: positive/negative monophasic, biphasic, and triphasic. Electrodermal activity storms are time windows with high concentrations of events. The definition of storms has changed through time (Burch, 1965; Sano et al., 2014), we used an equivalent

definition to the one given by Sano and colleagues, that is, a timespan of at least 1 min with a minimum of two EDA events. Particularly, we used the algorithm developed in Piccini et al. (2023) to detect EDA events and storms. Thereafter, we computed the normalized number of samples within either an EDA event or storm, together with their Euclidean norms. Additionally, we added sex as a categorical feature to complete the set of variables and normalized the features across individuals. The full feature set is shown in Table 3.

Finally, after training and testing the model on the complete variable set, we investigated whether we could reduce the feature set dimension by analyzing intra-variable correlation. We identified correlated features by computing the pairwise Pearson correlation coefficient  $r$ . We then reduced the dimension of the feature set by retaining only one of the correlated variables. We looked at the correlation matrix to identify the threshold value  $r_{th}$ .

## 2.6. Training procedure

Sleep stages are not equally distributed during the night, this asymmetry caused a significant imbalance in our dataset and affected model performance. To reduce the negative impact of this effect, we performed synthetic minority oversampling (SMOTE) (Chawla et al., 2002), that is, we generated artificial samples for the minority classes to alleviate the bias toward the most dominant class. We then trained models using the extreme gradient boosting (XGBoost) algorithm (Chen and Guestrin, 2016), since a gradient boosting algorithm was recently used in a similar application with promising results (Gashi et al., 2022).

We applied different validation methods. We either used leave-one-subject-out (LOSO) validation (Hastie et al., 2009), where we alternately left out one sample and used the other 59 samples as training data, or we did as previously and in addition, we trained the model using randomly selected 25% of the epochs from the left-out subject's night (Personalized). We always used the same seed for reproducibility. After this random sampling, we applied the SMOTE algorithm to the training data. We evaluated the OSA model only by means of the LOSO scheme. We did so, because of the way that we labeled the data for OSA detection, see Section 2.3.

## 2.7. Evaluation metrics

We computed different measures to evaluate the models' performances. All indices were obtained using scikit-learn (Pedregosa et al., 2011). F1 and recall scores were used to evaluate the sleep staging performances. While F1-score is a commonly used measure in ML applications, we used the recall score to account for the significant class imbalance (Gashi et al., 2022). Recall score is the ratio between true positives and the sum of true positives and false negatives and, thus, a measure for the number of relevant objects detected by the algorithm. The F1-score is the harmonic mean of precision and recall scores and is used in classification problems with imbalanced datasets, as the precision score on its own may be misleading. As we dealt with a multi-class classification problem, we used the macro version of

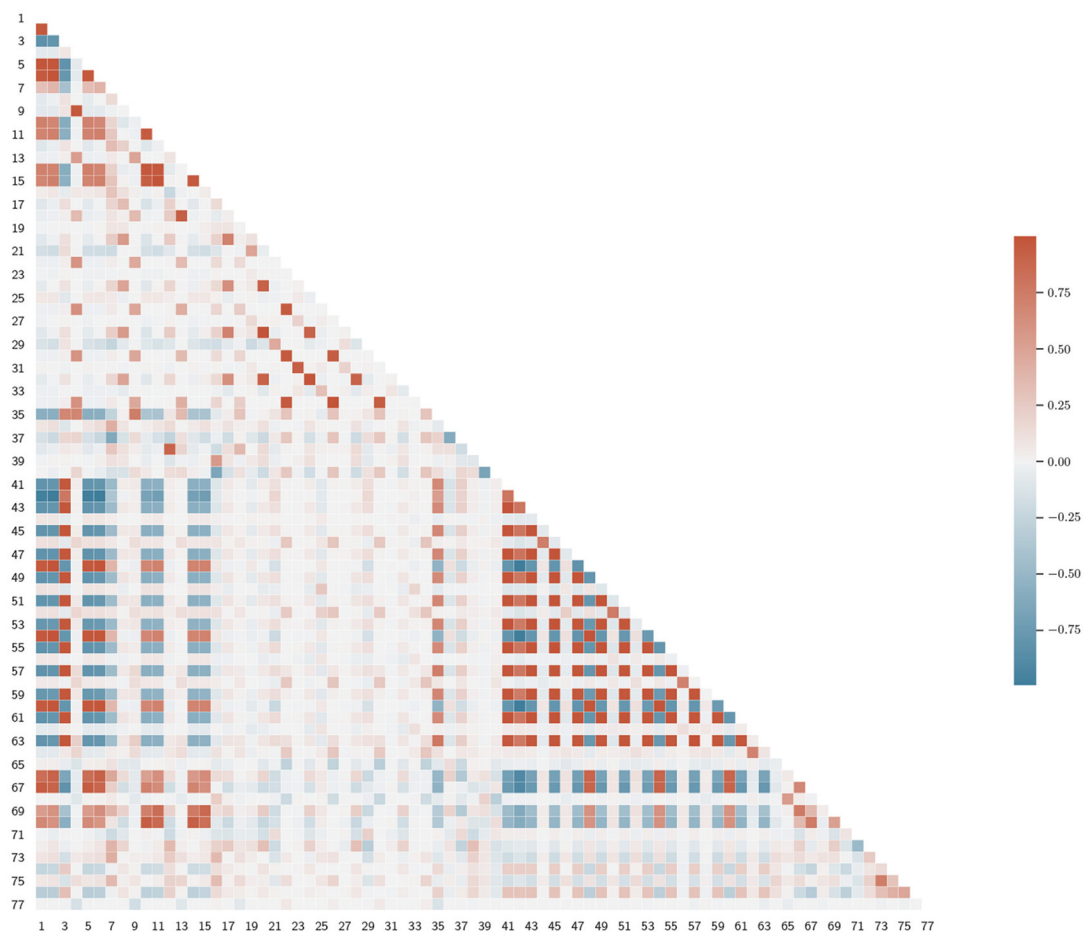


FIGURE 2

Representation of the lower triangular feature correlation matrix. We denoted the variables by index, as in Table 3.

both parameters; the macro F1-score is the average of all F1-scores, and the macro recall is the average of all recalls. For the remainder of the paper, we referred to the macro F1-score and macro recall value simply as F1-score and recall.

For the OSA model, we used the F1-score and accuracy values. Accuracy is the ratio between the number of correctly identified epochs and the total amount of epochs. In addition to these two measures, we evaluated the three-class algorithm's ability to distinguish between non-OSA participants and those with OSA. To do this, we considered all OSA epochs as equivalent, which made the classification problem a binary one; we then computed the accuracy score and referred to it as the adjusted accuracy score. We did not include the recall for OSA models' evaluations, as it deals with all misclassifications in the same way. Particularly, misclassifications between severe and mild OSA conditions and between non-OSA and OSA conditions have different clinical meanings.

## 2.8. SHapley Additive exPlanations

To evaluate the contribution of each training variable, we used SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017).

The technique was developed in game theory and only recently adapted to ML interpretability applications (Lundberg et al., 2018). To find the SHAP value of the  $i$ -th variable, we computed the predictions for all possible feature combinations with and without the  $i$ -th variable. The SHAP value is then the average of the contributions of the  $i$ -th variable to each prediction (Molnar, 2022).

## 3. Results and discussion

### 3.1. Feature reduction

Before presenting the models' performances, we offer an analysis of the feature set reduction process; for the sake of notation, we refer to variables by index, as in Table 3. We identified three main clusters of correlated variables by looking at the graphical representation of the correlation matrix (Figure 2). The first one involves features 1–15, which are statistical measures, in the time domain, of EDA and detEDA signals. The second one is a large cluster encompassing features 41–64, variables obtained in the time-frequency domain. Finally, non-linear features 65–70 also show meaningful correlation patterns.

After setting  $r_{th} = 0.8$ , we reduced the number of correlated variables. We decided which feature to eliminate as follows: first,



TABLE 4 Optimized feature set for the sleep staging models,  $r_{th} = 0.8$ .

Signal	Computed features
EDA	Mode, maximum of absolute value, line length, singular value decomposition (SVD) entropy, non-linear energy, Lyapunov exponent, maximum power spectral density (PSD) estimate, frequency of the maximum PSD estimate, Fisher's g (Posada-Quintero et al., 2016)
detEDA	Mode, maximum of absolute value, line length, singular value decomposition (SVD) entropy, non-linear energy, Lyapunov exponent, frequency of the maximum PSD estimate, Fisher's g (Posada-Quintero et al., 2016)
$\partial_t$ EDA	Mean, variance, median, number above zero
$\partial_t^2$ EDA	Mean, median
$\partial_t$ detEDA	Mean, median
$\partial_t^2$ detEDA	Median
EDA detail coefficients decomposition levels (DL) 1–4	Median, normalized numbers above zero
diffEDA	Sum of cross-correlation, maximum convolution value
EDA	Normalized number of event samples, normalized event energy, normalized number of storm samples, normalized storm energy
Individual	Sex

we computed the correlation coefficients between the  $i$ -th feature and the remaining ones, then we eliminated the  $j$ -th feature, if  $r_{i,j} > r_{th}$ , where  $r_{i,j}$  is the Pearson correlation coefficient between the  $i$ -th and the  $j$ -th variables and  $j > 1$ . We started at  $i = 1$ . In this way, we obtained a reduced set of 40 features, which we present in Table 4. We opted not to decrease further the  $r_{th}$ -value, as the resulting feature set did not present any significant clusters, see Figure 3. Also, lower values of  $r_{th}$  may result in worse classification performances.

### 3.2. Interpretation of sleep staging

We summarized the models' performances in Table 5, where the F1-scores and recall values are reported. Our results suggest a need for personalized models (Óskarsdóttir et al., 2022). A possible explanation for the relatively poor performance is that different brain regions can be in different sleep stages at the same time. For instance, sweat glands' activation signals and, thus, EDA, are generated in the hypothalamus (Rothhaas and Chung, 2021), while the EEG, used to manually label sleep stages, measures neocortex activity, and it is known that the two brain areas can be in different sleep stages (Guthrie et al., 2022). However, personalizing the LOSO-based model with a small number of epochs from the left-out sample dramatically improves the algorithm, see also Figures 4, 5, which show confusion matrices normalized such that the sum of each row equals one.

By looking at the confusion matrix in Figure 4B, we concluded that the personalized model cannot characterize the N1 sleep stage using only EDA. Furthermore, N1 detection appeared

to be a cumbersome task even when other ML methods and other signals were used, such as EEGs, electrooculograms (EOGs), and electromyograms (EMGs) (Chambon et al., 2018; Korkalainen et al., 2019). A similar disagreement in determining the sleep stage was also found when comparing different manual scorings (Magalang et al., 2013). However, the detection of slow wave sleep (SWS) phases, that is, deep sleep and N3 stage, and REM sleep phases worked well for both models. This was expected, since these are the phases with the most distinct EDA patterns. Notably, by looking at Figure 4B, we can conclude that, based on EDA, the N3 stage is more similar to the N2 stage than any other sleep stage.

Finally, we offer a graphical interpretation of the sleep staging model, trained on the reduced dataset, through the SHAP values of the 20 most influential variables, see Figure 6. It is worth noting that both models considered the number of EDA events to be highly relevant for N3 stage, see Figure 6. The models also predict a significant relationship between EDA storms and REM sleep. Indeed, it is known that EDA activity increases in the third cycle of REM sleep (Boucsein, 2012).

### 3.3. The need for personalization in sleep staging

Several physiological considerations support the need for personalization in EDA-based sleep staging. Nocturnal sweat, the principal cause of changes in skin electrical properties, is secreted to lower the core body temperature (CBT) (Baker, 2019). However, the thermoregulation process depends on a large number of factors, for example, age, BMI, sex, skin hydration, eccrine sweat gland concentration, and environmental conditions (Speakman, 2018; Grosiak et al., 2020; Yanovich et al., 2020). All the factors mentioned significantly impact sweat and, consequently, the EDA signal. Furthermore, the latter is also affected by subject-dependent brain dynamics.

It is not straightforward to decide which personal subset of epochs to choose, as different EDA patterns arise in different parts of the night; for example, EDA events are more frequent in REM sleep during the last sleep cycle (Boucsein, 2012), while rarer in other REM sleep periods. Furthermore, differences in sleep cycle duration caused by age and OSA condition, among other factors, may hinder the beneficial effect of the algorithm's personalisation. Because of this, we opted for a fixed-seed random-pick approach.

### 3.4. Interpretation of the OSA model

To evaluate the models' ability to distinguish between non-OSA persons and those with either mild or moderate to severe OSA, we used average values of the accuracy score, the F1-score, and the adjusted accuracy score. We also evaluated a binary classification problem, where participants either had OSA or not, for which we refrained from calculating the adjusted accuracy score. We present the results as we did for the sleep staging models in Table 6. They show that OSA severity determined through the EDA signal rather follows the classification obtained by using the ODI rather than the AHI. A possible explanation for this behavior is how the ODI

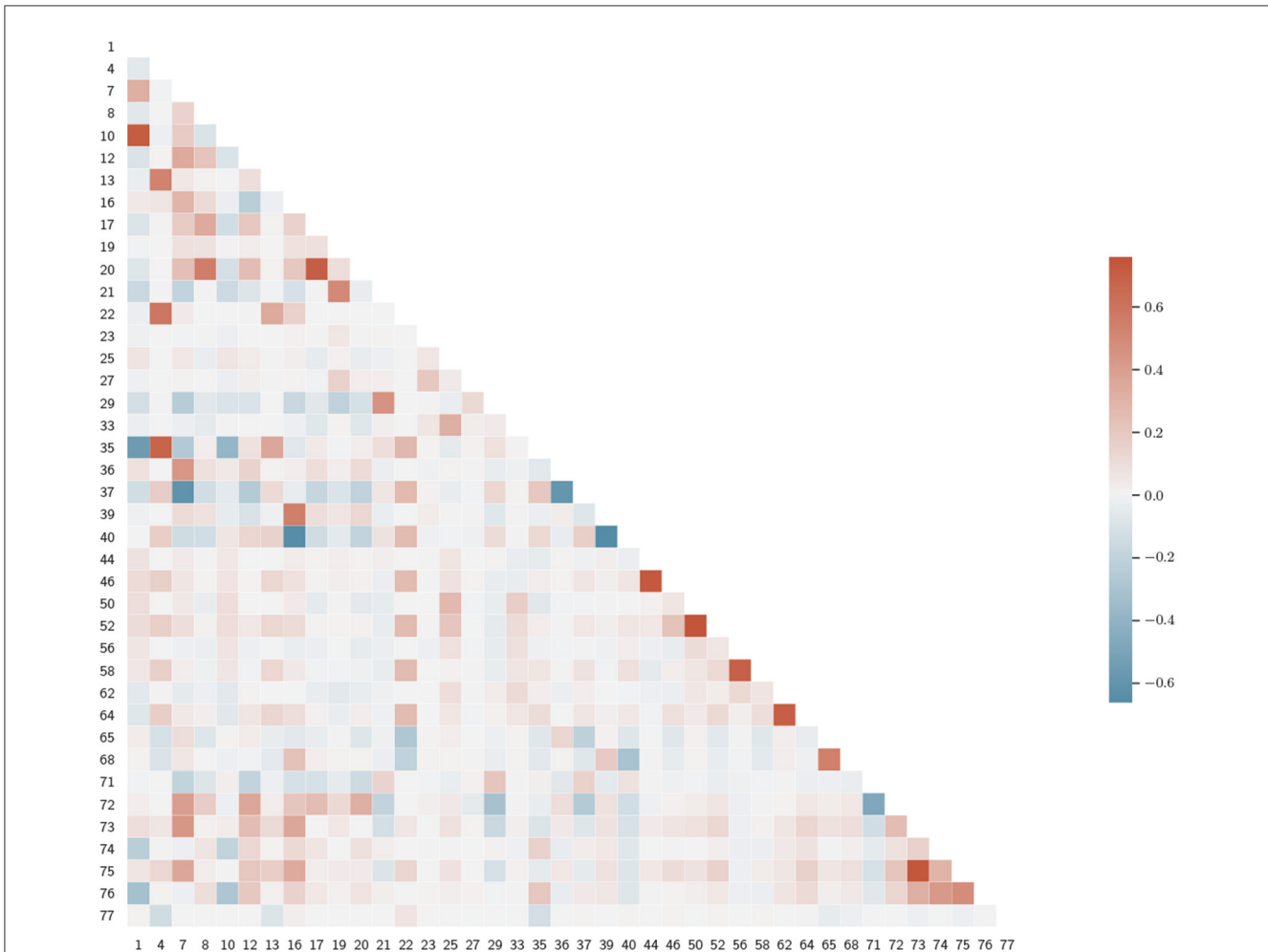


FIGURE 3 Correlation matrix for the reduced feature set. Variables are denoted by index, as in Table 3.

TABLE 5 Summary of sleep staging performance for, both, four stages, and five stages classification.

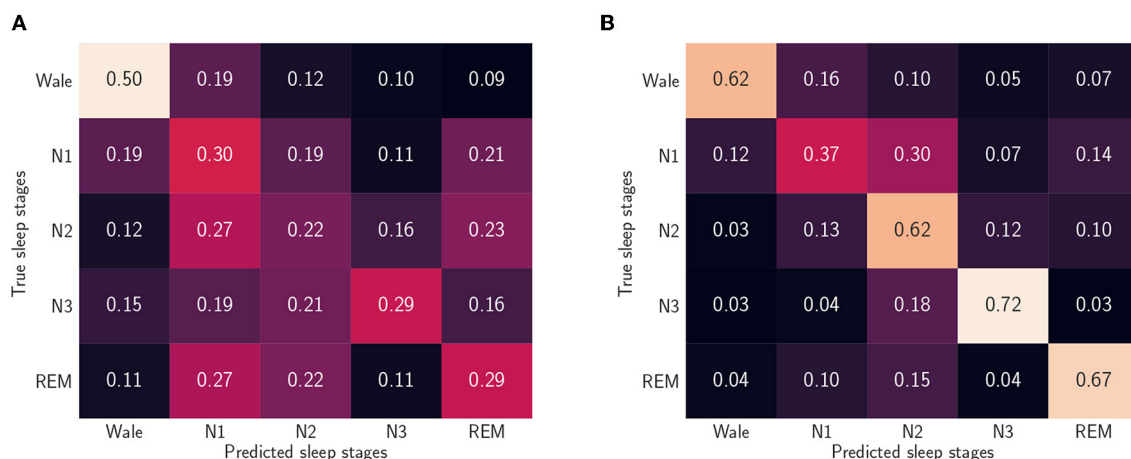
No. stages	r <sub>th</sub>	Leave-one-subject-out		Personalized	
		Macro F1-score	Macro recall score	Macro F1-score	Macro recall score
Five stages	0.8	27.3%	32.4%	57.5%	58.0%
Four stages	0.8	32.8%	39.7%	66.6%	66.9%

value divides the participants. Looking at Table 1, we observed that while both the indexes found the mean age to increase with the OSA severity, in ODI classification BMI values also increased with OSA severity. Lower BMI values have been associated with lower mean temperature values, (Walen and Buxbaum, 2011), which may result in less need for thermoregulation. Consequently, simpler sweating patterns may be observed, which are better learned by the algorithm.

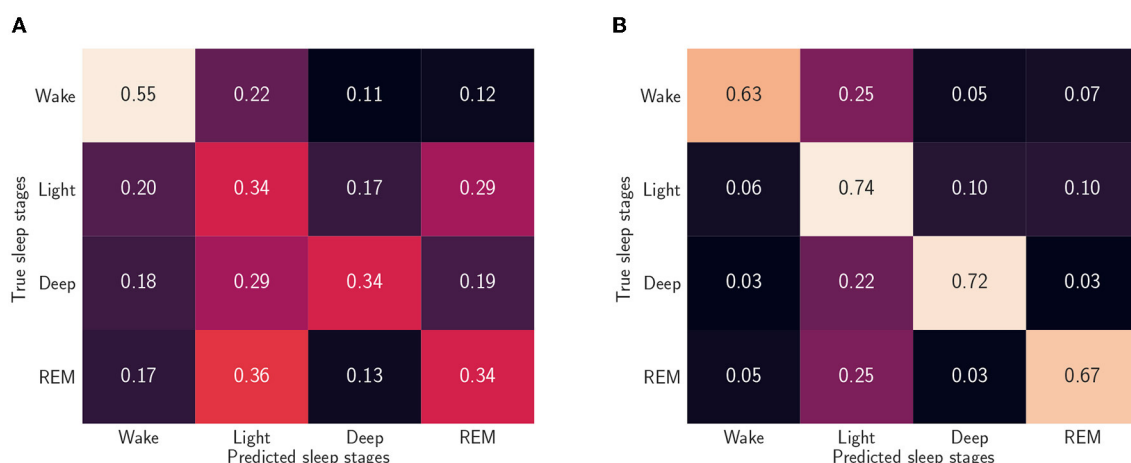
In Figures 7, 8, we used the SHAP values to present the effect of each variables on the different classification problem. Both three-class models choose normalized storm samples as one of the most significant variables, which relates well to the literature (Arnardottir et al., 2010).

### 3.5. Feature selections comparison

Out of the 77 extracted variables, only eight appear in all models' top 20 most important features. They are EDA mode,  $\partial_t$ EDA variance, detEDA mode, EDA maximum power spectral density (PSD) estimate, EDA frequency of the maximum PSD estimate,  $\partial_t$ EDA normalized numbers above zero, detEDA frequency of the maximum PSD estimate, and biological sex. The seven numerical variables are computed from two signals, that is, raw and de-trended EDA and the derivative of the raw signal; this subset is composed of variables spanning multiple domains, particularly time, frequency, and EDA-specific. This variety confirms the need to consider different dynamical behaviors



**FIGURE 4** Normalized confusion matrices when we consider five sleep stages and use the reduced feature set. **(A)** Leave-one-subject-out (LOSO). We trained the algorithm without including data from the left out participant. **(B)** Personalized model. In addition to the 59 participants training set, we used randomly picked 25% epochs of the test participants.



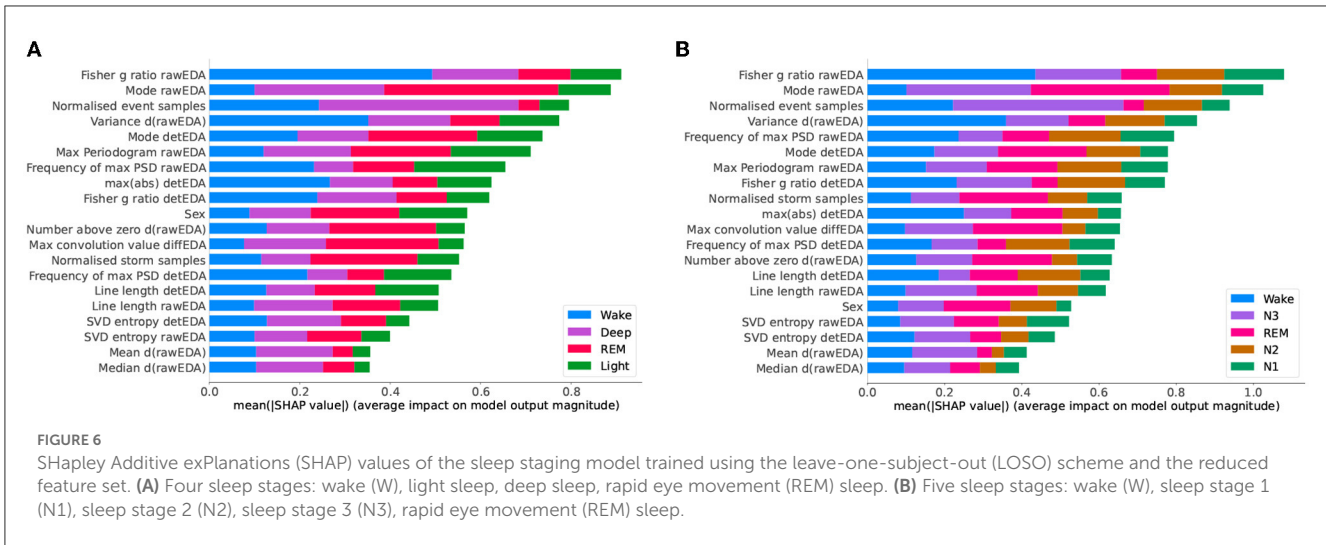
**FIGURE 5** Normalized confusion matrices when we consider four sleep stages and use the reduced feature set **(A)** Leave-one-subject-out (LOSO). We trained the algorithm without including data from the left out participant. **(B)** Personalized model. In addition to the 59 participants training set, we used randomly picked 25% epochs of the test participants.

and EDA-related phenomena when using this signal. The most common specific feature is the number of EDA storm samples, which is amongst the top 20 most important features for all models, except for the two-class ODI-based OSA classification problem. However, in the latter problem, normalized storm energy is considered a relevant feature. Works trying to relate EDA and OSA are scarce and based mainly on subjective night sweats reports (Nigro et al., 2022). Although it is well-established that OSA symptomatology includes abnormal sweating episodes (Arnardottir et al., 2010, 2013), there needs to be more understanding of the relationship between OSA and EDA events and storms. Our work concludes that evaluating EDA storms, their lengths or energies, is more decisive in detecting OSA, particularly severe expressions, than evaluating EDA events. This conclusion holds for OSA classifications based on both AHI and ODI severity.

### 4. Conclusion and future work

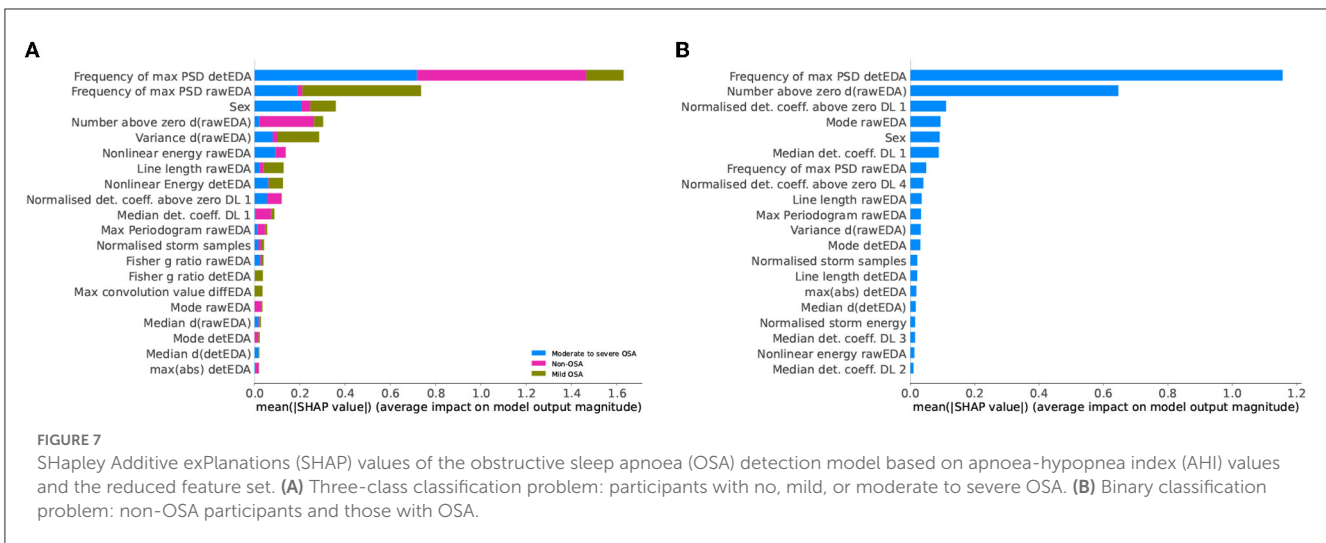
The presented work aimed at detecting sleep stages and OSA severity using only the EDA signal. Recently, Anusha and colleagues presented an ML algorithm for identifying the sleep stage of the hypothalamus, the brain region directly responsible for thermoregulation during sleep (Anusha et al., 2022), while, Gashi and colleagues presented a similar algorithm based on EDA that is able to detect wake/sleep stages and high/low sleep quality (Gashi et al., 2022). Latter algorithms are based on self-reported annotations. Despite these significant results, more research on the relationship between EDA and neocortex activity is needed. Our work is the first one, in which neocortex sleep stages are predicted solely based on EDA. In the first part of this research work, we presented a sleep staging algorithm that is





**TABLE 6** Results for obstructive sleep apnoea (OSA) detection, based on the apnoea-hypopnoea index (AHI) or on the oxygen desaturation index (ODI).

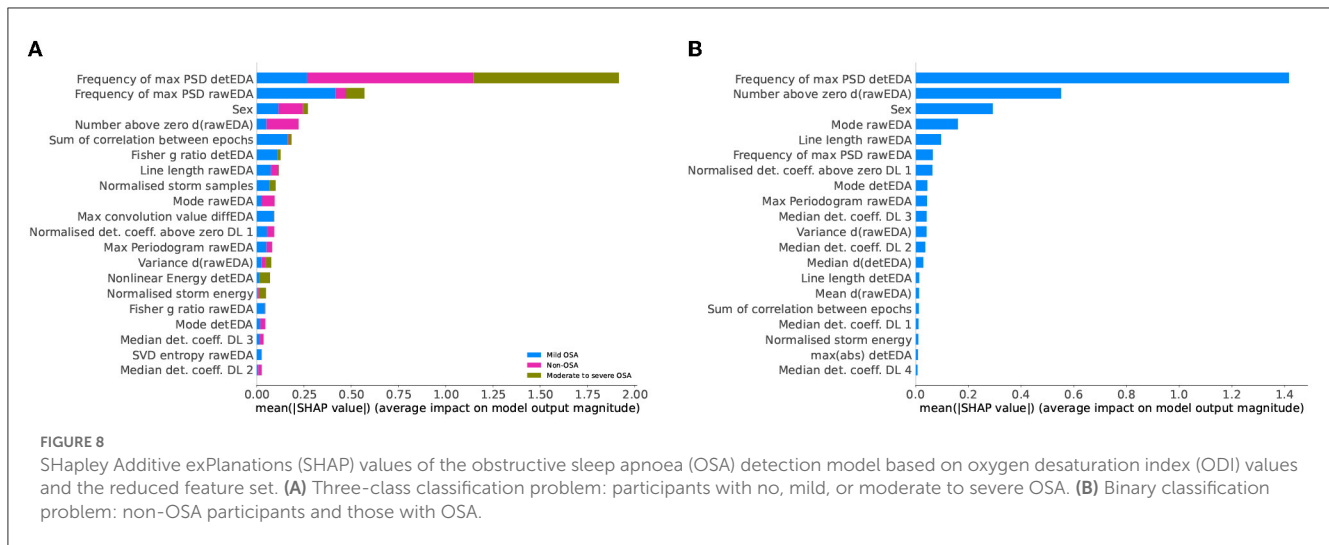
OSA structure	$r_{th}$	Mean accuracy score	Macro F1-score	Adj. accuracy score
AHI—non-OSA vs. OSA	0.8	75.7%	65.6%	–
ODI—non-OSA vs. OSA	0.8	82.0%	67.7%	–
AHI—Three groups	0.8	54.8%	32.9%	78.4%
ODI—Three groups	0.8	54.8%	32.9%	83.7%



particularly accurate in detecting those sleep stages, where specific EDA patterns are known to occur, which are N3 and REM sleep. In the second part, we focused on OSA detection. By using the EDA signal, we distinguished non-OSA participants from those with OSA with reasonable accuracy.

Our work has three main limitations. The first one is that the raw signal was recorded at 200 Hz, an unattainable sampling frequency for current wearables. However, the signal was significantly downsampled, to 35 Hz, before it was handled. Since EDA events occur in the frequency band [0.25–3 Hz] for endosomatic recordings, like the ones used in this study, further downsampling might potentially be performed without a

significant loss of information, which we leave for future work. The second limitation is that the sleep staging algorithm requires a certain amount of individual data manually scored by a sleep expert. While this prevents the sleep staging model from being user-independent and, thus, might limit its use in wearables, in clinical studies, requiring only a small part of the signal to be manually scored significantly saves time and cost. Moreover, our work adds to the body of evidence on how crucial it is to include knowledge about sleep processes in ML models. A final limitation is the participants' significant ranges in age and BMI within a relatively small sample size. While the participants' diversity ensured to obtain general models, it also prevented the algorithm



from learning patterns specific to a particular group, for example, individuals of the same biological sex and of similar age. Future studies may overcome this last limitation by using a more selected cohort or by considering the body temperature signal, therefore addressing the differences in mean body temperature due to various aspects such as age, sex, and BMI.

To improve on the reported results, in the future, we will also include additional signals obtainable through wearables, such as acceleration and skin temperature. Doing so might reduce the need for individual tuning of the algorithm and allow it to identify other sleep stages more accurately. More precise sleep staging based on data obtained from wearables will allow the estimation of more advanced sleep parameters used in sleep diagnostics, such as total sleep time and sleep efficiency. Finally, since the algorithm labels each epoch as “non-OSA” or “OSA-prone,” it will be possible to track a potential onset or worsening of sleep-disordered breathing. By adequately characterizing the development of OSA symptoms, it will be possible to define a threshold that will lead to suggesting to seek professional advice when exceeded.

## Data availability statement

The datasets used for this study is not publicly available due to General Data Protection Regulation (GDPR) reasons. Requests to access the datasets should be directed to [jacopop@ru.is](mailto:jacopop@ru.is).

## Ethics statement

The study received the approval of the National Bioethics Committee and the Data Protection Authority of Iceland (Sleep Revolution VSN-21-070). The patients/participants provided their written informed consent to participate in this study.

## Author contributions

JP: conceptualization, methodology, software, visualization, and writing. EA and MÓ: conceptualization, methodology,

supervision, reviewing, and editing. ESA: supervision, reviewing, editing, and funding acquisition. All authors contributed to the article and approved the submitted version.

## Funding

The study was carried out within the Sleep Revolution Project (Arnardottir et al., 2022). This project has received funding from European Union’s Horizon 2020 Research and Innovation Programme under grant agreement no. 965417.

## Acknowledgments

All authors would like to express their outmost gratitude to Tiina Siilak and Heiður Grétarsdóttir, expert sleep technologists, who painstakingly scored all sleep-related signals used in this work.

## Conflict of interest

ESA discloses lecture fees from Nox Medical, Philips, ResMed, Jazz Pharmaceuticals, Linde Healthcare, Alcoa-Fjordaral, and Wink Sleep. ESA is also a member of the Philips Sleep Medicine and Innovation Medical Advisory Board.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- AASM (1999). Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The report of an American Academy of Sleep Medicine Task Force. *Sleep* 22, 667–89.
- Anusha, A. S., Preejith, S. P., Akl, T. J., and Sivaprakasam, M. (2022). Electrodermal activity based autonomic sleep staging using wrist wearable. *Biomed. Signal Process. Control* 75, 103562. doi: 10.1016/j.bspc.2022.103562
- Arnardóttir, E. S., Islind, A. S., and Óskarsdóttir, M. (2021). The future of sleep measurements: a review and perspective. *Sleep Med. Clin.* 16, 447–464. doi: 10.1016/j.jsmc.2021.05.004
- Arnardóttir, E. S., Islind, A. S., Óskarsdóttir, M., Ólafsdóttir, K. A., August, E., Jónasdóttir, L., et al. (2022). The sleep revolution project: the concept and objectives. *J. Sleep Res.* 31, e13630. doi: 10.1111/jsr.13630
- Arnardóttir, E. S., Janson, C., Björnsdóttir, E., Benediksdóttir, B., Juliusson, S., Kuna, S. T., et al. (2013). Nocturnal sweating—a common symptom of obstructive sleep apnoea: the icelandic sleep apnoea cohort. *BMJ Open* 3, e002795. doi: 10.1136/bmjopen-2013-002795
- Arnardóttir, E. S., Thorleifsdóttir, B., Svanborg, E., Ólafsson, I., and Gíslason, T. (2010). Sleep-related sweating in obstructive sleep apnoea: association with sleep stages and blood pressure. *J. Sleep Res.* 19(1 Pt 2), 122–130. doi: 10.1111/j.1365-2869.2009.00743.x
- Baker, L. B. (2019). Physiology of sweat gland function: the roles of sweating and sweat composition in human health. *Temperature* 6, 211–259. doi: 10.1080/23328940.2019.1632145
- Berry, R. B., Quan, S. F., and Abreu, A. R. (2020). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Version 2.6*, Vol. 53. Darien, CT: American Academy of Sleep Medicine.
- Boucsein, W. (2012). *Electrodermal Activity*. Boston, MA: Springer US.
- Braithwaite, J. J., Watson, D. G., Jones, R., and Rowe, M. (2013). *A Guide for Analysing Electrodermal Activity (EDA) and Skin Conductance Responses (SCRs) for Psychological Experiments*. Technical Report, 2nd Version, Selective Attention and Awareness Laboratory (SAAL) Behavioural Brain Sciences Centre, University of Birmingham, UK.
- Broman, J.-E., and Hetta, J. (1994). Electrodermal activity in patients with persistent insomnia. *J. Sleep Res.* 3, 165–170. doi: 10.1111/j.1365-2869.1994.tb00124.x
- Burch, N. (1965). “Data processing of psychophysiological recordings,” in *Symposium on the Analysis of Central Nervous System Data Using Computer Methods*, 165–180.
- Chambon, S., Galtier, M. N., Arnal, P. J., Wainrib, G., and Gramfort, A. (2018). A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 758–769. doi: 10.1109/TNSRE.2018.2813138
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953
- Chen, T., and Guestrin, C. (2016). “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16 (New York, NY: Association for Computing Machinery), 785–794. doi: 10.1145/2939672.2939785
- Chung, F., Liao, P., Elsaid, H., Islam, S., Shapiro, C. M., and Sun, Y. (2012). Oxygen desaturation index from nocturnal oximetry. *Anesth. Analg.* 114, 993–1000. doi: 10.1213/ANE.0b013e318248f4f5
- Coifman, R. R., and Donoho, D. L. (1995). “Translation-invariant de-noising,” *Wavelets and Statistics. Lecture Notes in Statistics, Vol. 103*, eds A. Antoniadis and G. Oppenheim (New York, NY: Springer), 125–150.
- Di Lascio, E., Gashi, S., and Santini, S. (2018). Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1–21. doi: 10.1145/3264913
- Gashi, S., Alecci, L., Lascio, E. D., Debus, M. E., Gasparini, F., and Santini, S. (2022). The role of model personalization for sleep stage and sleep quality recognition using wearables. *IEEE Pervas. Comput.* 21, 69–77. doi: 10.1109/MPRV.2022.3164334
- Genzel, L., Kroes, M. C., Dresler, M., and Battaglia, F. P. (2014). Light sleep vs. slow wave sleep in memory consolidation: a question of global vs. local processes? *Trends Neurosci.* 37, 10–19. doi: 10.1016/j.tins.2013.10.002
- Grosiak, M., Koteja, P., Bauchinger, U., and Sadowska, E. T. (2020). Age-related changes in the thermoregulatory properties in bank voles from a selection experiment. *Front. Physiol.* 11, 576304. doi: 10.3389/fphys.2020.576304
- Guthrie, R. S., Ciliberti, D., Mankin, E. A., and Poe, G. R. (2022). Recurrent hippocampo-neocortical sleep-state divergence in humans. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2123427119. doi: 10.1073/pnas.2123427119
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. New York, NY: Springer.
- Hossain, M. B., Posada-Quintero, H. F., Kong, Y., McNaboe, R., and Chon, K. H. (2022). Automatic motion artifact detection in electrodermal activity data using machine learning. *Biomed. Signal Process. Control*, 74, 103483. doi: 10.1016/j.bspc.2022.103483
- Idiaquez, J., Casar, J. C., Arnardóttir, E. S., August, E., Santini, J., and Iturriaga, R. (2022). Hyperhidrosis in sleep disorders – a narrative review of mechanisms and clinical significance. *J. Sleep Res.* 32, e13660. doi: 10.1111/jsr.13660
- Jordan, A. S., McSharry, D. G., and Malhotra, A. (2014). Adult obstructive sleep apnoea. *Lancet* 383, 736–747. doi: 10.1016/S0140-6736(13)60734-5
- Korkalainen, H., Leppänen, T., Aakko, J., Nikkonen, S., Kainulainen, S., Leino, A., et al. (2019). Accurate deep learning-based sleep staging in a clinical population with suspected obstructive sleep apnea. *IEEE J. Biomed. Health Inform.* 99, 1–1. doi: 10.1109/JBHI.2019.2951346
- Lajos, L. E. (2004). *The Relation Between Electrodermal Activity in Sleep, Negative Affect, and Stress in Patients Referred for Nocturnal Polysomnography*. LSU Doctoral dissertations.
- Lundberg, S. M., Erion, G. G., and Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv Preprints*. arXiv, 1802.03888. doi: 10.48550/arXiv.1802.03888
- Lundberg, S. M., and Lee, S.-I. (2017). “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems 30*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Red Hook, NY: Curran Associates, Inc.), 4765–4774.
- Magalang, U. J., Chen, N.-H., Cistulli, P. A., Fedson, A. C., Gíslason, T., Hillman, D., et al. (2013). Agreement in the scoring of respiratory events and sleep among international sleep centers. *Sleep* 36, 591–596. doi: 10.5665/sleep.2552
- MATLAB (2022). *MATLAB version 9.12.0.1884302 (R2022a)*. Natick, MA: The MathWorks Inc.
- Molnar, C. (2022). *Interpretable Machine Learning*. 2nd Edn. Lean Pub.
- Nigro, C. A., Bledel, I., and Borsini, E. (2022). Independent association between hypoxemia and night sweats in obstructive sleep apnea. *Sleep Breath.* doi: 10.1007/s11325-022-02701-3
- Óskarsdóttir, M., Islind, A. S., August, E., Arnardóttir, E. S., Patou, F., Maier, A. M., et al. (2022). Importance of getting enough sleep and daily activity data to assess variability: longitudinal observational study. *JMIR Form. Res.* 6, e31807. doi: 10.2196/31807
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Piccini, J., August, E., Noel Aziz Hanna, S. L., and Arnardóttir, E. S. (2023). Automatic detection of electrodermal activity events during sleep. *Biomed. Signal Process. Control*.
- Poh, M.-Z., Loddenkemper, T., Swenson, N. C., Goyal, S., Madsen, J. R., and Picard, R. W. (2010). Continuous monitoring of electrodermal activity during epileptic seizures using a wearable sensor. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2010, 4415–4418. doi: 10.1109/IEMBS.2010.5625988
- Posada-Quintero, H. F., Florian, J. P., Orjuela-Cañón, A. D., Aljama-Corrales, T., Charleston-Villalobos, S., and Chon, K. H. (2016). Power spectral density analysis of electrodermal activity for sympathetic function assessment. *Ann. Biomed. Eng.* 44, 3124–3135. doi: 10.1007/s10439-016-1606-6
- Rothhaas, R., and Chung, S. (2021). Role of the preoptic area in sleep and thermoregulation. *Front. Neurosci.* 15, 664781. doi: 10.3389/fnins.2021.664781
- Sano, A., Picard, R. W., and Stickgold, R. (2014). Quantitative analysis of wrist electrodermal activity during sleep. *Int. J. Psychophysiol.* 94, 382–389. doi: 10.1016/j.ijpsycho.2014.09.011
- Schafer, R. (2011). What is a savitzky-golay filter? *IEEE Signal Process. Mag.* 28, 111–117. doi: 10.1109/MSP.2011.941097
- Speakman, J. R. (2018). Obesity and thermoregulation. *Handb. Clin. Neurol.* 156, 431–443. doi: 10.1016/B978-0-444-63912-7.break00026-6
- Subramanian, S., Tseng, B., Barbieri, R., and Brown, E. N. (2022). An unsupervised automated paradigm for artifact removal from electrodermal activity in an uncontrolled clinical setting. *Physiol. Meas.* 3, 43. doi: 10.1088/1361-6579/ac92bd

- Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., and Picard, R. (2015). Automatic identification of artifacts in electrodermal activity data. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2015, 1934–1937. doi: 10.1109/EMBC.2015.7318762
- Verdonck, T., Baesens, B., Óskarsdóttir, M., and vanden Broucke, S. (2021). Special issue on feature engineering editorial. *Mach. Learn.* doi: 10.1007/s10994-021-06042-2
- Waalén, J., and Buxbaum, J. N. (2011). Is older colder or colder older? the association of age with body temperature in 18,630 individuals. *J. Gerontol. A Biol. Sci. Med. Sci.* 66, 487–492. doi: 10.1093/geron/ glr001
- Yanovich, R., Ketko, I., and Charkoudian, N. (2020). Sex differences in human thermoregulation: relevance for 2020 and beyond. *Physiology (Bethesda)* 35, 177–184. doi: 10.1152/physiol.00035.2019
- Zontone, P., Affanni, A., Bernardini, R., Piras, A., and Rinaldo, R. (2019). “Stress detection through electrodermal activity (EDA) and electrocardiogram (ECG) analysis in car drivers,” in *2019 27th European Signal Processing Conference (EUSIPCO)* (ACoruna: IEEE), 1–5.