scientific reports

OPEN

Improved non-invasive detection of sleep stages when combining skin sympathetic nerve activity and heart rate variability analysis with AI

Md. Aktaruzzaman & Thomas H. Everett IV

Sleep is a cyclic physiological process that goes into different stages, and every stage has its' importance in the construction or recovery of physiological function. Sleep scoring is performed from polysomnography recordings which requires signals from multiple sensors in a specially equipped Sleep Lab. This is expensive, and normal sleep behavior may be affected due to the new environment and extensive testing equipment. Skin sympathetic nerve activity (SKNA) has been shown to vary between sleep stages, and it can be recorded simultaneously with the ECG on the skin using conventional ECG patch electrodes. In this study, we propose that sleep stages can be classified using features derived from the ECG and SKNA recordings. The study was performed on 21 subjects, and we initially extracted 14 heart rate variability (HRV) and 11 SKNA features, and then selected the 17 most relevant (12 HRV, 5 SKNA) features out of 25. We evaluated both individual and combined performance of HRV and SKNA for classification of 5 sleep stages, 3 stages and binary stages. Our study showed that the addition of SKNA information with HRV provides an improved recognition accuracy of 91.57%, 95.36%, and 96.14%, respectively for 5, 3, and binary sleep stages. The SKNA shows similar performance to the HRV for 2 and 3 sleep stages recognition. This AI-powered sleep classification system might provide an advancement in the development of a real-time sleep monitoring device with low computing power that can be used for screening sleep disorders in a large population.

Keywords Sympathetic nerve activity, Heart rate variability, Sleep stages recognition, Machine learning

Sleep is a complex physiological process and essential for building and recovery of the human body, and its physiological systems. According to the American Academy of Sleep Medicine[1] in normal sleep conditions, a person moves into five basic sleep stages: non-rapid eye movement (N1, N2, N3), rapid eye movement (REM), and awake (WK) during sleep. Every sleep stage plays a significant role in physiological recovery process. Deep sleep (N3) contributes to the insightful thinking, repair injuries, and reinforce the immune system[2]. On the other hand, REM sleep is believed to be essential to cognitive functions like memory, learning, and creativity[3]. Poor quality of sleep has been linked to several negative health outcomes, including increased risk of cardiovascular disease[4], obesity[5], and neurodegenerative disorders[6]. Despite these well-known risks, more than 70 million Americans experience sleep related problems and less than 20% of patients estimated to be properly diagnosed and treated for sleep disorders[7]. Sleep stage identification is determined during a sleep study with polysomnography (PSG), and it has been used as a gold standard for the diagnosis of several sleep disorders. The PSG method requires signals acquired from multiple sensors like electrocardiogram (ECG), electroencephalogram (EEG), electro-oculogram (EOG), and electromyogram (EMG) in a specially equipped laboratory. A sleep expert or neurologist then scores sleep from these signals for every 30 s recording epoch. Unfortunately, the use of PSG suffers from limitations as the equipment needed and the environment can negatively affect normal sleep behavior. Thus, alternative methods for long term monitoring of sleep behavior in a large population within the home environment in a noninvasive way have been attempted.

Many researchers have proposed sleep assessment from heart rate variability (HRV i.e., the variation in intervals between successive heart beats) which is noninvasive and inexpensive[8–12] using conventional machine

Division of Cardiovascular Medicine, Department of Medicine, Krannert Cardiovascular Research Center, Indiana University School of Medicine, Indianapolis, IN, USA. email: theveret@iu.edu

learning methods. However, the performance of sleep staging from HRV alone is still far from widespread practical implementation due to the low recognition rate of some sleep stages and the length of recording needed to get an accurate assessment of HRV. Some researchers have proposed identifying sleep stages from the combination of HRV and actigraphy (ACT) to improve the overall accuracy[13,14]. The ACT is useful for sleep quality assessment for binary sleep stages (SLEEP vs WAKE) rather than estimating each sleep stage. In addition, it can give a false positive of SLEEP when people are awake but without movement (e.g., watching movie). Thus, the combination of ACT with HRV has not improved the quality of assessment to the level of practical implementation.

Direct muscle sympathetic nerve activity (MSNA) using microneurography techniques has shown changes in MSNA in different sleep stages[15]. However, obtaining stable microneurography recordings over long periods of time such as during a sleep study is difficult. A new method has been developed for the noninvasive and simultaneous recordings of the electrocardiogram (ECG) and the skin sympathetic nerve activity (SKNA) that can provide long term recordings during a sleep study[16,17]. SKNA parameters (aSKNA, i.e., the average of nerve activity per sample, burst frequency) have been shown to differ between sleep stages[18]. However, SKNA parameters have not been used to non-invasively determine sleep stages.

Today's mobile health applications open the new paradigm of tele-home care monitoring that combines the standard telemedicine approach with the latest Internet of Things (IoT) concept[19]. In this context, the development of new generation sensors has inspired the combination of multiple sensor units in a single device[13]. In this regard, some researchers have proposed sleep stages recognition from the combination of signals. However, they performed classification from signals acquired from multiple sources of wearable sensors such as (HRV and Actigraphy, Photoplethysmogram and Respiratory)[13,20]. The use of different signals from multiple sources increases the power consumption and device size than the signals required from a single device.

Recent works[20,21] that employ deep learning-based algorithms achieved impressive results in sleep stage classification than the conventional machine learning algorithms. These include support vector machine (SVM), conventional neural networks, recurrent neural networks, deep belief networks, and autoencoders. Despite the improved performance of deep learning-based algorithms, a significant limitation lies in their black box behavior. A general concern when it comes to application of artificial intelligence in healthcare and medicine is the underlying mechanism of deep-learning algorithms. This is a major drawback of existing deep learning-based sleep staging algorithms that keeps them from being adopted towards clinical settings[22].

To address these issues, we evaluate the sleep quality of subjects by analyzing signals which could be easily integrated in a single wearable device. Specially in this work, we considered the combination of the information from ECG and SKNA signals which can be recorded simultaneously from the skin surface using a single lead traditional ECG patch electrodes for sleep staging using a machine learning (ML) approach. The main objectives of this study are (1) to study the feasibility of using SKNA signals for sleep staging, (2) to evaluate the individual performance of SKNA and HRV, and (3) to evaluate the combined performance of HRV and SKNA for determining the various stages of sleep.

Methods

The methodological steps involved in this study is shown by the block diagram in Fig. 1.

Study population and data acquisition

The present research was approved by the Institutional Review Board of the Indiana University School of Medicine (IRB Protocol #2009983991), and all research was performed in accordance with all relevant guidelines and regulations and in accordance with the Declaration of Helsinki. All patients provided informed consent for the SKNA recordings. Sleep studies and sleep stage scoring were conducted per American Academy Sleep Medicine (AASM) recommended protocol in 21 patients with suspected sleep apnea. Patient age ranged between 28 and 64 years; 13 males, 8 females; 13 were white, 8 were black or others; BMI: 24.2 to 36.7. Polysomnography was recorded using the Alice 6 Diagnostics Sleep System (Philips Respironics, Amsterdam, The Netherlands) as previously described[17]. Also as previously demonstrated[23], ECG and SKNA signals were recorded simultaneously from electrodes placed in an ECG Lead I configuration on the chest using a modified ME6000 Biomonitor (Mega Electronics Ltd, Kuopio, Finland) at a sampling rate of 10 KHz.

The ECG signal and SKNA was extracted from the raw signal by applying band pass filter of [0.05 41] Hz and [500 1000] Hz, respectively. The R-R series was extracted from the ECG using Labchart Pro 2008 (ADInstruments Ltd, Bella Vista, Australia). Data from these patients were included in previous studies[17,18], which showed the changes of SKNA in patients with obstructive sleep apnea and the effects of continuous positive airway pressure (CPAP) on SKNA parameters.

Preprocessing

Since the data was collected simultaneously but independently using separate devices, time synchronization was achieved by performing QRS detection on both sources and matching heart beats by inter-beat interval sequences. SKNA data was time shifted and scaled to match the sleep system timescale. The data was manually analyzed to exclude extreme artifacts in the ECG or SKNA due to motion or loose contract of ECG patch electrodes. These extreme artifacts have been discarded from the source signals, and we considered only R-R intervals during normal sinus rhythm, and the series of R-R intervals is termed as RR series. Some R-peaks are also mis-detected or missing as shown in Fig. 2a, so we considered only those R-R intervals that lie within the 95% CFI of the R-R values (RR series). RR and SKNA signals were segmented into a series of 30 s window called an epoch as shown in Fig. 2b.

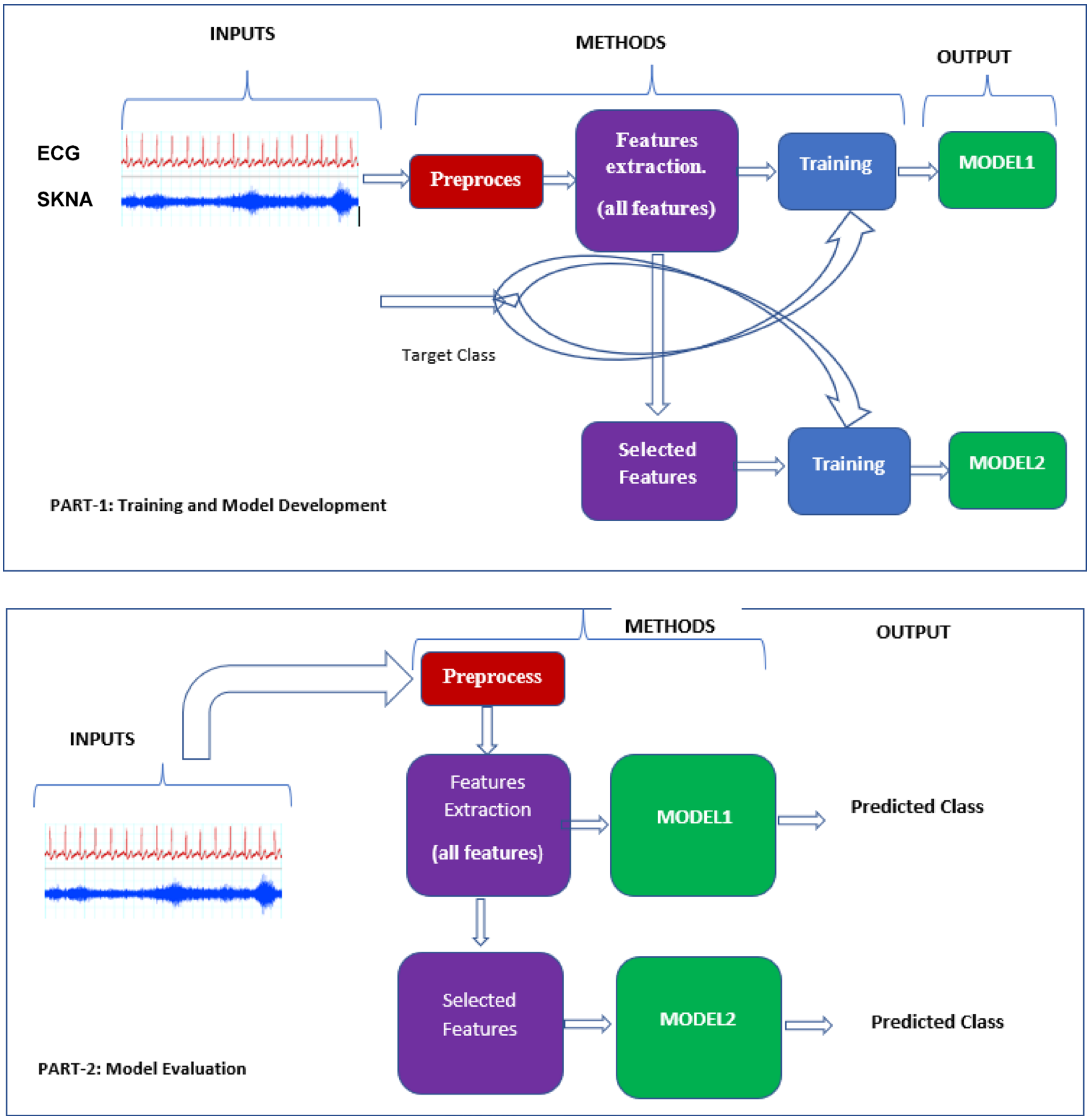

Fig. 1. Sleep stages classification steps from HRV and SKNA. The top panel of this diagram shows how the models MODEL1and MODEL2 are built using the full set of extracted features and the selected set of features, respectively. In the lower panel of the diagram, models MODEL1 and MODEL2 that were designed in the upper panel are now evaluated using the full set and selected set of features on the test data.

Features extraction

The raw SKNA and ECG signals were fed to the input of the classification model. After, preprocessing a set of features were extracted from both SKNA and HRV series derived from the ECG. Finally, a classification model (MODEL1) was developed based on the extracted features. To keep low both computing power and power consumption, we also applied features selection to reduce dimension of the features set, and another classification model (MODEL2) was developed using this reduced set of features. We developed and evaluated the performance of these models for HRV and SKNA features individually, and when these features were used in combination.

To extract features from RR series and SKNA, we considered a series of 5 min length consisting of 6 epochs before and 3 epochs after the current epoch as recommended previously[13]. We extracted 14 HRV parameters and 11 SKNA parameters from each 5 min series.

HRV parameters

1. MeanNN (s): The average of R-R intervals of the series measured in s.
2. SDNN (s): The standard deviation of the R-R intervals of the series in s.
3. RMSSD (s): The root mean square of the successive differences of R-R intervals in s.

Fig. 2. Preprocessing steps for features extraction. The arrows indicate the normal and abnormal ECG beats. The abnormal beats can be due to motion (Misdetected) or due to arrhythmia (Missing beat). Each colored block represents an epoch of 30 s. The green colored block marks the current epoch, and the blue and orange-colored epochs represent the 6 epochs before and 3 epochs after the current epoch, respectively.

4. $DFA_{\alpha 1}$ (a.u.): The short-term detrended fluctuation scaling exponent, a nonlinear measure of the R-R intervals and useful for the possible nonlinear and nonstationary characteristics in the series.
5. cvNN (a.u.): The relative dispersion of R-R intervals of the series.
6. SampEn (a.u.): A conditional differential entropy rate measures the complexity (or regularity in other words) of a time series. The estimation of sample entropy (SampEn)[24] depends on the selection of *m* (the embedding dimension) and *r* (the maximum tolerance of mismatch between the corresponding elements of small vectors) for a given series length (*N*). It is estimated as

   $\boldsymbol{S}ampEn\,(m, r, N) = \log P_m - \log P_{m+1}$, where P_m and P_{m+1}, are respectively, the probability of matching vectors of length *m* and *m* + 1.

7. bEn (a.u.): A newly proposed entropy metric called Bubble entropy[25], which does not require the selection of parameter *r* in SampEn estimation, and it is also less dependent on *m*. Unlike SampEn, it does not directly quantify conditional entropy instead it quantifies the entropy changes in ordering the small vectors of length *m* to *m* + *1* of a series of length *N*. Mathematically,

$$bEn\ \ (m, N) = \left(H_{swaps}^{m+1} - H_{swaps}^{m}\right) / \log\left(\frac{m+1}{m-1}\right) \tag{1}$$

   where H is the second order Rényi entropy[24].

8. SD1 (s): SD1 measured in (s) is the length of semi-minor of the ellipse around the 95% confidence of the RR series fitted to the Poincaré plots. It measures the short-term variability of RR series and is influenced by parasympathetic modulation.
9. SD2 (s): SD2 measured in second is the length of semi-major of the ellipse around the 95% confidence of the RR series fitted to the Poincaré plots. It measures the long–term variability of RR series and reflects the sympathetic activation.
10. SD1/SD2: The ratio of SD1 to SD2 provides information about the sympatho-vagal balance of the autonomic nervous system.

We extracted a set of 4 frequency domain parameters in addition to the time domain parameters. The frequency domain parameters were extracted in four frequency bands: [0.003 0.04] Hz, [0.04 0.15] Hz, and [0.15 0.4] Hz, respectively for VLF, LF, and HF.

1. VLF (%): Percentage of power in very low frequency [0.03 0.04] Hz band.
2. LF (%): Percentage of power in low frequency [0.04 0.15] Hz band.
3. HF (%): Percentage of power in the high frequency [0.15 0.4] Hz band.
4. LF/HF: The ratio of LF power to HF power.

SKNA parameters
We have derived a set of 11 features from SKNA signals:

(a) aSKNA (μV): The mean of sympathetic nerve activity value per sample, measured in micro-volt (μV).
(b) varSKNA (μV): The variance of SKNA (varSKNA) measures in (μV) the signal dispersion around it's mean.
(c) rmsSKNA(μV): The root mean square (RMS) of the amplitude of SKNA signals, measured in μV.
(d) skewSKNA: It measures skewness of the amplitude distribution of the SKNA signal.
(e) kurtSKNA: It mesures kurtosis i.e., the peakness of the distribution of signal's amplitude.
(f) wlSKNA: The waveform length (WL) of SKNA signal provides information about the shape and frequency content by measuring the cumulative length of the signal's waveform.

(g) zcSKNA: The zero crossing (ZC) of SKNA counts the number of times, the SKNA crosses zero axis, providing information about the signal's frequency.
(h) sscSKNA: It counts the number of times the sign of the signal's slope (SSC) changes, providing information about the signal's rate of change.
(i) wampSKNA: The Wilson amplitude (wamp) of SKNA computes the weighted sum of the absolute differences between adjacent samples that exceed a threshold value (Theta), providing information about the signal's amplitude changes over time.
(j) cfSKNA: The chrest factor of SKNA (cfSKNA) measures the peak-to-ratio of the SKNA signal, providing information about the signal's dynamic range.
(k) dfSKNA (Hz): The dominant frequency of SKNA (dfSKNA) provides the frequency with maximum power of the SKNA signal in (Hz). The dfSKNA was estimated from the integrated SKNA signals using the periodogram method.

Features selection

Features selection deals with reducing the number of variables to the input of a machine learning model to reduce computational cost and to improve classification performance. Statistical features selection evaluates the relationship between each input and the target variables, and select those features which have strong relationship, which is fast and effective. However, selecting appropriate statistical measures is sometimes challenging and critical. In addition, the data includes an unbalanced proportion of samples in the classes, such that the training of a model may be biased to the class with majority of samples. Since, Sleep includes 5 different sleep stages with majority of N2 (NREM2) and the proportion of REM and WAKE are very small compared to N3 (NREM3) and N2. For this reason, we used a performance measure called Cohen's Kappa that measures the reliability of the classification model irrespective of the overall accuracy to select most important features (or input variables). The pseudocode of features selection is given below:

1. Randomly divide the entire data samples into 80%: 20% for training: test.
2. *inputFeaturesList* = {All input variables}
3. *Set selectedFeaturesList* = ϕ, train*FeaturesList* = ϕ
4. For each feature F_i in the *inputFeaturesList*

   a. Set train*FeaturesList* = *selectedFeaturesList* ∪ {F_i }
   b. Build a classification model on the training set with 10-fold cross validation.
   c. Evaluate the model of test set and determine and store Cohen's Kappa of the model
   d. Set trainFeaturesList = *selectedFeaturesList*

5. Add the feature F_i for which Cohen's Kappa is maximum to the *selectedFeaturesList*
   i.e., selectedFeaturesList = selectedFeaturesList ∪ {F_i }

6. Exclude the selected feature from the *inputFeaturesList i.e.,*
   inputFeaturesList = inputFeaturesList - {F_i }

7. Repeat step 4 to step 6 until the *inputFeaturesList* is EMPTY.

Here, we refereed this features selection as "**Greedy Forward Approach**".

Classification

The manual sleep staging from PSG following the guidelines of AASM was considered as the gold standard in this study[1]. We used a support vector machine (SVM) with *"rbf"* kernel to classify sleep into 5 stages (N1, N2, N3, REM, and WK), 3 stages (NREM, REM, and WK), and binary classes (SLEEP and WK) using all features and best relevant features. We also evaluated the individual classification performance of HRV and SKNA parameters for binary to 5 stages classification problems. The classification model was trained and evaluated using leave one subject out (LOSO) scheme, where data from all subjects except one were used for training and the remaining subject was used for testing. The classification performance of the model was evaluated by Accuracy (Acc), Precision (Prec), and Recall (Rec), Sensitivity (Sens), Specificity (Spec), F1 score, and Cohen's Kappa (K). The values of Prec and Rec was reported as the average value of the Sens, Spec, respectively of each binary classification constructed from 3 and 5 stages classifications.

Since K-fold cross validation has less variance than the Leave One Subject Out (LOSO) cross validation technique, we used K-fold cross validation (K = 10) for best features selection. On the other hand, K-fold cross validation scheme may have the chance of data leakage from training /validation to testing. So, we evaluated our model using both K-fold cross validation and Leave One Subject Out (LOSO) techniques. In LOSO, one subject was kept out for testing and the remaining was used for training the model.

Statistical analysis

We compared both HRV and SKNA parameters to discriminate between sleep stages using the Wilcoxson Rank Sum test. In addition, if the addition of SKNA information with HRV parameters significantly improve the accuracy, we performed *McNemar* test on the sleep stages predicted by the classification models using HRV alone and for the combination of HRV and SKNA. Besides these, we also evaluated the power of both HRV and SKNA parameters to discriminate between sleep stages using Wilcoxon's statistical test.

Results

In 4 of the 21 patients, a split sleep study was performed in which CPAP therapy was introduced during the second half of the study. For these patients, we considered the recordings during the first part of the study, before CPAP therapy was initiated. A total of 16,233 epochs were analyzed.

From sleep studies that for each patient lasted an average of 7 h, there were 1827, 6720, 1743, 2226, 3717 epochs for N1, N2, N3, REM, and WK stages respectively, which corresponded to 11.25%, 41.40%, 10.74%, 13.71%, and 22.90% of the total epochs. The combination of information from HRV and SKNA improved the sleep stage recognition obtained using either HRV or SKNA only. The results found in this study contribute to the deeper understanding of SKNA for sleep study and opens the door for sleep monitoring using a single wearable device. The average values of the R-R intervals and the sympathetic nerve activity of a patient during a sleep study that included various sleep stages have been shown in Fig. 3. From Fig. 3, both heart rate and skin sympathetic nerve activity increases during WK and REM compared to NREM.

We extracted a set of 25 features (14 from HRV and 11 from the SKNA analysis) from RR series and SKNA signals of 5 min windows. Both HRV and SKNA parameters in different sleep stages and the *p* values of Wilcoxon's Ranksum test to distinguish between different sleep stages are represented in Figs. 4 and 5, respectively. The *p* values were statistically significant ($<<0.05$) for most of the parameters such as HRV parameters: SampEn, bEn, LF, HF, and LF/HF and SKNA parameters: aSKNA, SKEWNESS, KURTOSIS, WAMP, and DomFreq to distinguish between any pair of sleep stages as shown in Fig. 5. In most of the cases, the *p* values of the statistical test were found >0.05 to distinguish between the N1 and WK. Interestingly, for two SKNA parameters: zero crossing (ZC) and chrest factor (CF), the statistical test was significant to distinguish between the N1 and WK, however, they were not found significant between N2 and REM.

We applied a feature selection strategy to reduce the feature dimension by discarding the less important and redundant features from the extracted features set, and a subset of 12 HRV and 5 SKNA features were selected as the most relevant by the greedy forward approach based on Cohen's Kappa to distinguish sleep stages[26]. The selected features and the corresponding Cohen's kappa is shown in Fig. 6. Thus, we see from Fig. 6 that the dominant frequency (DF) of SKNA signal was found as the most important and the coefficient of variance (CV) as the lowest important features of the relevant features set. The addition of other features after CV does not increase the Cohen's kappa.

The values of the performance metrics obtained using the K-fold cross validation have been presented in Table 1. The performance metrics using the best relevant features remain almost same to the performance obtained using the full set of features. The values of Acc, Prec/ Sens, Rec/ Spec, F1, and Cohen's kappa using the selected features were respectively, 91.574%, 91.49%, 88.42%, 89.93%, and 0.88 for 5 stages classification.

The highest accuracy (96.14%) was found for binary classification using the combined set of features from both HRV and SKNA, and the performance measures were decreased with increasing the class size.

The values of the performance measures were also reduced when we trained and evaluated the classification model in LOSO scheme as shown in Table 2. The values of Acc have been reduced from 91.57%, 95.36%, and 96.14% to 71.80%, 79.82%, and 85.40%, respectively for 5 stages to binary sleep stages recognition. However,

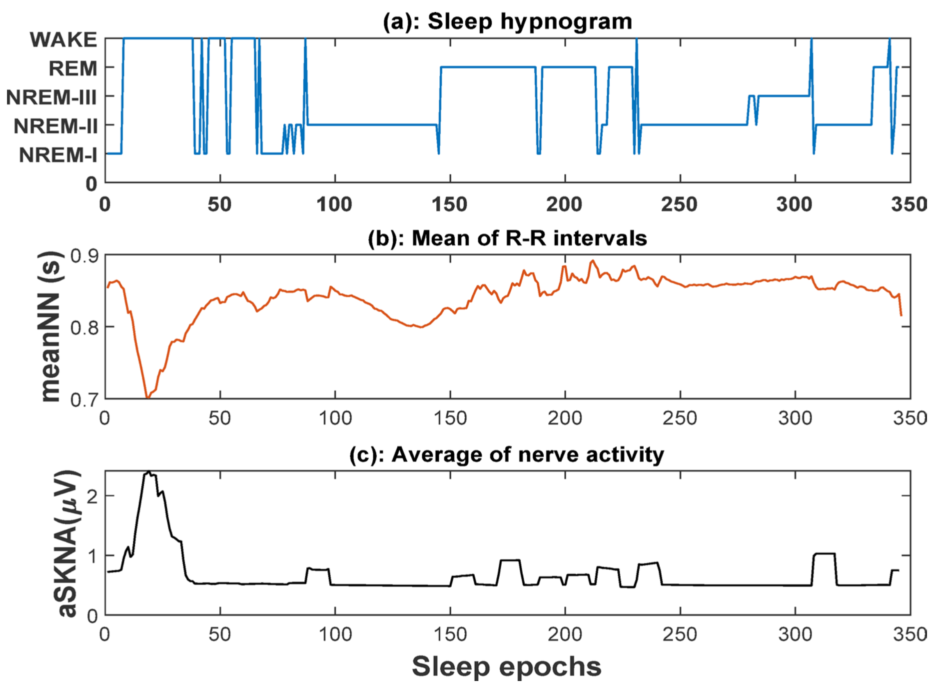

Fig. 3. The mean values of the R-R intervals and skin sympathetic nerve activity for successive 350 epochs of a patient. Panel (**a**) shows the sleep hypnogram, panels (**b**) and (**c**) show respectively the mean R-R intervals and aSKNA values.

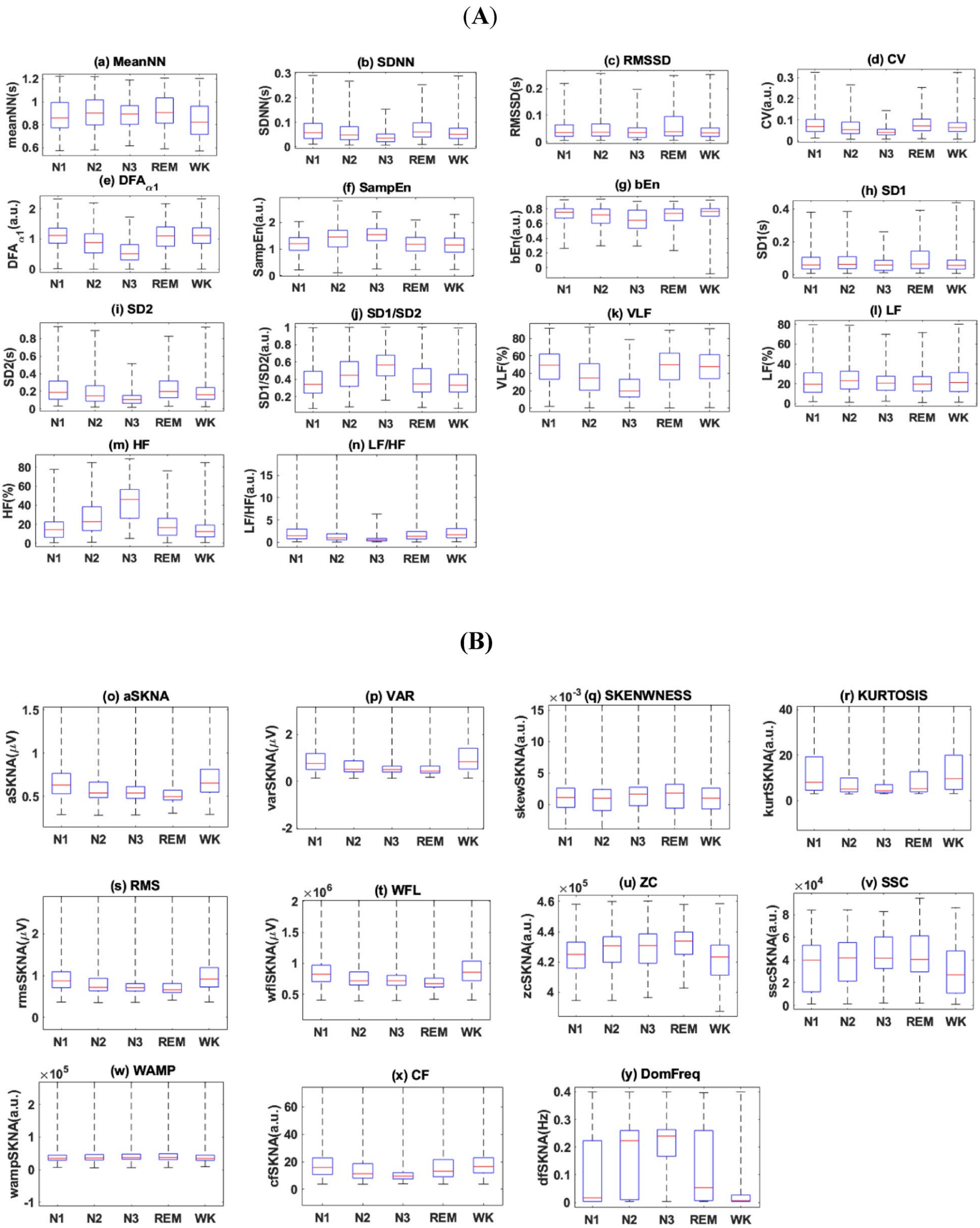

Fig. 4. Boxplot representation of HRV parameters panel (**A**), and SKNA parameters (**B**).

the model still provides about 85% average accuracy for binary sleep stages recognition with the reduced set of features, and the recognition accuracy of the model for some patients are more than 95%. The recognition rate of sleep stages of some patients (patient # 5, 13, and 18) is very poor (less than 70%) as shown in Fig. 7, because deep sleep stages (N2 and N3) with high recognition accuracy are either absent or very small in number during the entire recording in those patients, and hence the performance of the model is reduced when evaluated on these patients.

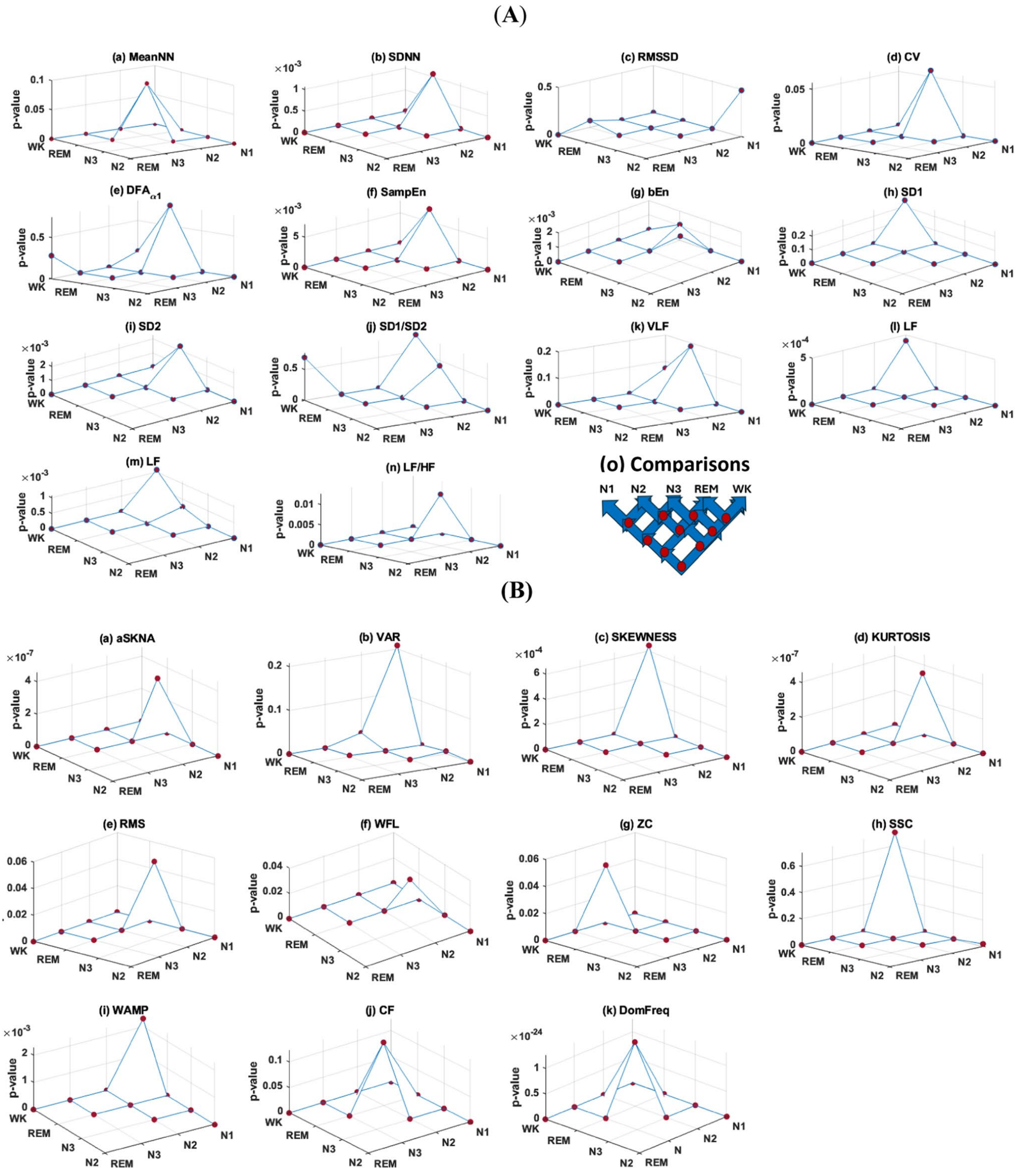

Fig. 5. *P* values of the Wilcoxon's RankSum test for HRV (**A**) and SKNA parameters (**B**) to distinguish between the sleep stages. As shown in panel (o) of (**A**), there are 10 comparisons between pairs of 5 sleep stages marked by the red solid circles. The *p* values of the statistical test were < 0.05 for all pairs of comparisons between the sleep stages except WK and N1 for the parameter VAR in panel (b) of (**B**).

In both training and evaluation methods, the best performance was found for binary sleep stages classification. The SLEEP vs WK recognition accuracy of each patient using either HRV or SKNA parameters alone and combinedly have been shown by the barograph Fig. 7. It is seen from Fig. 7 that the recognition accuracy of some patients was very high (> 95%), and some patients was low (< 70%) using any set of features. The SKNA parameters show very similar performance to the HRV parameters for sleep stages recognition, and

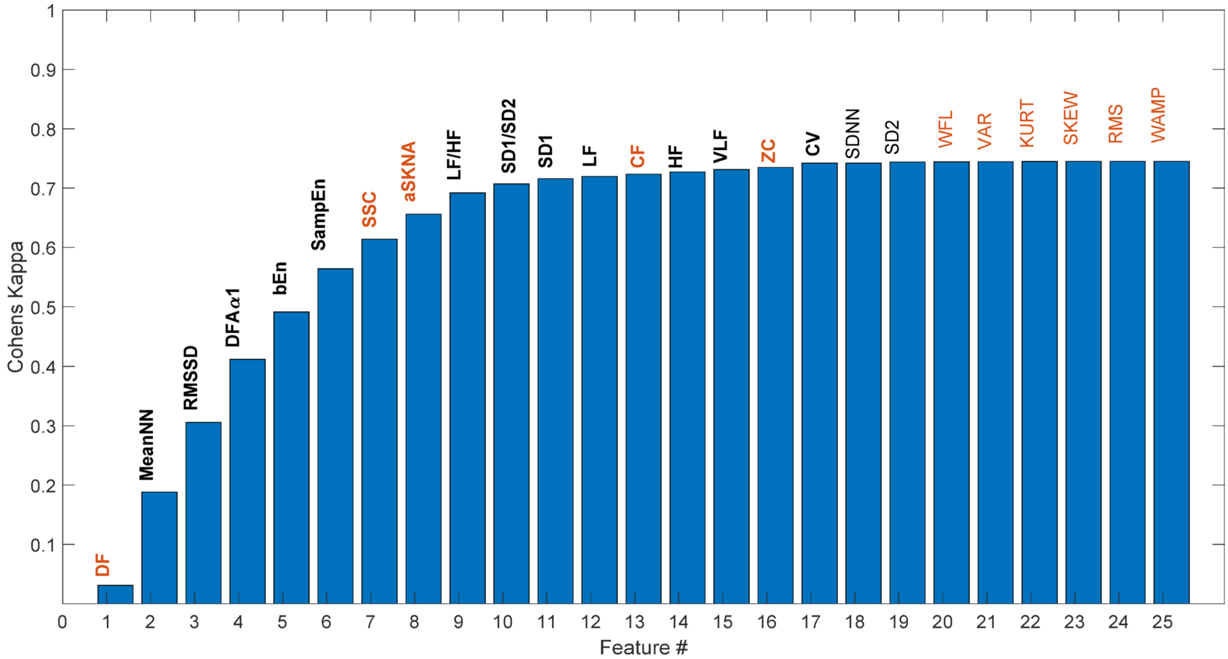

Fig. 6. The order of important features and the corresponding Cohen's kappa. SKNA features are marked by the orange color. The left most feature corresponds to the most important, and the features are ranked from left to right or most important to less important. The values of the performance metrics have been represented in Table 1.

Sleep stages	Full, reduced features	Acc (%)	Rec/Spec (%)	Prec./Sens (%)	F1 (%)	Cohen's Kappa
5	HRV + SKNA	92.02, 91.57	91.94, 91.49	89.11, 88.42	90.50, 89.93	0.78, 0.77
	HRV only	86.90, 85.95	87.15, 86.34	82.36, 81.06	84.68, 83.62	0.72, 0.76
	SKNA only	67.71, 64.21	68.34,63.25	57.28, 53.09	62.32, 57.73	0.53, 0.48
3	HRV + SKNA	95.59, 95.35	95.66, 95.36	94.10, 93.81	94.87, 94.58	0.79, 0.77
	HRV only	91.90, 91.49	91.84, 91.46	88.62, 88.61	90.03, 90.01	0.74, 0.74
	SKNA only	79.14, 76.73	80.27, 76.86	66.11, 62.41	72.51, 68.88	0.57, 0.52
2	HRV + SKNA	96.33, 96.14	96.11, 95.96	92.95, 92.53	94.51, 94.21	0.78, 0.79
	HRV only	94.15, 93.82	93.92, 93.69	88.52, 87.73	91.14, 90.61	0.71, 0.71
	SKNA only	86.67, 85.69	83.32, 81.69	74.59, 72.75	78.71, 76.96	0.55, 0.52

Table 1. Performance metrics of sleep stages recognition using 10-fold cross validation.

Sleep stages	Full, reduced features	Acc (%)	Prec./Spec (%)	Rec/Sens (%)	F1 (%)	Cohen's Kappa
5	HRV + SKNA	73.06, 71.80	77.19, 76.21	65.38, 64.11	70.79, 69.64	0.61, 0.59
	HRV only	67.70, 65.59	66.94,64.47	61.70, 59.44	64.21, 61.85	0.55,0.52
	SKNA only	53.95, 57.70	51.36,57.64	44.05, 48.85	47.43 ,52.88	0.34,0.39
3	HRV + SKNA	80.98, 79.82	83.71, 83.48	65.59, 63.39	73.55, 72.06	0.59, 0.56
	HRV only	77.62, 74.96	71.47, 66.70	64.44, 61.72	67.77, 64.11	0.55, 0.50
	SKNA only	70.73, 72.70	66.98, 70.44	51.65, 55.73	58.32, 62.23	0.35,0.41
2	HRV + SKNA	86.03, 85.40	97.99, 98.49	85.90, 84.12	91.55, 90.74	0.52, 0.45
	HRV only	83.44, 80.72	92.80, 89.61	86.70, 86.01	89.64, 87.78	0.49, 0.42
	SKNA only	79.27, 80.47	88.93, 89.78	84.94, 85.62	86.89, 87.65	0.38, 0.41

Table 2. Performance metrics for sleep stages recognition using LOSO training.

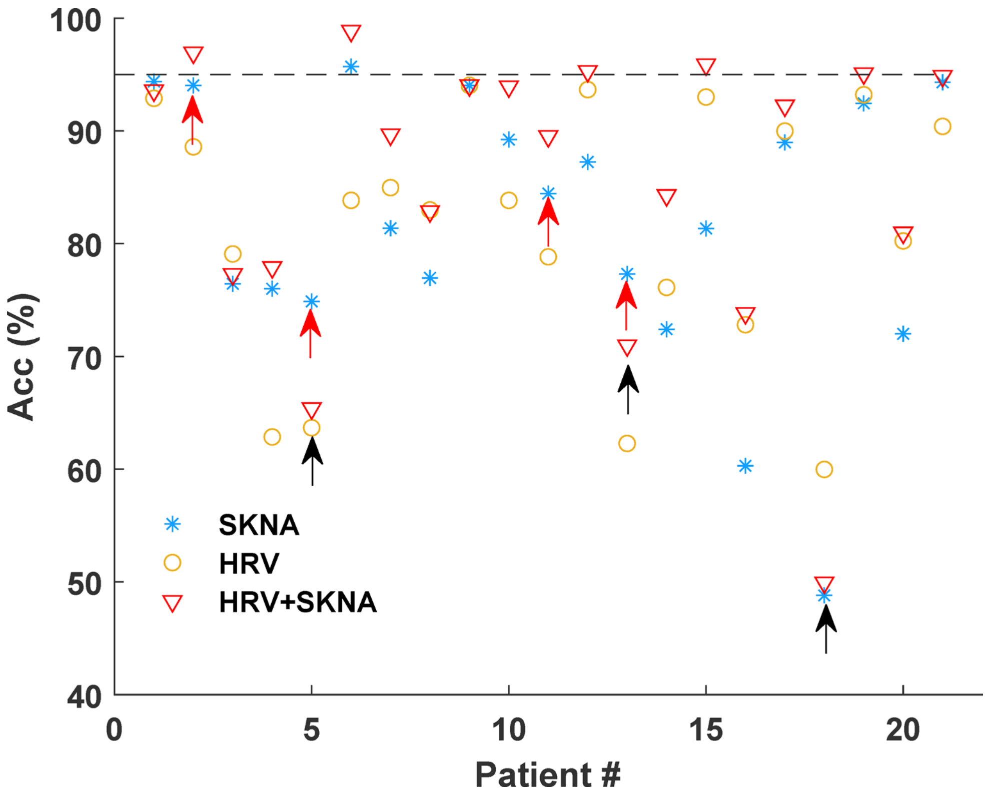

Fig. 7. The comparative recognition accuracy of SLEEP vs WK stages for each patient using the SKNA and HRV individually and for their combination. The black arrows mark the patients with poor accuracy. The dashed horizontal line labels to the 95% of accuracy. The red colored arrow indicates the patient number at which the SKNA parameters provided more Acc than the HRV parameters.

the combination of HRV with SKNA parameters increased the recognition accuracy. It was also observed from Fig. 7 that for some patients (patient # 2, 6, 11 and 13), the SKNA parameters provided better recognition accuracy than the HRV parameters. We see smaller values of Sens than the Spec for binary classification as shown in both Tables 1 and 2, which means that many WK epochs have been incorrectly recognized as the SLEEP epochs. This could be because the biased training of the model to the SLEEP class which has more samples, or this simple machine learning model cannot capture sufficient information about WK.

The *p* values of statistical tests (*Wilcoxon RankSum*) were found significant (p˂0.05) for these selected features to distinguish between SLEEP and WK, and most of them except RMSSD, SD1, and VLF [%] were also found significant to discriminate NREM, REM, and WK sleep stages. The *p* value of McNemar test was $<<0.05$ to discriminate between the predicted performance of the classification models obtained for HRV only vs the combination of HRV and SKNA parameters.

We ran the model on the 20% test set with 3516 samples for 5 sleep stages classification using the full set (25) features and the selected (17) features only. We observed that the total time required by the models for recognizing 3516 samples of the test using the full set of extracted features and the selected features are 0.386 s and 0.314 s, respectively. The throughput of the system increases from $9.108e^3$ to $1.119e^4$ for selected features set instead of the full set of extracted features.

Discussion

In this study, we used a single machine learning approach (SVM)[8] to study the feasibility of SKNA parameters for sleep stages recognition. We also investigated if the parameters obtained from SKNA and HRV could be used to together to improve the recognition accuracy of sleep stages obtained either by HRV or SKNA.

The features selection is a critical step in any classification problem, where the features are ranked based on a performance measure. A model can show better accuracy by chance with less important features than the robust ones. So, we considered Cohen's kappa instead of Acc as a standard for features selection. We found that dominant frequency (DF) of SKNA has been found as the most relevant one out of all features from HRV and SKNA as shown in Fig. 6. In addition, 5 other features of SKNA with 12 HRV features have been found as the relevant ones. Although both SampEn and bEn measures the complexity of a time series, but due to difference in the computational method their estimation can be different based on the presence of non-stationarity, upper and lower trends in the series. So, we have proposed both entropy measures, and both of them have been identified as important features. We proposed both linear, frequency, nonlinear, regularity measures of HRV analysis, nonlinear and regularity metrics except the mean and RMSSD have been selected as the most relevant features.

We evaluated individual and the combined power of HRV and SKNA parameters for sleep stages classification from binary to 5 sleep stages using the full set of features and the reduced set of features (Tables 1 and 2). The best selected features provide similar performance with the full set of features. However, the value of recall has been slightly reduced with reduce set of SKNA only. This means the discarded features of SKNA improves recognition of WK when using SKNA only, but their absence is compensated by the HRV parameters when used in combination. The HRV parameters provide better performance than the SKNA alone for the classification of 3 and 5 sleep stages as shown in Tables 1 and 2. However, their performance is very comparable for binary (SLEEP vs WK) stage classification. Thus, when interested only in sleep quality rather than staging each phase, SKNA can be used for long-term sleep monitoring in a noninvasive way. The combination of HRV and SKNA parameters significantly improves their individual performance for 3 and 5 sleep stages. This indicates that SKNA parameters captures some information to discriminate between the basic sleep stages that HRV parameters alone cannot and vice-versa.

For this study, we evaluated the classification performances for SVM models using both K-fold cross validation and LOSO training schemes. In LOSO training and testing, the resulting performance varied depending on which subject was used as the test case as shown in Fig. 7. The performance measures of HRV and SKNA were comparable for all classification of sleep stages except the 5-stage classification[8,11,13,24]. The value of ACC can also be affected by the series length, especially for SKNA parameters. Studies using HRV parameters for sleep stage recognition has previously been demonstrated, but to the best of our knowledge this is the first study using SKNA parameters for this purpose.

To get better performance of sleep stages detection, signals from multiple sensors have been incorporated to improve the distinction between individual sleep stages. However, the integration of multiple sensors increases device dimension, and the consumption of power. One important aspect of wearable devices is long-term monitoring of sleep. To maintain comfortability with long-term monitoring of sleep, researchers emphasize reducing device dimension with low power consumption.

There are works that propose sleep staging from the combination of multiple sensors such as ECG and respiratory, ECG with actigraphy, Photoplethysmography with accelerometer, HRV and actigraphy with the respiratory signals[13,27–29]. They collected signals by placing wearable devices at different locations of human body. In this study, two different signals (ECG and SKNA) are collected from the same source on the skin surface in noninvasive manner. If we compare the proposed study with existing studies that considered sleep stages classification from multiple sensors, we see our proposed method provides identical performance to the existing studies for sleep stages detection[13,27–29]. The major advantage of our proposed study is that it uses signals collected from the same sensor which helps to reduce device dimension and power consumption. One major drawback of our proposed method is that it requires signals acquired at high sampling rate which requires more memory. Fortunately, memory is not a great issue today due to the advancement of technology.

If we compare the performance of our proposed model with similar studies[11,13,24,27,30], that used features from HRV analysis and 4-fold cross validation for training a long short-term memory (LSTM) model, we obtained a similar performance using a very light SVM model with 17 features. In addition, our model also achieved similar performance to the deep learning-based models[20,30]. The model performance can be varied based on the dataset used for training and testing. However, our hypothesis has been supported by these preliminary studies, showing that there is information within the SKNA recordings which can be used to classify sleep stages.

Figures 4 and 5 show that both HRV and SKNA parameters have significant differences between sleep stages. As shown in Table 2, the combination of SKNA information with HRV improves sleep stages detection (the recognition of individual sleep stages) more than that obtained for the binary recognition of sleep stages using HRV or SKNA individually. When used individually, the detection of the individual sleep stages specially between (N1, WK, and REM) is not as accurate. However, the accuracy is improved when combining SKNA and HRV information in training the classification model.

While showing the improvement in combining SKNA with HRV for binary is important, we also think that the results for the 5 stages of sleep is also important to emphasize. If the 5 different stages of sleep can be accurately identified; it would help with diagnosing and managing sleep disorders and other chronic conditions.

For a majority of the sleep study, the patients are sleeping, however, artifacts can occur within the recordings. The criteria that we used for the classification of artifacts is the same as artifacts that occur during a regular ECG recording. With band-pass filtering from 500 – 1000 Hz, we can eliminate many of the artifacts as long as the artifacts did not produce significant alterations of the ECG morphology. As discussed previously[31], the most common types of artifacts that are encountered occur as a result of motion, muscle tension, poor electrode contact, electromagnetic interference, and implanted stimulators.

A limitation of the study is the small sample size of patients. In addition, these patients were referred for a sleep study for suspected sleep apnea. Patients with sleep apnea may not experience all of the stages that are associated with sleep, especially the stages of deep sleep. Thus, further reducing the amount of information used to train the models. In the present study, not all patients experienced all 5 stages of sleep during a sleep study which can reduce the accuracy of the models.

Conclusion

The sleep staging from skin sympathetic nerve activity is possible, and the combination of SKNA with heart rate variability parameters improve sleep stages classification which is statistically significant. The SKNA itself provides sleep scoring with similar a capability to HRV for binary. Therefore, this study supports that a single wearable device recording the ECG and SKNA signals from the standard ECG patch electrodes can be used to effectively monitor sleep patterns. These findings will open the possibilities of designing wearable devices for facilitating long-term monitoring of sleep within the home environment.

Data availability
The datasets generated during and/or analyzed during the current study are available from the corresponding author at Pennsieve.io.

Received: 2 July 2025; Accepted: 12 September 2025
Published online: 17 October 2025

References

1. Bonnet, M. H. et al. The scoring of arousal in sleep: Reliability, validity, and alternatives. *J. Clin. Sleep Med.* **3**, 133–145 (2007).
2. Yordanova, J., Kolev, V., Wagner, U. & Verleger, R. Differential associations of early- and late-night sleep with functional brain states promoting insight to abstract task regularity. *PLoS ONE* **5**, e9442. https://doi.org/10.1371/journal.pone.0009442 (2010).
3. Cai, D. J., Mednick, S. A., Harrison, E. M., Kanady, J. C. & Mednick, S. C. REM, not incubation, improves creativity by priming associative networks. *Proc. Natl. Acad. Sci. USA* **106**, 10130–10134. https://doi.org/10.1073/pnas.0900271106 (2009).
4. Grandner, M. A. et al. Sleep: Important considerations for the prevention of cardiovascular disease. *Curr. Opin. Cardiol.* **31**, 551–565. https://doi.org/10.1097/HCO.0000000000000324 (2016).
5. St-Onge, M. P. Sleep-obesity relation: Underlying mechanisms and consequences for treatment. *Obes. Rev.* **18**(Suppl 1), 34–39. https://doi.org/10.1111/obr.12499 (2017).
6. Malhotra, R. K. Neurodegenerative disorders and sleep. *Sleep Med. Clin.* **17**, 307–314. https://doi.org/10.1016/j.jsmc.2022.02.009 (2022).
7. Hossain, J. L. & Shapiro, C. M. The prevalence, cost implications, and management of sleep disorders: an overview. *Sleep Breath* **6**, 85–102. https://doi.org/10.1007/s11325-002-0085-1 (2002).
8. Aktaruzzaman, M. et al. The addition of entropy-based regularity parameters improves sleep stage classification based on heart rate variability. *Med. Biol. Eng. Comput.* **53**, 415–425. https://doi.org/10.1007/s11517-015-1249-z (2015).
9. Lewicke, A. et al. Sleep versus wake classification from heart rate variability using computational intelligence: Consideration of rejection in classification models. *IEEE Trans. Biomed. Eng.* **55**, 108–118. https://doi.org/10.1109/TBME.2007.900558 (2008).
10. Li, Q. et al. Deep learning in the cross-time frequency domain for sleep staging from a single-lead electrocardiogram. *Physiol. Meas.* **39**, 124005. https://doi.org/10.1088/1361-6579/aaf339 (2018).
11. Radha, M. et al. Sleep stage classification from heart-rate variability using long short-term memory neural networks. *Sci. Rep.* **9**, 14149. https://doi.org/10.1038/s41598-019-49703-y (2019).
12. Welch, A. J. & Richardson, P. C. Computer sleep stage classification using heart rate data. *Electroencephalogr. Clin. Neurophysiol.* **34**, 145–152. https://doi.org/10.1016/0013-4694(73)90041-2 (1973).
13. Aktaruzzaman, M. et al. Performance comparison between wrist and chest actigraphy in combination with heart rate variability for sleep classification. *Comput. Biol. Med.* **89**, 212–221. https://doi.org/10.1016/j.compbiomed.2017.08.006 (2017).
14. Song, T. A. et al. AI-Driven sleep staging from actigraphy and heart rate. *PLoS ONE* **18**, e0285703. https://doi.org/10.1371/journal.pone.0285703 (2023).
15. Hornyak, M., Cejnar, M., Elam, M., Matousek, M. & Wallin, B. G. Sympathetic muscle nerve activity during sleep in man. *Brain* **114**(Pt 3), 1281–1295. https://doi.org/10.1093/brain/114.3.1281 (1991).
16. Doytchinova, A. et al. Simultaneous noninvasive recording of skin sympathetic nerve activity and electrocardiogram. *Heart Rhythm* **14**, 25–33. https://doi.org/10.1016/j.hrthm.2016.09.019 (2017).
17. He, W. et al. Skin sympathetic nerve activity in patients with obstructive sleep apnea. *Heart Rhythm* **17**, 1936–1943. https://doi.org/10.1016/j.hrthm.2020.06.018 (2020).
18. Meng, G. et al. Successful continuous positive airway pressure treatment reduces skin sympathetic nerve activity in patients with obstructive sleep apnea. *Heart Rhythm* **19**, 127–136. https://doi.org/10.1016/j.hrthm.2021.09.018 (2022).
19. Silva, B. M., Rodrigues, J. J., de la Torre Diez, I., Lopez-Coronado, M. & Saleem, K. Mobile-health: A review of current state in 2015. *J. Biomed. Inform.* **56**, 265–272. https://doi.org/10.1016/j.jbi.2015.06.003 (2015).
20. Kazemi, K. et al. Improved sleep stage predictions by deep learning of photoplethysmogram and respiration patterns. *Comput. Biol. Med.* **179**, 108679. https://doi.org/10.1016/j.compbiomed.2024.108679 (2024).
21. Xu, S. et al. A review of automated sleep disorder detection. *Comput. Biol. Med.* **150**, 106100. https://doi.org/10.1016/j.compbiomed.2022.106100 (2022).
22. Pradeepkumar, J. et al. Toward interpretable sleep stage classification using cross-modal transformers. *IEEE Trans. Neural Syst. Rehabil. Eng.* **32**, 2893–2904. https://doi.org/10.1109/TNSRE.2024.3438610 (2024).
23. Everett, T. H. t., Doytchinova, A., Cha, Y. M. & Chen, P. S. Recording sympathetic nerve activity from the skin. *Trends Cardiovasc. Med.* **27**, 463–472 (2017). https://doi.org/10.1016/j.tcm.2017.05.003
24. Malik, J., Lo, Y. L. & Wu, H. T. Sleep-wake classification via quantifying heart rate variability by convolutional neural network. *Physiol. Meas.* **39**, 085004. https://doi.org/10.1088/1361-6579/aad5a9 (2018).
25. Manis, G., Aktaruzzaman, M. & Sassi, R. Bubble entropy: An entropy almost free of parameters. *IEEE Trans. Biomed. Eng.* **64**, 2711–2718. https://doi.org/10.1109/TBME.2017.2664105 (2017).
26. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Measur.* **20**, 37–46 (1960).
27. Somers, V. K., Dyken, M. E., Mark, A. L. & Abboud, F. M. Sympathetic-nerve activity during sleep in normal subjects. *N. Engl. J. Med.* **328**, 303–307. https://doi.org/10.1056/NEJM199302043280502 (1993).
28. Silva, F. B. et al. Sleep staging algorithm based on smartwatch sensors for healthy and sleep apnea populations. *Sleep Med.* **119**, 535–548 (2024).
29. Krauss, D. et al. Incorporating respiratory signals for machine learning-based multimodal sleep stage classification: A large-scale benchmark study with actigraphy and heart rate variability. *Sleep* **48**, zsaf091. https://doi.org/10.1093/sleep/zsaf091 (2025).
30. Morokuma, S. et al. Deep learning-based sleep stage classification with cardiorespiratory and body movement activities in individuals with suspected sleep disorders. *Sci. Rep.* **13**, 17730. https://doi.org/10.1038/s41598-023-45020-7 (2023).
31. Kumar, A. et al. Skin sympathetic nerve activity as a biomarker for syncopal episodes during a tilt table test. *Heart Rhythm* **17**, 804–812. https://doi.org/10.1016/j.hrthm.2019.10.008 (2020).

Acknowledgements
This study was supported in part by NIH Grant R01HL158952.

Author contributions
All authors contributed to the study conception and design. Md Aktaruzzaman: Writing- review & editing, validation, coding, and formal analysis. Thomas H. Everett IV: Writing-review and editing, conceptualization, and supervision. All authors read and approved the final manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to T.H.E.