Turkish Journal of Electrical Engineering and Computer Sciences

Volume 26 | Number 6 Article 3

1-1-2018

Human Sleep Scoring Based on K-Nearest Neighbors

SHAHNAWAZ QURESHI

SEPPO KARRILA

SIRIRUT VANICHAYOBON

Recommended Citation

QURESHI, SHAHNAWAZ; KARRILA, SEPPO; and VANICHAYOBON, SIRIRUT (2018) "Human Sleep Scoring Based on K-Nearest Neighbors," *Turkish Journal of Electrical Engineering and Computer Sciences*: Vol. 26: No. 6, Article 3. https://doi.org/10.3906/elk-1805-12
Available at: https://journals.tubitak.gov.tr/elektrik/vol26/iss6/3

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2018) 26: 2802 – 2818

doi:10.3906/elk-1805-12

Human Sleep Scoring Based on K-Nearest Neighbors

Shahnawaz QURESHI[1], **Seppo KARRILA**[2], **Sirirut VANICHAYOBON**[3*]

[1]Information Systems Technology and Applied Research Laboratory (*i*Star), Department of Computer Science, Faculty of Science, Prince of Songkla University, Songkhla, Thailand

[2]Faculty of Science and Industrial Technology, Prince of Songkla University, Surat Thani, Thailand

Received: 02.05.2018 • **Accepted/Published Online:** 13.08.2018 • **Final Version:** 29.11.2018

Abstract: Human sleep is one of the essential indicators that gauge the overall health and well-being. Presently, it is common for people to face issues related to sleep. Various biomedical signals including electroencephalogram (EEG), electrooculography (EMG), and electrooculography (EOG) are utilized in the diagnosis and during the treatment of sleep disorder cases. An automatic classification to diagnose sleep problems can help in the analysis of sleep EEG data. In this current study, an effort is made to classify the sleep stages from a single EEG channel (C4-A1) based on K-nearest neighbors (K-NN) with three alternative distance metrics. The Euclidean distance is the most commonly used distance measure in K-NN, and no prior study of sleep EEG data has inspected the classification performance of K-NN with various distance measures. Therefore, this study aimed to investigate whether the distance function affects the performance of K-NN in the classification of sleep data. Euclidean, Manhattan and Chebyshev distance measures were individually tested with K-NN classification, and their performances were compared based on accuracy, sensitivity, specificity, F-measure, Kappa statistic and computation time for both Rechtschaffen & Kales and American Academy of Sleep Medicine standard labelings of the sleep stages. The experimental results show that the Manhattan distance function with K = 5 was the best choice for classification of the sleep stages, achieving 98.46% and 98.77% correct rates for the two labelings with comparatively rapid computations.

Key words: K-nearest neighbors, electroencephalogram, sleep stages

1. Introduction

Sleep is an integral part of our life, and it plays a pivotal role in our daily activities. According to a survey, sleep disorder causes on average 16%-20% [1] of road accidents, involving the death of many people annually. Sleep analysis by monitoring the activities of the human by electroencephalogram (EEG), electromyography (EMG), and electrooculography (EOG) helps in identifying sleep-related problems [2]. The EEG method is considered one of the best techniques to record the human brain activity [3–5]. Sleep comprises a sequence of distinct physiological stages that can be distinguished from EEG related features. Neuroscientists have identified different types of brain signals such as alpha, beta, delta, and theta, associated with the sleep stages. According to Rechtschaffen & Kales (R&K) rule [6], there are six different stages of sleep namely: awake (W), non-rapid eye movement (N-REM) stage1, N-REM stage 2, N-REM stage 3, N-REM stage 4, and the rapid eye movement (REM). The American Academy of Sleep Medicine (AASM) proposed a new standard for the sleep stages [7], replacing N-REM stages 1 to 4 with N1, N2, and N3 respectively [2, 7, 8]. The terminologies of R&K and of

*Correspondence: siriut.v@psu.ac.th

AASM for the distinct sleep stages are compared in Table 1, and these are the standard labels that we seek to identify from EEG data.

Table 1. The comparison of sleep stages in R&K and AASM.

R&K	AASM
Stage W	Stage W
N-REM stage 1	N1
N-REM stage 2	N2
N-REM stage 3 N-REM stage 4	N3
REM stage 5	Stage REM

As per AASM, stages 3 and 4 are considered identical and merged under the name deep sleep, also known as slow wave sleep (SWS) stage [9]. N-REM stage 1 represents the transition to sleep, when wakefulness is replaced by the more deliberate alpha waves, and soon the slower theta waves begin to emerge. N-REM Stage 2 is a light sleep stage, during which the EEG signals further decrease but their amplitude increases. The theta waves having frequency 4–7 Hz are interrupted by bursts of activity know as sleep spindle [9]. During sleep stage 2, fast and high amplitude waves are observed, called the K-complex [9]. Delta waves appear in the EEG during N-REM stage 3. Sleep spindle and K-complex continue to arise, although being comparatively less than in stage 2. The Delta waves dominate, and overall the neuronal activity is lowest in N-REM stage 4. The frequency range is below 2 Hz. Different sleep stages with EEG waves are shown in Figure 1.

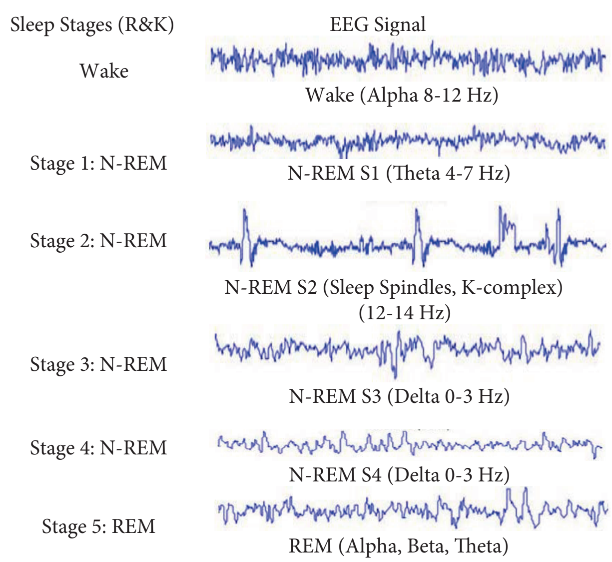


Figure 1. Representation of sleep stages associated with brain signal.

Based on the figure above, reliable and appropriate methods are required to achieve efficient and speedy sleep EEG classification. This paper proposes to identify, by comparative testing, the best K-NN classifier and distance function used by it, for automatic labeling of the sleep stages.

The paper is organized as follows: Section 2 is a brief literature review regarding the classification of sleep stages by various machine learning methods in sleep studies. Section 3 presents methodology and procedures, including the data acquisition process and data decomposition, the distinct statistical features extracted from EEG signals, and the K-NN algorithm for the classification tested here. Section 4 is results and discussion. Finally, conclusions and potential future work are presented in Section 5.

2. Related work

Many feature extraction techniques and classification algorithms are available in the literature to analyze sleep EEG data. The EEG signal provides recordings of sleep information for classification in sleep studies [4, 10]. Time domain statistical features are commonly used to extract meaningful information from EEG, EMG, and EOG signals. Seventy-three characteristics (zero-crossing rate, variance, skewness, kurtosis, etc.) in sleep staging have been identified and used with discriminant analysis and Fisher quadratic classifier [11]. The results show that the classification of sleep data improves with suitable feature selection. Statistical features that are commonly extracted from sleep EEG data are shown in Table 2.

Table 2. Commonly used statistical features to analyse sleep data.

Time Domain Statistical Feature	Studies (REF #)
Standard Deviation	[5, 12]
Median	[5]
Arithmetic Mean	[13]
Skewness	[14, 15]
Kurtosis	[15]
Zero Crossing	[3]
Variance	[16]
Cross correlation	[17]

Chen et al. [18] proposed Hopfield neural network (HNN) classifier to call the different sleep stages during daytime naps using EEG signals, and attained 80.60% accuracy. Phan Huy, et al. [10] presented K-NN technique for classifying sleep EEG data and achieved 98.32% and 94.49% accuracies. However, that study classified the four sleep stages that are awake, stage1 + REM, stage 2 and slow wave stage (SWS) from sleep recording data on four subjects, by using the common Euclidean metric. Another prior study [19] used a support vector machine (SVM) classifier to differentiate between awake and drowsy states. The results show high precision and accuracy, namely 98.01% and 97.91% respectively, of drowsiness detection. Moreover, Kempfner et al. [20] proposed SVM for detecting the sleep stages using EEG and EOG signals in healthy and elderly patients. The proposed algorithm achieved 91.00% success rate.

Özşen et al. [14] proposed NN classification of EEG, EMG and EOG sleep data, achieving 90.93% success rate. Similarly, Hsu et al. [21] presented a recurrent neural network (RNN) using energy features to classify EEG sleep data. The proposed classifier was found to be efficient and accurate with 87.20% success rate. Sen et al. [2] deployed five machine learning algorithms, decision tree (DT), feedforward neural network (FFNN), radial basis network (RBN), SVM and random forest (RF) for classification of sleep stages based on EEG signals. The proposed method obtained 97.03% accuracy rate and the DT algorithm was the fastest among those tested. According to [22], graph domain features such as visibility graph (VG) and horizontal graph (HVG) were tested with an SVM to classify sleep EEG signals. The proposed method was found efficient for single channel sleep classification and obtained 87.50% accuracy. For automatic classification of sleep EEG data, Obayya et al. [23] deployed wavelet transform and fuzzy clustering algorithm (FA). The accuracy obtained was 92.27%. Similarly, Aboalayon et al. [4] proposed five classification techniques SVM, DT, NN, K-NN and naive bayes (NB) with statistical features for classification of sleep stages from EEG signals. The DT classifier achieved the high accuracy of 97.30%. Tsinalis et al. [24] proposed NN model to classify five sleep stages and achieved an overall 78.00% accuracy. Shahnawaz et al. [25] applied RF, bagging and SVM classifiers to classify six sleep stages. RF showed the best performance and achieved 97.73% overall accuracy. Karimzadeh et al. [26] proposed a distributed classification method based on EEG phase and feature envelope by two standards (R&K and AASM). They achieved overall accuracies 88.97% and 83.17%, respectively. To summarize, the previous studies in the literature have proposed various feature extraction and machine learning algorithms for classifying sleep EEG data. Most studies discussed feature extraction and classifier algorithms to label sleep EEG data. This current study is part of our ongoing research. According to systematic literature review conducted we found that different studies have applied K-NN for classification of sleep stages. Besides, based on our previous study [25] we utilized different classifiers and found that K-NN with alternative distance metrics shows good performance for classification of sleep stages among other classifiers. However, this study seeks to select a combination of K-NN with one of the alternative distances measures for classifying sleep EEG data to sleep stages. The alternative measures tested are Euclidean, Manhattan and Chebyshev, which are applied to single channel EEG data.

3. Methodology and procedure

This study aims to evaluate the K-NN classifier with alternative distance measures for classification of sleep stages using EEG data to assist sleep experts in the diagnosis of sleep problems. The objective is to identify the best distance measure based on high classification accuracy in available sleep EEG datasets. The proposed methodology comprises four steps: (1) data preprocessing to remove redundancy and artifacts, (2) feature extraction, (3) K-NN for classification, and (4) performance evaluation. An overview of the proposed process is shown in Figure 2.

Step 1: Data preprocessing

The current study used one single EEG channel (C4-A1), applied filter to remove the artifacts and made a segmentation. The dataset used in this study was obtained from the public St.Vincent dataset (UCD) with full-night recordings from 25 subjects from the PhysioNet database [27]. The subjects included 21 males and 4 females (age: 50 ± 10 years, range 28–68 years, BMI: 31.6 ± 4.0 kg/m2, range 25.1–42.5 kg/m2, and AHI: 24.1 ± 20.3, range 1.7–90.9).

A dataset may include many discrepancies such as incomplete or inconsistent data, or can be lacking in specific actions or trends. Sampling, denoising, normalization, detrending, and calibration [28] are common “data scrubbing techniques” in signal processing applications. However, here we focus on particular aspects of such preprocessing, specifically on artifact removal and segmentation that are common in EEG analysis problems. The EEG sleep recordings suffer from various types of artifacts, such as eye movements (slow and fast), chest movements, head, and body movement, ECG interference, line-noise artifacts, sweat, etc.[28]. Frequency selective (low pass, high pass, band-pass, band-stop) filters have been used in artifact processing to

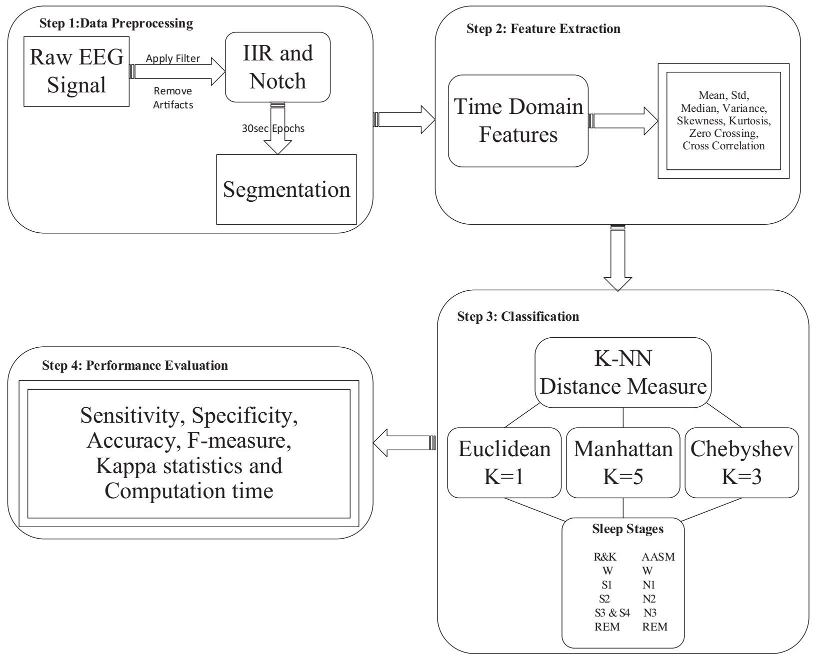


Figure 2. Overview of the steps in this study.

remove muscle artifacts from sleep recordings. Further, regression analysis, Wiener filtering [28], and adaptive filtering [29] are used to eliminate EMG and EOG artifacts from EEG signals. Tenth order infinite impulse response (IIR) bandpass filter and $12th$ order stop band notch filter were utilized to eliminate the baseline interference from the EEG signals. After filtering the data, the signal of a single EEG channel was segmented into 30-sec-long epochs for scoring by the R&K and AASM standards. The sampling rate determined the number of samples in each epoch: at 100 Hz for the EEG channel, the number of samples per epoch was 3.840. The number of epochs varied by length of the sleep recording. The preprocessing steps are summarized as algorithm 1.

Algorithm 1 EEG Sleep Data Preprocessing

Input: Raw EEG Signal
Output: Sleep Stages
1: **for all** *subject* **do**
2: Set sampling frequency (f_s)
▷ *Artifacts' Removing*
3: Apply IIR and Notch
4: Apply sleep expert scores
5: Set samples per epoch = f_s * 30
6: **for** i = 1 **to** number of epochs **do**
7: Get current epoch with samples per epoch
8: **end for**
9: **end for**

Sleep staging

Following the R&K manual, each subject's data were scored and annotated for sleep stages by well-trained technicians. In this study, the EEG signal was divided into 30 sec length samples (epochs) fitting within a single sleep stage. Table 3 documents the epoch counts by sleep stages and their total number based on R&K and AASM.

Table 3. Epoch distribution across the sleep stages.

Sleep Stages R&K							
	W	S1	S2	S3	S4	REM	Total
epochs	773	645	1,181	166	429	646	3,840

Sleep Stages AASM						
	W	N1	N2	N3	REM	Total
epochs	773	645	1,181	595	646	3,840

Step 2: Feature extraction

For the best performance of a classifier with high-dimensional input data, it is necessary to reduce the dimensionality by feature selection or extraction. In machine learning, [30] the original signals are often not used by a classifier; some signals are noisy or irrelevant, and only make the classification more difficult. Different categories of feature extraction can be identified, such as time domain, non-linear, frequency-based, and entropy-based features [2, 31]. These categories have different characteristics, for example, the time-based features tend to be statistical measures.

Table 4. A set of features extracted from EEG data.

Feature Name	Formula
Arithmetic Mean (AM) [13]	$AM = \frac{1}{N}\sum_{i=1}^{N} x_i$
Standard Deviation (SD) [5, 12]	$SD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - AM)^2}{N-1}}$
Median (M) [5]	$M = \lceil \frac{N}{2} \rceil^{th}$
Variance (Var) [16]	$Var = \frac{\sum_{i=1}^{N}(x_i - AM)^2}{N-1}$
Skewness (S) [14, 15]	$S = \frac{\sum_{i=1}^{N}(x_i - AM)^3}{(N-1)SD^3}$
Kurtosis (K) [15]	$K = \frac{\sum_{i=1}^{N}(x_i - AM)^4}{(N-1)SD^4}$
Zero Crossing (ZC) [3]	A ZC count is incremented by one every time the signal crosses from positive to negative or back to positive.
Cross Correlation (CC) [17]	CC function indicates the similarity of a signal to another signa with a given lag.
Here, x_i is a time series and N is the number of data points	

In this study, we focused on time domain features of the EEG signal, since the EEG waveforms vary from one sleep stage to another. Furthermore, this type of features are interpretable and suited for real-time applications. They also require little computation and the baseline manual scoring is done from time traces. The statistical attributes of the EEG epochs that were used as candidate features are presented in Table 4.

Step 3: K-NN for classification

The K-NN is a simple algorithm that stores all cases available for training, and classifies new instances based on their similarity to the learning samples as measured by a distance function, which is often the common Euclidean distance. This is a nonparametric classification technique that completely avoids assumptions about probability densities [4, 5, 32], and it is usually considered when there is no prior knowledge on the distribution of data. Sleep staging has been investigated by use of K-NN classifiers [4, 22].

K-NN is widely used for both classification and regression. Various important benefits of this approach, from the classification point of view, include ease of interpretation, rapid calculations, suitability for large datasets, competitive results, multi-class ability instead of only binary calls, and useful predictions from noisy data. Based on the characteristics of K-NN classifier, we selected it as the classifier for sleep stage classification in this current study. The following steps are involved in applying the K-NN Algorithm 2 to our dataset.

Algorithm 2 Classification K-NN

Input: Sleep EEG Dataset, distanceFunction
Output: Sleep Stages Prediction

```
max = 0
for i=1 to 10 do
    accuracy = K-NN(SleepEEGdataset, distanceFunction, i )
    if (accuracy > max) then
        max = accuracy
        K = i
        store SleepStagesPrediction
    end if
end for
return K, SleepStagesPredicition
```

Distance measures

In data mining, the distance means a concrete way to quantify the similarity/dissimilarity of data points. The distance of two cases is calculated from their feature vectors. We aim to select the best of alternative common distance metrics for the classification of sleep EEG data with K-NN, based on classification performance. For this purpose, the following distances' measures were tested as alternatives.

Case 1: Euclidean distance (ED)

This metric is generally considered the best proximity measure and is efficient and productive with continuous or dense data. The Pythagorean theorem is the basis of this metric, making the distance between two points independent of rotation of coordinate axes. The Euclidean distance of two feature vectors is

$$ED(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

Here x and y are represented by the feature vectors x = $(x_1, x_2 ... x_n), y = (y_1, y_2, ... y_n)$ with n being the dimensionality of feature space. In this study, we applied different K values ranging from 1 to 10 to investigate how K affects the performance of sleep stage classification from EEG data. Tests including K values ranging from 1 to 10 showed the best classification accuracy with K = 1; using more neighbors degraded classifier performance. K = 1 was the best choice with both R&K and AASM standards giving the "true labels". The effects of K value with both standards are shown in Figure 3(a).

Case 2: Manhattan distance (MD)

This metric is the sum of absolute differences by coordinate, and is also known as the L1 norm, taxicab norm, rectilinear distance, or city block distance. The MD is calculated as follows:

$$MD(x,y) = \sum_{i=1}^{n} |x_i - y_i| \tag{2}$$

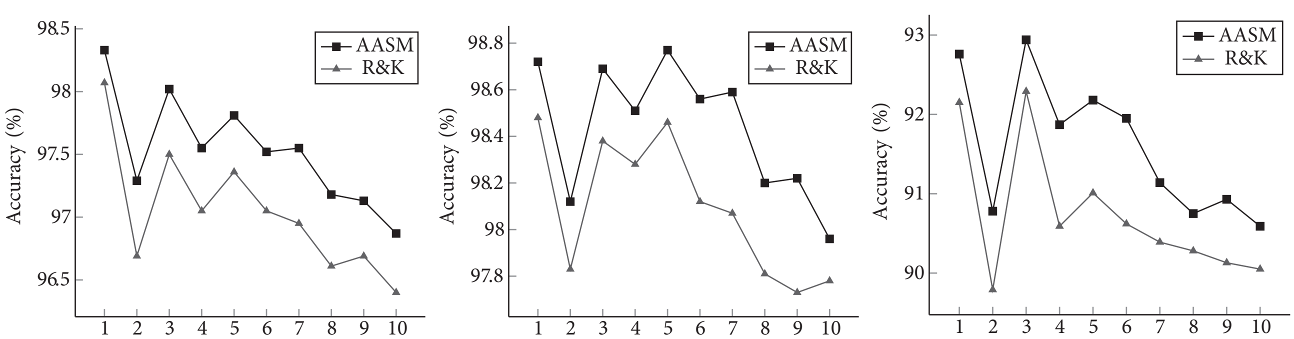
In this study, we applied K values ranging from 1 to 10 to the classification of sleep EEG data, and K = 5 gave the best performance according to both R&K and AASM standards. The effects of K value with both standards are shown in Figure 3(b).

Case 3: Chebyshev distance (CD)

This metric is also known as the maximum value distance, and is suitable in cases when two entities are distinguished by large difference in any of the coordinate dimensions. The Chebyshev distance is defined by

$$CD(x,y) = max_i |x_i - y_i| \tag{3}$$

On testing K range from 1 to 10, we found that K = 3 gave the best results relative to both R&K and AASM standard calls for the sleep EEG data. The effects of K value with both standards are shown in Figure 3(c).



(a) K = 1 to 10 with Euclidean Distance (b) K = 1 to 10 with Manhattan Distance (c) K = 1 to 10 with Chebyshev Distance

Figure 3. Comparison for K values with Euclidean, Manhattan and Chebyshev distance on sleep EEG data.

Step 4: Performance evaluation

The statistical tools used included different statistical parameters as predictive features and assessment of classification by sensitivity, specificity, accuracy, F-measure, kappa statistic, computation time and comparisons between alternative distance measures, with "true labels" based on both R&K and AASM standards. The current study employed the confusion matrix in assessing the performance of K-NN algorithm. Terminology related to the confusion matrix is now discussed.

Sensitivity (Se)

This measures the ability to correctly call cases from a class defined by a fixed correct label.

$$Se = \frac{(TP)}{(TP + FN)}(\%) \tag{4}$$

Here TP and FN are the true positive and false negative counts of the considered class.

Specificity (Sp)

This measures the ability to correctly call cases NOT in a class defined by its correct label. In effect sensitivity and specificity are measures of the purity of separation by a classifier, and are traded off in the sense that one can be improved at a cost to the other, e.g. by adjusting the threshold level in some types of classifier.

$$Sp = \frac{(TN)}{(TN + FP)}(\%) \tag{5}$$

Accuracy (Acc)

This describes the overall correctness of the model and is the sum of correct classification calls divided by the total number of cases.

$$Acc = \frac{(TP + TN)}{(TP + FN + TN + FP)}(\%) \tag{6}$$

F-measure

It is also called F-score and describes the weighted average of precision and recall.

$$F - measure = \frac{(2 * TP)}{(2 * TP + FP + FN)}(\%) \tag{7}$$

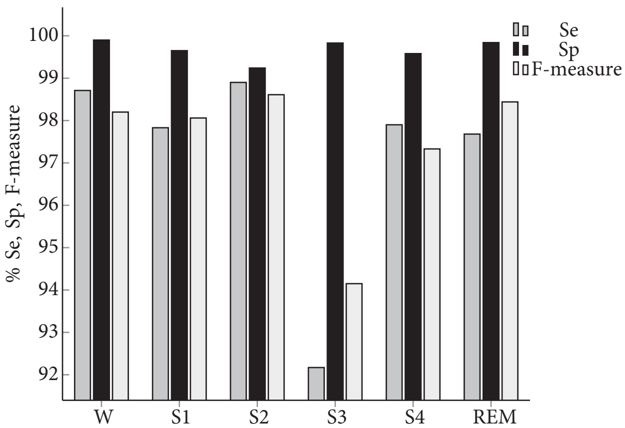
4. Experimental setup

A personal computer with a 3.3 GHz Intel Core ^{TM} i5-4590 processor and 4.00GB RAM was used in the computational experiments. The true labels were determined according to R&K and AASM standards; there were two sets of true labels tested. The raw EEG data was preprocessed to segregate the artifacts by IIR and Notch, as mentioned in step 1. The eight time-domain features of the EEG signals, used as candidate predictive features, were discussed in step 2. The current study used only one single EEG channel (C4-A1). To avoid the class-bias problem [33] from a large number of samples (total of 96,000) with 30 sec epochs of 25 subjects unequally distributed over the six and five classes, stratification was employed to balance the data. Moreover, standardization was used to normalize the features that serve as spatial coordinates of the data instances in the K-NN classification procedure. The differences in sleep patterns between subjects are larger than the variations within one subject [34, 35], so training and testing with only one subject at a time would not support system robustness or generalization. Developing the system with data from all subjects is obligatory to yield a good predictive system that can generalize.

In the K-NN classifier design, the K was set from 1 to 10 for comparison of these alternatives. Moreover, ten-fold cross-validation was utilized by dividing the dataset into 90% training, and 10% testing sets to both train and test the K-NN classifiers. The three distance functions, namely Euclidean, Manhattan, and Chebyshev metrics, were employed in the K-NN classifiers.

4.1. Results and discussion for Euclidean distance with K = 1

The experiment was carried out on the sleep dataset to check the performance in terms of sensitivity, specificity, accuracy, F-measure, and computation time. In the results, K-NN with Euclidean distance (ED) and K = 1 shows good accuracy taking 1.59 sec of training time and produced good classification of the 6 sleep stages at 98.07% overall accuracy. To access the analysis among sleep stages, epoch by epoch analysis was performed by using sensitivity, specificity, and F-measure.



(a) Se, Sp and F-measure with Euclidean Distance K=1

	Predicted						
	Stage	W	S1	S2	S3	S4	REM
	W	763	4	3	0	0	3
	S1	8	631	6	0	0	0
Actual	S2	5	5	1168	2	1	0
	S3	0	0	6	153	7	0
	S4	0	0	3	4	420	2
	REM	5	2	2	0	6	631

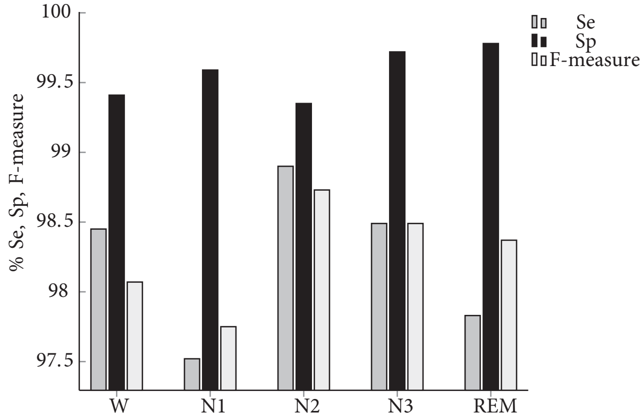
(b) Confusion Matrix with Euclidean Distance K=1

Figure 4. Performance metrics with Euclidean distance K=1 for R&K standard labeling.

Sensitivity for sleep epochs scored shows that this K-NN classified sleep stages correctly. The awake epoch was accurately called as awake epoch. However, the classification of W, S2, and REM achieved good F-measure values with K = 1 and Euclidean distance. The sensitivity, specificity, and F-measure for each sleep stage with R&K set of true labels are shown in Figure 4(a). Also, in the performance evaluation by confusion matrix across all sleep stages based on R&K, the K-NN classifier with Euclidean distance and K = 1 had the best classifier performance of stages W and S2, while the sleep stages S1, S3, S4, and REM had poorer classifier calls. The analyses are summarized in Figure 4(b) .

In the AASM standard, stages S3 and S4 are merged into the one sleep stage N3, known as SWS. When the true labels were provided by this standard, it was found that the K-NN algorithm showed good prediction results with 98.33% overall accuracy, and training time of 1.43 sec in the classification of 5 sleep stages.

The sensitivity, specificity, and F-measure for each sleep stage based on AASM rules are shown in Figure 5(a). However, in performance evaluation by the confusion matrix, the stages W, N2 and N3 had the best performance, while staging N1 and REM was comparatively poorer as shown in Figure 5(b). To summarize, the K-NN classifier with Euclidean distance showed good performance in classification of stages W and S2/N2, being more accurate in the calls than in the other sleep stages.



(a) Se, Sp and F-measure with Euclidean Distance K=1

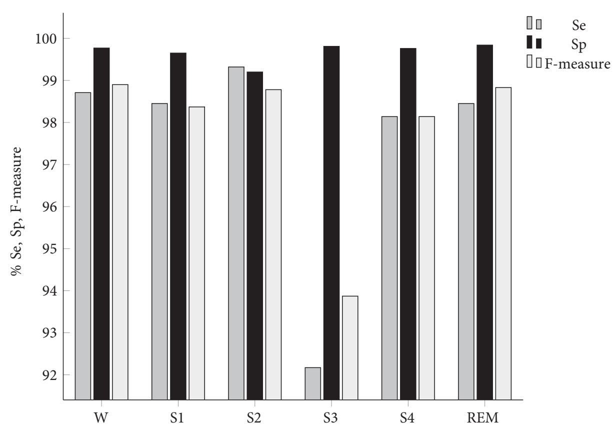
		Predicted				
	Stage	W	N1	N2	N3	REM
Actual	W	761	5	2	1	4
	N1	8	629	8	0	0
	N2	4	7	1168	2	0
	N3	0	0	6	586	3
	REM	6	1	1	6	632

(b) Confusion Matrix with Euclidean Distance K=1

Figure 5. Performance metrics with Euclidean distance K=1 for AASM standard labeling.

4.2. Results and discussion for Manhattan distance with K = 5

The MD metric with K = 5 was employed with the true labels from R&K and AASM standards, and this outperformed the results with the other distance alternatives in terms of sensitivity, specificity, F-measure and computation time. The sensitivity for sleep epochs scored shows that this K-NN classified sleep stages more correctly than K = 1 (ED). Particularly, the W epoch was called more accurately than with K = 1 (ED). F-measure shows the best performance to classify the sleep stages, indicating that this model is better than the K = 1 (ED).



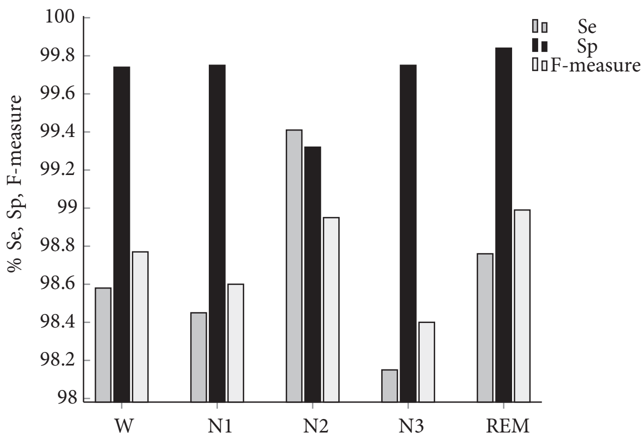
(a) Se,Sp and F-measure with Manhattan Distance K=5

		Predicted					
	Stage	W	S1	S2	S3	S4	REM
Actual	W	763	6	1	0	0	3
	S1	2	635	8	0	0	0
	S2	2	3	1173	3	0	0
	S3	0	0	10	153	3	0
	S4	0	0	2	4	421	2
	REM	3	2	0	0	5	636

(b) Confusion Matrix with Manhattan Distance K=5

Figure 6. Performance metrics with Manhattan distance K=5 for R&K standard labeling.

According to the results in terms of sensitivity, specificity and F-measure shown in Figure 6(a), the Manhattan distance with K = 5 shows the best performance in classifying the sleep stages. Based on the R&K standard, the algorithm achieved 98.46% overall accuracy with 0.42 sec of training time. Further, the classifier called W stage, S1, S2, and REM stage accurately, while S4 was ranked second, but the calls of S3 were not as good as those shown in Figure 6(b). Per the AASM standard, the K-NN classifier with Manhattan distance and K = 5 achieved approximately 98.77% overall accuracy with less training time (0.18 sec). The sensitivity, specificity, and F-measure for each sleep stage based on AASM rules are shown in Figure 7(a). On analyzing the performance for the different sleep stages, it was found that the classifier achieved its highest correct prediction rates for stages W, N1, N2 and REM, while stage N3 had poorer prediction rate. The analyses of K-NN classifier for the different sleep stages, based on AASM standards, are shown in Figure 7(b).



(a) Se, Sp and F-measure with Manhattan Distance K=5

	Predicted					
	Stage	W	N1	N2	N3	REM
	W	762	6	1	0	4
	N1	3	635	7	0	0
Actual	N2	2	1	1174	4	0
	N3	0	0	10	584	1
	REM	3	1	0	4	638

(b) Confusion Matrix with Manhattan Distance and K=5

Figure 7. Performance metrics with Manhattan distance K=5 for AASM standard labeling.

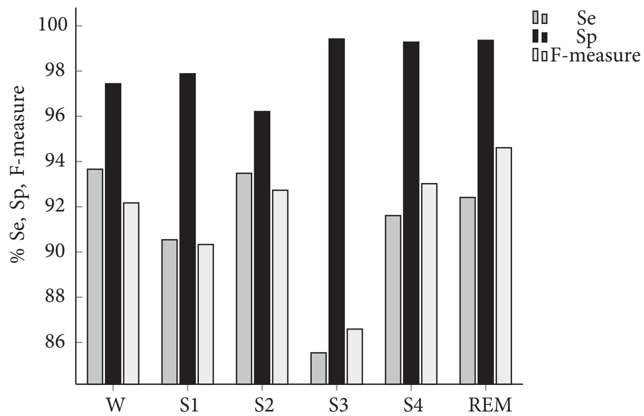
To conclude the discussion, we discovered that the K-NN classifier is a good choice for the prediction of sleep stages W, S1/N1, S2/N2 and REM, relative to both labeling standards.

4.3. Results and discussion for Chebyshev distance with K = 3

The CD metric with K = 3 was applied to our dataset having both R&K and AASM standard labels, for classification calls of those labels. The performance of this combination was not as good as in the previous two cases with different distance metrics. Based on the R&K standard, 92.29% overall accuracy was achieved with 4.25 sec of training. As shown in Figure 8(a), the sensitivity for sleep epoch shows that this K-NN classified sleep stages relatively poorly compared to both K = 1 (ED) and 5 (MD), while the specificity for W sleep epochs was poor. This model does not show promising F-measure values compared to the Euclidean and Manhattan alternatives. Regarding calls of the different stages, the classifier predicted W and S2 stages accurately, but not the other sleep stages. Stages S4 and REM are ranked poorer. The classifier shows poor performance in calls of stages S1 and S3. The analysis of call accuracy for the different stages is shown in Figure 8(b).

When the classifier with CD measure and K = 3 was applied with AASM standard labels, it achieved 92.94% overall accuracy with 4.02 sec of training. The sensitivity, specicity, and F-measure for each sleep stage based on AASM rules are shown in Figure 9(a). It is seen that the classifier called the N2 and N3 stages accurately, but not the other stages, as shown in Figure 9(b). REM ranked second, and in W and N1 prediction the K-NN performance was not good. To summarize our discussion regarding both labeling standards, the overall performance of K-NN with distance measure CD is not as good as the other two distance measures.

Finally, a performance comparison between the three alternative distance' measures regarding the training time is shown in Figure 10. The apparent misclassification rates are also affected by the significant inconsistencies in the sleep patterns of the subjects [32]. Because the data for training and testing were randomly selected for 25 subjects, there is a high likelihood of inconsistency in these samples. For instance, the W stage features could be diverse for the first and second subjects, while training on these feature samples could mislead when trying to identify the awake stage of the third and fourth subjects.

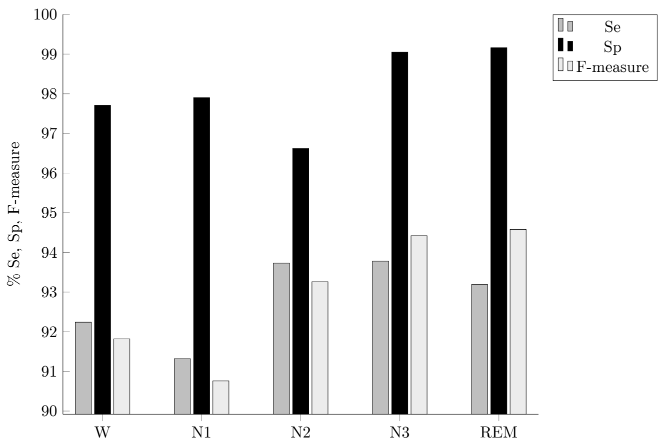


(a) Se, Sp and F-measure with Chebyshev Distance K=3

	Predicted						
	Stage	W	S1	S2	S3	S4	REM
	W	724	19	19	0	1	10
	S1	22	584	36	0	3	0
	S2	26	31	1104	10	5	5
Actual	S3	2	0	18	142	4	0
	S4	4	4	14	10	393	4
	REM	20	10	9	0	10	597

(b) Confusion Matrix with Chebyshev Distance K=3

Figure 8. Performance metrics with Chebyshev distance K=3 for R&K standard labeling.



(a) Se,Sp and F-measure with Chebyshev Distance K=3

	Predicted					
	Stage	W	N1	N2	N3	REM
	W	713	25	21	1	13
	N1	22	589	30	3	1
Actual	N2	21	28	1107	19	6
	N3	4	4	24	558	5
	REM	20	7	11	6	602

(a) Confusion Matrix with Chebyshev Distance K=3

Figure 9. Performance metrics with Chebyshev distance K=3 for AASM standard labeling.

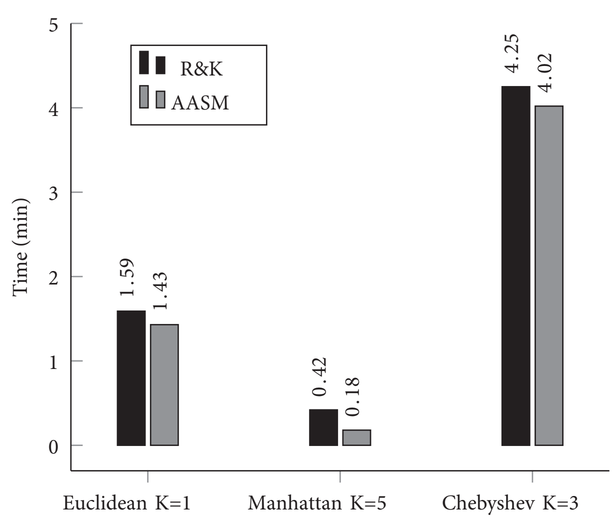


Figure 10. Training times with the alternative distance measures for both R&K and AASM labels.

To sum up, from the confusion matrix of Manhattan alternative with K = 5, it can be seen that this model not only gives the best accuracy but also is good at distinguishing the sleep stages. For instance, W stage is correctly distinguished from stages S3 and S4 for R&K and N3 for AASM labeling. Thus the K-NN classifier K = 5 with Manhattan distance measure outperformed the other alternatives in the classification of sleep stages labeled by both R&K and AASM standards, and with the least training time. However, Landis and Koch [36] have presented ranges for interpreting Kappa values: 0.0–0.20 for slight agreement, 0.21–0.40 for fair agreement, 0.41–0.60 for moderate agreement, 0.61–0.80 for significant agreement, and 0.81–1.00 for perfect agreement. Cohen's kappa statistics were also calculated for the three alternative metrics to make a performance comparison. The best Kappa value (0.98) was for the model with K = 5 and Manhattan distance,

Table 5. Comparison of the proposed method to prior published studies.

Authors	Classifiers	Signal	Sleep Stages	Accuracy %
Chen, Xi, et al	HNN	EEG	Day time Nap Sleep	80.60
Phan, Huy, et al	K-NN	EEG	Awake / Sleep	98.32, 94.49
Yu, Shaoda, et al	SVM	EEG	Awake / Drowsy	97.91
Kempfner, Jacob, et al	SVM	EEG and EOG	Awake,REM,NREM	91.00
Özşen, Seral	NN	EEG, EMG, EOG	AASM	90.93
Hsu, Yu-Liang, et al	RNN	EEG	AASM	87.20
Şen, Baha, et al	DT, FFNN, RBN SVM, RF	EEG	R&K	97.30
Zhu, Guo-hun	SVM	EEG	AASM	87.50
Obayya, Marwa	FA	EEG, EMG, ECG and EOG	R&K	92.27
Aboalayon et al	SVM, DT, NN K-NN, NB	EEG	Awake,REM,NREM	97.30
Tsinalis et al	NN	EEG	AASM	78.00
Shahnawaz et al	RF, Bagging, SVM	EEG	R&K	97.73
Karimzadeh, Foroozan, et al	Distributed classification	EEG	R&K, AASM	88.97, 83.17
Proposed Method	K-NN((K=5) with Manhattan Distance)	EEG	R&K, AASM	98.46, 98.77

with both labeling standards. Based on these results, the K-NN classifier with Manhattan distance and K = 5 could be applied in an automatic sleep classification system.

5. Conclusion

Classification of sleep stages from observed EEG data is a complex task. The process demands a sophisticated controlled experiment setup: keen observation of the patient, critical analysis of the EEG data, and a candid sleep expert's opinion. Despite the multiple barriers faced, yet researchers continue testing specific methods in sleep EEG studies to aid in the diagnosis and treatment of various sleep-related problems. This study focused on evaluating K-NN classifiers with alternative distance metrics to classify EEG data obtained during sleep. For this purpose, we tested K-NN with Euclidean, Manhattan, and Chebyshev distances. The classification performance was carefully evaluated with sleep EEG data collected from 25 subjects. The data were preprocessed to remove artifacts/noise by using independent component analysis, and time domain features were extracted as candidate predictive features.

The K-NN with Manhattan distance and K = 5 showed the best overall performance with accuracies 98.46% and 98.77%, with the least 0.42 and 0.18 sec execution times and Kappa value 0.98 for both R&K and AASM standard labelings of the sleep stages. The K-NN with Manhattan distance and K = 5 could be an excellent choice to classify sleep stages from EEG, but there is still room for improvement. Future work could apply signal processing techniques to extract more unexplored features to achieve better identification of the distinct sleep stages from experimental data. Finally, the performance of the proposed method is compared to prior published studies in Table 5.

Acknowledgements

This work was supported by the Higher Education Research Promotion and the Thailand's Education Hub for Southern Region of ASEAN Countries Project Office of the Higher Education Commission, Prince of Songkla University,Thailand. We thank Publication Clinic, Research and Development Office, and Prince of Songkla University for their help in manuscript proofreading.

References

[1] De Mello MT, Narciso FV, Tufik S, Paiva T, Spence DW, BaHammam AS, Verster JC, Pandi-Perumal SR. Sleep disorders as a cause of motor vehicle collisions. Int J Prev Med 2013; 4: 246-257.

[2] Şen B, Peker M, Çavuşoğlu A, Çelebi FV. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. J Med Syst 2014; 38: 18.

[3] Şen B, Peker M. Novel approaches for automated epileptic diagnosis using FCBF selection and classification algorithms. Turk J Elec Eng Co 2013; 21: 2092-109.

[4] Aboalayon KA, Almuhammadi WS, Faezipour M. A comparison of different machine learning algorithms using single channel EEG signal for classifying human sleep stages. In: IEEE 2015 Systems, Applications and Technology Conference; 1-1 May 2015; Long Island, NY, USA. pp. 1-6.

[5] Chai TY, Woo SS, Rizon M, Tan CS. Classification of human emotions from EEG signals using statistical features and neural network. International Journal of Integrated Engineering 2010; 1: 1-6.

[6] Rechtschaffen A. A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. Clin Neurophysiol 1968; 26: 644.

[7] Iber C, Ancoli-Israel S, Chesson A, Quan SF. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Westchester, IL: American Academy of Sleep Medicine; 2007.

[8] Güneş S, Polat K, Yosunkaya S, Dursun M. A novel data pre-processing method on automatic determining of sleep stages: K-means clustering based feature weighting. In: Complex Systems and Applications (ICCSA) 2009, France. pp. 112-117.

[9] Silber MH, Ancoli-Israel S, Bonnet MH, Chokroverty S, Grigg-Damberger MM, Hirshkowitz M, Kapen S, Keenan SA, Kryger MH, Penzel T, Pressman MR. The visual scoring of sleep in adults. J Clin Sleep Med 2007; 3: 121-131.

[10] Phan H, Do Q, Vu DL. Metric learning for automatic sleep stage classification. In: IEEE 2013 International Conference of the Engineering in Medicine and Biology Society; 3-7 July 2013; Osaka, Japan. pp. 5025-5028.

[11] Šušmáková K, Krakovská A. Discrimination ability of individual measures used in sleep stages classification. Artif Intell Med. 2008; 44: 261-277.

[12] Fathima T, Bedeeuzzaman M, Farooq O, Khan YU. Wavelet based features for epileptic seizure detection. MES Journal of Technology and Management 2011; 2: 108-112.

[13] Subasi A. EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst Appl 2007; 32: 1084-1093.

[14] Özşen S. Classification of sleep stages using class-dependent sequential feature selection and artificial neural network. Neural Comput Appl 2013; 23: 1239-1250.

[15] Gandhi TK, Chakraborty P, Roy GG, Panigrahi BK. Discrete harmony search based expert model for epileptic seizure detection in electroencephalography. Expert Syst Appl 2012; 39: 4055-4062.

[16] Mohseni HR, Maghsoudi A, Shamsollahi MB. Seizure detection in EEG signals: A comparison of different approaches. In: IEEE 2006 Engineering in Medicine and Biology Society International Conference; 30 Aug-3 Sept 2006; New York, NY, USA. pp. 6724-6727.

[17] Cajochen C, Khalsa SB, Wyatt JK, Czeisler CA, Dijk DJ. EEG and ocular correlates of circadian melatonin phase and human performance decrements during sleep loss. Am J Physiol 1999; 277: 640-649.

[18] Chen X, Wang B, Wang X. Automatic sleep stage classification for daytime nap based on hopfield neural network. In: IEEE 2013 Control and Decision Conference; 25-27 May 2013; Guiyang, China. pp. 2671-2674.

[19] Yu S, Li P, Lin H, Rohani E, Choi G, Shao B, Wang Q. Support vector machine based detection of drowsiness using minimum EEG features. In: IEEE 2013 Social Computing International Conference; 8-14 Sept 2013; Alexandria, VA, USA. pp. 827-835.

[20] Kempfner J, Jennum P, Sorensen HB, Christensen JA, Nikolic M. Automatic sleep staging: From young aduslts to elderly patients using multi-class support vector machine. In: IEEE 2013 International Conference Engineering in Medicine and Biology Society; 3-7 July 2013; Osaka, Japan. pp. 5777-5780.

[21] Hsu YL, Yang YT, Wang JS, Hsu CY. Automatic sleep stage recurrent neural classifier using energy features of EEG signals. Neurocomputing 2013; 104: 105-114.

[22] Zhu G, Li Y, Wen PP. Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal. IEEE J Biomed Health 2014; 806: 803.

[23] Obayya M, Abou-Chadi FE. Automatic classification of sleep stages using EEG records based on Fuzzy c-means (FCM) algorithm. In: IEEE 2014 National Radio Science Conference; 28-30 April 2014; Cairo, Egypt. pp. 265-272.

[24] Tsinalis O, Matthews PM, Guo Y. Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders. Ann Biomed Eng 2016; 44: 1587-1597.

[25] Qureshi S, Vanichayobon S. Evaluate different machine learning techniques for classifying sleep stages on single-channel EEG. In: IEEE 2017 Computer Science and Software Engineering International Joint Conference; 12-14 July 2017; Nakhon Si Thammarat, Thailand. pp. 1-6.

[26] Karimzadeh F, Boostani R, Seraj E, Sameni R. A distributed classification procedure for automatic sleep stage scoring based on instantaneous electroencephalogram phase and envelope features. IEEE T Neur Sys Reh 2018; 26: 362-370.

[27] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000; 101: 215-220.

[28] Motamedi-Fakhr S, Moshrefi-Torbati M, Hill M, Hill CM, White PR. Signal processing techniques applied to human sleep EEG signals-A review. Biomed Signal Proces 2014; 10: 21-33.

[29] He P, Wilson G, Russell C, Gerschutz M. Removal of ocular artifacts from the EEG: a comparison between time-domain regression method and adaptive filtering method using simulated data. Med Bio Eng Comput 2007; 45: 495-503.

[30] Chandrashekar G, Sahin F. A survey on feature selection methods. Computers and Electrical Engineering 2014; 40: 16-28.

[31] Subha DP, Joseph PK, Acharya R, Lim CM. EEG signal analysis: a survey. J Med Syst 2010; 34: 195-212.

[32] Yılmaz B, Asyalı MH, Arıkan E, Yetkin S, Özgen F. Sleep stage and obstructive apneaic epoch classification using single-lead ECG. Biomedical Engineering Online 2010; 9: 39.

[33] Lewicke A, Sazonov E, Corwin MJ, Neuman M, Schuckers S. Sleep versus wake classification from heart rate variability using computational intelligence: consideration of rejection in classification models. IEEE Trans Biomed Eng. 2008; 55: 108-118.

[34] Van Dongen HP, Vitellaro KM, Dinges DF. Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. Sleep 2005; 28: 479-498.

[35] Tucker AM, Dinges DF, Van Dongen HP. Trait interindividual differences in the sleep physiology of healthy young adults. J Sleep Res 2007; 16: 170-180.

[36] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-174.

Appendix

Confusion matrices from cross-validation of classifiers using the same features. For a fair comparison, we employed two different algorithms with the same predictive candidate features as used by the K-NN classifier. The multilayer perceptron (MLP) and decision tree (DT) have their best performances and confusion matrices shown in Table 6–13. Furthermore, our previous study [25] is based on different machine learning techniques to classify the sleep stages.

Table A1. Se, Sp, F-measure and overall Acc with R&K standard.

Classifier	Stage	W	S1	S2	S3	S4	REM
MLP	S_e	0.87	0.83	0.93	0.75	0.74	0.89
	S_p	0.96	0.96	0.92	0.98	0.98	0.97
	F-measure	0.87	0.84	0.90	0.75	0.81	0.90
Overall Acc(84.13%)							

Table A2. MLP confusion matrix with R&K standard.

Predicted							
	Stage	W	S1	S2	S3	S4	REM
Actual	W	679	36	30	3	1	24
	S1	41	539	52	3	5	5
	S2	19	25	1106	13	8	10
	S3	1	3	27	125	6	4
	S4	12	14	47	21	320	15
	REM	27	13	8	2	19	577

Table A3. S, Sp, F-measure and overall Acc with AASM standard.

Classifier	Stage	W	N1	N2	N3	REM
MLP	S_e	0.86	0.84	0.91	0.86	0.86
	S_p	0.96	0.96	0.94	0.97	0.97
	F-measure	0.87	0.84	0.90	0.87	0.87
Overall Acc (87.68%)						

Table A4. MLP confusion matrix with AASM standard.

Predicted						
	Stage	W	N1	N2	N3	REM
Actual	W	669	36	22	17	29
	N1	37	546	48	5	9
	N2	24	34	1083	26	14
	N3	5	7	52	513	18
	REM	30	29	9	22	556

Table A5. Se, Sp, F-measur and overall Acc with R&K standard.

Classifier	Stage	W	S1	S2	S3	S4	REM
DT	S_e	0.82	0.82	0.88	0.77	0.74	0.88
	S_p	0.94	0.95	0.93	0.98	0.97	0.96
	F-measure	0.82	0.80	0.88	0.78	0.78	0.87
Overall Acc(84.19%)							

Table A6. DT confusion matrix with R&K standard.

Predicted							
	Stage	W	S1	S2	S3	S4	REM
Actual	W	637	58	37	1	8	32
	S1	47	531	44	3	9	11
	S2	38	44	1047	13	22	17
	S3	2	4	16	129	10	5
	S4	21	13	42	15	319	19
	REM	29	19	12	0	16	570

Table A7. Se, Sp, F-measure and overall Acc with AASM standard.

Classifier	Stage	W	N1	N2	N3	REM
DT	S_e	0.81	0.8	0.88	0.84	0.87
	S_p	0.94	0.95	0.93	0.97	0.96
	F-measure	0.81	0.80	0.87	0.85	0.86
Overall Acc (84.73%)						

Table A8. DT confusion matrix with AASM standard.

Predicted						
	Stage	W	N1	N2	N3	REM
Actual	W	630	41	45	16	41
	N1	53	516	49	11	16
	N2	34	52	1041	32	22
	N3	17	12	45	503	18
	REM	36	10	20	16	564