A CLASSIFICATION OF SLEEP DISORDERS WITH OPTIMAL FEATURES USING MACHINE LEARNING TECHNIQUES

Thakerng Wongsirichot*, Anantaporn Hanskunatai

Department of Computer Science, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, 10520, Thailand

ABSTRACT:

Background: Sleep disorders become one of the early warnings of potential Non-Communicable Diseases (NCDs). Polysomnography (PSG) or sleep test is a formal method to diagnose sleep disorders. However, the PSG is limited in many hospitals due to its high costs. It also requires various sensors attached to a patient, which may cause inconvenience. Moreover, trained sleep specialists are required to interpret the gigantic PSG data. Researchers attempt to identify sleep disorders using alternative techniques.

Method: This study proposed an alternative technique for sleep-related syndrome and sleep disorder classification with optimal features. Patient PSG datasets were retrieved from a hospital in the south of Thailand. In the data preprocessing stage, the datasets were analyzed and normalized using feature extraction and selection mechanisms. Optimal feature selection using the average information gain values was evaluated with the 10-fold cross validation. Four Machine Learning (ML) techniques, *k*MC, *k*NN, SVM and MLP, were used in our experiments. The selected ML techniques have been performed and evaluated with the 10-fold cross validation in data preprocessing and model construction phases.

Results: The *k*NN achieved the highest overall classification results. The optimal features with *k*NN (*opf-k*NN) was proposed. The selected features were PULSE, SAO2, CANR and CHEST. With the selected optimal features, only the ordinary oxygen oximeter and the ECG machine were required. Overall classification result of the *opf-k*NN achieved at 95.17%±3.91.

Conclusion: Although the PSG is the formal sleep disorder diagnosis, alternative diagnostic techniques are beneficial especially to patients. Our study proposed the *opf-k*NN technique to classify sleep disorders with two concerns, the limited access to high-priced medical equipment and patient comfortability. Finally, sleep specialists also obtain benefits in optimizing bio-signal interpretations with only four optimal features.

Keywords: Sleep disorders; Polysomnography; Machine learning; Classification model

DOI: 10.14456/jhr.2017.26 *Received: July 2016; Accepted: November 2016*

INTRODUCTION

A number of people living with Non-Communicable Diseases (NCDs) or chronic diseases rose up to 133 million by the end of 2013 in the United States especially hypertension and cardiovascular diseases [1]. The WHO reported that 17.5 million people die of the cardiovascular diseases or 31% of all worldwide deaths in 2012 [2]. In Thailand, there were more than 400,000 NCD deaths and 29% of those were below the age of 60 in 2010 [3]. One of the early sign symptoms of the NCD is sleep disorders. Sleep disorders are found in both human and animals. There are a number of sleep disorder categories including parasomnias, dyssomnias, and circadian rhythm [4]. One common type of dyssomnias is Obstructive Sleep Apnea (OSA). The OSA is caused by the obstructions of the

* *Correspondence to: Thakerng Wongsirichot*
E-mail: thakerng.w@gmail.com

Cite this article as: Wongsirichot T, Hanskunatai A. A classification of sleep disorders with optimal features using machine learning techniques. J Health Res. 2017; 31(3): 209-17. DOI: 10.14456/jhr.2017.26

upper airway and series of repetitive pauses of breathing (apnea) during sleep. The American Academy of Sleep Medicine (AASM) stated the definition of the OSA when there are five or more detected apnea events per hour during sleep. The detections can be conducted in partial or full night sleep [4, 5]. Other kinds of early warning sleep-related syndromes are hypoventilation, hypoxemic, hypopnea syndromes, Restless Leg Syndrome (RLS), etc. Since 1930s, researchers attempted to describe hidden implications of sleep. The first full night Electroencephalogram (EEG) recording in human was performed [6]. In 1957, two phases of sleep were identified, the Non-Rapid Eye Movement (NREM) and the Rapid Eye Movement (REM).

Rechtschaffen and Kales (R&K) introduced a new scoring manual since 1968. Practically, the NREM stage consisted of four sub sleep stages and only one stage is the REM stage [7]. The most recent sleep stage classification standard is defined by the AASM. Specifically, the NREM consists of N-1, N-2 and N-3 sub-stages and there is only one stage in the REM. Sleep specialists interpret the sleep stages by reading bio-signals in the EEG. The N-1 or NREM-1 stage is a transition step from wakefulness and sleep. Slower heart rates are found in this stage with normal breathing parameters. A person can be easily awaked from external stimulus in the N-1 such as high volume noise, increasing of temperature, etc. Theta wave patterns are found in the EEG. The N-2 or NREM-2 stage is defined as a starting point of actual sleep. The overall muscle activity substantially decreases. A person gains less conscious awareness to external stimulus in the N-2 stage. Theta wave, sleep spindles and K-complexes wave patterns are presented in the EEG. The N-3 or NREM-3 stage is called a deep sleep stage. External stimulus have trivial effect on a person in the N-3 stage. Delta wave and a few sleep spindles are presented in the EEG. The R or REM stage occurs in latter half of sleep cycle before gaining consciousness. A person breathes more rapid with apparently eye movements in the REM stage. A mixed frequency of brain waves mostly low amplitude waves can be found in the EEG. In the sleep disorder diagnosis, Polysomnography (PSG) was used to classify sleep stages. Sleep stages occur in cycles and repeat during sleep. The PSG or sleep test records bio-signals during sleep. The bio-signals are collected by a full range of sensors attached to a patient's body. The sensors include a combination of Electrocardiographic (ECG), Electroencephalographic (EEG), Electrooculographic (EOG), and Electromyographic (EMG) [6, 8, 9].

RELATED WORK

Researchers in related fields proposed techniques that can possibly detect sleep disorders more efficiently. Most of the proposed techniques were constructed based on a combination of mathematical theories and machine learning techniques. A wide range of researches were conducted with different aspects. The identification of sleep stages is used to classify sleep disorders. A formal sleep stage classification framework was proposed for separating sleep stages automatically. It included a feature selection process and a classification process using MLP and flexible decision rules. The results showed 82% accuracy in the deep and paradoxical sleep [10]. A study of the effects of apnea duration in infants on changes of heart rates and oxygen desaturation levels was conducted. 236 apnea epochs were collected from multichannel recordings. The result showed both heart rates and oxygen desaturation levels were significantly related to the apnea durations [11]. An application of ambulatory ECG to identify OSA was studied. Time and frequency domains of Heart Rate Variability (HRV) analysis techniques had been engaged. The results reached 81.25% sensitivity, 46.81% specificity and 64.21% positive predictive value [12]. A Neural Network (NN) system was used to validate the oxygen desaturation levels from oximeters to predict OSA. The overall classification result was promisingly at 93.30% accuracy [13]. A screening OSA test on mobile phones was initiated. A group of researchers built a mobile application so called SleepAp that records Photoplethysmogram (PPG) signals via mobile oximeters. The classification using the Support Vector Machine (SVM) reached 88.40% accuracy in the first experiment and 92.30% accuracy in the second experiment [14].

MATERIALS AND METHODS

Polysomnography (PSG) or sleep test is a formal sleep disorder diagnosis recording a full range of signals from brain, heart, muscle, etc. [15]. A set of full PSG multichannel recordings have been collected from full night sleep (8-12 hours) of patients. The recordings were extracted from the Sleepscan VISION software [16] at the Songklanagarin Hospital, Hat Yai, Songkhla,

Table 1 The PSG signal attributes

Category	Features	Description
Electroencephalogram (EEG)	C3-A2, C4-A1	Monopolar EEG in a position C3-A2, C4-A1
	F3-A2, F4-A1	Bipolar EEG in a position F3-A2, F4-A1
	O1-A2, O2-A1	Monopolar EEG in a position O1-A2, O2-A1
Body and muscle movement	LOC, ROC	Left / right outer canthus
	CHEST	Chest movement
	ABDO	ABDO movement
	CHIN	Chin movement
	LAT, RAT	Left / right anterior tibialis
Electrocardiography (ECG)	ECG	Electrocardiography
	PULSE	Pulse
Thoracic respiratory efforts	CANR	Nasal airflow
	FLOW	Mouth airflow
	SAO2	Oxygen desaturation
	SNR	Amplitude of snoring sounds

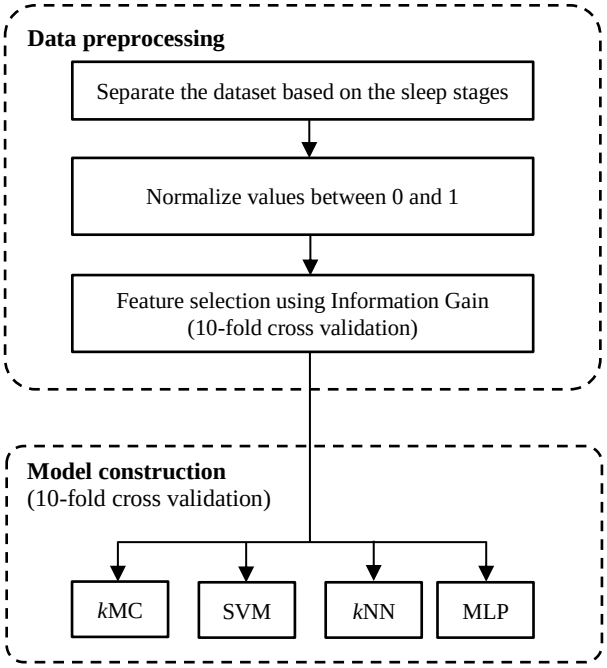

Figure 1 The proposed classification model

Thailand. This study was approved by the hospital's director with explicit attentions to the patient confidentiality.

The collected dataset, Table 1, was initially retrieved from five subjects. The subjects were two males and three females (age 39±11.2 years, BMI 27.75±4.5 kg/m^2, AHI 8.2±2.0). It contained 440,593 records. The dataset consisted of four classes of sleep disorders, including Oxygen Desaturation (D), Hypopnea (H), Isolated Limb Movement (I) and Periodic Limb Movement (P) that were sequentially indicated in every 30-second epochs by professional sleep technicians. The objective of this study was to classify sleep-related syndromes and sleep disorders with optimal features using machine learning techniques. In order to achieve a common classification model, the retrieved datasets, which were retrieved from different patients, were combined as one, without subjective concerns. The proposed classification model consists of two main steps, the data preprocessing and the model construction as shown in Figure 1.

Data preprocessing

The dataset was divided into sleep stages, N-1, N-2, N-3, and R. Each record represented in a form of discrete signal that was a sequence of values in specific time series. Three consecutive data records

Time-stamp	C3-A2	C4-A1	…	Sleep disorder
10478.75	-0.1874	7.4874	…	D
10478.75	-1.7853	7.7363	…	D
10478.75	-3.7690	7.3395	…	D
…	…	…	…	…
16032.01	-18.449	0.9919	…	P
16032.01	-17.258	6.9429	…	P
16032.01	-17.227	7.2478	…	P

Figure 2 An example of the original dataset

Table 2 Information gain values of features in sleep stages

Features	Information gain				
	N-1	N-2	N-3	R	Avg
PULSE	0.297	0.300	0.066	0.675	0.335
SAO2	0.197	0.192	0.087	0.594	0.268
CANR	0.050	0.042	0.015	0.402	0.127
CHEST	0.043	0.040	0.018	0.266	0.092
CHIN	0.030	0.004	0.016	0.272	0.081
ABDO	0.062	0.052	0.020	0.156	0.073
ROC	0.170	0.027	0.037	0.054	0.072
FLOW	0.040	0.030	0.016	0.142	0.057
F3-A2	0. 150	0.018	0.008	0.015	0.048
ECG	0.140	0.010	0.002	0.020	0.043
O1-A2	0.029	0.043	0.030	0.044	0.037
LOC	0.025	0.027	0.031	0.047	0.033
O2-A1	0.006	0.038	0.025	0.055	0.031
C3-A2	0.017	0.034	0.017	0.013	0.020
C4-A1	0.007	0.011	0.010	0.015	0.014
LAT	0.005	0.011	0.004	0.026	0.012
F4-A1	0.007	0.007	0.010	0.019	0.011
RAT	0.004	0.007	0.001	0.008	0.005
SNR	0.003	0.006	0.000	0.000	0.002

with the same timestamp represented one second in time, as shown in Figure 2.

The attribute original values in the dataset were in different scales. Therefore, all of the numeric attributes were normalized between 0 and 1. The 10-fold cross validation method is used to evaluate models especially in classification and prediction problems. In practice, an original dataset is randomly partitioned into ten portions (P_1, P_2, …, P_{10}). A portion (P_i) of the ten portions is selected as a test set. Explicitly, every data portion is tested exactly once and becomes a training set nine times. The training sets are used to construct the classification model and the test sets are used to measure the performance of the classification models. The process is iteratively running until each of the portions has been tested [17]. The Information Gain (IG) is a heuristic function that quantifies abilities of features or attributes in classifying data. A calculation of IG values is to identify high IG features that are used as a set of selected features in a classification model [17-19]. In each of the sleep stages, IG was used to statistically measure each of the features with the 10-fold cross validation method. Table 2 showed the feature selection result using IG. Specifically, in a specific sleep stage, each feature was evaluated with the 10-fold cross validation method. Finally, each feature has ten IGs from the 10-fold cross validation method. The average value of the ten IGs was calculated. Moreover, the average IGs of a feature in all of sleep stages are averaged into one averaged IG values as shown in "avg" column. The average of IGs were used as a ranker of the attributes.

Model construction

The model construction was designed to compare classification performances of four selected ML techniques including *k*-Mean Clustering (*k*MC), *k*-Nearest Neighbor (*k*NN),

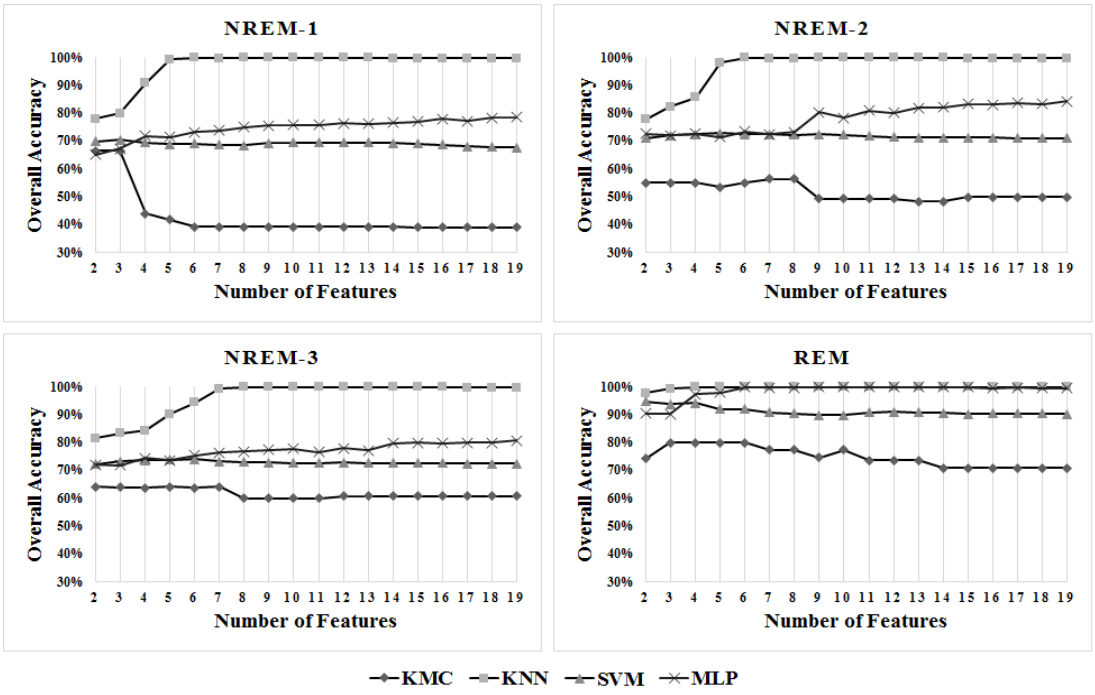

Figure 3 Overall classification accuracy results of the selected ML techniques

Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). The 10-fold cross validation method was used to evaluate the performances of the selected ML techniques. The *k*MC is an unsupervised machine learning technique. It assigns each of the instances in a dataset into exactly one cluster by measuring its similarity without consideration of a predefined class. For classification problems, the classified instance is assigned to a cluster in which the distance between the instance and the cluster is minimum. The *k*NN is a supervised machine learning technique. The number of nearest neighbors (*k*) is set. The *k* nearest neighbors are selected by calculating the distances between all instances in the training set and the classified instance. If the *k* nearest neighbors is set more than one, the instance is classified into a class by a majority vote of the *k* nearest neighbors [17]. The SVM is a supervised machine learning technique. It classifies data into classes by using a hyperplane. The hyperplane is an optimal separator, which calculated from a training set. The original SVM is based on a binary classification category [17]. The MLP is a supervised machine learning technique in the Artificial Neuron Network (ANN). It stimulates the concept of human brain that comprises of a number of neurons or nodes connected together. Each of the neurons contain a computational function so called activation function. It adjusts weights of nodes to improve classification or prediction performance according to a specified learning rate [17].

In our study, there are a number of predetermined conditions and settings. In the *k*MC, the number of *k* clusters was set based on the number of sleep disorder classes in each sleep stages in our experiment. Additionally, the *k*MC was solely used in one of our previous research works to identify the sleep disorders [20, 21]. In the *k*NN, the *k* value was set to 1. In the SVM, the kernel function was selected to the Radial Basis Function (RBF) with the Euclidean Distance function. The RBF was recommended in similar research works [14]. In multi-class problems, more than one SVM binary classifiers are used. The ensemble of SVM classifiers plays an important role with error correcting mechanisms to improve overall classification accuracy. In the MLP, different learning rates were recommended and tested. The suitable learning rate was set to 0.03.

EXPERIMENTAL RESULTS

Figure 3 represented the overall classification accuracy results of the selected ML techniques in each of the sleep stages. The graphs can be classified into three common patterns. Firstly, the accuracies increase when the number of features increase. Secondly, the accuracies substantially decrease

Table 3 A summary of overall classification results of the machine learning (ML) techniques

Sleep stage	ML technique	Overall accuracy	Weighted sensitivity	Weighted F measure	Weighted ROC
N-1	*k*MC	42.79%±0.09	0.50±0.06	0.52±0.03	0.55±0.01
	*k*NN	96.85%±0.07	0.97±0.07	0.97±0.07	0.98±0.05
	SVM	69.13%±0.01	0.69±0.01	0.62±0.01	0.59±0.01
	MLP	74.39%±0.04	0.74±0.04	0.70±0.05	0.81±0.04
N-2	*k*MC	51.62%±0.03	0.52±0.03	0.55±0.02	0.59±0.00
	*k*NN	96.81%±0.07	0.97±0.07	0.97±0.07	0.98±0.04
	SVM	71.78%±0.01	0.72±0.01	0.68±0.01	0.65±0.01
	MLP	78.31%±0.05	0.78±0.05	0.78±0.05	0.77±0.06
N-3	*k*MC	61.64%±0.02	0.62±0.02	0.62±0.01	0.58±0.00
	*k*NN	96.24%±0.07	0.96±0.07	0.96±0.07	0.97±0.06
	SVM	72.90%±0.01	0.73±0.01	0.60±0.01	0.57±0.01
	MLP	77.09%±0.03	0.77±0.03	0.46±0.06	0.74±0.04
R	*k*MC	77.95%±0.02	0.79±0.02	0.79±0.02	0.79±0.04
	*k*NN	99.64%±0.01	1.00±0.00	1.00±0.00	1.00±0.00
	SVM	92.28%±0.02	0.92±0.02	0.92±0.02	0.94±0.01
	MLP	96.95%±0.04	0.97±0.03	0.97±0.03	0.97±0.03

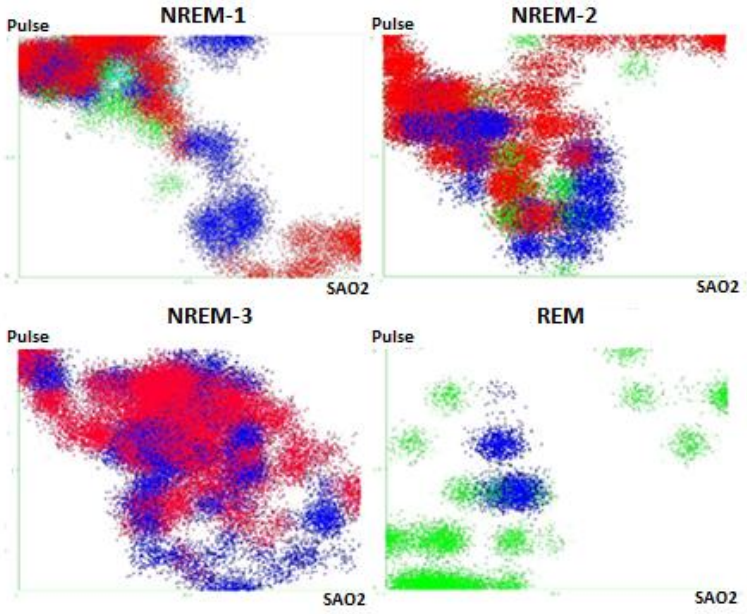

Figure 4 Data visualization of PULSE and SAO2 in all of the sleep stages

when the number of features increase. Finally, the number of features have no effect the accuracies. The *k*NN is classified into the first category. It reveals that the increasing of the number of features reflects the accuracies. On the other hand, the *k*MC classification performance results decrease when the number of features increase, which is categorized in the second case.

Table 3 summarized overall classification results of the selected ML techniques. The measurements were in a form of $\bar{X}$ ± SD. There are four main measurements presented including overall accuracy, weighted sensitivity, weighted specificity and weighted ROC. The calculation of the measurements were conducted based on the number of selected features and their ranked orders. Specifically, the first calculation was conducted with 19 features based on the ranked order as shown in Table 2. The second calculation omitted the lowest IG feature, SNR so there were 18 selected for the calculation. The calculation process continued until the last two highest IG features, Pulse and SAO2, were used in the calculation. All of the measurements were weighted due to the differences

Table 4 A comparison of classification results of the *k*NN(*2f*), *k*NN(*5f*), *opf-k*NN, *best-k*NN

Sleep stage	ML technique	Overall accuracy	Weighted sensitivity	Weighted F measure	Weighted ROC
N-1	*k*NN*(2f)*	77.91%±0.03	0.78±0.03	0.32±0.01	0.77±0.01
	*k*NN*(5f)*	90.82%±0.02	0.91±0.03	0.91±0.03	0.91±0.02
	*opf-k*NN	92.67%±0.05	0.93±0.03	0.93±0.02	0.93±0.02
	*best-k*NN	99.98%±0.02	1.00±0.00	1.00±0.00	1.00±0.00
N-2	*k*NN*(2f)*	77.81%±0.04	0.78±0.03	0.23±0.04	0.78±0.04
	*k*NN*(5f)*	85.73%±0.04	0.86±0.04	0.86±0.02	0.86±0.03
	*opf-k*NN	96.64%±0.03	0.97±0.03	0.97±0.03	0.97±0.03
	*best-k*NN	99.99%±0.01	1.00±0.00	1.00±0.00	1.00±0.01
N-3	*k*NN*(2f)*	81.51%±0.04	0.82±0.03	0.36±0.04	0.80±0.04
	*k*NN*(5f)*	84.33%±0.03	0.84±0.04	0.84±0.04	0.82±0.05
	*opf-k*NN	91.40%±0.03	0.91±0.03	0.91±0.03	0.91±0.03
	*best-k*NN	99.96%±0.02	0.99±0.01	1.00±0.01	0.92±0.02
R	*k*NN*(2f)*	97.75%±0.03	0.98±0.03	0.98±0.02	0.98±0.03
	*k*NN*(5f)*	99.98%±0.02	1.000±0.01	1.000±0.01	1.000±0.01
	*opf-k*NN	99.98%±0.02	1.000±0.01	1.000±0.01	1.000±0.01
	*best-k*NN	100.00%±0.00	1.000±0.00	1.000±0.00	1.000±0.00
AVERAGE	*k*NN*(2f)*	83.75%±9.49	0.84±0.10	0.23±0.16	0.83±0.10
	*k*NN*(5f)*	90.22%±7.08	0.90±0.07	0.90±0.07	0.90±0.08
	*opf-k*NN	95.17%±3.91	0.95±0.04	0.95±0.04	0.95±0.04
	*best-k*NN	99.98%+0.02	0.99±0.01	1.00±0.00	0.98±0.04

of sleep disorder classes.

The best classification results were *k*NN, MLP, SVM and *k*MC, respectively. The overall accuracies of the *k*MC were not promising. The accuracies of the *k*MC reduced dramatically at specific number of features. Theoretically, the *k*MC partitions observations into clusters with the nearest mean. An instance is classified into a cluster based on the calculation of distance between the instance and a nearest cluster. Overlapping between instances in different classes lead to higher misclassification rates. On the other hand, the *k*NN classifies clusters based on the *k* nearest points. Multiple *k* nearest points can be used to classify an instance into a cluster [17] so that it is a possible reason that the *k*NN outperformed the *k*MC in our study. Theoretically, the SVM and the MLP are more robust in classification problems. Finally, the SVM and the MLP seems to have steady classification performance trends. The overall accuracies of the SVM and the MLP are not strongly affected by the alteration of number of features.

Figure 4 showed a data visualization of two axes, namely PULSE and SAO2, which obtained the top two highest average IGs in Table 2. It revealed the *k*MC classified incompetently in the dataset that contained a lot of overlapping instances especially in N-1, N-2 and N-3. In this case, the data distribution in REM was relatively round and less overlapping so it performed reasonably better as shown in Table 3. In addition, the *k*NN confirmed the classification performance was relatively high in the REM. Specifically, there were two D (blue) groups that were located separately to each other and the P (green) group in the REM. The data points were uniformly distributed so that the *k*NN was a suitable classification technique, especially in REM [22].

The first two highest average IG features were PULSE and SAO2, 0.335 and 0.268 in Table 2. The rest obtained the average IGs less than 0.13. Therefore, the features, PULSE and SAO2, were selected as the first two nominated features. The PULSE and SAO2 are a subset of the PPG, which are measured by an ordinary oximeter [14]. Other selected features after PULSE and SAO2 were CANR, CHEST and CHIN, respectively. Technically, CANR and CHEST are monitored using an ordinary ECG. CHIN is recorded by the EMG sensor attached to a patient's chin, which is less patient comfortability. Therefore, the selected optimal features were PULSE, SAO2, CANR and CHEST with considerations of less equipment and patient comfortability. A further experiment was used to measure the *k*NN classification performances with a different number of features. Firstly, the *k*NN(*2f*) consisted of two features, PULSE and SAO2. Secondly, the *k*NN(*5f*) selected only top five ranked features by the average IG values. Thirdly, the *opf-k*NN, the proposed model,

used only four features including PULSE, SAO2, CANR and CHEST. Finally, the *best-k*NN achieved the maximum classification results regardless of a number of selected features. The number of selected features in the *best-k*NN were 12, 12, 16 and 5 in N-1, N-2, N-3, and R, respectively.

Table 4 showed a comparison of overall classification results of *k*NN(*2f*), *k*NN(*5f*), *opf-k*NN and *best-k*NN. The performance measures were evaluated using the 10-fold cross validation method. In average, the *best-k*NN obtained the highest average accuracy at 99.98%+0.02. The second highest was the *opf-k*NN at 95.17%±3.91. The *k*NN*(5f)* was ranked as the third position from the best one with five selected features followed by the *k*NN*(2f)*. Our main intention targeted to the *opf-k*NN that employed appropriate features. Therefore, a paired-sample *t* test was conducted to compare the classification accuracies between each of the *k*NN algorithms, which employed different numbers of features. Significance level was defined as a *P* value of less than 0.05. Firstly, there was no statistically significant difference between the mean of correctly classified classes ($p = 0.18$) of the *best-k*NN and the *opf-k*NN. Secondly, there was statistically significant difference between the mean of correctly classified classes ($p = 0.0085$). According to the paired-sample *t* test results, the *opf-k*NN achieved the overall classification accuracy was comparable to the *best-k*NN with only four optimal features. Moreover, the *opf-k*NN also statistically outperformed the *k*NN(*2f*). The paired-sample *t* tests were conducted based on all of the gathered datasets in order to confirm the generalization of the classification model.

DISCUSSION

In our study, there were a number of factors required for considerations. The detection of sleep disorders normally requires a combination of oxygen oximeter, ECG, EEG, and PSG. The PSG alone includes all of the essential signal sensors for sleep disorder diagnoses. However, the PSG is expensive due to the setup cost and its hardware and software. Some patients feel unconformable during the sleep test due to the attaching sensors. More economical medical equipment is also available separately that can be used for the diagnosis. The main purpose was to technically reduce a number of features to the sleep-related syndrome and sleep disorder classifications. There were two main reasons to obtain the optimal feature set, a reduction of high-priced medical equipment and an improvement of patient comfortability during the diagnosis. Specifically, the statistical test showed that the *best-k*NN and the *opf-k*NN is no statistically significant difference in terms of classification performances. However, the *best-k*NN required the maximum number of features. The proposed *opf-k*NN, required only 4 features including PULSE, SAO2, CANR and CHEST in which only required the ordinary oxygen oximeter and the ECG machine. It can be concluded that the *opf-k*NN overcame the *best-k*NN in terms of costs and patient comfortability with no statistically significant difference. The discovery of the optimal features and the *opf-k*NN technique can assist hospitals or medical centers that are not equipped with the PSG. The proposed model can be used as a screening test. In addition, sleep technicians gain advantages to optimize bio-signal interpretations with only four optimal features. This initial study aids in determining minimum features required for sleep disorder classification. In the case of developing countries, our study can benefit small or medium hospitals that are not equipped with the PSG. There are several limitations to our study. Our subjects were diagnosed with positive OSA symptoms, which do not mimic the diversity of entire population. Another limitation is the selected PSG machine and setting used in the hospital was only from one manufacturer. A comparison to other PSG machines is an advantage.

CONCLUSION

Sleep disorders become one of the early warning signs of the NCDs. A number of researchers applied ML techniques in order to classify the sleep disorders, especially the OSA. The objective of our study was to classify sleep-related syndromes and sleep disorders with optimal features. The average IG was used to rank the optimal features. The optimal features included PULSE, SAO2, CANR and CHEST. A comparison of different ML techniques was thoroughly conducted. The highest classification accuracy was the *k*NN followed by MLP, SVM and *k*MC, respectively. An advanced step was performed to compare the *k*NN with different number of features. Our proposed model, *opf-k*NN, adapted the *k*NN to analyze with the optimal features. The overall classification results were at 95.17%±3.91. The *opf-k*NN achieved the overall classification accuracy is statistically comparable to the *best-k*NN, the best classifier with

maximum number of features. This study had no intention to discover the technique that achieved the highest classification scores but also concern about the minimum use of high-priced medical equipment and maintaining patient comfortability. In common, this study was a systematic investigation to identify effective features for archiving classification goals. The investigation targeted into in-depth data analyses. This study attempted to discover a generalized sleep disorder classification model. The population size could be one of the concerns. However, this study initially reviewed more than 400,000 records of bio-signal data from subjects. The dataset covered both males and females and widely spreading of age 39±11.2. This study benefits in two aspects. Firstly, it can be designed as a screening test with minimum medical equipment and sensors used. Secondly, sleep technicians can apply this technique to optimize bio-signal interpretations with only four optimal features.

REFERENCES

1. Centers of Disease Control and Prevention [CDC]. [cited 2015 September 7]. Available from: http://www.cdc.gov/
2. World Health Organization [WHO]. Cardiovascular disease, Fact sheet N°317; 2009 [cited 2015 September 7. Available from: http://www.who.int/mediacentre/factsheets/fs317/en/index.html
3. Institute of Medical Research and Technology Assessment. Literature review: the current situation and care model of non-communicable diseases. Nonthaburi: Ministry of Public Health, Institute of Medical Research and Technology Assessment; 2014.
4. American Academy of Sleep Medicine [AASM]. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. Sleep. 1999 Aug; 22(5): 667-89.
5. Azarbarzin A, Moussavi Z. Snoring sounds' statistical characteristics depend on anthropometric parameters. Journal of Biomedical Science and Engineering. 2012; 5(5): 245-54. doi: 10.4236/jbise.2012.55031
6. Jafari B, Mohsenin V. Polysomnography. Clinics in Chest Medicine. 2010; 31(2): 287-97. doi: 10.1016/j.ccm.2010.02.005
7. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects. Washington DC: Washington Public Health Service, US Government Printing Office; 1968.
8. Haba-Rubio J, Krieger J. Evaluation instruments for sleep disorders: a brief history of polysomnography and sleep medicine. In: Chiang RPY, Kang SC, editors. Introduction to modern sleep technology. Netherlands: Springer; 2012. p.19-31.
9. Berry R. Fundamentals of sleep medicine. Philadelphia: Elsevier Saunders; 2012.
10. Chapotot F, Becq G. Automated sleep-wake staging combining robust feature extraction, artificial neural network classification, and flexible decision rules. International Journal of Adaptive Control and Signal Processing. 2010; 24:409-23. doi: 10.1002/acs.1147
11. Carbone T, Marrero LC, Weiss J, Hiatt M, Hegyi T. Heart rate and oxygen saturation correlates of infant apnea. J Perinatol. 1999 Jan; 19(1): 44-7.
12. Sun J, Li X, Guo J, Han F, Zhang H. Identification of obstructive sleep apnea syndrome by ambulatory electrocardiography: clinical evaluation of time-domain and frequency-domain analyses of heart rate variability in Chinese patients. Cell Biochem Biophys. 2011 Apr; 59(3): 165-70. doi: 10.1007/s12013-010-9128-6
13. Almazaydeh L, Faezipour M, Elleithy K. A neural network system for detection of obstructive sleep apnea through SpO2 signal features. International Journal of Advanced Computer Science and Applications. 2012; 3(5):7-11.
14. Behar J, Roebuck A, Shahid M, Daly J, Hallack A, Palmius N, et al., SleepAp: an automated obstructive sleep apnoea screening application for smartphones. Computing in Cardiology. 2013; 40: 257-60.
15. Robertson B, Marshall B, Carno MA. Polysomnography for the sleep technologist: instrumentation, monitoring, and related procedures. 1st ed. St. Louis: Elsevier Mosby; 2014.
16. Natus Medical Incorporated. Natus medical incorporated. 2012 [cited 2015 June 12]. Available from: http://www.natus.com/
17. Han J, Kamber M, Pei J. Data mining: concepts and techniques. Waltham: Morgan Kaufmann; 2011.
18. Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. In: ICML '97 Proceedings of the Fourteenth International Conference on Machine Learning; 1997. San Francisco, CA, USA: Morgan Kaufmann; 1997. p. 412-420.
19. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explorations. 2009; 11: 10-8.
20. Wongsirichot T, Iad-ua N, Wibulkit J. A snoring sound analysis application using K-Mean clustering method on mobile devices. In: Kacprzyk J, editor. Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015. Wroclaw, Poland: Springer; 2015. p. 789-796.
21. Wongsirichot T, Hanskunatai A. A comparative investigation of PSG signal patterns to classify sleep disorders using machine learning techniques. In: Intelligent computing theories and methodologies. Fuzhou, China: Springer International Publishing Switzerland; 2015. p. 510-21.
22. Domeniconi C, Peng J, Gunopulos D. Locally adaptive metric nearest-neighbor classification. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002 Sep; 24(9): 1281-5. doi: 10.1109/Tpami.2002.1033219