Selecting Features of Single Lead ECG Signal for Automatic Sleep Stages Classification using Correlation-based Feature Subset Selection

Ary Noviyanto[1], Sani M. Isa[2], Ito Wasito[3] and Aniati Murni Arymurthy[4]

[1] Computer Science, Universitas Indonesia
Depok 16424/Jawa Barat, Indonesia

[2] Computer Science, Universitas Indonesia
Depok 16424/Jawa Barat, Indonesia

[3] Computer Science, Universitas Indonesia
Depok 16424/Jawa Barat, Indonesia

[4] Computer Science, Universitas Indonesia
Depok 16424/Jawa Barat, Indonesia

Abstract

Knowing about our sleep quality will help human life to maximize our life performance. ECG signal has potency to determine the sleep stages so that sleep quality can be measured. The data that used in this research is single lead ECG signal from the MIT-BIH Polysomnographic Database. The ECG's features can be derived from RR interval, EDR information and raw ECG signal. Correlation-based Feature Subset Selection (CFS) is used to choose the features which are significant to determine the sleep stages. Those features will be evaluated using four different characteristic classifiers (Bayesian network, multilayer perceptron, IB1 and random forest). Performance evaluations by Bayesian network, IB1 and random forest show that CFS performs excellent. It can reduce the number of features significantly with small decreasing accuracy. The best classification result based on this research is a combination of the feature set derived from raw ECG signal and the random forest classifier.

Keywords: *ECG features, Correlation-based Feature Subset Selection, RR interval, EDR, Raw ECG Signal, Sleep stages.*

1. Introduction

The quality of sleep directly affects the quality of life. Using a particular measure [1], we can calculate the sleep quality of somebody by knowing the composition of his/her sleep stages. Sleep experts analyze polysomnogram data as the standard technique to determine the sleep stages. Polysomnogram is a simultaneous recording of physiological variables during sleep that include brain activity (*electroencephalogram*, EEG), eye movements (*electroocculogram*, EOG), and chin muscle activity (*electromyohram*, EMG) [2].

Based on the previous works [3, 4, 5, 6], ECG as a substitute of the standard technique to determine the sleep stages has promising results. The main reason of using ECG is the expensiveness of recording process of the polysomnogram data. The data gathering has to do in a sleep laboratory that is expensive with uncomfortable processes for patients and also require trained staff. Another issue is that the sleep study (polysomnography) is costly. It means that the manual determination of the sleep stages in a long sequence of the polysomnogram data is a work that requires endurance and high accuracy. The manual determination of the sleep stages can also trigger lack standard of the sleep stages determination (i.e. every sleep expert may have different results in the sleep stages determination). The automatic processes in the polysomnography are necessary to handle the issue in the manual determination of sleep stages.

There are two groups of sleep in sleep architecture; non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep [7]. The NREM sleep can be divided into NREM 1, NREM 2, NREM 3 and NREM 4. The graphic that represents the sleep stages sequence, called *hypnogram*, is shown in Figure 1. ECG (*Electrocardiogram*), or sometimes called EKG, simply is

a signal from the heart during it is beating [8]. The ECG's waveform and its attributes are shown in Figure 2.
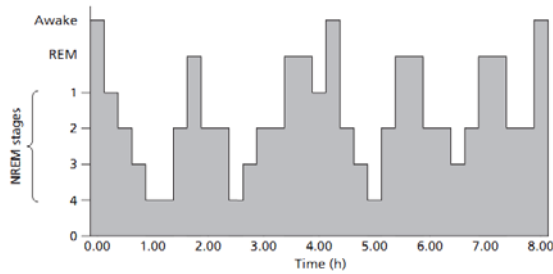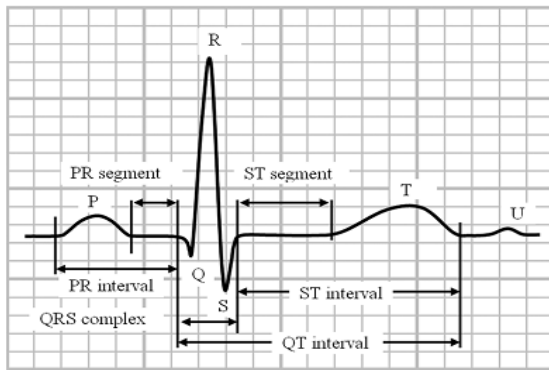

Fig. 1 Normal adult hypnogram [7].

Fig. 2 ECG's waveforms and intervals identified [9].

Many features can be derived from only a single lead ECG signal. Several previous researches used various different features. We will define several feature sets that are constructed from single lead ECG's features and determine the best performance feature set for the automatic sleep stages classification.

2. Previous Works

Several previous researches have given promising result that ECG can be a substitute of the standard data to determine the sleep stages. Shinar *et al.* [3], in 2001, have used only ECG signal to detect Slow Wave Sleep (SWS) with 80% correct identification. They have used time dependent spectral component (Very Low Frequency, Low Frequency and High Frequency) that decomposed from RR interval using wavelet transform as the features.

Lewicke *et al.* [4], in 2008, have determined sleep and wake only using ECG signal of infants. Using fuzzy C-means (FCM) clustering algorithm, they determined that the best feature from standard Heart Rate Variability measure derived from RR interval is the mean. Using rejection approach, the model has achieved 85%-87%

correct classification while rejecting 30% of the data with a kappa statistic of 0.65-0.68.

Yilmaz *et al.* [5], in 2010, have investigated features that derived only from single-lead ECG to classify sleep stages and obstructive apneaic epochs. Using quadratic discriminant analysis (QDA) and support vector machines (SVM) methods have 60% or 70% accuracy for specific sleep stage on healthy subject and 89% accuracy for five obstructive sleep apnea (OSA). They have used the median, inter-quartile range, and mean absolute deviation values that derived from RR interval as the features.

Bsoul *et al.* [6], in 2010, have done research in sleep quality assessment based on ECG measurements. They have used a multi-stage Support Vector Machines (SVM) classifier. Using a binary decision tree (BDT) technique for four classes, three-stage multi-class SVMs are needed and perform good result with high accuracy. This research uses a lot of features. For every ECG segment (i.e. 30 seconds), 112 feature measures can be extracted; 60 for RR time series and 52 for EDR time series.

3. Dataset and Features

In our research, we use ECG signal from the MIT-BIH Polysomnographic Database that can be downloaded from http://www.physionet.org [10]. The MIT-BIH Polysomnographic Database contains multiple physiologic signals during sleep. We have total 18 records of ECG signal with various recorded length from 16 subjects.

From 18 records of ECG signal, we extract several features of single lead ECG signal that potentially can be used to classify the sleep stages. We have three categories of ECG's features; the features that derived from RR interval, EDR information and raw ECG signal. We can get 39 features in total with composition: 12 features from RR interval, 12 features from EDR information and 15 features from raw ECG signal. Those features will be selected and evaluated in order to determine the optimal features for the sleep stages classification.

3.1 Features Derived from RR Interval

RR interval is a distance of two successive top R waves. If the R waves are in normal beats, we can call it as NN interval (Normal to Normal interval). To describe variations of RR interval, we used Heart Rate Variability (HRV) [11]. HRV is divided into time domain measures and frequency domain measures. The common time domain measures consist of [12],
1.   Average of all NN intervals (AVNN);

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

141

2. Standard deviation of all NN intervals (SDNN);
3. Square root of the mean of the squares of differences between adjacent NN intervals (rMSSD);
4. Percentage of differences between adjacent NN intervals that are greater than 50 ms; a member of the larger pNNx family (pNN50);

and the common frequency domain measures consist of [12],

1. Total spectral power of all NN intervals up to 0.04 Hz (TOTPWR);
2. Total spectral power of all NN intervals between 0.003 and 0.04 Hz (VLF);
3. Total spectral power of all NN intervals between 0.04 and 0.15 Hz. (LF);
4. Total spectral power of all NN intervals between 0.15 and 0.4 Hz (HF);
5. Ratio of low to high frequency power (LF/HF).

Based on Yilmaz *et al.* [8], we can also derived three features that belong to time domain measures; they are median, Inter-quartile range (IQR) and Mean absolute deviation (MAD). All of the features that derived from RR interval are coded using prefix "RR-" (e.g. RR-AVNN, RR-SDNN, and so on).

3.2 Features Derived from EDR

ECG signal has been related to respiratory signal [5]. Respiratory information can be derived from ECG signal, called ECG-derived respiration (EDR) information. EDR information is obtained by calculating regions under the QRS segments of ECG signal. The region is a fixed window around R point (100ms centered in R point [13]). Before we can calculate EDR from ECG signal, we have to correct ECG signal by subtracting original ECG signal with the baseline of the ECG signal. The baseline can be calculated by filtering the original ECG signal using median filter of 200 ms and 600 ms respectively [13].

From EDR information, we can also extract same features as RR interval. All of the features that derived from EDR information are coded using prefix "EDR-" (e.g. EDR-AVNN, EDR-SDNN, and so on).

3.3 Features Derived from Raw ECG Signal

Raw ECG signal is the original ECG signal without transformation into other forms (e.g. RR Interval or EDR information). They are 15 features that can be derived from raw ECG signal. The features are listed in Table 1 [14, 15].

Table 1: List of Features Derived from Raw Signal ECG [14, 15]

Features	Equations
Energy	$\sum_i x_i^2$
4th Power	$\sum_i x_i^4$
Curve Length	$\sum_i \|x_i - x_{i-1}\|$
Nonlinear Energy	$\sum_i -x_i \cdot x_{i-2} + x_{i-1}^2$
Peak Power (Max PSD)	$max(PSD)$
Peak Frequency	$index(max(PSD))$
Mean PSD	$mean(PSD)$
Median PSD	$median(PSD)$
Spectral Entropy	$\sum_j PSD_j \cdot \log PSD_j$
Katz Fractal Dimension	$\dfrac{\log N}{\log N + \log \frac{max(\sqrt{(x_i-x_0)^2+i^2})}{\sum_i \sqrt{(x_i-x_{i+1})^2+1}}}$
Detrended Fluctuation Analysis (DFA)	the slope of the line relating log of root-mean-square fluctuation to log n.
Higuchi Fractal Dimension (HFD)	based on Higuchis algorithm
Hjorth Mobility	$\sqrt{\dfrac{\sum_i \frac{x_i-x_{i-1}}{N}}{\sum_i \frac{x_i}{N}}}$
Hjorth Complexity	$\sqrt{\dfrac{\sum_i \frac{(x_i-2x_{i-1}+x_{i-2})^2}{N} \cdot \sum_i \frac{x_i}{N}}{(\sum_i \frac{x_i-x_{i-1}}{N})^2}}$
Petrosian Fractal Dimension (PFD)	$\dfrac{\log N}{\log N + \log(\frac{N}{N+0.4N_\delta})}$

3.4 Dataset Construction

We construct dataset into two groups. In the first group, all of the records (18 records) are combined into a large data. In the second group, each of the record is treated as a separate data; we can call it as *subject-based dataset*. It is intended to find out the robustness of the classifiers and the features. Every group consists of:

1. Feature set *a*: it contains features that are derived from RR interval;
2. Feature set b: it contains features that are derived from EDR information;
3. Feature set *c*: it contains features that are derived from raw ECG signal; and
4. Feature set *d*: it is a combination of the features that are derived from RR interval, EDR information and from raw ECG signal.

4. Methodology

To investigate the best feature set for the sleep stages classification, we evaluate the performance of the feature sets using four different characteristics classifier; they are Bayesian Network learning, Multilayer perceptron, IB1, and Random Forest Classifier. The feature sets can be full set features or selected features. The correlation-based feature subset selection is used to get the best selected features in every feature set.

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

142

4.1 The Classifiers

A Bayesian network is a probabilistic based classifier. A Bayesian network contains a set of variables and a directed acyclic graph (DAG) as the structure that is constructed using K2 algorithm. A Bayesian network represents a probabilities distribution as Equation 1 [16] where U is a set of variable and $pa(u)$ is parents of $u \in U$.

$$P(U) = \prod_{u \in U} p(u \mid pa(u))$$

(1)

MLP (Multilayer Perceptron) is one of the neural network approaches. MLP is a generalization of single layer perceptron that consists of an input layer, one or more hidden layer and an output layer [17]. The visualization of MLP is depicted as Figure 3. The MLP uses the back-propagation learning as the learning algorithm. For this research, the number of neuron in the hidden layer is sum of the number of input neuron and the number of output neuron divided by 2.
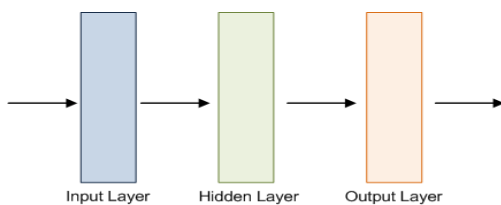

Input Layer   Hidden Layer   Output Layer

Fig. 3 The Multilayer Perceptron.

IB1 is the simplest instance based learner that known as K-nearest neighbors with K=1. In this algorithm, the similarity measure is defined as a negative value of a Euclidean distance, according to Equation 2. How IB1 works can be described in Algorithm 1 [18].

$$sim(x, y) = -\sqrt{\sum_i (x_i - y_i)^2}$$

(2)

Algorithm 1: Instance Based Learning [18]

```
1:  CD ← ∅
2:  for all x ∈ TrainningSet do
3:      y_max ← argmax_{y∈CD}(sim(x, y))
4:      if class(x) == class(y_max) then
5:          classification ← correct
6:      else
7:          classification ← incorrect
8:      end if
9:      CD ← CD ∪ {x}
10: end for
```

Random forest [19] is a decision tree based classifier that contains many classification trees (in this research, we generate 10 trees). The basic idea of random forest classifier is that every single input will be classified using each classification tree in the forest and the final classification result is done by selecting the most votes in the forest.

4.2 The Feature Selection Method

The Correlation-based Feature Subset Selection (CFS) is a feature selection method based on correlation. Correlation is a degree of dependence or predictability of one variable with another [20]. Based on this technique, the feature subsets that can be signatures have two criterions: correlated with the class and uncorrelated one feature with the other. CFS calculate an evaluation measure of feature subset, called *merit*, using an evaluation function according to Equation 3 [20], where S is a feature subset, M_S is the heuristic *merit* of S, k is the feature size, $\overline{r_{cf}}$ is the average of feature-class correlation, and $\overline{r_{ff}}$ is the average of feature-feature inter-correlation.

$$M_S = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}$$

(3)

To get the best feature subset, there are three possible searching techniques [20];

1. Forward selection.
    This searching technique greedily adds one feature until the feature set has no higher value of the evaluation measure.
2. Backward elimination.
    This searching technique greedily eliminates one by one feature until degrading value of the evaluation measure.
3. Best first.
    Best first searches through the search space, either forward or backward. Because of this pure technique is exhaustive; this searching technique has a stopping criterion. The stopping criterion is no improvement value of the evaluation measure over the current best subset in five consecutive fully expanded subsets. In this research, we use the Best first with forward direction as the searching technique.

5. Experimental Results

5.1 Evaluation Measures

We use two evaluation measures; there are percentage of instances correctly classified and Kappa statistic [21]. Kappa statistic gives us a numerical rating of the degree of agreement between the ground truths and the predictions. Table 2 is the interpretation of Kappa statistic. To

convince about the results of evaluation measures, we use k-fold validation method with $k = 10$.

Table 2: Interpretation of Kappa Statistic [21]

Kappa	Agreement
< 0	Less than chance agreement
0.01 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Almost perfect agreement

5.2 Feature Selection Results

Using CFS, we can take the best selected features for the feature set that derived from RR interval, EDR information and Raw ECG signal.

1. Features derived from RR interval.
   From 12 features in total, there are 6 selected features; RR-rMSSD, RR-pNN50, RR-TOTPWR, RR-VLF, RR-LF/HF and RR-MAD.
2. Features derived from EDR information.
   From 12 features in total, there are 5 selected features; EDR-TOTPWR, EDR-VLF, EDR-LF, EDR-HF and EDR-LF/HF.
3. Features derived from Raw ECG signal.
   From 15 features in total, there are 5 selected features; 4th Power, Katz Fractal Dim, Hjorth Mobility, Hjorth Complexity and PFD.
4. Features derived from RR interval, EDR information and Raw ECG signal.
   From 39 features in total, there are 14 selected features; RR-rMSSD, RR-TOTPWR, RR-LF/HF, RR-MAD, EDR-VLF, EDR-LF, EDR-HF, EDR-LF/HF, 4th Power, Katz Fractal Dim, HFD, Hjorth Mobility, Hjorth Complexity and PFD.

Based on these results, the features that are selected in the features selection process of the feature set d (i.e. combination of RR interval, EDR information and Raw ECG signal) are also selected in the features selection process of the feature set a (i.e. RR interval), b (i.e. EDR information) or c (i.e. Raw ECG signal) but not vice versa. For example, RR-rMSSD is selected in the features selection process of the feature set d and also selected in the features selection of the feature set a; RR-pNN50 is selected in the features selection process of the feature set a but not selected in features selection process of the feature set d.

5.2 Result and Discussion

The classification results of the first group dataset are presented in Table 3 and Table 4. We can depict them into line chart in Figure 4. Based on Figure 4, the graphic of dash line and continues line in Bayesian network, IB1 and random forest have shape same pattern. We can see in Figure 4b; multilayer perceptron does not have same shape pattern. Reducing features in the feature set c with MLP as the classifier has decreased the accuracy for 10.65%.

Table 3: The classification result using the full set features; %C is percent correct classification and K is the Kappa statistic

Feature Set	BN		MLP		IB1		RF	
	%C	K	%C	K	%C	K	%C	K
a	41.01	0.24	48.87	0.23	40.00	0.17	50.33	0.28
b	42.69	0.26	54.10	0.30	48.35	0.29	55.73	0.36
c	59.01	0.45	59.32	0.41	79.34	0.71	79.80	0.72
d	60.53	0.48	68.12	0.55	70.56	0.59	76.24	0.66

Table 4: The classification result using the best selected features; %C is percent correct classification and K is the Kappa statistic

Feature Set	BN		MLP		IB1		RF	
	%C	K	%C	K	%C	K	%C	K
a	44.60	0.24	47.55	0.19	38.75	0.15	49.10	0.26
b	41.42	0.24	52.35	0.27	46.99	0.27	52.59	0.33
c	58.52	0.43	48.66	0.24	74.87	0.65	78.85	0.71
d	60.65	0.48	62.40	0.45	67.82	0.55	76.50	0.67

Table 5: Comparison of decreasing accuracy and number of features; DA is Decreasing of Accuracy in percent and DN is Decreasing of Number of Features in percent.

Feature set	BN		MLP		IB1		RF	
	DA	DN	DA	DN	DA	DN	DA	DN
a	-3.59	50.00	1.32	50.00	1.25	50.00	1.22	50.00
b	1.27	58.33	1.75	58.33	1.36	58.33	3.14	58.33
c	0.49	66.67	10.65	66.67	4.47	66.67	0.95	66.67
d	-0.13	64.10	5.72	64.10	2.74	64.10	-0.26	64.10

Table 5 shows the comparison of decreasing accuracy and number of features. The negative value in Table 5 (e.g. BN with feature set a and feature set d, RF with feature set d) means that the accuracy is increase. For Bayesian network and random forest classifiers, selecting appropriate features will increase the accuracy. For multilayer perceptron and IB1, reducing the number of features through features selection process generally will always reduce accuracy.
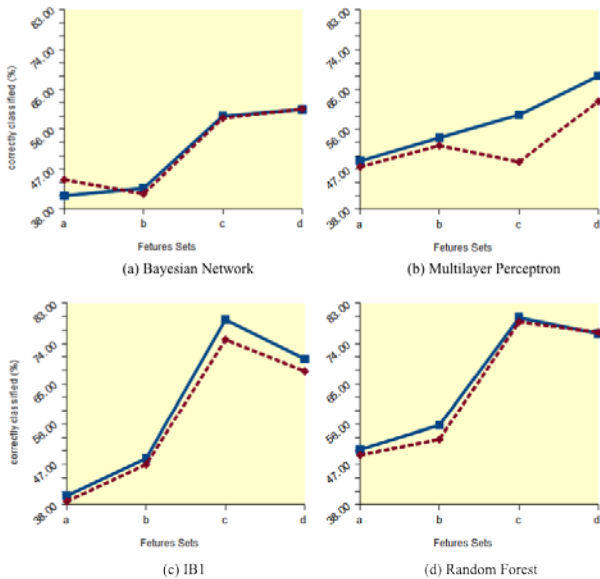
Fig. 4 The effect of feature selection in accuracy over the classifiers; y-axis is correctly classified in percent and x-axis is the feature set; dash line is using selected features and continues line is using full set features.
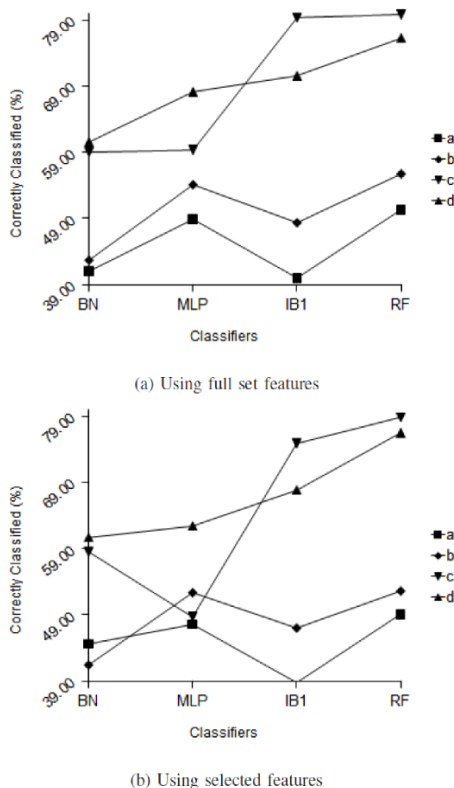

Fig. 5 Comparison of the feature sets; a, b, c, and d are the feature set; y-axis is correctly classified in percent and x-axis is the classifiers.

The line chart in Figure 5 that derived from Table 3 and Table 4 shows the performance of the feature sets. Overall,

feature set c and d are better than feature set a and feature set b for all classifiers. Based on this experiment, the feature set c and feature set d which include raw ECG signal show good performance to determine the sleep stages for the all classifiers.

According to the result in Table 3 and Table 4, the best performance of Bayesian network and Multilayer perceptron are achieved when using the feature set d (i.e. combination of features that derived from RR interval, EDR information and raw ECG signal) in both cases; with and without features selection. Whereas the best performance of IB1 and Random forest are achieved when using the feature set c (i.e. the features that derived from raw signal ECG) in both cases.

In the experiment using combination all of records (i.e. first group dataset); the best accuracy is 79.80% with 0.72 kappa statistic. This performance is achieved using the random forest as the classifier and 15 features that derived from raw ECG signal (i.e. the feature set c without feature selection). Using CFS, the best accuracy is 78.85% with 0.71 kappa statistic. This performance is achieved using the random forest as the classifier and 5 selected features that derived from raw ECG signal. CFS can reduce 66.67% number of features in the feature set c, from 15 features becomes 5 features. The accuracy is reduced by 0.95% and the kappa statistic is reduced by 0.01.

The comparison results of subject based dataset (i.e. second group dataset) are show in Table 6, 8, 10, and 12 for full set features and Table 7, 9, 11, and 13 for selected features. To simplify our analysis we can resume Table 6, 8, 10, and 12 into Table 14 and Table 7, 9, 11, and 13 into Table 15. Using subject based dataset, the random forest classifier and the feature set c that derived from raw ECG signal do the best performance with overall average percent correct classification is 80.95% and the average of kappa statistic is 0.70 in full set features. The random forest classifier and the feature set c do the best performance also in the selected features using CFS; overall average percent correct classification is 79.78% and the average kappa statistic is 0.68. CFS can decrease 66.67% number of features, from 15 features becomes 5 features with only 1.17% decreasing of average percent correct classification and 0,02 decreasing of average kappa statistic. This best result is close to IB1. IB1 using the feature set c can achieve 80.54% average percent correct classification and 0.70 average kappa statistics in full set features; 79.29% average percent correct classification and 0.68 average kappa statistics in selected features.

Table 6: Percent correct comparison of every records using feature set a without feature selection; %C is percent correct classification and K is the Kappa statistic

Rec.	BN %C	K	MLP %C	K	IB1 %C	K	RF %C	K
slp01a	60.40	0.42	58.10	0.38	55.21	0.35	61.09	0.42
slp01b	56.31	0.25	60.18	0.31	59.39	0.33	63.93	0.37
slp02a	64.02	0.46	77.18	0.60	73.73	0.56	76.60	0.58
slp02b	72.30	0.57	77.80	0.64	76.04	0.62	77.26	0.63
slp03	44.99	0.29	54.96	0.33	45.38	0.24	57.37	0.37
slp04	58.32	0.36	66.48	0.32	64.12	0.36	69.48	0.40
slp14	45.06	0.23	52.74	0.28	45.97	0.23	53.94	0.30
slp16	62.58	0.46	64.90	0.47	58.26	0.39	66.33	0.49
slp32	72.23	0.52	77.27	0.57	72.78	0.51	78.21	0.58
slp37	77.33	0.37	89.59	0.55	84.50	0.41	89.08	0.51
slp41	43.23	0.23	49.15	0.30	45.66	0.26	52.10	0.34
slp45	52.14	0.18	58.18	0.15	53.83	0.21	61.22	0.18
slp48	51.96	0.32	57.95	0.38	53.79	0.33	58.70	0.40
slp59	54.20	0.42	56.34	0.44	50.08	0.36	57.92	0.45
slp60	71.15	0.51	71.06	0.49	58.95	0.33	74.52	0.55
slp61	47.33	0.29	54.74	0.33	46.27	0.25	54.00	0.33
slp66	65.61	0.48	67.00	0.50	58.71	0.38	67.32	0.50
slp67x	51.77	0.25	55.04	0.30	55.64	0.32	61.08	0.38
Avg.	**58.39**	**0.37**	**63.81**	**0.41**	**58.80**	**0.36**	**65.56**	**0.43**

Table 7: Percent correct comparison of every records using feature set *a* with feature selection; %C is percent correct classification and *K* is the Kappa statistic

Rec.	BN %C	K	MLP %C	K	IB1 %C	K	RF %C	K
slp01a	61.56	0.40	60.32	0.41	57.40	0.38	62.95	0.45
slp01b	60.96	0.28	60.87	0.30	60.64	0.35	64.62	0.39
slp02a	67.12	0.47	76.63	0.59	72.42	0.53	77.01	0.59
slp02b	72.50	0.57	78.13	0.65	75.66	0.61	77.71	0.64
slp03	48.79	0.30	53.95	0.31	44.86	0.24	56.15	0.36
slp04	61.00	0.37	67.63	0.32	65.56	0.37	70.57	0.42
slp14	48.02	0.22	52.65	0.27	45.10	0.21	53.98	0.30
slp16	60.94	0.42	63.72	0.46	58.70	0.40	65.94	0.49
slp32	73.94	0.51	78.83	0.59	71.86	0.49	78.19	0.58
slp37	86.06	0.47	90.82	0.58	84.45	0.41	88.80	0.50
slp41	45.48	0.26	48.63	0.29	47.59	0.29	50.99	0.33
slp45	55.20	0.19	59.99	0.12	51.61	0.18	60.28	0.17
slp48	51.01	0.30	56.54	0.36	51.61	0.30	55.51	0.35
slp59	54.16	0.41	58.58	0.46	45.29	0.30	58.32	0.46
slp60	72.82	0.52	74.61	0.55	58.98	0.33	74.43	0.55
slp61	49.99	0.28	55.03	0.30	47.98	0.27	54.39	0.33
slp66	64.21	0.46	69.75	0.54	64.40	0.46	66.84	0.50
slp67x	52.50	0.24	58.12	0.34	54.99	0.31	58.71	0.35
Avg.	**60.35**	**0.37**	**64.71**	**0.41**	**58.84**	**0.36**	**65.30**	**0.43**

Table 8: Percent correct comparison of every records using feature set *b* without feature selection; %C is percent correct classification and *K* is the Kappa statistic

Rec.	BN %C	K	MLP %C	K	IB1 %C	K	RF %C	K
slp01a	52.23	0.34	51.74	0.27	45.53	0.20	50.56	0.27
slp01b	65.89	0.46	72.13	0.53	69.76	0.51	75.21	0.58
slp02a	45.20	0.27	72.99	0.53	61.32	0.35	67.18	0.39
slp02b	72.60	0.59	71.79	0.54	73.19	0.57	74.48	0.58
slp03	47.91	0.31	55.91	0.35	52.97	0.34	60.12	0.41
slp04	59.28	0.37	72.65	0.44	67.84	0.42	72.68	0.47
slp14	54.51	0.35	57.20	0.36	54.36	0.33	60.32	0.40
slp16	60.38	0.43	63.33	0.46	59.52	0.42	66.41	0.50
slp32	68.77	0.47	77.88	0.58	73.59	0.53	77.93	0.57
slp37	74.76	0.29	86.63	0.42	83.88	0.32	86.56	0.39
slp41	49.03	0.33	55.63	0.39	53.63	0.37	58.88	0.43
slp45	57.79	0.24	62.74	0.16	59.89	0.30	67.03	0.32
slp48	52.19	0.34	57.82	0.39	55.76	0.36	61.02	0.44
slp59	51.98	0.39	50.78	0.37	45.65	0.31	54.58	0.41
slp60	72.28	0.54	75.82	0.59	68.01	0.47	76.62	0.60
slp61	48.33	0.35	57.49	0.38	57.73	0.41	61.09	0.44
slp66	59.82	0.39	62.43	0.43	59.01	0.38	61.64	0.42
slp67x	66.66	0.48	62.90	0.42	62.00	0.40	66.41	0.47
Avg.	**58.87**	**0.39**	**64.88**	**0.42**	**61.32**	**0.39**	**66.59**	**0.45**

Table 9: Percent correct comparison of every records using feature set *b* with feature selection; %C is percent correct classification and *K* is the Kappa statistic

Rec.	BN %C	K	MLP %C	K	IB1 %C	K	RF %C	K
slp01a	51.04	0.30	50.09	0.26	48.06	0.25	53.16	0.31
slp01b	63.27	0.43	71.00	0.50	75.53	0.60	77.23	0.62
slp02a	44.13	0.26	59.34	0.24	70.09	0.51	69.13	0.46
slp02b	73.49	0.58	72.36	0.55	70.67	0.54	72.73	0.56
slp03	47.78	0.30	57.33	0.36	54.84	0.37	59.86	0.42
slp04	69.70	0.42	74.57	0.46	67.95	0.43	73.43	0.49
slp14	53.08	0.31	53.99	0.26	53.57	0.33	57.25	0.36
slp16	54.46	0.35	62.09	0.42	61.41	0.44	65.03	0.48
slp32	78.78	0.58	79.21	0.59	74.75	0.55	77.39	0.57
slp37	79.23	0.28	86.85	0.30	83.54	0.36	85.55	0.37
slp41	49.87	0.33	55.66	0.39	52.49	0.36	53.94	0.37
slp45	59.98	0.23	61.61	0.05	61.79	0.33	70.64	0.44
slp48	56.81	0.39	61.14	0.44	58.37	0.39	60.72	0.43
slp59	53.66	0.41	53.14	0.39	54.67	0.42	57.70	0.45
slp60	75.33	0.59	77.32	0.61	69.59	0.49	75.71	0.59
slp61	48.91	0.34	57.50	0.36	59.29	0.43	63.12	0.47
slp66	60.58	0.40	66.31	0.49	60.81	0.41	65.23	0.47
slp67x	62.90	0.41	64.30	0.44	56.16	0.32	67.45	0.49
Avg.	**60.17**	**0.38**	**64.66**	**0.39**	**62.98**	**0.42**	**66.96**	**0.47**

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

146

Table 10: Percent correct comparison of every records using feature set *c* without feature selection; %C is percent correct classification and *K* is the Kappa statistic

Rec.	BN		MLP		IB1		RF	
	%C	K	%C	K	%C	K	%C	K
slp01a	69.64	0.58	80.84	0.71	82.14	0.74	84.37	0.77
slp01b	74.97	0.61	83.11	0.72	84.80	0.75	85.29	0.75
slp02a	66.49	0.54	89.18	0.81	86.70	0.78	87.46	0.78
slp02b	81.83	0.72	89.74	0.84	91.44	0.87	90.01	0.84
slp03	67.61	0.57	70.73	0.58	74.51	0.65	80.15	0.72
slp04	68.75	0.52	85.32	0.73	88.52	0.79	87.75	0.77
slp14	61.77	0.45	69.39	0.56	73.07	0.61	72.06	0.59
slp16	73.69	0.63	83.72	0.76	84.52	0.78	83.59	0.76
slp32	61.67	0.40	81.25	0.65	80.56	0.65	82.57	0.68
slp37	78.67	0.43	94.03	0.75	93.83	0.76	94.14	0.76
slp41	58.16	0.44	63.06	0.50	67.16	0.55	68.47	0.57
slp45	67.88	0.46	80.09	0.62	84.83	0.73	83.91	0.70
slp48	63.38	0.47	70.49	0.57	70.35	0.57	69.55	0.56
slp59	64.64	0.56	77.55	0.71	78.49	0.73	76.13	0.69
slp60	71.33	0.53	79.07	0.64	81.09	0.68	83.09	0.71
slp61	61.99	0.51	76.73	0.67	77.94	0.69	77.65	0.69
slp66	73.14	0.60	78.19	0.67	76.00	0.64	75.86	0.64
slp67x	70.61	0.55	74.70	0.60	73.76	0.59	75.07	0.62
Avg.	**68.68**	**0.53**	**79.29**	**0.67**	**80.54**	**0.70**	**80.95**	**0.70**

Table 11: Percent correct comparison of every records using feature set *c* with feature selection; %C is percent correct classification and *K* is the Kappa statistic

Rec.	BN		MLP		IB1		RF	
	%C	K	%C	K	%C	K	%C	K
slp01a	72.27	0.59	73.70	0.60	81.95	0.73	80.91	0.71
slp01b	73.66	0.58	79.03	0.63	83.63	0.73	83.71	0.73
slp02a	83.48	0.72	85.43	0.75	86.52	0.77	86.37	0.76
slp02b	81.32	0.71	87.27	0.80	91.59	0.87	89.27	0.83
slp03	68.77	0.56	62.10	0.41	73.03	0.63	77.70	0.69
slp04	73.68	0.57	81.56	0.65	87.58	0.78	87.77	0.78
slp14	61.34	0.44	64.51	0.47	69.66	0.56	70.19	0.56
slp16	73.45	0.62	78.61	0.68	83.29	0.76	82.05	0.74
slp32	75.40	0.54	80.49	0.63	80.38	0.65	81.60	0.66
slp37	88.55	0.55	91.80	0.62	93.20	0.74	92.61	0.69
slp41	57.39	0.43	60.01	0.45	68.35	0.57	66.97	0.55
slp45	77.35	0.55	75.04	0.50	84.23	0.72	82.96	0.68
slp48	68.72	0.55	69.35	0.56	66.93	0.52	68.99	0.55
slp59	65.44	0.56	70.50	0.62	75.19	0.68	75.44	0.69
slp60	72.84	0.54	70.76	0.49	77.10	0.62	82.26	0.70
slp61	61.80	0.49	65.72	0.50	75.19	0.65	75.60	0.66
slp66	71.44	0.57	76.42	0.65	74.63	0.62	74.95	0.62
slp67x	71.97	0.55	74.91	0.61	74.71	0.61	76.71	0.64
Avg.	**72.16**	**0.56**	**74.84**	**0.59**	**79.29**	**0.68**	**79.78**	**0.68**

Table 12: Percent correct comparison of every records using feature set *d* without feature selection; %C is percent correct classification and *K* is the Kappa statistic

Rec.	BN		MLP		IB1		RF	
	%C	K	%C	K	%C	K	%C	K
slp01a	66.59	0.53	70.75	0.55	70.63	0.56	79.85	0.69
slp01b	79.47	0.68	84.03	0.74	81.56	0.70	86.01	0.77
slp02a	69.52	0.58	85.43	0.76	84.66	0.75	86.90	0.78
slp02b	83.02	0.75	88.40	0.82	86.31	0.79	88.56	0.82
slp03	61.05	0.48	68.59	0.55	70.30	0.59	79.09	0.70
slp04	68.93	0.54	82.42	0.68	82.81	0.69	86.89	0.75
slp14	64.53	0.48	67.41	0.53	67.12	0.53	69.96	0.56
slp16	75.57	0.66	80.94	0.72	78.63	0.69	83.25	0.76
slp32	67.96	0.48	77.58	0.60	79.46	0.64	81.66	0.66
slp37	73.01	0.37	93.18	0.73	92.49	0.70	92.68	0.67
slp41	57.99	0.44	64.68	0.52	67.02	0.55	68.83	0.58
slp45	66.55	0.45	74.54	0.54	77.25	0.60	81.90	0.65
slp48	62.35	0.47	69.21	0.55	71.44	0.59	71.47	0.59
slp59	66.73	0.58	70.21	0.62	66.52	0.58	75.15	0.68
slp60	78.81	0.65	80.48	0.67	77.17	0.62	82.97	0.71
slp61	60.10	0.49	72.50	0.61	76.52	0.66	76.36	0.66
slp66	72.12	0.58	69.46	0.54	69.78	0.55	73.68	0.60
slp67x	74.13	0.61	75.70	0.62	75.74	0.62	74.88	0.61
Avg.	**69.36**	**0.55**	**76.42**	**0.63**	**76.41**	**0.63**	**80.00**	**0.68**

Table 13: Percent correct comparison of every records using feature set *d* with feature selection; %C is percent correct classification and *K* is the Kappa statistic

Rec.	BN		MLP		IB1		RF	
	%C	K	%C	K	%C	K	%C	K
slp01a	68.60	0.55	72.97	0.59	73.82	0.60	79.33	0.68
slp01b	78.21	0.66	82.91	0.72	82.46	0.72	83.87	0.73
slp02a	68.25	0.57	85.78	0.77	84.20	0.74	85.32	0.75
slp02b	82.80	0.74	84.66	0.76	86.66	0.79	86.39	0.79
slp03	63.02	0.49	68.71	0.55	71.68	0.61	78.43	0.69
slp04	73.98	0.59	84.49	0.71	82.98	0.69	87.19	0.76
slp14	62.92	0.46	65.33	0.50	66.17	0.52	69.40	0.55
slp16	75.72	0.65	80.57	0.72	80.66	0.72	82.55	0.75
slp32	71.16	0.51	79.41	0.63	78.53	0.62	82.10	0.67
slp37	86.24	0.55	93.27	0.72	93.80	0.74	92.93	0.70
slp41	58.52	0.45	64.19	0.51	65.21	0.53	67.39	0.56
slp45	70.70	0.50	74.96	0.53	77.41	0.61	81.41	0.65
slp48	64.64	0.49	68.73	0.55	65.67	0.50	69.88	0.56
slp59	64.60	0.55	70.81	0.63	69.03	0.61	73.97	0.67
slp60	82.33	0.70	80.89	0.68	77.56	0.63	84.09	0.73
slp61	63.10	0.51	71.26	0.58	74.87	0.64	77.09	0.67
slp66	73.88	0.61	72.87	0.59	73.17	0.60	75.14	0.63
slp67x	71.85	0.56	73.98	0.59	70.10	0.54	74.44	0.60
Avg.	**71.14**	**0.56**	**76.43**	**0.63**	**76.33**	**0.63**	**79.50**	**0.67**

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

147

Table 14: Average percent correct comparison of every feature set without feature selection; %C is percent correct classification and K is the Kappa statistic

Feature set	BN		MLP		IB1		RF	
	%C	K	%C	K	%C	K	%C	K
a	58.39	0.37	63.81	0.41	58.80	0.36	65.56	0.43
b	58.87	0.39	64.88	0.42	61.32	0.39	66.59	0.45
c	68.68	0.53	79.29	0.67	80.54	0.70	80.95	0.70
d	69.36	0.55	76.42	0.63	76.41	0.63	80.00	0.68

Table 15: Average percent correct comparison of every feature set with feature selection; %C is percent correct classification and K is the Kappa statistic

Feature set	BN		MLP		IB1		RF	
	%C	K	%C	K	%C	K	%C	K
a	60.35	0.37	64.71	0.41	58.84	0.36	65.30	0.43
b	60.17	0.38	64.66	0.39	62.98	0.42	66.96	0.47
c	72.16	0.56	74.84	0.59	79.29	0.68	79.78	0.68
d	71.14	0.56	76.43	0.63	76.33	0.63	79.50	0.67

4. Conclusions

The CFS shows good result for Bayesian network, IB1 and random forest with relatively small amount in decreasing accuracy and significantly reducing the number of the features but it is not too good for multilayer perceptron. Increasing accuracy in Table 5 indicates that not all features will be useful and adding inappropriate features may decrease the accuracy of some classifiers. For example Bayesian network classifier with selected feature set *a* and feature set *d*, random forest classifier with selected feature set *d*. Multilayer perceptron and IB1 always show better results when not reducing the number of features using the features selection method.

Overall, random forest classifier is better than Bayesian network, multilayer perceptron or IB1 for any feature sets in the first group dataset. On the other hand, based on our experiment and setting, the features that derived from raw ECG signal have more potency to be the signature of sleep stages than the features that derived from RR interval or EDR information. This feature set has higher accuracy and kappa statistic than the others. The combination of the random forest classifier and the features that derived from raw ECG signal shows the best performance. Using full set features from raw ECG signal, we can get 79.80% correctly classified instances and 0.72 kappa statistic. CFS as feature selection method shows good result. We can reduce 66.67% number of features that derived from raw ECG signal and only reduce 0.95% accuracy and 0.01 kappa statistic. The accuracy becomes 78.85% and the kappa statistic becomes 0.71.

The experiments using subject based dataset support our conclusion that the combination of the random forest as the classifier and the features that derived from raw ECG signal perform better result than the others. Using full set features from raw ECG signal, we can get 80.95% of average percent correct classification and 0.70 average kappa statistic. CFS also performs well; using selected features, we can get 79.78% average percent correct classification and 0.68 average kappa statistics; it means only reduce 1.17% in the average percent correct classification and 0.02 in the average kappa statistics. This results show almost the same with the first group dataset. It means the classifiers and the feature set are robust.

References

[1] American Academy of Sleep Medicine, *International classification of sleep disorders, revised: Diagnostic and coding manual,* Chicago, Illinois: American Academy of Sleep Medicine, 2001.

[2] C. Pollak, M.J. Thorpy, and J. Yager. *The Encyclopedia of Sleep and Sleep Disorders. Facts on File library of health and living*. New York: Facts on File, 2009.

[3] Z. Shinar et al., "Automatic detection of slow-wave-sleep using heart rate variability", *Computers in Cardiology 2001*, 2001, pp. 593-596.

[4] A. Lewicke, et al., "Sleep versus wake classification from heart rate variability using computational intelligence: consideration of rejection in classification models", *IEEE Trans Biomed Eng*, Vol. 55, No. 1, 2008, pp. 108–18.

[5] B. Yilmaz et al., "Sleep stage and obstructive apneaic epoch classification using single-lead ecg", *BioMedical Engineering OnLine*, Vol. 9, No. 1, 2010, pp. 39.

[6] M. Bsoul et al., "Real-time sleep quality assessment using single-lead ECG and multi-stage SVM classifier", in *the International Conference of IEEE Engineering in Medicine and Biology Society*, 2010, Vol 2010, pp 1178–1181.

[7] J.M. Shneerson. *Sleep medicine: a guide to sleep and its disorders*, Oxford: Blackwell Publishing, 2005.

[8] G. D. Clifford, F. Azuaje, and P. McSharry, *Advanced Methods And Tools for ECG Data Analysis*, USA:Artech House, Inc., 2006.

[9] R. V. Andreo, B. Dorizzi, and J. Boudy, "Ecg signal analysis through hidden markov models", *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 8, 2006, pp. 1541–1549.

[10] A. L. Goldberger et al., "Physiobank, physiotoolkit, and physionet : Components of a new research resource for complex physiologic signals", *Circulation*, Vol. 101, No. 23, 2000, pp. e215–220.

IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011
ISSN (Online): 1694-0814
www.IJCSI.org

148

[11] "Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and lectrophysiology", *Circulation*, Vol. 93, No. 5, 1996, pp. 1043–1065.

[12] J. E. Mietus. *Time domain measures:from variance to pnnx.* Internet: http://physionet.org/events/hrv-2006/mietus-1.pdf. June 2011.

[13] P. de Chazal, T. Penzel, and C. Heneghan, "Automated detection of obstructive sleep apnoea at different time scales using the electrocardiogram", *Physiological Measurement*, Vol. 25, No. 4, 2004, pp. 967.

[14] F. S. Bao, X. Liu, and C. Zhang, "PyEEG:An open source python module for EEG/MEG feature extraction. *Comp. Int. and Neurosc.*, Vol. 2011, 2011.

[15] M. Wiggins et al., Evolving a bayesian classifier for ecg-based age classification in medical applications. *Applied Soft Computing*, Vol. 8, No. 1, 2008, pp. 599 – 608.

[16] R. R. Bouckaert, "Bayesian network classifiers in weka", Technical report, University of Waikato, May 2008.

[17] S. Haykin, *Neural Networks: A Comprehensive Foundation (2nd edition)*, Upper Saddle River, NJ: Prentice Hall, 1999.

[18] D. W. Aha and D. Kibler. "Instance-based learning algorithms", In *Machine Learning*, 1991, pp 37–66.

[19] L. Breiman, "Random forests", *Mach. Learn.*, Vol. 45, 2001, pp. 5–32.

[20] M. A. Hall, "Correlation-based Feature Subset Selection for Machine Learning", PhD thesis, University of Waikato, Hamilton, New Zealand, 1998.

[21] A. J. Viera and J. M. Garrett, "Understanding Interobserver Agreement: The Kappa Statistic", *Family Medicine*, Vol. 37, No. 5, 2005, pp. 360–363.

Ary Noviyanto holds Bachelor degree in Computer Science from Department of Computer Science, Universitas Gadjah Mada, Indonesia. He is a research assistant for image processing and pattern recognition laboratory in faculty of computer science, Universitas Indonesia.

Sani M. Isa holds Bachelor degree in Mathematics from Faculty of Natural Science, Padjadjaran University, Indonesia; and Master degree from Faculty of Computer Science, University of Indonesia.

Ito Wasito holds Ph.D in Computer Science, School of Computer Science and Information Systems, Birkbeck College, University of London, United Kingdom.

Aniati Murni Arymurthy holds Bachelor degree in Electrical Engineering from Faculty of Engineering, University of Indonesia, Master degree in Computer and Information Science, The Ohio State University, United States of America, and Doctor degree in Dept. Of Optoelectronics and Laser Application, University of Indonesia.