IEIE Transactions on Smart Processing and Computing, vol. 6, no. 1, February 2017
https://doi.org/10.5573/IEIESPC.2017.6.1.021

Automatic Detection of Sleep Stages based on Accelerometer Signals from a Wristband

Minsoo Yeo[1], Yong Seo Koo[2], and Cheolsoo Park[1]**

[1] Department of Computer Engineering, Kwangwoon University / Seoul, Korea
{minsooyeo119112@gmail.com, parkcheolsoo@kw.ac.kr}
[2] Department of Neurology, Korea University Medical Center, Anam Hospital, Korea University College of Medicine / Seoul, Korea yo904@naver.com

***** Corresponding Author: parkcheolsoo@kw.ac.kr

Received January 16, 2017; Accepted February 6, 2017; Published February 28, 2017

** Regular Paper*

***Abstract*:** In this paper, we suggest an automated sleep scoring method using machine learning algorithms on accelerometer data from a wristband device. For an experiment, 36 subjects slept for about eight hours while polysomnography (PSG) data and accelerometer data were simultaneously recorded. After the experiments, the recorded signals from the subjects were preprocessed, and significant features for sleep stages were extracted. The extracted features were classified into each sleep stage using five machine learning algorithms. For validation of our approach, the obtained results were compared with PSG scoring results evaluated by sleep clinicians. Both accuracy and specificity yielded over 90 percent, and sensitivity was between 50 and 80 percent. In order to investigate the relevance between features and PSG scoring results, information gains were calculated. As a result, the features that had the lowest and highest information gain were skewness and band energy, respectively. In conclusion, the sleep stages were classified using the top 10 significant features with high information gain.

***Keywords*: Artificial Intelligence, Pattern recognition and classification, Signal processing**

1. Introduction

Polysomnography (PSG) is one of the most accurate ways to diagnose sleep disorders. It includes electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), and many other physiological signals. However, there are some disadvantages with PSG. First, it is difficult to take a PSG test due to the complexity of the procedure and its high cost. Additionally, patients can feel stressed and uncomfortable owing to the unfamiliar environment of sleep laboratories and hospitals, which can affect the physiological signals. Although there have been studies into replacing PSG [1], these studies were not suitable for completely replacing it, since the complexity of the procedures were still problematic. In this study, we propose a method using a wearable wristband to score sleep stages automatically, which is a comfortable way for patients to undergo testing in their own homes. Furthermore, it is not expensive to use an accelerometer sensor to estimate sleep stages.

2. Methods

Fig. 1 shows the procedure of our proposed method to score sleep stages using an accelerometer sensor. First, PSG data from the subjects were obtained while simultaneously recording accelerometer data from the wrist. Then, we preprocessed the accelerometer data using a fifth-order Butterworth filter to remove movement artifacts [2, 3]. The preprocessed signals were classified into four sleep stages using five different machine learning algorithms. Next, the information gains were calculated to choose the optimal number of features for the classification of sleep stages, and 10 significant features with high information gain were selected.

2.1 Experiment

Thirty-six subjects took part in the experiment, sleeping for about eight hours in a sleep laboratory, where all subjects underwent standard PSG. Signals including EEG, EOG, bi-lateral tibials EMG, ECG, airflow (nasal

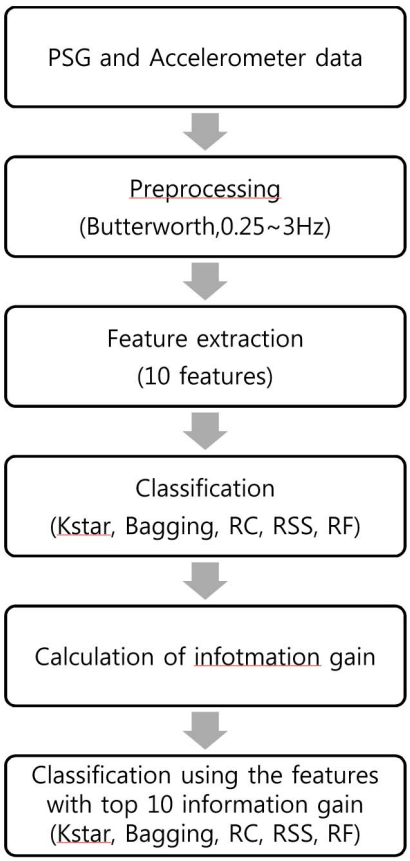

Fig. 1. The procedure of the proposed method.

Table 1. The sleep information of all subjects, including sleep duration, BMI, height, and weight.

	Time (h)	BMI	Height	Weight
Average	8.353889	22.63472	172.4605	68.15789
Std. Dev.	0.56882	2.674178	8.779877	13.27164

thermistor), chest and abdominal excursion (piezo bands), and oxyhemoglobin saturation were recorded. Additionally, movement data from the subjects were recorded using a three-axis accelerometer sensor placed on the non-dominant wrist. The accelerometer data were recorded during the eight-hour sleep with 100 sampling frequencies sufficient to measure the movement of the subjects during sleep. Table 1 shows the average and standard deviation for the sleep durations across all subjects, including BMI, height, and weight.

2.2 Dataset

As mentioned earlier, accelerometer data at the N1 and N2 stages were similar in nature and merged into a light stage, while the N3 stage was considered a deep sleep stage [4]. Hence, four sleep stages (wake, rapid eye movement [REM], light, and deep sleep) were estimated in the experiment. Fig. 2 illustrates the distribution of each sleep stage epoch number. The average number of epochs of the wake, REM, and deep sleep stages were about 100. In addition, the average epochs of the light sleep stage

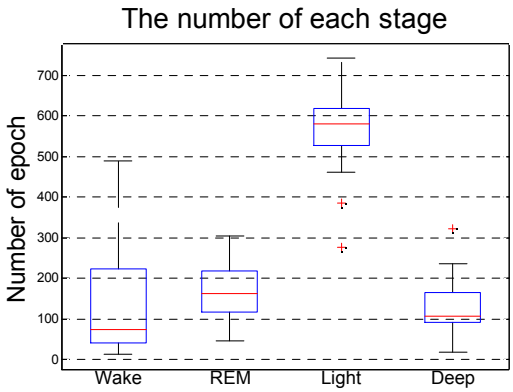

Fig. 2. The distribution in the number of epochs in each sleep stage. While the average number in wake, REM, and deep sleep stages was 100, the average number in the light stage was over 550.

Table 2. The description of features. Each feature was calculated for x-axis, y-axis, z-axis, and the square root of the three axes, yielding 40 features in total.

Feature	Equation	Simple description
Mean	$\bar{a} = \frac{1}{N}\sum_{n=1}^{N} a(n)$	The mean represents the total amount of human movement. Where N is data length of an epoch and $a(n)$ is the accelerometer data.
Standard Deviation	$Std(a) = \sqrt{(\frac{1}{N-1}\sum_{n=1}^{N}(a(n)-\bar{a})^2)}$	Standard deviation of the accelerometer data.
Correlation	$Corr(x,y) = \frac{1}{N-1}\sum_{n=1}^{N}\frac{(x_n-\bar{x})(y_n-\bar{y})}{std(x)std(y)}$	The correlation coefficient between two accelerometer data from different axis.
Kurtosis	$Krt(a) = \frac{E[(a-\bar{a})^4]}{std(a)^2} - 3$	The kurtosis is a measure of the peakedness of the signal, which is computed using the 4th central moment and the standard deviation .
Crest Factor	$Crest(a) = \frac{max(a(n))}{\sqrt{\frac{1}{N-1}\sum_{n=1}^{N} a(n)^2}}$	The crest represents signal impulsiveness. Where the max($a(n)$) is shows the maximum value of accelerometer data.
Skewness	$Sk(a) = E\left[\left(\frac{a-\bar{a}}{std(a)}\right)^3\right]$	The skewness of the distribution of the accelerometer data.
Zero Crossing	$ZCR(a) = \lvert\{n \in \mathbb{N} \mid (2 \le n \le N) \wedge (a(n)\cdot a(n-1) < 0)\}\rvert$	The number of zero crossing in the accelerometer data.
Entropy	$EP(a) = -\sum_{k} p_a(k)\ln(p_a(k))$	The entropy of the accelerometer data p_a is a probability mass function of the accelerometer data.
Band Energy	$BE(t) = \frac{\sum_{n=1}^{N} a'(n)}{\sum_{n=1}^{N} a(n)}$	The energy of normalized subband normalized by total energy of the signals, where a' shows a subband signal (0.25-3Hz)
Spectral Flux	$SF(t) = \sum_{n=2}^{N}(a_t(n) - a_{t-1}(n))^2$	The magnitude of difference between the successive data, where a_t represents the accelerometer data at t' epoch.

were about 600.

2.3 Preprocessing

In order to remove movement artifacts, the accelerometer signals from subjects were filtered using a fifth-order Butterworth filter, designed with MATLAB R2014a. The cut-off frequency of the bandpass filter was from 0.25 Hz to 3 Hz.

2.4 Feature Extraction

The preprocessed data were collected with 30-second epochs for comparison with the PSG result, which was used as ground truth. Then, the data were changed into features. Table 2 shows a detailed description of each feature. There were 40 features in our study, since each feature was extracted for the x-, y-, and z-axis, and the

intensity of three axes.

2.5 Classification

Accelerometer data were classified into the sleep stages using five supervised machine learning algorithms, that is, KStar, bagging, random committee, random subspace, and random forest. Weka tools were utilized to implement the algorithms [4]. For validation of performance, 10-fold cross validation was applied to the separation results among the sleep stages.

The KStar classifier is an instance-based classifier. The class of a test instance is based upon the class of those training instances determined by some similarity function. It is different from other instance-based classifiers since it uses an entropy-based distance function [5].

Bagging is the ensemble classifier that consists of many classifiers traded by different training subsets. The results are decided by majority vote of the classifiers [6].

Random committee is an ensemble classifier that consists of many base classifiers. Each base classifier is constructed using a different random number seed. The final prediction results are a simple average of the results estimated by each base classifier [7].

Random subspace (RS) is an attribute bagging method that can avoid overfitting with a small amount of data. This machine learning algorithm consists of weak classifiers that compensate for the small size of the data. It can improve performance by using a random sample of features, instead of using all the features. Random subspace is different from other ensemble classifiers, such as Bagging, since it uses random subsets of the feature space [8].

Random forest (RF) is a supervised machine learning algorithm consisting of many decision trees. Each decision tree is made by bootstrap samples of the same dataset, and it uses a random variable selection. The results are decided by majority vote of decision trees that comprise the random forest [9].

2.6 Feature study

We calculated the information gain of each feature to look into the optimal features for automatic sleep scoring. Information gain is defined as follows [10]:

$$\begin{aligned} &InfoGain(Class, Feature) \\ &= H(Class) - H(Class \mid Feature) \end{aligned} \tag{1}$$

Based on investigations of the information gain, the band energy (BE) of the x-axis had the largest information gain, yielding 0.32. More than half of the features had information gain over 0.2. The feature with the lowest information gain was skewness of the y-axis: 0.05. Fig. 3 illustrates the information gain of each feature.

Sleep stages were classified by increasing the number of the features based on high information gain to investigate the optimal number of sleep-related features. As can be seen in Fig. 4 (classification performance corresponding to the number of features), around 10 significant features could yield meaningful performance.

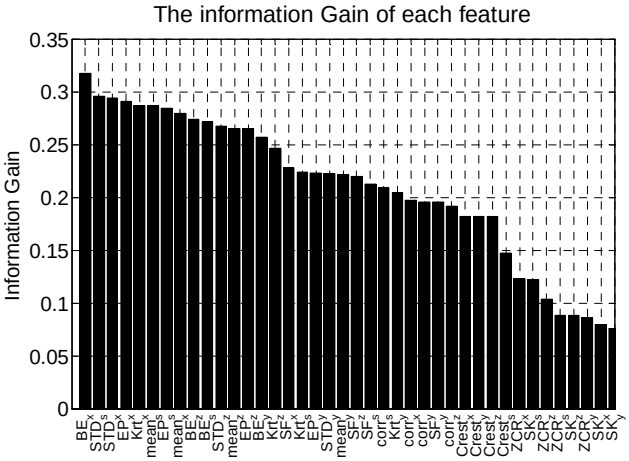

Fig. 3. The information gain of each feature.

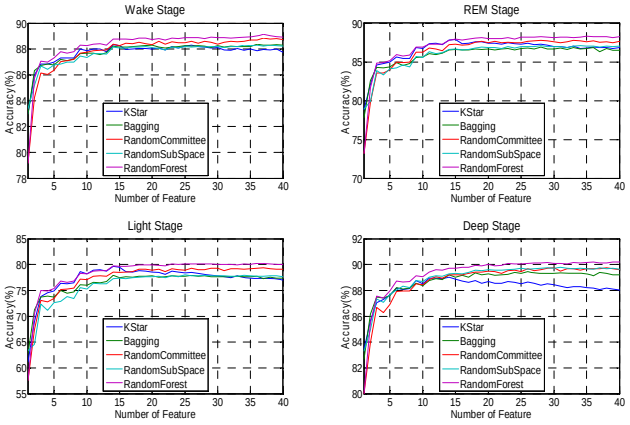

Fig. 4. The average accuracies of the 36 subjects depending on features with high information gain.

3. Result

The following performance indicators (accuracy, sensitivity, and specificity) were computed for the assessment of the classification results.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively. PSG estimation results were used as ground truth. Fig. 5 shows the accuracies of sleep scoring results using machine learning algorithms. Overall, the average accuracies of wake, REM, and deep sleep stages were greater than 90%, and accuracy of the light stage was over 80%.

Fig. 6 displays the sensitivities of the sleep scoring

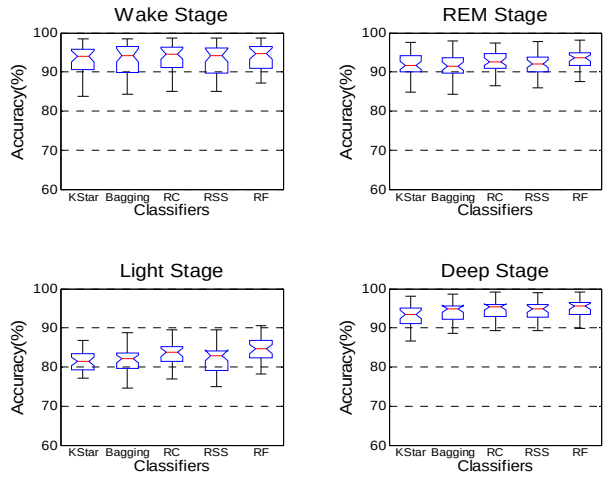

Fig. 5. Classification accuracies of four sleep stages across the 36 subjects using five classification algorithms.

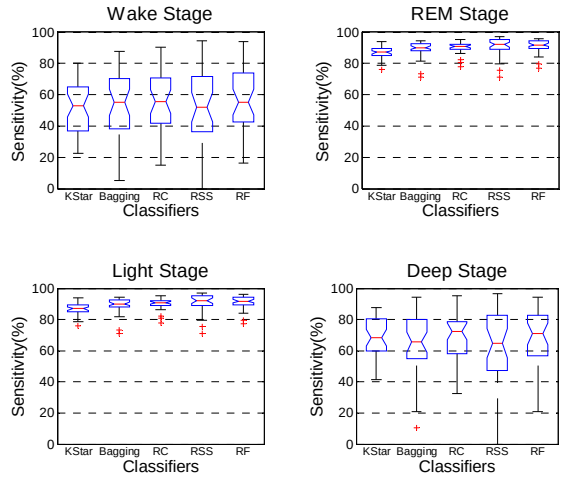

Fig. 6. Classification sensitivities of the four sleep stages across the 36 subjects using five machine learning algorithms.

results using machine learning algorithms. Overall, the average sensitivity for the light and REM stages was greater than 90%, and for the deep sleep and wake stages, around 50%. The small number of epochs in the wake and deep sleep stages caused relatively lower sensitivity, compared to the other stages.

Fig. 7 shows the specificity in the sleep scoring results. Overall, the average specificity for REM, deep sleep, and wake stages was greater than 90% and specificity of the light stage was around 70%. The larger number of epochs for the light stage caused relatively lower specificity, compared to the other stages.

We classified the sleep stages using the top 10 features with high information gain for the optimal number of features in order to estimate the sleep stages. Fig. 8 illustrates the accuracy of the sleep scoring results using the top 10 significant features with high information gain. The results were similar to previous results.

The sensitivities of the sleep estimation results using the top 10 features with high information gain are shown

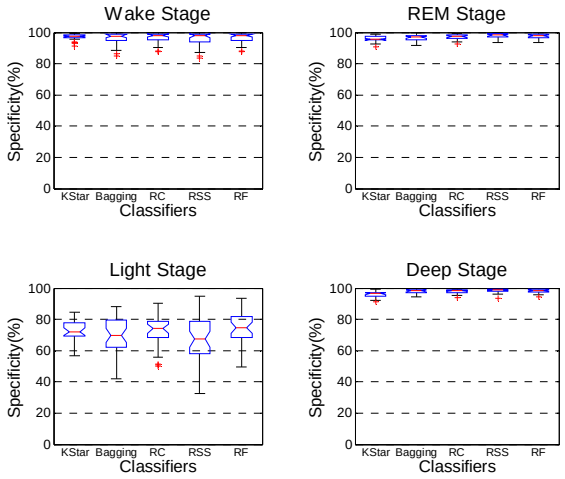

Fig. 7. Classification specificities of the four sleep stages across the 36 subjects using five classification algorithms.

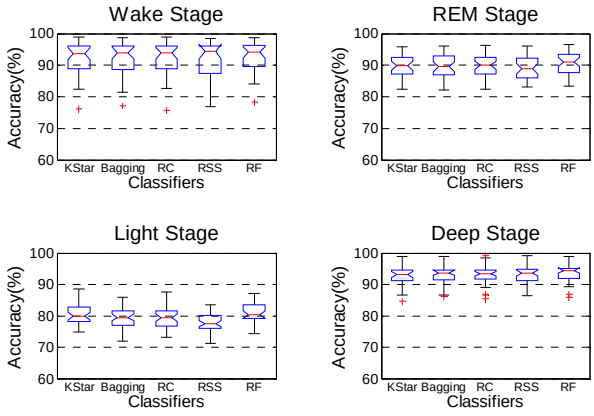

Fig. 8. The accuracies of the four stages using the 10 significant features with high information gain.

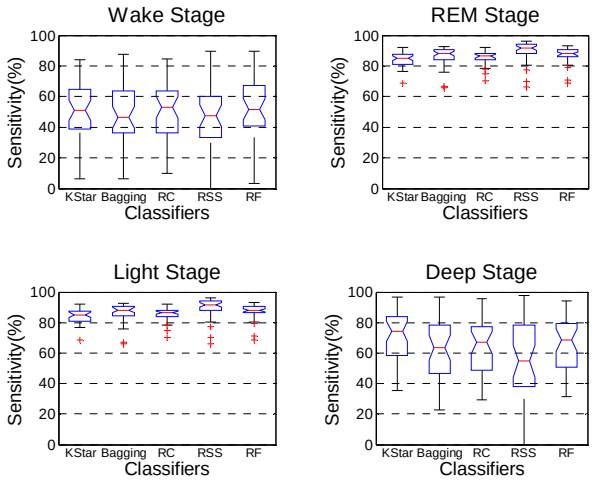

Fig. 9. The sensitivities of the four sleep stages using 10 significant features with high information gain.

in Fig. 9. The results are similar to the previous results with all 40 features.

Fig. 10 shows the specificities of the sleep estimation

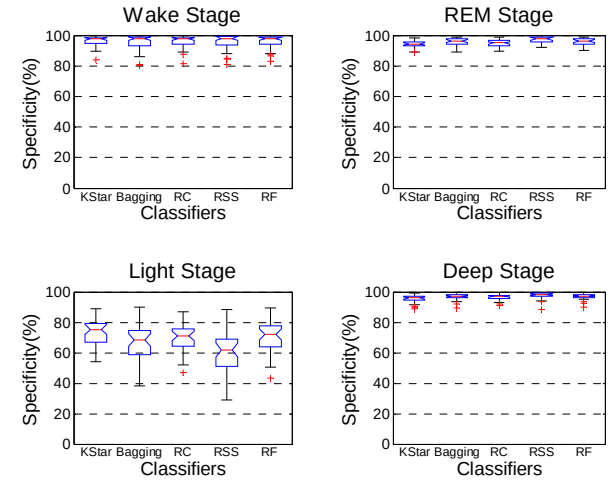

Fig. 10. The specificities of the four stages using 10 significant features with high information gain.

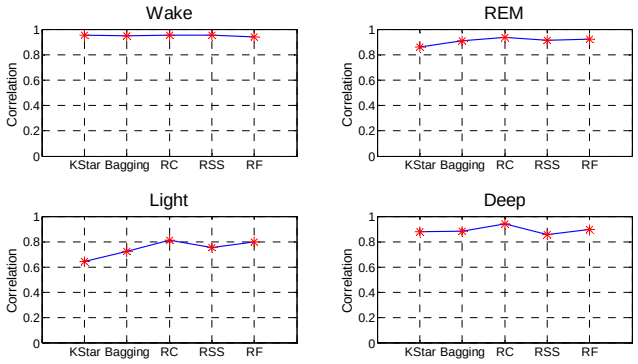

Fig. 11. The correlation between results using different feature numbers.

results using the top 10 features with high information gain. The results are similar to the previous results with all 40 features.

For rigorous validation of the results using only 10 significant features, the correlation between two previously referred to methods, using all 40 features and 10 significant features only, was calculated. Fig. 11 displays the correlations between two methods corresponding to the classifiers. Overall correlations between the two methods yielded over 0.9.

4. Conclusion

The results presented in this paper show that our proposed method is efficient at scoring sleep stages automatically. While PSG measurements include EEG, EOG, EMG, ECG, and various biomedical signals, our method measures only wrist movement using an accelerometer sensor. Furthermore, our approach could be more natural and convenient for subjects who undergo a sleep test, compared to conventional PSG recording owing to the possibility of using it in the homes of the subjects, which would not cause stress or discomfort. The information gain of each feature was computed for optimal feature selection, and 10 significant features were selected for estimation of sleep stages. The small number of features produced performance similar to the results from using all 40 features, which was validated by correlation results between the two methods.

Acknowledgement

This work was supported by the Research Program for Automatic Sleep State Detection funded by Naver Corporation and by a 2015 Institute for Information, communications Technology Promotion (IITP) grant funded by the Korea government (MISP) (No. B0184-15-1003, The Development of oneM2M Conformance Testing Tool and QoS Technology).

References

[1] Morillo, Daniel Sánchez, "An accelerometer-based device for sleep apnea screening." IEEE transactions on information technology in biomedicine, vol. 14, no. 2, pp. 491-499, 2010. Article (CrossRef Link)

[2] Karantonis, Dean M., "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring." IEEE transactions on information technology in biomedicine, vol 10, No. 1, pp. 156-167, 2006. Article (CrossRef Link)

[3] Jean-Louis, Girardin, "Sleep estimation from wrist movement quantified by different actigraphic modalities." Journal of neuroscience methods, vol 105, No. 2, pp. 185-191, 2001. Article (CrossRefLink)

[4] Witten, Ian H., and Eibe Frank. "Data Mining: Practical machine learning tools and techniques." Morgan Kaufmann 2005. Article, vol 2005. Article (CrossRef Link)

[5] Cleary, John G., Leonard E. Trigg. "K*: An instance-based learner using an entropic distance measure." Proceedings of the 12th International Conference on Machine learning, vol 5, pp. 108-114, 1995 Article (CrossRefLink)

[6] Dahiya, Shashi, S. S. Handa, and N. P. Singh. "Impact of bagging on MLP classifier for credit evaluation." Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on. IEEE, 2016, vol. 2016, pp. 3794-3800, 2016 Article (CrossRefLink)

[7] Lira, Milde MS. "Combining multiple artificial neural networks using random committee to decide upon electrical disturbance classification." 2007 International Joint Conference on Neural Networks. IEEE, 2007, vol. 2007, pp. 2863-2868, 2007. Article (CrossRef Link)

[8] Hosseini, Mohammad-Parsa, Abolfazl Hajisami, and Dario Pompili. "Real-time Epileptic Seizure Detection from EEG Signals via Random Subspace Ensemble Learning." Autonomic Computing (ICAC),

2016 IEEE International Conference on. IEEE, 2016, vol. 2016, pp. 209-218, 2016. Article (CrossRef Link)

[9] Liaw, Andy, and Matthew Wiener. "Classification and regression by randomForest." R news, vol. 2, no. 3, pp. 18-22, 2002. Article (CrossRef Link)

[10] Dai, Jianhua, and Qing Xu. "Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification." Applied Soft Computing, vol. 13, no. 1, pp. 211-221, 2013. Article (CrossRef Link)

Minsoo Yeo received his BSc in Oriental Bio-Medical Engineering from Sangji University, Gangwon-do, South Korea, in 2016. Currently, he is a graduate student in the Bio-Medical Computing Laboratory (BMCL), Department of Computer Engineering, Kwangwoon University, Seoul, South Korea. His research interests include bio-medical signal processing, machine learning, and statistical analysis.

Yong Seo Koo, MD, PhD, is a clinical assistant professor in the Department of Neurology at the Korea University College of Medicine. Dr. Koo received an MD (2005), an MSc (2008), and a PhD (2014) in Medical Science from Korea University in Seoul, South Korea. He also received a BSc (2013) in Information Statistics from Korea National Open University. He completed his education in neurology residency at Korea University Medical Center between 2006 and 2010. His research interests are in epilepsy, intraoperative neurophysiologic monitoring, and sleep disorders. He is also interested in computational analyses of electroencephalography, event-related potentials, and polysomnography.

Cheolsoo Park is an assistant professor of Computer Engineering at Kwangwoon University, Seoul, South Korea. He received a BEng in electrical engineering from Sogang University, Seoul, South Korea, and an MSc from the biomedical engineering department at Seoul National University, Seoul, South Korea. In 2012, he received his PhD in adaptive nonlinear signal processing from Imperial College London, London, UK, and worked as a postdoctoral researcher in the bioengineering department at the University of California, San Diego, USA. His research interests are mainly in the areas of machine learning and adaptive and statistical signal processing, with applications in brain computer interfaces, computational neuroscience, and wearable technology. He is a member of the IEEE.