Optimal Classifier for Detection of Obstructive Sleep Apnea Using a Heartbeat Signal

Erdenebayar Urtnasan\textsuperscript{1,2}, Jong-Uk Park\textsuperscript{2}, SooYong Lee\textsuperscript{3} and Kyoung-Joung Lee\textsuperscript{2}
\textsuperscript{1}Department of Medical Engineering, Huree University, Ulaanbaatar, Mongolia
\textsuperscript{2}Department of Biomedical Engineering and \textsuperscript{3}Liberal Education, Yonsei University, Wonju, Korea

Abstract
This study is to find the optimum classifier that can be easy and robust diagnostic method of the obstructive sleep apnea (OSA) using a heartbeat signal. The heartbeat signal was acquired from the 92 patients with OSA. The dataset consists 98,060 epochs, from them the training sets contained 68,642 epochs from the 63 OSA patients and test sets contained 29,418 epochs from the 29 OSA patients, respectively. The heartbeat signal was analyzed in the time and frequency domain and six features were extracted (normal-to-normal [NN], standard deviation of mean NN [SDNN], root mean square of successive differences [rMSSD], low-frequency [LF], high-frequency [HF], and LF/HF ratio). All extracted features were used to train the following classifiers: linear discriminant analysis (LDA), decision tree (DT), logistic regression (LR), k-nearest neighbor (KNN) and support vector machine (SVM). The top three classifiers (SVM, DT, and LDA) showed the accuracy of 93.2%, 93.2%, and 93.2% for test sets, respectively. Then, the top three classifiers could be effective on sleep studies and OSA detections.

Keywords: Obstructive sleep apnea, Machine learning, SVM, LR, DT, LDA, KNN

1. Introduction
Quality of sleep is affected by the quality of life. Obstructive sleep apnea (OSA) is a most common sleep-disordered breathing (SDB) which decrease the quality of sleep. OSA is defined as upper airway collapse at least 10 seconds during sleep. It can cause repetitive shortness of breath and sleep fragmentation, which degrade the quality of sleep and life\textsuperscript{[1]}. In addition, OSA can lead to excessive fatigue, sleepiness, and even drowsy driving, which can result in traffic accidents and other more serious tragic consequences such as heart attacks and sudden death\textsuperscript{[2,5]}.

Polysomnography (PSG) is a standard method for objectively evaluating the SDB. An objective diagnosis of the SDB can be provided based on bio-signals given by polysomnography. However, polysomnography also has several drawbacks, such as the need for expensive diagnostic equipment, attachment of multiple sensors, and manual reading by experts\textsuperscript{[6]}. Manual reading by sleep specialists, in particular, is time-consuming and labor-intensive. In addition, different results can be produced or errors can occur depending on the experience...
and subjective judgment of the specialist. A computerized SDB detection scheme can solve these problems. ECG signals are widely used for computerized OSA detection. Last decades, there were several algorithms proposed in the literature for automatic detection of OSA using a not only ECG signal but also oxygen saturation and respiratory effort signals. In those studies, various machine learning methods such as k-nearest neighbor (KNN) [7, 8], support vector machine (SVM) [9], AdaBoost [10], and neural network [11] were used for OSA detection. For instance, Xie and Minn [10] used classifier that combine AdaBoost with Decision Stump and Bagging with REPTree, and employed features extracted from ECG and saturation of peripheral oxygen (SpO2) signals. Khandoker et al. [9] used SVM with 28 features extracted from heart rate variability signal and ECG-derived respiration data to detect OSA. Varon et al. [11] also used principal components of the QRS complexes as features and classified using least-squares support vector machine. Nguyen et al. [12] performed heart rate complexity based on recurrence quantification analysis of heart rate variability to classify OSA episodes. However, since above methods used a number of features obtained from the various bio-signals, they were very complex and time consumed.

In this study, we used only few features from the one channel heartbeat signal for easy implementation. In addition, we found an optimal classifier using the feature set that we extracted for accurate classification of OSA. Only six features were extracted from the heartbeat signal and the following five methods were used as classifier to evaluate the performance: linear discriminant analysis (LDA), decision tree (DT), logistic regression (LR), KNN, and SVM.

2. Method

2.1 Study population

Ninety-two subjects (male: 73, female: 19) with OSA participated in this study. The data set was split into training and test sets for the machine learning algorithms. The training sets contained 68,642 epochs from the 63 OSA patients and test sets contained 29,418 epochs from the 29 OSA patients, respectively. The data set splits were the same for each of the methods used.

The apnea hypopnea index (AHI) was calculated by the number of occurrences of the apnea and hypopnea during total hours of sleep [13]. A subjects demographics and sleep-related variables were not significantly different between the training set and validation sets, as shown in Table 1. This study was authorized by the Institutional Review Board (No. IRB-2012-01-063) of Samsung Medical Center (SMC). All subjects provided written informed consent for participating.

### Table 1. OSA patients’ information in detail

| Measures                                | Training | Validation |
|-----------------------------------------|----------|------------|
| Gender (male/female)                    | 50/13    | 23/6       |
| Age (yr)                                | 57.76±10.59 | 57.54±11.57 |
| BMI (kg/m^2)                            | 26.37±3.16 | 25.28±2.51  |
| AHI (per hour)                          | 31.45±12.55 | 20.85±11.43 |
| Total recording time (min)              | 442.86±38.55 | 441.66±55.41 |
| Total sleep time (min)                  | 351.80±76.89 | 346.97±56.78 |
| Sleep efficiency (%)                    | 79.55±15.67 | 78.74±9.74  |

BMI: body mass index, AHI: apnea hypopnea index.

2.2 Data recording

The standard full-night PSG data were analyzed and conducted using N7000 (Embla System Inc., Denver, Co, USA) in the Sleep Center of SMC. The average recording time was 7.37 hours and recorded channels are as following: EEG, EMG, EOG, ECG, chest and abdominal volume changes, nasal-oral airflow, body position, snoring, and blood oxygen saturation. The ECG were recorded at 200 samples/sec and stored with 16-bit resolution. According to an American Academy of Sleep Medicine (AASM) guideline [13], a trained sleep specialist scored PSG data from each subject every 30 seconds (1 epoch) using RemLogic 2.0 software.

2.3 Feature extraction

All six features were extracted from a heartbeat signal for OSA diagnosis. An ECG signal was processed with a bandpass filter (0.5-32Hz) to remove the noises. The heartbeat signals were found from filtered ECG using a Pan and Tompkins algorithm [14] and manually corrected.

The heartbeat signal was analyzed in the time and frequency domain to extract features. In the time domain, the features were calculated as the mean of normal-to-normal (NN) intervals, standard deviation of mean NN (SDNN), and root mean square of successive differences of last NN (rMSSD). The welch method was applied for the frequency domain analysis in every 30 seconds and extracted features were low-frequency (LF) power range (0.04-0.15 Hz), high-frequency (HF) power range
Table 2. The real value of the heartbeat features

| Features | Normal   | OSA      | p-value |
|----------|----------|----------|---------|
| NN       | 74.2±11.2| 99.4±24.1| 0.61    |
| SDNN     | 1,113±114.2| 1,392±346.4| 0.66    |
| rMSSD    | 1,577±162.6| 1,974±491.1| 0.66    |
| LF       | 47.2±671.0| 4.7±33.9 | 0.54    |
| HF       | 101.5±161.1| 8.5±71.1 | 0.62    |
| LF/HR    | 0.8±0.7 | 0.6±0.4 | 0.57    |

(0.15-0.40 Hz), and LF/HF ratio.

All selected features are well known in sleep studies, also been used in the various studies which detects or classifies the OSA [8–11, 15, 16]. So that, they were selected and applied to five different methods of the machine learning to find the optimal one. The real value of the extracted features from a heartbeat signal are represented in Table 2.

2.4 Linear discriminant analysis

The LDA is the representative of the dimensionality reduction models that aims to summarize or describe data using less information. The LDA finds the direction vector such that projected data have the largest possible between-class separation while the within-class is kept as small as possible [15].

2.5 Logistic regression

The LR is an algorithm that constructs a separating hyperplane between two data sets, using the logistic function to express distance from the hyperplane as a probability of class membership. LR is widely used in medical applications for the ease with which is analyze the relationship between predictors, and an outcome that is dichotomous responses such as the presence or absence of an apnea event [17].

2.6 k-Nearest neighbor

The KNN method is considered as the instance-based classifier, which typically build up and instance of training data and compare new data to the instance using a similarity measure in order to find the best match and make a prediction. KNN is popular density estimation algorithm for numerical data. The density estimation uses a distance measure (usually Euclidean, or Manhattan). For a given distance measure, the only parameter of the algorithm is $k$, the number of neighbors. The parameter $k$ determines the smoothness of the density estimation: larger values consider more neighbors, and therefore smooth over local characteristics. Smaller values consider only limited neighborhoods. Generally, the choice of $k$ can only be determined empirically [18].

2.7 Decision Tree

The DT paradigm constructs classifiers by dividing the data set into smaller and more uniform groups, based on a measure of disparity or entropy. It does this by identifying a variable and a threshold in the domain of this variable that can be used to divide the data set into two groups. The best choice of variable and threshold is the one that minimizes the disparity measures in the resulting groups. The advantages of DTs over many of the other methods used here is that small DTs can be interpreted by humans as decision rules. Therefore, they offer a way to extract decision rules from a database. This makes them especially well suited for medical applications.

2.8 Support vector machine

SVMs calculates separating hyperplanes that maximize the margin between two sets of data points. By using Lagrange multipliers, the problem can be formulated in such a way that the only operations on the data points are the calculation of scalar products. While the basic training algorithm can only construct linear separators, kernel functions can be used to calculate scalar products in higher dimensional spaces. If the kernel functions are nonlinear, the separating boundary in the original space will be nonlinear. Because there are many different kernel functions, there is a wide variety of possible SVM models. In this study, we applied a single binary SVM classifier with a radial basis function employed as the kernel function. The multiplier coefficient $a$ and regularization parameter $c$ were determined empirically ($a = 0.5$, $c = 1$). All classifiers were trained and tested on the classification learner toolbox of MATLAB (Mathworks, Inc., Natick, MA, USA).

2.9 Performance analysis

To evaluate the performance of OSA detection, the accuracy ($\%$), sensitivity ($\%$), and positive predictive value ($\%$) were calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

(1)
where true positive (TP), true negative (TN), false positive (FP), and false negative (FN) refer to the number of events in which normal is classified as normal, normal as abnormal, abnormal as normal, and abnormal as abnormal, respectively. All statistical values were compared between the training and test sets.

3. Experiment Result

Each of the five methods presented above was run on each epoch of the data set of training and test set. Five methods obtained OSA accuracy of 95.2% SVM, 95.2% DT, 95.2% LDA, 94.9% LR, and 93.1% KNN for the training sets, respectively. The performance of the accuracy was shown as 93.2% SVM, 93.2% DT, 93.2% LDA, 92.6% LR, and 90.8% KNN for test sets, respectively (Figure 1).

Three of the five methods (SVM, DT, and LDA) showed the robust performance for the whole dataset. Other two methods (LR and KNN) showed slightly lower results than top three methods. As shown in Table 3, the top three methods detected OSA with the sensitivity and positive predictive value (PPV) were 95.2% and 100% in SVM, 95.2% and 100% in DT, 95.2% and 100% in LDA for the training set, respectively.

As shown in Table 4, the top three methods detected OSA with the sensitivity and PPV and were 93.2% and 100% in SVM, 93.2% and 100% in DT, 93.2% and 100% in LDA for the test set, respectively.

Finding the optimal model for a given OSA classification task depends on not only discriminant power, but also another factors such as cost of model construction and computational power of model. However, we focused on determining the classification performance along and disregarded another two points. Because we used the same dataset for all algorithms, so that the cost of the collecting data is the same for each method. In addition, the rapid advancements in hardware power and CPU speeds are not a problem in shallow learning anymore, so that computational power of the model is not an issue.

Five methods were investigated in this paper, the top three (SVM, DT, and LDA) showed robust classification results, whereas the other two (LR and KNN) drop off considerably on some of the classification task. Even the worst of the five methods achieved sensitivity and PPV values that are comparable to advanced studies [8–11]. The top three (SVM, DT, and LDA) obtained excellent results above the accuracy of 93%. The results showed relatively good performance even though we used only a heartbeat signal and simple feature sets. The top three methods showed the possibility of accurate OSA detection based on a heartbeat signal.
4. Conclusions

We investigated a simple and accurate method to detect OSA using a heartbeat signal. Only six features were extracted from a heartbeat signal and five different machine-learning methods were compared to find an optimal on the problem of OSA classification. Our results shows the top three classifiers (SVM, DT, and LDA) were with accuracy of 93.2%, 93.2%, and 93.2%, respectively. Top three SVM, DT and LDA classifiers showed strong performance, they could be effective on sleep studies and OSA detections.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgements

This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea (No. NRF-2014H1C1A1063845).

References

[1] S. P. Patil, H. Schneider, A. R. Schwartz, and P. L. Smith, “Adult obstructive sleep apnea: pathophysiology and diagnosis,” Chest, vol. 132, no. 1, pp. 325-337, 2007. https://doi.org/10.1378/chest.07-0040

[2] V. Kapur, K. P. Strohl, S. Redline, C. Iber, G. O’Connor, and J. Nieto, “Underdiagnosis of sleep apnea syndrome in U.S. communities.” Sleep Breath, vol. 6, no. 2, pp. 49-54, 2002. https://doi.org/10.1055/s-2002-32318

[3] K. Banno and M. H. Kryger, “Sleep apnea: clinical investigations in humans,” Sleepmedicine, vol. 8, no. 4, pp. 400-426, 2007. https://doi.org/10.1016/j.sleep.2007.03.003

[4] H. M. Engleman and N. J. Douglas, “Sleep-4: sleepiness, cognitive function, and quality of life in obstructive sleep apnoea/hypopnoea syndrome,” Thorax, vol. 59, no. 7, pp. 618-622, 2004. https://doi.org/10.1136/thx.2003.015867

[5] F. Lopez-Jimenez, F. H. S. Kuniyoshi, A. Gami, and V. K. Somers, “Obstructive sleep apnea: implications for cardiac and vascular disease,” Chest, vol. 133, no. 3, pp. 793-804, 2008. https://doi.org/10.1378/chest.07-0800

[6] N. A. Collop, “Scoring variability between polysomnography technologists in different sleep laboratories,” Sleepmedicine, vol. 3, no. 1, pp. 43-47, 2002. https://doi.org/10.1016/S1389-9457(01)00115-0

[7] C. Maier and H. Dickhaus, “Recurrence analysis of nocturnal heart rate in sleep apnea patients,” Biomedizinische Technik, vol. 51, no. 4, pp. 224-228, 2006. https://doi.org/10.1515/BMT.2006.042

[8] M. O. Mendez, A. M. Bianchi, M. Matteucci, S. Cerutti, and T. Penzel, “Sleep apnea screening by autoregressive models from a single ECG lead,” IEEE Transactions on Biomedical Engineering, vol. 56, no. 12, pp. 2838-2850, 2009. https://doi.org/10.1109/TBME.2009.2029563

[9] A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, “Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings,” IEEE Transactions on Information Technology in Biomedicine, vol. 13, no. 1, pp. 37-48, 2009. https://doi.org/10.1109/TITB.2008.2004495

[10] B. Xie and H. Minn, “Real-time sleep apnea detection by classifier combination,” IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 3, pp. 469-477, 2012. https://doi.org/10.1109/TITB.2012.2188299

[11] C. Varon, A. Caicedo, D. Testelmans, B. Buyse, and S. Van Huffel, “A novel algorithm for the automatic detection of sleep apnea from single-lead ECG,” IEEE Transactions on Biomedical Engineering, vol. 62, no. 9, pp. 2269-2278, 2015. https://doi.org/10.1109/TBME.2015.2422378

[12] L. Chen, X. Zhang, and C. Song, “An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram,” IEEE Transactions on Automation Science and Engineering, vol. 12, no. 1, pp. 106-115, 2015. https://doi.org/10.1109/TASE.2014.2345667

[13] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, R. M. Lloyd, C. L. Marcus, and B. V. Vaughn, “The AASM manual for the scoring of sleep and associated events,” Rules, Terminology and Technical Specifications, Version 2.2, American Academy of Sleep Medicine, Ed. Darien, IL: American Academy of Sleep Medicine, 2015.
[14] J. Pan and W. J. Tompkins, “A real-time QRS detection algorithm,” *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230-236, 1985. https://doi.org/10.1109/TBME.1985.325532

[15] U. Erdenebayar, H. K. Lee, H. Kim, and K. J. Lee, “Sleep apnea detection using a piezo snoring sensor: a pilot study,” *Journal of Biomedical Engineering Research*, vol. 35, no. 4, pp. 75-80, 2014. https://doi.org/10.9718/JBER.2014.35.4.75

[16] U. Erdenebayar, H. K. Lee, H. Kim, and K. J. Lee, “Sleep apnea detection using a piezo snoring sensor: a pilot study,” *Journal of Biomedical Engineering Research*, vol. 35, no. 4, pp. 75-80, 2014. https://doi.org/10.9718/JBER.2014.35.4.75

[17] H. D. Nguyen, B. Wilkins, Q. Cheng, and B. A. Benjamin, “An online sleep apnea detection method based on recurrence quantification analysis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1285-1293, 2014. https://doi.org/10.1109/JBMI.2013.2292928

[18] H. D. Nguyen, B. Wilkins, Q. Cheng, and B. A. Benjamin, “An online sleep apnea detection method based on recurrence quantification analysis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1285-1293, 2014. https://doi.org/10.1109/JBMI.2013.2292928