Detecting Slow Wave Sleep via One or Two Channels of EEG/EOG Signals

Liang-Wen Hang
Sleep Medicine Center, Department of Internal Medicine, China Medical University Hospital, Taichung, Taiwan
Department of Healthcare Administration, Asia University, Taichung, Taiwan
Email: lungwen.hang@gmail.com

Bo-Lin Su and Chen-Wen Yen
Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-Sen University, Kaohsiung, Taiwan
Email: {bolin.su.2013, cmurobot}@gmail.com

***Abstract*—This work develops a number of automatic slow wave sleep (SWS) detection methods that employ only one or two channels of EOG/EEG signals. In addition to the reduction of signal channels, a distinct feature of the proposed approach is the introduction of a new feature set that can make the proposed approach insensitive to interpersonal differences of the physiological signals. The tested subjects include 265 and 947 persons underwent full overnight polysomnography from two different sleep centers. With 265 subjects from one center as the training set and 147 subjects from the other center as the validation set, the first part of our experiments yields SWS detection results of Kappa coefficients 0.72-0.78, sensitivity 0.77-0.90 and positive predictive value 0.73-0.82. Using the 947 subject dataset, the second part of the experiments compares the relative merits of the tested methods and investigates the impacts of SWS ratio and severity of sleep apnea on the performances of the proposed methods. Finally, our results suggest that the quality of the training set is of great importance for the development of accurate SWS detection methods.**

***Index Terms*—slow wave sleep, EEG, EOG, automatic sleep staging**

I. Introduction

Slow wave sleep (SWS) is a very important part of human sleep and has been found to be associated to a variety of clinical problems including diabetes risk [1], memory consolidation [2], psychiatric disorders [3] and hypertension [4].Another factor that distinguishes the significance of SWS is the availability of the SWS-enhancing drugs. By increasing the time spent in SWS, these drugs open a new direction for remedying problems caused by sleep loss [5].

A possible strategy to comprehensively study the interaction between SWS and the SWS-related problems is to perform large human population studies for SWS. However, this requires a very intensive amount of work for signal acquisition and sleep staging. Typically, after an overnight polysomnography (PSG), sleep stages are scored based on electroencephalogram (EEG), electromyogram (EMG) and electrooculogram (EOG) signals. Measuring and recording these multiple channels of physiological signals are costly and complicated. In addition, manual scoring of the sleep stages is tedious and time consuming.

These difficulties have been tackled from two directions. First, in responding to the inefficiency of manual scoring, many attempts have been made to automate the sleep staging process by using artificial intelligence or pattern recognition techniques [6]-[11]. Second, the complexity of the signal acquisition and processing tasks can be simplified by using fewer channels of signals. Such simplifications typically use one or two channels of EOG or EEG signals [12]-[17].

It should be noted that both EOG and EEG activities are propagated to the head surface. As a result, the signals picked up by the EEG or EOG electrodes are actually the combinations of both signals. Separating the EOG and EEG signals is often a challenging task [18]. Another problem of EEG/EOG signal processing is the interpersonal differences. Typically, human physiological signals are qualitatively similar but not quantitatively identical. For instance, one of unique properties of SWS is its large-amplitude low-frequency EEG waveform pattern. However, the actual amplitude and frequency band of such a slow wave activity vary from person to person and are difficult to model.

The goal of this study is to introduce automatic SWS detection methods by using one to two channels of EEG/EOG signals. Compared to previous work, the proposed methods have the several distinct features. First, without the need of EEG-EOG separation, this work introduces a number of new signal features that make the proposed approaches relatively insensitive to interpersonal differences. Second, this study systematically compares the effectiveness of different combinations of EEG and EOG signals for detecting SWS. Third, this study investigates the influences of SWS ratio and AHI value on the accuracy of sleep staging since such influences have rarely been investigated.

II. Materials and Methods

Manuscript received April 25, 2013; revised July 10, 2013

doi: 10.12720/ijsps.1.1.84-88

A. Study Subjects and Polysomnography

Patients were recruited from China Medical University Hospital and Sheng Mei Hospital Sleep Centers. These two sleep centers will be referred to as center C and center S thereafter. Clinical data were collected retrospectively. The 265 subjects from center C were enrolled from 2005 to 2008. The 947 subjects from center S were enrolled from 2006 to 2008.PSG data were recorded using computerized polysomnographic systems. The stored data were digitized for computer analysis by data analysis software (Matlab; MathWorks Inc; Natick, MA). The raw data were reviewed by an experienced doctor and scored by a sleep technician certified by the Taiwan Sleep Medicine Society. Every 30 seconds epoch was scored according to the criteria of Rechtschaffen and Kales [19].

In the first part of the experiments, the 265 subjects from center C were used as the training set and a 147 subject subset from center S was chosen as the validation set. Table I shows the clinical features and sleep stage composition of these subjects. Studied in the second part of the experiments, the clinical features and sleep stage composition of the 947 subjects from center S are given in Table II. Note that one of the critical differences between these two tables is the percentages of SWS.

TABLE I. CLINICAL CHARACTERISTICS OF THE TESTED SUBJECTS FOR THE FIRST PART OF THE EXPERIMENTS

Clinical Characteristics	Learning Set	Validation Set
Number of epochs	207937	122455
Age, yr	34.3 ±15.5	27.9 ±12.5
BMI, kg/m2	25.5 ±5.2	24.5 ±5.2
AHI event/hour	17.6 ±18.6	11.6 ±13.1
Male/Female, gender	165/100	95/52
Sleep efficiency %	84.71 ±11.2	89.77 ±8.7
Wake %	14.92 ±11.2	11.69 ±8.5
REM %	17.05 ±7.1	17.85 ±6.6
S1%	10.5 ±6.3	7.76 ±4.5
S2%	42.38 ±10.3	47.74 ±9.0
SWS%	15.13 ±7.73	14.96 ±6.5

TABLE II. CLINICAL CHARACTERISTICS OF THE TESTED SUBJECTS FOR THE SECOND PART OF THE EXPERIMENTS

Clinical Characteristics	
Number of epochs	779661
Age, yr	41.7 ±13.3
BMI, kg/m2	26.3 ±4.6
AHI event/hour	25.3 ±24.8
Male/Female, gender	728/225
Sleep efficiency %	86.85 ±10.0
Wake %	14.52 ±9.8
REM %	16.63 ±6.8
S1%	12.65 ±9.9
S2%	52.89 ±11.8
SWS%	3.3 ±5.9

B. Features for Sleep Staging

This study formulates the SWS detection as a binary classification problem whose goal is to distinguish SWS and NSWS (non-SWS) on an epoch-by-epoch basis by using information provided by a number of signal features. In this work, the feature set is divided into three groups each of which consists of three features. The first group is introduced to characterize the waveform of the signals. Derived from the first feature group, the second and third feature groups are designed to deal with the interpersonal difference problem of the signals. For the sake of convenience, we will use symbol x, y and z to represent the features of the first, second and third feature groups, respectively.

The first feature group is originated from the zero-crossing points. To generate these features, an epoch is first divided into thirty one-second intervals. For each of such intervals, after subtracting the interval mean from the signal, the zero-crossing points (ZCPs) can be located by finding the points where the signal changes sign. With these ZCPs, the time history of a signal can be divided into a number of segments which will be referred as ZCP segments hereafter. Apparently, the speed of the signal can be characterized by the lengths of the ZCP segments. Hence, the first two features employed in this work are the sample mean and standard deviation of the ZCP segment lengths.

The third feature can be written as:

$$x_3 = \sum_{i=1}^{I} (z_{i+1} - z_i) \int_{z_i}^{z_{i+1}} | s(t) | dt \qquad (1)$$

where I is the number of ZCP segments for the epoch under consideration, z_i denotes the time associated with the ith ZCP, and the $s(t)$ denotes the signal being processed. Note that the integral appears in (1) is essentially the absolute value of the area under the response curve of signal $s(t)$. Apparently, increasing the amplitude of $s(t)$ makes this area larger and reducing the speed of $s(t)$ leads to larger z_{i+1}-$z_i$difference. Together, these two properties can characterize the high-amplitude low-frequency EEG signal pattern frequently appear in SWS.

In developing features to overcome the influences of interpersonal differences of the physiological signals, we first assume that the probability density function (PDF) of the feature variable x can be written as $p(x, \mu)$ where μ is the mean of x. By assuming C_1 and $C_2$to be the ratios of SWS and NSWS for a particular night of PSG recording, the PDF of the feature for this PSG recording can be written as:

$$p(x, \mu) = C_1 p_S(x, \mu_S) + C_2 p_N(x, \mu_N) \qquad (2)$$

where p_S and p_N are the PDFs of x associated with the SWS and NSWS, respectively. We further assume that the functional forms of p_S and p_N to be fixed and their only unknown parameters are their means μ_S and μ_N. Clearly, μ_S and μ_N as well as the weighting coefficients C_1 and C_2 are individually dependent.

To account for the uncertainty associated with μ_S and μ_N, for every first group feature x, the corresponding second group feature y is chosen as the percentile rank of x within the PSG recording. The rationale behind this design is based on the fact that the mean of a random variable can be determined by a sample whose value and

percentile rank are known as long as the PDF of this random variable is completely specified except for the value of its mean.

As shown in (2), another possible source of interpersonal differences come from the weighting coefficients C_1 and C_2 which represent the relative amount time spent in SWS and NSWS, respectively. According to the Bayes theorem, the optimal decision rule for a classification problem depends on the posterior probabilities as well as the prior probabilities of the patterns to be classified [20].For our SWS detection problem, the prior probabilities are simply C_1 and C_2 which typically vary from person to person and from night to night.

This study develops the proposed SWS detection methods by learning from a training set. Since the staging results of the training set are completely known, we can readily derive p_S and p_N for this training set and then determine the corresponding $p(x)$ via (2). Similar to (2), the PDF $q(x)$ of an arbitrary PSG recording that awaits sleep staging can be expressed as:

$$q(x) = D_1 p_S(x) + D_2 p_N(x) \quad (3)$$

To develop features that are adaptable to this new PSG recording, we first acknowledge the fact that if the value of the feature variable x is known to be a, then we can always determine an Δx to satisfy

$$\int_{a-\Delta x}^{a+\Delta x} p(x)dx = \varepsilon \quad (4)$$

provided that $0 \leqq \varepsilon \leqq 1$. Note that the upper (lower) limit of this definite integral should be replaced by the maximum (minimum) value of x if $a + \Delta x (a - \Delta x)$ is larger (smaller) than maximum (minimum) value of x. In this study, the value of ε is chosen as 0.05.

Once Δx has been determined, the values of the three definite integrals appear in the following equality can all be determined.

$$\int_{a-\Delta x}^{a+\Delta x} q(x)dx = D_1 \int_{a-\Delta x}^{a+\Delta x} p_S(x)dx + D_2 \int_{a-\Delta x}^{a+\Delta x} p_N(x)dx \quad (5)$$

With this equality and the fact that the sum of D_1 and $D_2$is one, we have enough information to determine D_1 and D_2.Therefore,for every first subset feature x, this study generates $\int_{a-\Delta x}^{a+\Delta x} q(x)dx$ as a third subset feature z.

III. Results

A. A Comparative Study for the Proposed Method

The goal of the first part of the experiments is to test the proposed approach by using different combinations of EEG/EOG signals. In specific, for the one-channel configuration, C3M2, C4M1, REOG and LEOG had been independently tested for SWS detection. For the two-channel combinations, the tested methods included C3M2-C4M1, REOG-LEOG and C3M2-REOG. In performing these tests, the one-hidden-layer multilayered perceptron (MLP) was chosen as the classifier. The connection weights of the MLP were determined by the conventional backpropagation algorithm.

Based on a large dataset (265 PSGs, 136150 epochs) and two channels of EOG signals, the method proposed by Virkkala *et al.* [15] was also designed for SWS detection. The parameters of their SWS detection rule were determined by using a training set of 133 subjects. Applying this SWS detection rule to the other 132 subject validation set yielded the following SWS detection results: Kappa coefficient 0.7, accuracy (epoch-by-epoch agreement) 0.93, sensitivity 0.75, PPV (positive predictive value) 0.71. The ratios of SWS in their training and validation sets are both about 15%.

In order to compare our methods to that of [15], the datasets used in the first part of the experiments were purposely arranged such that their SWS ratios to be approximately 15%. With three hidden neurons, the MLP was trained by the 265 subject training set and then tested on the 147 subject validation set. As demonstrated by the validation results summarized in Table III, all of the tested methods achieve higher Kappa value, sensitivity and PPV. With the exception of the method that uses only LEOG, the accuracy of all tested methods are also higher.

TABLE III. Results of the First Part of the Experiments

Channels	Kappa	Accuracy	Sensitivity	PPV
LEOG	0.72	0.927	0.77	0.75
REOG	0.73	0.932	0.78	0.76
C4M1	0.76	0.935	0.89	0.73
C3M2	0.77	0.938	0.87	0.76
REOG-LEOG	0.76	0.936	0.83	0.76
C3M2- C4M1	0.77	0.936	0.90	0.74
C3M2- REOG	0.78	0.943	0.80	0.82

B. The Influences of the SWS Ratio

The goal here is to study two other factors that can influence the performances of the proposed approaches and possibly many, if not all, other automatic sleep staging methods. The first actor is the SWS ratio. The second factor is AHI (apnea-hypopnea index) which is often used to characterize the severity of sleep apnea. For this part of the experiments, we used the dataset of Table II to develop the proposed methods. By using an eight hidden-neuron MLP to perform classification, the training results are summarized in Table IV. Note that since the number of training examples is significantly larger than the number of MLP parameters (779661 epochs versus 90 MLP connection weights), the overfitting problem is unlikely to occur [21]. Therefore, there is little need to set up a validation set to test the generalization capability of the MLP.

The results of Table IV are apparently inferior to those of Table III. This discrepancy can be explained by the variation of the SWS ratio. Compared to the 15% SWS in the first part of the experiments, the SWS ratio of the second study is only 3.3%. To investigate the difficulty caused by small SWS ratio, we selected six different SWS% subsets from the sleep center S dataset. With a 5% SWS ratio interval, these subgroups were arranged such

that the differences of their AHI means are statistically insignificant at 5% level.

TABLE IV. RESULTS FOR THE SECOND PART OF THE EXPERIMENTS

Channels	Kappa	Accuracy	Sensitivity	PPV
LEOG	0.54	0.973	0.49	0.63
REOG	0.54	0.975	0.47	0.67
C4M1	0.62	0.977	0.61	0.66
C3M2	0.65	0.978	0.63	0.68
REOG-LEOG	0.61	0.976	0.60	0.64
C3M2- C4M1	0.67	0.979	0.66	0.69
C3M2- REOG	0.66	0.979	0.66	0.69

The results of Table V demonstrate that the SWS detection performances deteriorate with the decline of SWS ratio. Note that the results of Table V(as well as Table VII which will be presented later) were obtained by using C3M2 signal to perform SWS detection. However, other SWS detection methods proposed in this work yielded similar results.

TABLE V. THE INFLUENCES OF SWS RATIO

SWS ratio%	0	0-5	5-10	10-15	15-20	>20
Subjects	8	33	26	28	26	9
AHI	5.04	5.25	5.17	5.22	5.26	5.23
Kappa	0.00	0.33	0.61	0.70	0.75	0.78
Sensitivity	0.00	0.50	0.57	0.74	0.72	0.76
PPV	0.00	0.27	0.72	0.73	0.87	0.90

C. The Falsely Detected SWS

Since the false negative error (the error that incorrectly classify SWS as NSWS) can be directly derived from sensitivity, to gain more insight into the test results, Table VI summarizes the false positive errors of the tested methods.

Table VI shows the percentages of the misclassified epochs for every NSWS stage including the rapid-eye movement (REM), stages 1 (S1), stage 2 (S2) and wake period. The results of Table VI clearly show that, among all NSWS stages, S2 was mostly likely to be falsely classified as SWS. In contrast, relatively few wake and REM sleep epochs were incorrectly classified as SWS.

TABLE VI. SUMMARY OF THE FALSE POSITIVE ERROR PERCENTAGES IN DETECTING SWS

Channels	Wake	REM	S1	S2
LEOG	1.33	1.84	0.59	6.87
REOG	0.95	1.11	0.47	6.85
C4M1	1.06	0.11	1.00	9.60
C3M2	0.74	0.15	0.68	8.45
REOG-LEOG	0.22	0.03	0.67	7.93
C3M2- C4M1	0.64	0.06	0.75	9.80
C3M2- REOG	0.33	0.10	0.39	5.34

D. The Influences of AHI Value

To demonstrate the influences of the severity of sleep apnea, Table VII presents training results of four subsets of patients (healthy AHI < 5, mild AHI, 5<AHI< 15, moderate 15<AHI< 30, severe AHI>30) from the 947 subject dataset. Note that the mean differences of the SWS ratios of these four subsets are statistically insignificant at the level of 0.05. As shown in Table VII, the tested method performs better for the healthy persons than for the apnea patients.

TABLE VII. THE INFLUENCES OF AHI VALUE

AHI	<5	5-15	15-30	>30
Subjects	107	101	73	64
SWS%	5.11%	5.12%	5.13	5.12%
Kappa	0.68	0.57	0.61	0.59
Sensitivity	0.72	0.55	0.59	0.55
PPV	0.67	0.64	0.66	0.69

IV. DISCUSSIONS

The results of the first part of our experiments show that the performances of the proposed methods compare favorably with those of the previous approaches. In addition, as reported in Virkkala *et al.*'s study [15], the human inter-rater epoch-by-epoch agreement for one of their datasets is approximately 94% which is very close to the classification accuracy obtained by the proposed approach in the first part of the experiments.

Based on single channel results of Table II, it is found that EEG (Kappa coefficient 0.76-0.77) outperforms EOG (Kappa coefficient 0.72-0.73) in detecting SWS. In comparison, the performances of the four tested two channel methods are relatively close. As a consequence, the REOG-LEOG combination seems to be method of choice since it avoids the difficulty associated with EEG signal measurement.

The results of Table V show that lower SWS ratio leads to worse SWS detection results. This phenomenon can be explained by the training set imbalance problem [22]-[24]. For a binary classification problem, training set imbalance occurs when the number of examples of one class (majority class) is significantly larger than the number of examples of the other class (minority class). As the ratio of the numbers of examples of these two classes becomes more extreme, the classifier places more emphasis on the majority class in order to minimize the overall error. Unfortunately, this is often accomplished at the expenses of scarifying the classification accuracy of the minority class. As a result, the classifier can easily become one-sidedness and thus be rather ineffective in detecting minority class examples. In the present study, SWS is apparently the minority class. Consequently, as the SWS ratio becomes smaller, accurate SWS detection becomes more difficult to achieve.

Based on the PSG recordings of three healthy subjects and three OSA patients, it was shown that automatic sleep staging methods can be less accurate for the sleep apnea patients than for healthy subjects [6]. However, to the best our knowledge, this potential problem has never been more extensively studied. Therefore, the last part of this study examines the influences of AHI. As demonstrated by Table VII, the proposed approach is indeed less effective in dealing with sleep apnea patients.

This trend of performance degradation also appears in human scoring. In specific, several inter-scorer reliability studies have found that human sleep staging results of OSA (obstructive sleep apnea) patients have a comparatively lower degree of agreement than those of the healthy subjects [25]-[28].

For the apnea patients, the signal infidelity caused by the increased movement artifacts, arousals and other apnea symptoms may lead to performance degradation of the proposed methods. In addition, imperfect training set that contains incorrect human staging decisions may also result in performance decline.

With an imperfect training set, building a perfect automatic sleep staging method does not seem to be a realistic expectation. Apparently, for the successful development of automatic sleep staging methods, we need large and accurate PSG data sets. A possible approach for developing such datasets is via an iterative process of human-computer interactions. Initially, the automatic staging method is first designed by using a standard dataset. Next, epochs that are misclassified by the automatic sleep staging method should be reexamined by human experts to make sure that the human sleep stage scoring for each of such epochs is indeed correct. Once such an inspection process has been completed, the classifier should be redesigned by using the updated dataset. This iterative process should be repeated until no manual scoring result need to be altered.

In conclusion, based on newly developed signal features, this work proposes and tests a number of one or two channel EEG/EOG based SWS detection methods. In developing the proposed methods, special attention has been given to resolve the difficulties associated with the interpersonal differences of the physiological signals. In addition, in examining the effectiveness of the proposed methods, this study also studies the influences of SWS ratio and AHI values. As demonstrated by the experimental results, despite being overlooked by many previous studies, these two factors can degrade the performances of the automatic sleep staging methods. The results of this work also show that large and accurate PSG datasets are of great importance for the development of high performance automatic sleep staging methods.

REFERENCES

[1] E. Tasali, R. Leproult, D. A. Ehrmann, and E. Van Cauter, "Slow-wave sleep and the risk of type 2 diabetes in humans," in *Proc. Natl Acad Sci USA*, vol. 105, 2008, pp. 1044-1049.

[2] S. Diekelmann and J. Born, "The memory function of sleep," *Nat Rev Neurosci*, vol. 11, pp. 114-126, 2010.

[3] E. K. Lee and A. B. Douglass, "Sleep in psychiatric disorders: Where are we now?" *Can J Psychiatry*, vol. 55, pp. 403-412, 2010.

[4] M. M. Fung, *et al.* "Decreased slow wave sleep increases risk of developing hypertension in elderly men," *Hypertension*, vol. 58, pp. 596-603, 2010.

[5] J. K. Walsh, "Enhancement of slow wave sleep: Implications for insomnia," *J Clin Sleep Med*, vol. 15, S27-32, 2009.

[6] H. J. Park, J. S. Oh, D. U. Jeong, and K. S. Park, "Automated sleep stage scoring using hybrid rule- and case-based reasoning," *Comput Biomed Res*, vol. 33, pp. 330-349, 2000.

[7] R. Agarwal and J. Gotman, "Computer-assisted sleep staging," *IEEE Trans Biomed Eng*, vol. 48, pp. 1412-1423, 2001.

[8] P. Anderer, G. Gruber, *et al*. "An E-health solution for automatic sleep classification according to Rechtschaffen and Kales: validation study of the Somnolyzer 24 x 7 utilizing the Siesta database," *Neuropsychobiology*, vol. 51, pp. 115-133, 2005.

[9] J. Caffarel, G. J. Gibson, J. P. Harrison, C. J. Griffiths, and M. J. Drinnan, "Comparison of manual sleep staging with automated neural network-based analysis in clinical practice," *Med Biol Eng Comput*, vol. 44, pp. 105-110, 2006.

[10] F. Poré, A. Kachenoura, H. Gauvrit, C. Morvan, G. Carrault, and L. Senhadji, "Blind source separation for ambulatory sleep recording," *IEEE Trans Inf Technol Biomed*, vol. 10, pp. 293-301, 2006.

[11] M. E. Tagluk, N. Sezgin, and M. Akin, "Estimation of sleep stages by an artificial neural network employing EEG," *EMG and EOG. J Med Syst*, vol. 34, pp. 717-725, 2010.

[12] R. Agarwal, T. Takeuchi, S. Laroche, and J. Gotman, "Detection of rapid-eye movements in sleep studies," *IEEE Trans Biomed Eng*, vol. 52, pp. 1390-1396, 2005.

[13] C. Berthomier, X. Drouot, M. Herman-Sto ca, P. Berthomier, J. Prado, D. Bokar-Thire, O. Benoit, J. Mattout, and M. P. d'Ortho, "Automatic analysis of single-channel sleep EEG: Validation in healthy individuals," *Sleep*, vol. 30, pp. 1587-1595, 2007.

[14] J. Virkkala, J. Hasan, A. Värri, S. L. Himanen, and K. Müller, "Automatic sleep stage classification using two-channel electro-oculography," *J Neurosci Methods*, vol. 15, pp. 109-115, 2007.

[15] J. Virkkala, J. Hasan, A. Värri, S. L. Himanen, and K. Müller, "Automatic detection of slow wave sleep using two channel electro-oculography," *J Neurosci Methods*, vol. 15, pp. 171-177, 2007.

[16] U. Malinowska, H. Klekowicz, A. Wakarow, S. Niemcewicz, and P. J. Durka, "Fully parametric sleep staging compatible with the classical criteria," *Neuroinformatics*, vol. 7, pp. 245-253, 2009.

[17] S. Güneşa, K. Polata, and S. Yosunkayab, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting," *Expert Syst App*, vol. 37, pp. 7922-7928, 2010.

[18] R. J. Croft and R. J. Barry, "Removal of ocular artifact from the EEG: A review," *Neurophysiol Clin*, vol. 30, pp. 5-19, 2000.

[19] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring Systems for Sleep Stages of Human Subject*, Washington DC: National Institutes of Health 1968, Pub. No. 204.

[20] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, San Diego, Academic Press, 2008.

[21] S. Amari, N. Murata, and K. R. Müller, "Statistical theory of overtraining-Is cross-validation asymptotically effective?" in *Advances in Neural Information Processing Systems 8*, MIT Press, 1995, pp. 176-82.

[22] N. V. Chawla, N. Japkowicz, and A. Kolcz, "Editorial: Special issue on learning from imbalanced data sets. SIGKDD," *Explor*, vol. 6, pp. 1–6, 2004.

[23] H. He and E. A. Garcia, "Learning from imbalance data," *IEEE Trans Knowledge and Data Eng*, vol. 21, pp. 1263-1284. 2009.

[24] T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Supervised neural network modeling: An empirical investigation into learning from imbalanced data with labeling errors," *IEEE Trans Neural Netw*, vol. 21, pp. 813-30, 2010.

[25] R. G. Norman, I. Pal, C. Stewart, J. A. Walsleben, and D. M. Rapoport, "Interobserver agreement among sleep scorers from different centers in a large dataset," *Sleep*, vol. 23, pp. 901-908, 2000.

[26] N. A. Collop, "Scoring variability between polysomnography technologists in different sleep laboratories," *Sleep Med*, vol. 3, pp. 43-47, 2002.

[27] H. Danker-Hopfe, D. Kunz, *et al*. "Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders," *J Sleep Res*, vol. 13, pp. 63-69, 2004.

[28] M. Basner, B. Griefhahn, and T. Penzel, "Inter-rater agreement in sleep stage classification between centers with different background," *Somnologie*, vol. 12, pp. 75–84, 2008.

