Sleep/wake state prediction and sleep parameter estimation using unsupervised classification via clustering

Yasser EL-Manzalawy
Information Sciences and Technology
Pennsylvania State University
University Park, USA
yme2@psu.edu

Orfeu Buxton
Biobehavioral Health
Pennsylvania State University
University Park, USA
orfeu@psu.edu

Vasant Honavar
Information Sciences and Technology
Pennsylvania State University
University Park, USA
vhonavar@ist.psu.edu

Abstract— **Sleep quality impacts virtually all aspects of life, including health, mood, emotions, cognition, memory, behavior, and performance. Actigraphy offers a lower-cost alternative to conventional polysomnography (PSG), the gold standard for measuring sleep quality. Effective use of actigraphy for assessing sleep quality requires reliable methods for detecting sleep/wake states from actigraphy measurements. Machine learning offers a promising approach to building sleep/wake state detectors from actigraphy data. However, current machine learning approaches rely on expert labeled training data that can be expensive and laborious to acquire. In this work, we introduce a novel approach for integrating unsupervised learning algorithms and domain knowledge heuristics, based on statistical properties of clustered sleep and wake epochs, to develop reliable sleep/wake state prediction models using unlabeled wrist actigraphy data. Experimental results using a dataset of 37 participants and covering 282 sleeping periods demonstrate the viability of the proposed approach on developing sleep/wake state detection models from unlabeled actigraphy data with a predictive performance that is comparable with the performance of models developed using some state-of-the-art supervised learning algorithms applied to labeled actigraphy data. Our results lay the groundwork for developing fully automated machine learning models for sleep/wake state prediction and sleep parameters estimations by eliminating the need for costly and labor-intensive expert annotations of PSG recordings for labeling actigraphy data.**

Keywords- polysomnography, actigraphy, unsupervised learning, classification via clustering, sleep/wake state detection

I. INTRODUCTION

Polysomnography (PSG), the gold standard for measuring sleep [1], uses multiple sensing devices to measure physiologic parameters of sleep including brain dynamics of electroencephalography (EEG), eye movements (EOG), muscle activity (EMG), heart physiology (ECG), and respiratory function [1, 2]. Many sleep specialists use PSG in conjunction with clinical assessments as a sleep diagnostic tool. This technique suffers from several drawbacks [3]: i) Because of the limited portability of the PSG device, the participant needs to sleep in a laboratory for the recording to be done or ambulatory equipment can be used for unattended recording with the participant at home; ii) PSG recordings are expensive and rarely obtained for more than a night or two participant. iii) PSG recordings are typically manually examined and scored on an epoch-per-epoch basis, often by registered polysomnographic technologists. These restrictions can limit the applicability of this technique to large-scale, population-level sleep research studies for more than a night or two [2].

Portable techniques for measuring sleep parameters (e.g., wearable accelerometry devices, perhaps with additional sensors [4]) would allow for recording to be done in the home, decrease the cost and inconvenience, improve the evaluation by observing subjects in their natural settings [5], and facilitate large-scale research studies conducted over large number of participants and for extended time periods [2]. Unfortunately, algorithms for detecting sleep/wake state detection from actigraphy data suffer from poor sensitivity (i.e., wake state detection rates) [6, 7]. To address this limitation, a number of machine learning based models have been proposed for automatic detection of sleep/wake state (e.g., [8, 9]). However, these models are often developed using supervised machine learning algorithms, which require labeled training data (i.e., actigraphy data in which each epoch is assigned a sleep/wake label using PSG gold standard). Obtaining such labeled data is cumbersome, time-consuming, and expensive because it requires participants to have simultaneous PSG and actigraphy recordings. Moreover, the learned models are more likely to be dataset specific and might not generalize well on test data from other studies [7].

In this study, we propose a novel *unsupervised* machine learning approach for constructing sleep/wake state detection models from *unlabeled* actigraphy data. The approach uses unsupervised learning algorithms for clustering training data into two groups and then leverages some statistical properties of clustered sleep and wake epochs in order to determine the appropriate label of each cluster. This research enables the development of reliable sleep/wake state predictors using *unlabeled* actigraphy data only and advances the process of generating fully automated models by eliminating the need for PSG data in the model training phase. This approach could be adapted for physical activity recognition tasks from accelerometer data [10] and generally opens up the possibility for further developing individualized models from unlabeled data.

II. CLASSIFICATION VIA CLUSTERING

Given a training dataset consisting of n labeled samples, $\{< x_i, f(x_i) >\}_{i=1}^{n}$ where $x_i \in R^d$ is a feature vector in the d-dimensional space and its class label, $f(x_i)$, is assigned

using an unknown function $f: R^d \rightarrow C$, where $C = \{C_1, \ldots, C_m\}$ is the set of class labels. Classification via clustering (CVC) [11, 12], as the name suggests, offers an approach to building a classifier using clustering. CVC first clusters a set of instances $U = \{x_i\}_{i=1}^n$, into k clusters where k$\geq m$. Then the resulting set of clusters is mapped to a set of class labels. When the class labels for the training data are available, a simple way to associate clusters with class labels is to assign to each cluster, the label that represents the majority of the instances assigned to that cluster. An unlabeled sample to be classified is first assigned to one of the clusters, and the class label associated with the cluster is returned as the predicted label for the sample. Algorithm 1 describes a meta classifier for CVC available in the WEKA machine learning workbench [13]. Note that although the clustering is carried out in unsupervised fashion on the unlabeled training data, CVC relies on the labels of the training data in associating the resulting clusters with classes. In the next section, we propose a novel unsupervised CVC algorithm (UCVC) that eliminates the need for labeled training data.

Algorithm 1: Training CVC

Input: labeled data $L = \{< x_i, y_i >\}_{i=1}^n$, Clustering algorithm A, Number of clusters k, set of class labels $C = \{C_1, \ldots, C_m\}$

1. Let $U = \{x_i\}_{i=1}^n$ and apply A to cluster U into k groups
2. Associate each cluster with a class label that represents the majority of instances in the cluster

III. Materials And Methods

A. Datasets

Actigraphy and PSG data were retrieved for 37 participants from the following studies: insomnia [14]; baseline sleep in healthy participants from a pilot study and published studies [15-17]; older adults [2]; and sleep restriction in healthy participants [18] such that each subject has recordings for at least 3 sleeping periods. Detailed experimental settings of these studies including conditions of the study, characteristics of actigraphy devices, temporal alignment of actigraphy and PSG clocks, and PSG scoring are summarized in [2]. In this data, the number of recordings per subject varies between 3 to 11 recordings. To split the data into training and test sets, we categorized the subjects by their number of recordings and split them equally into train and test sets. This data partitioning procedure aims at eliminating bias by equally distributing subjects from different studies into training and test sets and by ensuring that data from each subject is either in training or test set. The final training dataset is composed of recordings for 19 subjects including 145 sleeping periods with 90,060 and 19,653 sleep and wake 30 second epochs, respectively. Similarly, the final test dataset is composed of recordings for 18 subjects including 137 sleeping periods with 97,205 and 22,572 sleep and wake 30 second epochs, respectively.

B. Feature Extraction

We represent each 30-second epoch using its contextual information in the form of activity counts of the target epoch as well as 10 epochs preceding and 10 epochs following the target epoch. In other words, each training instance consists of a contiguous window of 21 epochs, labeled with a binary (sleep/wake) label that indicates whether the target center epoch corresponds to the sleep state or the wake state as determined by the annotated PSG data.

We experimented with the following data representations of each 21-epoch window: i) Binarized activity counts (BAC) corresponding to the 21 activity counts in each window after binarizing them using a cutoff of 15 (i.e., activity counts greater than 15 map to 1 (wake) and activity counts less than 15 map to 0 (sleep)); [2]; ii) Normalized activity counts (NAC) corresponding to the 21 activity counts in each window after normalizing activity counts in the entire sleeping period to fall within the interval [0,1]; iii) Normalized activity counts plus summary statistics (NAC+), a total of 30 numeric features obtained by concatenating NAC and 9 summary statistics features based on three statistics, mean, lag-one autocorrelation, and number of non-zero epochs, extracted from the entire 21-epoch window, left side of the window (epochs 1 to 11), and right side of the window (epochs 11 to 21).

C. Proposed Method

We propose a novel *unsupervised* CVC algorithm (UCVC) for training a CVC model using unlabeled data. The basic idea is to use domain knowledge heuristics (e.g., some statistical property that discriminates clusters of sleep epochs from clusters of wake epochs) to associate a class label with each of the clusters. In this work, we experimented with three heuristics, based on simple observations regarding the number of sleep and wake samples in the data and the fact that wake samples often have higher activity counts, for assigning sleep/wake labels to clusters:

- H1: assign 'wake' label to the cluster with smaller number of instances.
- H2: assign 'wake' label to the cluster including samples with higher average accelerometer activity. Accelerometer activity of a sample is determined by the activity score of the central epoch (i.e., epoch 11 in each 21-epoch window).
- H3: assign 'wake' label to the cluster including samples with higher average accelerometer activity. Accelerometer activity of a sample is determined as the sum of its 21-epoch activity scores.

We implemented and tested UCVC (Algorithm 2) using four clustering methods, namely, *k*-means (KM) [19, 20], fuzzy c-means (FCM) [21], Gaussian mixture [22] using full covariance matrix (GMF), and Gaussian mixture using diagonal covariance matrix (GMD). We will refer to the models constructed using Algorithm 2 and any of these

Algorithm 2: Training ClassificationViaClustering from unlabeled actigraphy data

Input: unlabeled data $U = \{x_i\}_{i=1}^{n}$, Clustering algorithm A, Number of clusters $k = 2$, set of class labels $C = \{sleep, wake\}$

1. Apply A to cluster U into k groups
2. Associate each cluster with the class label determined by applying heuristic *H1, H2,* or *H3*

clustering methods as UCVC_KM, UCVC_FCM, UCVC_GMF, and UCVC_GMD.

D. Supervised Learning Models

We compared the UCVC sleep/wake state predictors trained on unlabeled actigraphy data with sleep/wake state predictors trained using supervised learning algorithms on labeled actigraphy data where the labels were obtained using expert scored PSG data. We considered four commonly used supervised machine learning algorithms: Gaussian Naïve Bayes (GNB) [23], Logistic Regression (LR) [24]; Random Forest (RF) [25] with 100 decision trees; and Extreme Gradient Boosting (XGB) [26] with 100 trees.

E. Performance Evaluation Metrics

We assessed the performance of different models for predicting sleep/wake state using four widely used threshold-dependent metrics, namely sensitivity (Sn), specificity (Sp), accuracy (ACC), and Matthew's correlation coefficient (MCC) [27]. These four metrics depend on the classification threshold used to convert predicted class probabilities into binary class labels. In our experiments, the optimal threshold was determined (for each classifier) such that MCC on training data is maximized. For comparing predictors using all possible thresholds, we reported the area under Receiver Operating Characteristic (ROC) curve [28].

F. Sleep Parameter Estimation

Accurate estimation of sleep parameters (e.g., sleep efficiency) could provide significant information about health conditions [29]. Here, we assess the performance of UCVC and supervised learning models in estimating five standard sleep parameters: i) Total Sleep Time (TST), which is the amount of actual sleep time (in minutes) in a sleep period; ii) Sleep Onset Latency (SOL), defined as the length of time (in minutes) that it takes to accomplish the transition from full wakefulness to sleep; iii) Sleep Efficiency (SE), defined as the ratio between actual sleep time and time spent in bed; iv) Wake After Sleep Onset (WASO), which determine the amount of time (in minutes) a person spends awake, starting from when they first fell asleep to when they become fully awake and do not attempt to go back to sleep; v) Number of Awakenings (NA), NA is the number of transitions from sleep to wakefulness lasting more than 15 seconds (given PSG-determined sleep stage).

IV. Results And Discussions

A. UCVC vs. Supervised Learning Models

Table 1 reports the performance of four UCVC models estimated using the independent test set. Using BAC data representation, both *k*-means based methods, UCVC_KM and UCVC_FCM, outperform Gaussian Mixture (GM) based methods, UCVC_GMF and UCVC_GMD, in terms of ACC, Sp, MCC, and AUC. Using NAC representation, GM based models have slight improvements in AUC as well as ACC and Sp. On the other hand, the performance of UCVC_KM predictor substantially drops compared to its performance using BAC representation. Interestingly, concatenating statistical and NAC features (NAC+ representation) allows UCVC_KM classifier to reach the highest observed performance in terms of AUC and MCC. Thus, our results suggest that using *k*-means clustering on the NAC+ representation yield the best performing UCVC model.

Table 2 reports predictive performance estimates of four supervised models on the independent test set. We note that, like UCVC models, supervised models prefer NAC+ data representation. For all supervised models the highest AUC is observed using NAC+ data representation. Switching from NAC to NAC+ representation yields an increase in AUC in the range 0.0-0.03. The zero improvement is reported using XGB whereas the 0.03 improvements are obtained using GNB.

B. Estimation of Sleep Parameters using Machine Learning Models

We report the mean summary statistics of five standard sleep parameters determined for each sleeping period in our test data using PSG and predicted sleep/wake state by the eight machine learning models considered in this study.

Due to space limitation, we omit the results obtained using BAC and NAC features and report only the results obtained using NAC+, the best performing data representation, in Table 3. The first row in Table 3 shows the PSG ground truth estimates of the five sleep parameters. The remaining rows report the sleep parameters estimated using our UCVC and supervised learning models. We note that there is no single model that comes closest to all five gold standard estimates. Gaussian Naive Bayes (GNB) supervised learning model yields the best estimates of TST, SE, and WASO. In the case of average SOL, we observe that all models under-estimate it and the closest estimate is obtained using UCVC_GMF. In the case of mean NA, the closest estimates are obtained using RF and XGB predictors.

Figure 1 shows the scatter plot for SE (left) and WASO (right) estimates of top two performing UCVC and supervised learning models, respectively. The lines show a linear best fit for the four models and the line of identity. In both cases, the best linear fit lines are close to each other. The visualized results show that for SE estimates, the four models tend to over-estimate the true SE values. For WASO estimates, the four models tend to underestimate the true WASO scores (except for sleeping periods with true WASO scores less than 40 minutes).

TABLE 1: PERFORMANCE COMPARISONS OF UCVC MODELS FOR PREDICTING SLEEP/WAKE STATE USING UNLABELED ACTIGRAPHY DATA. EXACT RESULTS ARE OBTAINED USING ANY OF THE THREE HEURISTICS.

Representation	Method	ACC	Sn	Sp	MCC	AUC
BAC	UCVC_KM	**0.85**	0.30	**0.98**	0.42	0.78
	UCVC_FCM	0.84	0.43	0.93	0.41	0.78
	UCVC_GMF	0.67	**0.76**	0.64	0.32	0.71
	UCVC_GMD	0.69	0.73	0.69	0.33	0.72
NAC	UCVC_KM	0.75	0.41	0.83	0.22	0.71
	UCVC_FCM	0.77	0.52	0.86	0.35	0.79
	UCVC_GMF	0.74	0.54	0.78	0.29	0.73
	UCVC_GMD	0.73	0.54	0.78	0.29	0.74
NAC+	UCVC_KM	**0.85**	0.39	0.96	**0.44**	**0.80**
	UCVC_FCM	**0.85**	0.38	0.96	**0.44**	0.76
	UCVC_GMF	0.66	0.74	0.64	0.30	0.72
	UCVC_GMD	0.63	0.70	0.61	0.24	0.69

TABLE 2: PERFORMANCE COMPARISONS OF SUPERVISED LEARNING MODELS FOR PREDICTING SLEEP/WAKE STATE USING LABELED ACTIGRAPHY DATA.

Representation	Method	ACC	Sn	Sp	MCC	AUC
BAC	GNB	0.83	**0.48**	0.92	0.42	0.78
	LR	**0.85**	0.40	0.96	0.45	0.79
	RF	**0.85**	0.26	**0.98**	0.39	0.76
	XGB	**0.85**	0.37	0.97	0.45	0.79
NAC	GNB	0.81	0.42	0.90	0.34	0.76
	LR	0.83	0.40	0.94	0.39	0.78
	RF	0.84	0.29	0.96	0.36	0.77
	XGB	**0.85**	0.39	0.96	0.45	**0.81**
NAC+	GNB	0.83	0.43	0.93	0.40	0.79
	LR	0.84	0.39	0.95	0.41	0.79
	RF	0.84	0.30	0.97	0.39	0.78
	XGB	**0.85**	0.42	0.96	**0.46**	**0.81**

In summary, our results show that the UCVC_KM predictor has predictive performance that is comparable to the best performing supervised learning models and can provide accurate estimates of three standard sleep parameters: TST, SE, and WASO.

V. CONCLUSIONS

We presented a novel approach for developing sleep/wake state detectors using unlabeled actigraphy data. Specifically, we introduced a variant of the classification via clustering (CVC) algorithm [13] that can be used for unsupervised training of sleep/wake state detectors from unlabeled data.

Our results demonstrate that the UCVC models using *k*-means clustering have predictive performance that is comparable with the performance of the machine learning models trained using supervised machine learning algorithms and labeled actigraphy data. By eliminating the need for costly and cumbersome expert-annotated PSG recordings for labeling actigraphy data for developing sleep/wake state detectors and estimating sleep parameters, our results make it possible to conduct large-scale sleep research studies in naturalistic settings and over extended time periods. The resulting unsupervised classification via clustering framework is broadly applicable in scenarios where labeled data are cumbersome or costly to obtain but cluster properties and domain knowledge can be used to devise

TABLE 3: PERFORMANCE COMPARISONS OF UCVC AND SUPERVISED MODELS (USING NAC+ REPRESENTATION) IN TERMS OF AVERAGE SLEEP PARAMETER. PSG REPRESENTS THE GOLD STANDARD ESTIMATES. HIGHLIGHTED SCORES REPRESENT CLOSEST ESTIMATES TO THE GOLD STANDARD

Method	Total Sleep Time (TST;min)	Sleep Onset Latency (SOL; min)	Sleep Efficiency (SE; %)	Wake After Sleep Onset (WASO; min)	Number of Awakenings (NA)
PSG	355.02	18.17	80.55	53.94	21.56
Actigraphy-derived estimates					
UCVC_KM	390.65	9.04	87.75	41.13	8.42
UCVC_GMD	243.04	13.98	54.96	181.10	44.84
UCVC_GMF	250.16	**17.59**	56.42	167.50	48.19
UCVC_FCM	393.79	8.87	88.43	38.17	8.36
GNB	**376.25**	8.40	**84.51**	**56.98**	15.88
LR	388.20	6.98	87.06	46.77	31.34
RF	404.07	6.05	90.54	31.80	**19.21**
XGB	387.26	7.93	86.90	46.53	**23.95**

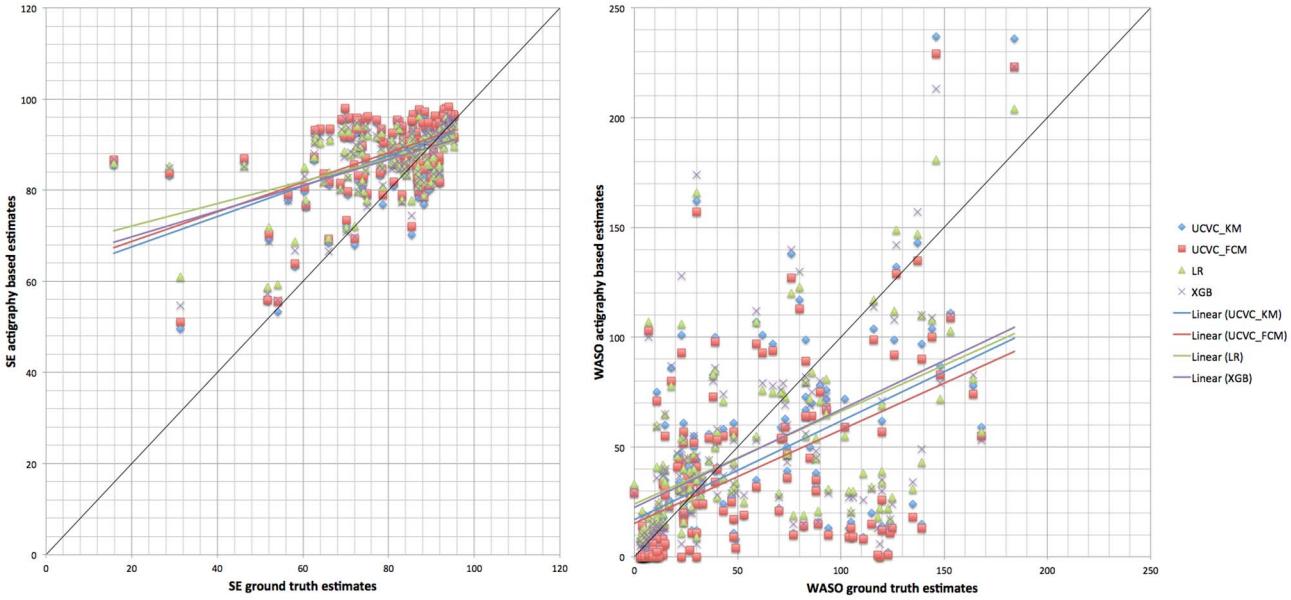

Figure 1: Comparisons of SE (left) and WASO (right) ground truth estimates and actigraphy based estimates derived using UCVC_KM, UCVC_FCM, LR, and XGB models for 137 test sleeping periods. The lines show a linear best fit for the four models and the line of identity.

effective heuristics for associating clusters with class labels. Work in progress is aimed at: developing personalized sleep/wake state predictors and sleep parameter estimators that better account for variability across individuals; developing variants of the proposed unsupervised classification via clustering method for predicting sleep stages [30] and recognizing physical activity from accelerometer and other wearable sensor data [10].

ACKNOWLEDGMENT

This project was supported in part by the Edward Frymoyer Endowed Professorship in Information Sciences and Technology at Pennsylvania State University and the Sudha Murty Distinguished Visiting Chair in Neurocomputing and Data Science at the Indian Institute of Science [both held by Vasant Honavar] and the Pennsylvania

State University's Center for Big Data Analytics and Discovery Informatics which is co-sponsored by the Institute for Cyberscience, the Huck Institutes of the Life Sciences, and the Social Science Research Institute at the university. The project was also supported by the National Heart, Lung and Blood Institute (R01HL107240) and General Clinical Research Center (M01-RR02635).

REFERENCES

[1] M. Hirshkowitz, "The history of polysomnography: tool of scientific discovery," *Sleep Medicine: A Comprehensive Guide to Its Development, Clinical Milestones, and Advances in Treatment,* pp. 91-100, 2015.

[2] M. Marino, Y. Li, M. N. Rueschman, J. Winkelman, J. M. Ellenbogen, J. Solet*, et al.*, "Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography," *Sleep,* vol. 36, p. 1747, 2013.

[3] M. Hirshkowitz and A. Sharafkhaneh, "Comparison of portable monitoring with laboratory polysomnography for diagnosing sleep-related breathing disorders: scoring and interpretation," *Sleep Medicine Clinics,* vol. 6, pp. 283-292, 2011.

[4] N. Stanley, "Actigraphy in human psychopharmacology: a review," *Human Psychopharmacology: Clinical and Experimental,* vol. 18, pp. 39-49, 2003.

[5] M. Bruyneel, S. Van den Broecke, W. Libert, and V. Ninane, "Real-time attended home-polysomnography with telematic data transmission," *International journal of medical informatics,* vol. 82, pp. 696-701, 2013.

[6] J. Paquet, A. Kawinska, and J. Carrier, "Wake detection capacity of actigraphy during sleep," *Sleep,* vol. 30, pp. 1362-1369, 2007.

[7] D. F. Kripke, E. K. Hahn, A. P. Grizas, K. H. Wadiak, R. T. Loving, J. S. Poceta*, et al.*, "Wrist actigraphic scoring for sleep laboratory patients: algorithm development," *Journal of sleep research,* vol. 19, pp. 612-619, 2010.

[8] A. Domingues, T. Paiva, and J. M. Sanches, "Sleep and wakefulness state detection in nocturnal actigraphy based on movement information," *IEEE Transactions on Biomedical Engineering,* vol. 61, pp. 426-434, 2014.

[9] G. Orellana, C. Held, P. Estevez, C. Perez, S. Reyes, C. Algarin*, et al.*, "A balanced sleep/wakefulness classification method based on actigraphic data in adolescents," presented at the Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE, 2014.

[10] Y. Zheng, W. K. Wong, X. Guan, and S. Trost, "Physical Activity Recognition from Accelerometer Data Using a Multi-Scale Ensemble Method," presented at the Proceedings of the 27th AAAI Conference on Artificial Intelligence, 2013.

[11] M. I. Lopez, J. Luna, C. Romero, and S. Ventura, "Classification via clustering for predicting final marks based on student participation in forums," *International Educational Data Mining Society,* pp. 148-151, 2012.

[12] B. Abdullateef, M. Modinat, and S. Shakirat, "Enhanced classification via clustering techniques using decision tree for feature selection," *Algorithms,* vol. 9, 2015.

[13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter,* vol. 11, pp. 10-18, 2009.

[14] J. W. Winkelman, O. M. Buxton, J. E. Jensen, K. L. Benson, S. P. O'Connor, W. Wang*, et al.*, "Reduced brain GABA in primary insomnia: preliminary data from 4T proton magnetic resonance spectroscopy (1H-MRS)," *Sleep,* vol. 31, pp. 1499-1506, 2008.

[15] O. M. Buxton, J. M. Ellenbogen, W. Wang, A. Carballeira, S. O'connor, D. Cooper*, et al.*, "Sleep disruption due to hospital noises: A prospective evaluation," *Annals of internal medicine,* vol. 157, pp. 170-179, 2012.

[16] T. T. Dang-Vu, S. M. McKinney, O. M. Buxton, J. M. Solet, and J. M. Ellenbogen, "Spontaneous brain rhythms predict sleep stability in the face of noise," *Current Biology,* vol. 20, pp. R626-R627, 2010.

[17] S. M. McKinney, T. T. Dang-Vu, O. M. Buxton, J. M. Solet, and J. M. Ellenbogen, "Covert waking brain activity reveals instantaneous sleep depth," *PLoS One,* vol. 6, p. e17351, 2011.

[18] O. M. Buxton, M. Pavlova, E. W. Reid, W. Wang, D. C. Simonson, and G. K. Adler, "Sleep restriction for 1 week reduces insulin sensitivity in healthy men," *Diabetes,* vol. 59, pp. 2126-2133, 2010.

[19] J. MacQueen, "Some methods for classification and analysis of multivariate observations: UCLA, Western Manag. Sci," Inst. Working Paper 961966.

[20] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*: Prentice-Hall, Inc., 1988.

[21] J. Warner. *Scikit-fuzzy: A fuzzy logic toolbox for scipy*. Available: http://pythonhosted.org/scikit-fuzzy/

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological),* pp. 1-38, 1977.

[23] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach*: Springer Science & Business Media, 2013.

[24] S. Le Cessie and J. C. Van Houwelingen, "Ridge estimators in logistic regression," *Applied statistics,* pp. 191-201, 1992.

[25] L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.

[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785-794.

[27] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview," *Bioinformatics,* vol. 16, pp. 412-424, 2000.

[28] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters,* vol. 27, pp. 861-874, 2006.

[29] D. W. Jung, Y. J. Lee, D. U. Jeong, and K. S. Park, "New predictors of sleep efficiency," *Chronobiology international,* vol. 34, pp. 93-104, 2017.

[30] T. Willemen, D. Van Deun, V. Verhaert, M. Vandekerckhove, V. Exadaktylos, J. Verbraecken*, et al.*, "An evaluation of cardiorespiratory and movement features with respect to sleep-stage classification," *IEEE journal of biomedical and health informatics,* vol. 18, pp. 661-669, 2014.