

# Daytime Sleepiness Level Prediction Using Respiratory Information

Kazuhiko Shinoda, Masahiko Yoshii, Hayato Yamaguchi and Hirotaka Kaji

Frontier Research Center, Toyota Motor Corporation

{kazuhiko\_shinoda, masahiko\_yoshii, hayato\_yamaguchi\_aa, hirotaka\_kaji}@mail.toyota.co.jp

## Abstract

Daytime sleepiness is not only the cause of productivity decline and accidents, but also an important metric of health risks. Despite its importance, the long-term quantitative analysis of sleepiness in daily living has hardly been done due to time and effort required for the continuous tracking of sleepiness. Although a number of sleepiness detection technologies have been proposed, most of them focused only on driver's drowsiness. In this paper, we present the first step towards the continuous sleepiness tracking in daily living situations. We explore a methodology for predicting subjective sleepiness levels utilizing respiration and acceleration data obtained from a novel wearable sensor. A class imbalance handling technique and hidden Markov model are combined with supervised classifiers to overcome the difficulties in learning from an imbalanced and time series dataset. We evaluate the performance of our models through a comprehensive experiment.

## 1 Introduction

Inadequate sleep causes daytime sleepiness, and consequently, leads to productivity decline, accidents and health deterioration. Huge economic loss caused by sleepiness was reported in the recent studies [Hafner *et al.*, 2016; Hillman *et al.*, 2018]. To design an effective strategy to reduce daytime sleepiness, the association between sleepiness and lifestyle habits (e.g. sleep duration, activities and food) must be revealed. Subjective measures are widely employed to investigate the cause and effect of sleepiness [Åkerstedt *et al.*, 2014; Shimura *et al.*, 2018]. However, there have been a quite limited number of long-term and large-scale quantitative studies on sleepiness in daily living. If daytime sleepiness can be measured automatically, it will bring a huge impact on the field of health and wellness research from the aspects of time resolution and burdens for both subjects and researchers.

A wide variety of physiological signals including heart-beat, respiration, pupil dilation, electrodermal activity are affected by the autonomic nervous system (ANS) [Robertson *et al.*, 2012]. The relation between sleepiness and ANS has also been confirmed [Pressman and Fry, 1989]. Therefore, a

number of technologies have been proposed to detect sleepiness from the physiological information by using this principle. In this paper, we attempt to predict daytime subjective sleepiness by using a wearable respiration sensor. Respiration is one of the promising physiological information to capture various mental states and activities such as sleep and speaking. This approach allows us to track daytime sleepiness automatically for a long period of time.

In order to develop a high performance predictor, we need to tackle two difficulties: The first technical challenge is the appropriate design of an experiment to collect sleepiness data under a natural condition. Most of the previous studies focused on driver's drowsiness and did not consider daytime sleepiness. We conduct a data collection procedure which contains both active and passive tasks representing the natural situations in daily life. The second technical challenge is to train a classifier from a time series data with the class imbalance problem. Sleepiness data collection under the natural situation causes the bias in sleepiness levels. To cope with this problem, we combine a class imbalance handling technique and a time series processing method with a supervised machine learning classifier. To the best of our knowledge, this is the first study that predicts daytime sleepiness by using a wearable respiration sensor.

## 2 Related Work

Various measurement methods are carried out to evaluate sleepiness. Subjective measures such as the visual analog scale [Monk, 1989], the Stanford sleepiness scale [Hoddes *et al.*, 1973], the Epworth sleepiness scale [Johns, 1991] and the Karolinska sleepiness scale (KSS) [Åkerstedt and Gillberg, 1990] are widely employed to measure sleepiness in laboratory and clinical settings. Kaida *et al.* demonstrated the validity and reliability of KSS against electroencephalogram (EEG) and behavioral indicators of sleepiness [Kaida *et al.*, 2006]. Facial and behavioral expression scores annotated by experts were also used in a lot of automotive drowsiness detection studies [Hachisuka *et al.*, 2011]. As a physiological approach, EEG-based methods are frequently used in laboratory settings [Sforza *et al.*, 2002], although it is difficult to install in daily living situations.

There have been plenty of studies with regard to drowsiness detection for drivers. Physiological and behavioral information such as heart rate variability based on electrocar-

diagram [Vicente *et al.*, 2016], eye blink [Häkkinen *et al.*, 1999], driver’s behavior captured by an accelerometer and a gyroscope in a smart watch [Lee *et al.*, 2016] were employed to detect driver’s drowsiness.

Respiration reflects voluntary and involuntary motions, hence it contains the various information about mental states. Sleep stage prediction is one of the promising application [Tataraidze *et al.*, 2015]. Plarre *et al.* investigated a wearable stress monitoring which utilized respiration and electrocardiogram data obtained in a daily life environment [Plarre *et al.*, 2011]. Furthermore, Igasaki *et al.* reported that subjective sleepiness while driving can be predicted by respiration [Igasaki *et al.*, 2016]. These previous studies utilized a thoracic respiration band to collect respiratory data.

### 3 Data Collection

#### 3.1 Experimental Setting

Eighteen individuals (nine males and nine females) who belonged to our affiliation participated in the experiment. The mean of their age was 33.8 and the standard deviation was 9.8. The design of the experiment was approved and conducted according to the ethical guidelines of Toyota Motor Corporation. We sufficiently explained the details and obtained the informed consents from all the subjects.

We did not include the driving task in the experiment as it may undesirably affect subjects’ sleepiness. More specifically, they try hard to keep awake because making mistakes while driving may lead to fatal accidents, while they often accept sleepiness and fall sharply into sleep under daily living situations. It has been reported in the literature that this resistance to sleepiness can affect physiological responses [Horne and Reyner, 1999], and consequently, respiratory data while driving and daily tasks may reflect sleepiness in different ways. In order to collect natural sleepiness data in daily living, the experiment was composed of the following two main parts with a five-minute rest in between:

- **Typing task:** The subjects were instructed to type the content of a book as much as possible for five minutes.
- **Video watching task:** The subjects were instructed to sit on a sofa and watch a movie for 90 minutes. They had a choice of which one to watch from several dramas and animated movies without stimulating scenes.

We started the experiment at around 14:00 every time to make sure that the subjects gradually got sleepy from the alert state during the experiment. Before the experiment, we instructed them to take enough sleep the day before, not to take caffeine, not to eat too much on the day of the experiment and go to the restroom beforehand. Furthermore, we added the following two procedures to make the subject feel comfortable during and after the experiment: 1) Provide bottled water. The subject can drink it anytime they want. 2) At the end of the experiment, increase the luminance of the room to reduce the subject’s sleepiness. All the subjects executed this protocol twice in different days. A quiet room in our laboratory was used as the experimental venue. The overview of the room is shown in Figure 1.



Figure 1: The experiment room.



Figure 2: The wearable respiration sensor.

In order to measure their respiratory signal and body movement, we employed the newly developed wearable respiration sensor [Kaji *et al.*, 2018]. The sensor was attached to a subject’s abdomen as shown in Figure 2 to capture the small abdominal movement as a respiratory waveform. It is easier and more comfortable to wear than a thoracic band, hence suitable for the data collection in the daily living situation. Three-axis accelerometer built into the sensor was used to measure body movement. We set the sampling frequency to 20 Hz.

As the ground truth of sleepiness, KSS, a 9-point Likert scale rated from “extremely alert (1)” to “very sleepy (9)” [Åkerstedt and Gillberg, 1990], was employed. We programmed it on an Android tablet to periodically ask sleepiness with the notification sound and vibration. The subjects were asked their sleepiness before and after the typing task and during the video watching task every three minutes. We chose the shorter interval of asking KSS than one used in the previous study on driver’s drowsiness [Åkerstedt *et al.*, 2005] to increase the number of the available data. However, we confirmed it hardly affected the subject’s sleepiness in the preliminary experiment. In fact, we observed the subject sometimes did not notice the sound and vibration, and therefore could not answer KSS due to strong sleepiness. Missing values were replaced with KSS 9 in such a case.

The collected KSS scores were then divided into two sleepiness levels, namely, Alert (KSS 1 to 5) and Sleepy (KSS 6 to 9) as shown in Table 1. We reframed the sleepiness prediction as a binary classification problem because the physiological data were heavily noisy and the subjective measure could have a bias depending on a subject. Binary classification is much simpler in terms of resolution than multiclass classification and regression, but still useful for capturing the

Sleepiness level	KSS score	Definition
Alert	1	Extremely alert
	2	Very alert
	3	Alert
	4	Rather alert
	5	Neither alert nor sleepy
Sleepy	6	Some signs of sleepiness
	7	Sleepy, but no effort to keep awake
	8	Sleepy, some effort to keep awake
	9	Very sleepy, great effort to keep awake, fighting sleep

Table 1: The definition of the subjective sleepiness levels.

change points and the trends of sleepiness.

Those data were synchronized and stored in the data server. As a consequent, we collected approximately 3,420 minutes of data in total.

## 3.2 Data Processing

### Preprocessing

The collected respiratory data were first preprocessed with a fourth-order butterworth bandpass filter for noise reduction. We set the lower bound of the passband frequency to 0.05 Hz, which was sufficient to capture respiratory signals given that the breathing rate of healthy adults ranges from 12 to 20 per minute, i.e. 0.2 - 0.33 Hz [Yuan *et al.*, 2013]. The upper bound frequency was chosen so that the filtered waveforms reflect important features of raw waveforms. We found from the preliminary experiment that 2.5 Hz was the optimal frequency for balancing noise reduction and the preservation of the original waveform features.

Both of the filtered respiratory data and the raw acceleration data were then segmented into 60-second windows with 30-second overlaps. A sleepiness level is associated with fluctuations in a breathing pattern, therefore each window has to be long enough to contain as many breathing cycles as required for the accurate evaluation of breathing fluctuations. The 60-second window can contain at least six breathing cycles, which were found to be necessary from our preliminary experiment. We annotated each 60-second window with the nearest value of the collected subjective sleepiness levels.

### Feature Extraction

We calculated the mean, standard deviation, max and min of nine respiration parameters described as follows for each 60-second window. The number shows the unit's place of the feature ID.

- 0. Inspiratory Duration:** The duration between the bottom and the next peak of a respiratory waveform.
- 1. Expiratory Duration:** The duration between the peak and the next bottom of a respiratory waveform.
- 2. Duration Ratio:** The ratio of Expiratory duration to Inspiratory duration.
- 3. Pause Duration:** The duration of the rest between expiration and inspiration.

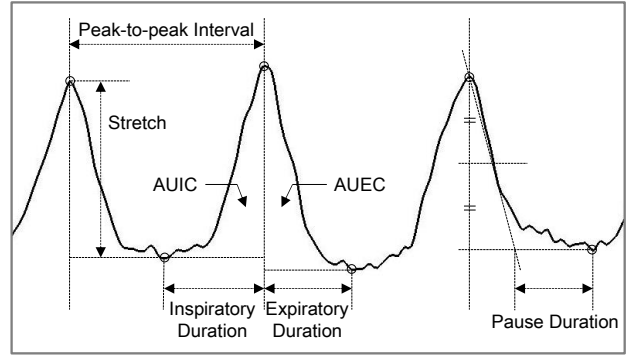


Figure 3: The illustration of the respiration parameters.

- 4. AUIC:** The area under an inspiratory curve.
- 5. AUEC:** The area under an expiratory curve.
- 6. Area Ratio:** The ratio of AUEC to AUIC.
- 7. Peak-to-Peak Interval:** The interval between the peak and the next peak of a respiratory waveform.
- 8. Stretch:** The amplitude from the peak to the bottom.

Figure 3 illustrates the respiration parameters. Pause Duration is geometrically defined as shown in Figure 3. To be more specific, find the point which is as high as half of the Stretch on an expiratory curve, and let HD denote the duration between the peak and this point. Pause Duration (PD) can be defined as the following equation using Expiratory Duration (ED):  $PD = ED - 2HD$ . We did not use frequency-domain features in order to suppress the computational cost as we aim to develop the online sleepiness prediction and embed it into the wearable sensor.

We also computed the mean, standard deviation, max and min of the  $\ell_2$  norm of the first-order difference of three-axis accelerations  $\|\mathbf{a}_t - \mathbf{a}_{t-1}\|$  as the body-movement features, where  $\mathbf{a}_t = (a_t^x, a_t^y, a_t^z)^T$  is the  $t$ -th output. We numbered this body-movement parameter 9. All the extracted features were normalized to have zero mean and unit standard deviation to reduce the inter- and intra- personal differences of breathing patterns and body motion.

### Feature Selection

We took a two-step method for the feature selection. The first step was to eliminate strong correlation among the features. We calculated the correlation coefficients of all the pairs, and identify the pairs/groups of the features which strongly correlated with each other (if the absolute value of the correlation coefficients exceeded 0.7). All the features in the pairs/groups except one with the highest correlation with the sleepiness level were removed. It resulted in removal of 19 features.

In the second step, we employed the aggregated feature importance computed by multiple random forests (RFs) [Breiman, 2001] as a selection criterion because it was reported to be able to well balance the robustness of selected features and prediction performance [Saeys *et al.*, 2008]. We created 100 RFs and simply aggregated the feature importance derived from each RF. Undersampling was used before creating each RF to take into account the imbalanced

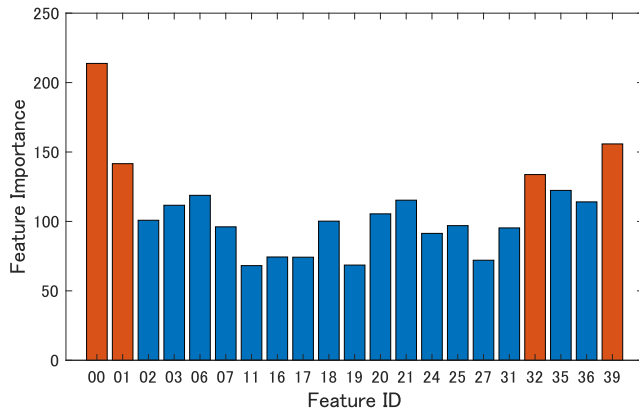


Figure 4: The aggregated feature importance. The selected features are shown as red bars. The unit's place of the ID shows the parameter. The ten's place is associated with the function for calculating features; 0X: mean, 1X: standard deviation, 2X: max, 3X: min.

dataset (see the next section for the detailed explanation of the class imbalance problem). The candidate features were ranked based on the importance and we included the best four features in our experimental models. The aggregated importance of the 21 candidate features is shown in Figure 4.

As a result of the above selection method, the mean of Inspiratory Duration, the mean of Expiratory Duration, the min of Duration Ratio and the min of the body-movement parameter were selected. The selected set included three features related to Inspiratory Duration and Expiratory Duration. They were also selected in the feature selection algorithm performed in the previous study which built models for stress prediction in a natural environment [Plarre *et al.*, 2011]. These results indicate Inspiratory Duration and Expiratory Duration are of great importance in predicting human mental states in an everyday life situation. On the other hand, none of the features related to the amplitude (AUIC, AUEC and Stretch) were effective mainly because of the measuring principle of the sensor. It measures respiration by sensing the abdominal movements, which leads to large variability in the amplitude depending on the subject's posture and body size.

## 4 Method

Let  $\mathbf{x} = (x_1, \dots, x_d)^\top$  be a  $d$ -dimensional respiration and body-movement feature vector and  $y \in \{\text{Alert}, \text{Sleepy}\}$  be the class label of the sleepiness levels determined by KSS. There are various problems when predicting human mental states from physiological data by using supervised classification methods. We explored a procedure to improve the performance of the sleepiness prediction by focusing on the two major problems, i.e. class imbalance and the i.i.d. assumption. The effectiveness of the proposed procedure was examined through careful model validation.

### 4.1 Class Imbalance Handling

Class imbalance often becomes a serious problem in the domain of human mental state prediction because mental states are not completely controllable even if experiments are designed to induce desired states to some extent. Sleepiness is

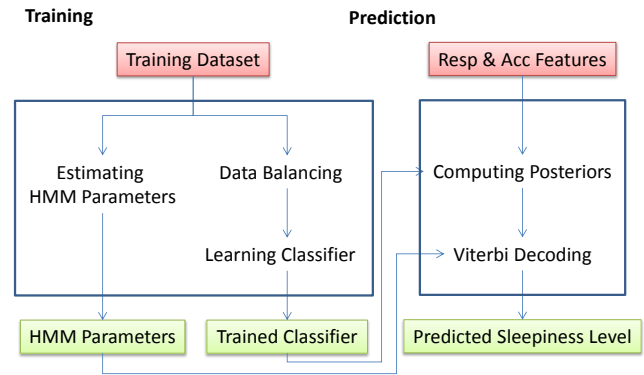


Figure 5: The block diagram of the sleepiness prediction procedure.

not an exception to this and the collected data were imbalanced (Alert: 30.6%, Sleepy: 69.4%). Learning a classifier from imbalanced datasets without proper techniques can result in biased prediction towards the majority class.

A number of methods for handling the class imbalance problem have been developed and proposed. Some experimental studies demonstrated that undersampling bagging worked better than other types of methods [Galar *et al.*, 2012; Khoshgoftaar *et al.*, 2011]. In this paper, we used two types of undersampling bagging i.e. exactly balanced bagging (EBB) [Chang *et al.*, 2003] and roughly balanced bagging (RBB) [Hido and Kashima, 2008] in addition to the other basic methods i.e. undersampling (US), oversampling (OS) and synthetic minority oversampling technique (SMOTE) [Chawla *et al.*, 2002] to compare the effectiveness.

### 4.2 Time Series Processing

Although supervised classification methods such as support vector machine (SVM) [Vapnik, 1998] and RF can well represent the relation between feature vectors and classes, the sequence of prediction results as a whole can be unnatural due to the i.i.d. assumption. Especially in the case of using a wearable sensor, data can be easily affected by body motion, making it difficult to obtain noiseless waveforms even after filtering. Consequently, supervised classifiers predict the uninterpretable sequence of sleepiness which is unlikely to occur in real life.

We combined classification methods and hidden Markov model (HMM) [Rabiner, 1989] to overcome this problem like [Han and Wang, 2014; Richard *et al.*, 2018] who combined a neural network and HMM. In their procedure, the posterior probability  $p(y|\mathbf{x})$  of each class is computed by a classifier, and converted to the likelihood which is proportional to a conditional probability:

$$p(\mathbf{x}|y) \propto \frac{p(y|\mathbf{x})}{p(y)},$$

where  $p(y)$  is a class prior. The likelihood is used in the Viterbi decoding [Rabiner, 1989] to predict the sequence of the classes. In this experiment,  $p(y)$  was empirically computed from the ratio of classes contained in a training set.

	- norm.	+ norm.	US	OS	SMOTE	EBB	RBB
	-	.5684(.0000)	.5688(.0000)	<b>.5853</b> (.0021)	<b>.5849</b> (.0011)	.5843(.0012)	.5844(.0011)
NB	MED	.5892(.0000)	.5878(.0000)	.6133(.0029)	<b>.6139</b> (.0018)	<b>.6142</b> (.0012)	.6134(.0014)
	HMM	.5481(.0000)	.6223(.0000)	.6227(.0039)	<b>.6234</b> (.0024)	<b>.6237</b> (.0025)	.6234(.0022)
	-	.5508(.0008)	.5483(.0009)	.5825(.0041)	.5785(.0035)	.5774(.0027)	.5852(.0014)
SVM	MED	.5645(.0023)	.5536(.0009)	.6168(.0064)	.6164(.0049)	.6150(.0047)	.6175(.0025)
	HMM	.5551(.0032)	.6404(.0035)	<b>.6669</b> (.0159)	.6507(.0121)	.6418(.0128)	.6580(.0042)
	-	.4339(.0000)	.5330(.0000)	.5822(.0019)	.5827(.0014)	.5820(.0013)	<b>.5830</b> (.0006)
LR	MED	.4352(.0000)	.5333(.0000)	.6166(.0035)	.6175(.0026)	.6168(.0023)	<b>.6189</b> (.0012)
	HMM	.5755(.0000)	.7044(.0000)	.7002(.0085)	.6998(.0083)	.7015(.0054)	<b>.7059</b> (.0061)
	-	.5754(.0046)	.5944(.0041)	.5940(.0052)	<b>.6037</b> (.0036)	.5874(.0040)	.5968(.0028)
RF	MED	.5912(.0064)	.6114(.0058)	.6376(.0079)	.6298(.0049)	.6280(.0065)	.6394(.0048)
	HMM	.6609(.0078)	.6813(.0074)	.6880(.0113)	.6257(.0066)	.6491(.0108)	<b>.6926</b> (.0074)
	-	.5941(.0106)	.5739(.0034)	.6060(.0069)	.6079(.0058)	.5984(.0088)	.6134(.0018)
NN	MED	.6050(.0127)	.5727(.0038)	.6424(.0095)	.6451(.0080)	.6304(.0130)	<b>.6513</b> (.0028)
	HMM	.6298(.0266)	.6798(.0142)	.6672(.0212)	.6706(.0176)	.6508(.0229)	<b>.6832</b> (.0069)

Table 2: Comparison of the mean and standard deviation of the F-measure over 100 trials. The first two columns show the results without any class imbalance methods.

### 4.3 Procedure for Sleepiness Prediction

Considering the aforementioned problems, we established the procedure for the sleepiness prediction illustrated in Figure 5. Both a class imbalance method and HMM are combined with a supervised classifier at the same time. For training, we first applied a class imbalance method to obtain a balanced dataset, and then built an unbiased classifier. The parameters of HMM such as the state transition probabilities are estimated from the sequence of the subjective sleepiness levels. Prediction is done in two steps using the trained classifier and the HMM parameters.

### 4.4 Experiment

Using the selected four features, we conducted an experimental analysis of the sleepiness prediction with respect to the classifiers, the class imbalance handling and the time series processing. All the combinations were evaluated based on the macro F-measure expressed as the equation below since the dataset was imbalanced;

$$\text{macro F-measure} = \frac{1}{c} \sum_{i=1}^c \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i},$$

where  $c = 2$  is the number of the classes. The related works employed the  $k$ -fold cross validation to evaluate the performance of their models, which means the training set and the test set included the data obtained from the same subject [Igasaki *et al.*, 2016; Plarre *et al.*, 2011]. However, physiological data with corresponding labels are often quite hard to collect from all system users in practice. We used the leave-one-subject-out cross validation (LOSO-CV) [Bao and Intille, 2004] to ensure the evaluation under more practical conditions. In our LOSO-CV, the samples of 17 subjects were used for training the model and the remaining one subject’s samples were used as the test set.

The baseline classifiers we selected were naive Bayes (NB), SVM, logistic regression (LR), RF and a three-layer

neural network (NN). The hidden layer of NN consisted of four nodes with the sigmoid function. Note that all of these classifiers can produce the posterior probabilities of the sleepiness levels. Here, the method proposed in [Platt, 2000] was employed in order to yield the posterior probabilities from SVM. We also compared a median filter (MED) of a 150-second window with HMM as a technique for smoothing time series data. Since random sampling was used in the class imbalance handling, we executed 100 trials with different random seeds to evaluate the performances.

## 5 Result

Table 2 presents the mean and standard deviation of an F-measure for each combination over 100 trials. Bold face denotes the best and comparable combinations in terms of the class imbalance handling techniques according to the paired  $t$ -test at the significance level of 5%. We first compared the results with and without the feature normalization (the first two columns). Although there were some combinations which showed the better F-measure without the normalization, it generally worked well. All the baseline classifiers with HMM got the significant improvement by the feature normalization. This result indicates that the normalization was especially helpful for the accurate prediction of the posterior probabilities. In light of this, we used the normalized features to compute the F-measure of the combinations with the class imbalance handling.

The results shown in Table 2 proved that the proposed procedure was effective for all the five classifiers. It was able to improve the F-measure by 0.1109 on average compared to the single classifier. The best performing combination was LR-EBB-HMM with the mean F-measure of 0.7059, which was rather high considering the ground truth of sleepiness was based on the self-report ratings.

The improvement made by adding a class imbalance handling technique and HMM also can be confirmed visually.

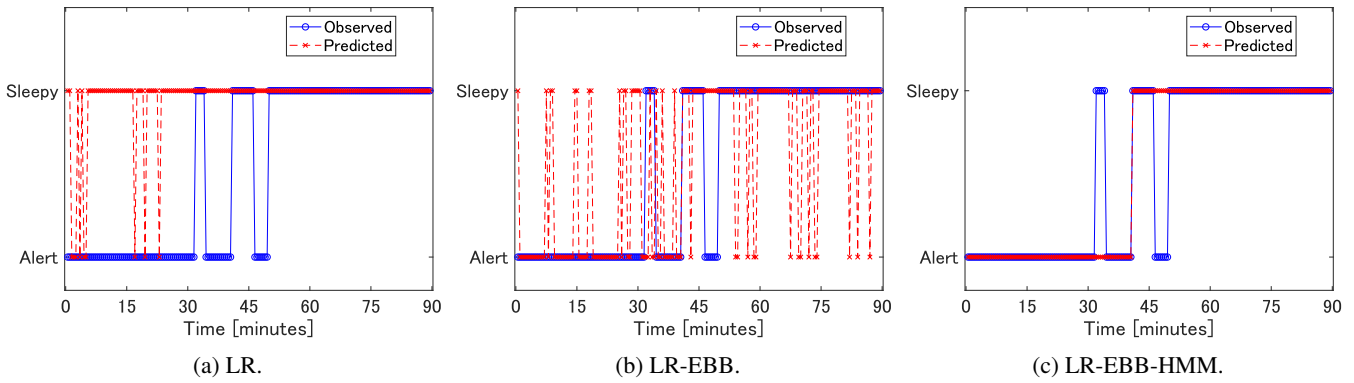


Figure 6: The transitions of the observed and the predicted sleepiness levels of a subject by three different combinations. Blue solid line with circle: observed sleepiness level, red dotted line with cross: predicted sleepiness level.

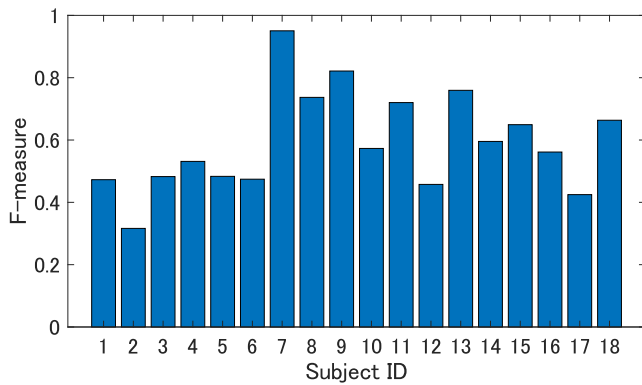


Figure 7: F-measure of LR-EBB-HMM by subject. Subject 1 to 9 are males and 10 to 18 are females.

The transitions of the predicted sleepiness by three different combinations are shown in Figure 6. LR-EBB-HMM in Figure 6c correctly predicted the trend of the sleepiness transition while LR in Figure 6a was strongly biased towards the Sleepy class. However, LR-EBB-HMM seems to have a tendency to ignore relatively small changes in sleepiness. LR-EBB in Figure 6b is still not enough because its prediction unrealistically changed in a short period of time, making it difficult to figure out how the actual sleepiness changed over time.

We observed that adding HMM alone significantly improved the F-measure. This may be because posteriors are divided by priors when put into HMM, which is substantially equal to adjusting the class distribution of a dataset. In this respect, the combination of supervised classifiers and HMM also acts like a class imbalance method. Indeed, LR-HMM, RF-HMM and NN-HMM outperformed some other combinations with the class imbalance method.

When focusing on the class imbalance handling, we found EBB and RBB were included in many of the best or comparable combinations, which was consistent with the previous studies arguing that undersampling bagging was an effective approach. We confirmed these bagging-based methods worked well also with HMM and a median filter.

Figure 7 shows the mean F-measure of LR-EBB-HMM

by subject. The F-measure varies across the subjects and it seems hard to build a common high-performance predictor for all. This result shows that there were huge interpersonal differences in breathing patterns like many other physiological indicators. One effective approach to this problem might be to gradually adapt a model to each user as accumulating his/her data.

## 6 Conclusion

In this study, we collected a set of respiration and acceleration data associated with subjective sleepiness levels using a novel wearable sensor. Considering the characteristics of the obtained dataset, we proposed combining a class imbalance technique and HMM with a supervised machine learning classifier. The performance of the models was examined through the experiment in a careful manner.

The major contribution of this study lies in our methodological approach for predicting the subjective sleepiness levels in daily living. We collected the physiological data of the subjects doing both active and passive tasks which represent activities in everyday life, unlike the prior studies which concentrated on driver’s sleepiness. Therefore, our dataset and derived models more precisely reflected the actual association of the physiological information and sleepiness in daily living. We found from our experiment that combining a class imbalance technique and HMM significantly enhance the prediction performance of a machine learning classifier. The F-measure of our best model was 0.7059, which was sufficient to capture the trend of sleepiness transition. Our procedure also has the potential to be utilized in a broad range of applications in the field of human mental state prediction because class imbalance and time series are the common properties of data in the field.

This study also has some limitations. First of all, only 18 subjects participated in the data collection. More samples are necessary for more accurate sleepiness prediction as well as better understanding of the association between physiological features and sleepiness. Secondly, there remain a lot of activities to consider other than typing and watching a movie. We need to investigate how breathing patterns and sleepiness

shift during different tasks in order to create a robust model as respiration can be affected by actions and feelings.

The future work may include a study on the relation between sleepiness and various performance indicators such as productivity, health, etc. since the wearable-sensor-based approach enables the long-term tracking of sleepiness without burdening users. We also need to establish a methodology for individual adaptation of a sleepiness prediction model because we found the interpersonal difference in the respiratory information and the self-report ratings made it difficult to build one high-performance predictor for all users. However, collecting the sufficient amount of labeled data is generally hard due to the high annotation cost. Some methods from weakly supervised learning [Zhou, 2017] and online learning will be helpful for this purpose.

## References

- [Åkerstedt and Gillberg, 1990] Torbjörn Åkerstedt and Mats Gillberg. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52 1-2:29–37, 1990.
- [Åkerstedt *et al.*, 2005] Torbjörn Åkerstedt, Björn Peters, Anna Anund, and Göran Kecklund. Impaired alertness and performance driving home from the night shift: a driving simulator study. *Journal of Sleep Research*, 14(1):17–20, 2005.
- [Åkerstedt *et al.*, 2014] Torbjörn Åkerstedt, Anna Anund, John Axelsson, and Göran Kecklund. Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function. *Journal of Sleep Research*, 23(3):242–254, 2014.
- [Bao and Intille, 2004] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In *International conference on pervasive computing*, pages 1–17. Springer, 2004.
- [Breiman, 2001] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.
- [Chang *et al.*, 2003] Edward Y. Chang, Beita Li, Gang Wu, and Kingshy Goh. Statistical learning for effective visual information retrieval. In *ICIP 2003*, volume 3, pages III–609, Sept 2003.
- [Chawla *et al.*, 2002] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(1):321–357, June 2002.
- [Galar *et al.*, 2012] Mikel Galar, Alberto Fernandez, Eudurne Barrenechea, Humberto Bustince, and Francisco Herrera. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 42(4):463–484, July 2012.
- [Hachisuka *et al.*, 2011] Satori Hachisuka, Kenji Ishida, Takeshi Enya, and Masayoshi Kamijo. Facial expression measurement for detecting driver drowsiness. In Don Harris, editor, *Engineering Psychology and Cognitive Ergonomics*, pages 135–144, 2011.
- [Hafner *et al.*, 2016] Marco Hafner, Martin Stepanek, Jirka Taylor, Wendy M. Troxel, and Christian Van Stolk. Why sleep matters – the economic costs of insufficient sleep: A cross-country comparative analysis. Research report, RAND Corporation, Santa Monica, CA, 2016.
- [Häkkinen *et al.*, 1999] Helinä Häkkinen, Heikki Summala, Markku Partinen, Mikko Tiihonen, and Jouni Silvo. Blink duration as an indicator of driver sleepiness in professional bus drivers. *Sleep*, 22(6):798–802, 1999.
- [Han and Wang, 2014] Kun Han and DeLiang Wang. Neural networks for supervised pitch tracking in noise. In *ICASSP 2014*, pages 1488–1492, May 2014.
- [Hido and Kashima, 2008] Shohei Hido and Hisashi Kashima. Roughly balanced bagging for imbalanced data. In *SDM08*, pages 143–152, 2008.
- [Hillman *et al.*, 2018] David Hillman, Scott Mitchell, Jared Streatfeild, Chloe Burns, Dorothy Bruck, and Lynne Pezzullo. The economic cost of inadequate sleep. *Sleep*, 41(8):zsy083, 2018.
- [Hoddes *et al.*, 1973] E. Hoddes, Vincent Zarcone, Hugh Smythe, Roger Phillips, and William C. Dement. Quantification of sleepiness: A new approach. *Psychophysiology*, 10(4):431–436, 1973.
- [Horne and Reyner, 1999] Jim Horne and Louise Reyner. Vehicle accidents related to sleep: a review. *Occupational and environmental medicine*, 56(5):289–294, 1999.
- [Igasaki *et al.*, 2016] Tomohiko Igasaki, Kazuki Nagasawa, Izzat Aulia Akbar, and Nao Kubo. Sleepiness classification by thoracic respiration using support vector machine. In *BMEiCON 2016*, pages 1–5, Dec 2016.
- [Johns, 1991] Murray W. Johns. A new method for measuring daytime sleepiness: the epworth sleepiness scale. *Sleep*, 14(6):540–545, 1991.
- [Kaida *et al.*, 2006] Kosuke Kaida, Masaya Takahashi, Torbjörn Åkerstedt, Akinori Nakata, Yasumasa Otsuka, Takashi Haratani, and Kenji Fukasawa. Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, 117(7):1574 – 1581, 2006.
- [Kaji *et al.*, 2018] Hirotaka Kaji, Hayato Yamaguchi, Kazuhide Shigeto, and Hirokazu Kikuchi. Wearable respiration sensor platform using ultrasound transducer. In *Adjunct Proceedings of ACM Ubicomp/ISWC*, 2018.
- [Khoshgoftaar *et al.*, 2011] Taghi M. Khoshgoftaar, Jason Van Hulse, and Amri Napolitano. Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 41(3):552–568, May 2011.
- [Lee *et al.*, 2016] Boon-Leng Lee, Boon-Giin Lee, and Wan-Young Chung. Standalone wearable driver drowsiness detection system in a smartwatch. *IEEE Sensors Journal*, 16(13):5444–5451, July 2016.
- [Monk, 1989] Timothy H. Monk. A visual analogue scale technique to measure global vigor and affect. *Psychiatry research*, 27:89–99, Feb 1989.

- [Plarre *et al.*, 2011] Kurt Plarre, Andrew Raji, Syed Monowar Hossain, Amin Ahsan Ali, Motohiro Nakajima, Mustafa Al'absi, Emre Ertin, Thomas Kamarck, Santosh Kumar, Marcia Scott, Daniel Siewiorek, Asim Smailagic, and Lorentz E. Wittmers. Continuous inference of psychological stress from sensory measurements collected in the natural environment. In *IPSN 2011*, pages 97–108, Apr 2011.
- [Platt, 2000] John Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In Alexander J. Smola, Peter Bartlett, Bernhard Schölkopf, and Dale Schuurmans, editors, *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, June 2000.
- [Pressman and Fry, 1989] Mark R. Pressman and June M. Fry. Relationship of autonomic nervous system activity to daytime sleepiness and prior sleep. *Sleep*, 12(3):239–45, July 1989.
- [Rabiner, 1989] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [Richard *et al.*, 2018] Alexander Richard, Hilde Kuehne, Ahsan Iqbal, and Juergen Gall. Neuralnetwork-viterbi: A framework for weakly supervised video learning. In *CVPR*, 2018.
- [Robertson *et al.*, 2012] David Robertson, Italo Biaggioni, Geoffrey Burnstock, Phillip A. Low, and Julian F. R. Paton, editors. *Primer on the Autonomic Nervous System, Third Edition*. Elsevier, 2012.
- [Saeys *et al.*, 2008] Yvan Saeys, Thomas Abeel, and Yves Van de Peer. Robust feature selection using ensemble feature selection techniques. In Walter Daelemans, Bart Goethals, and Katharina Morik, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 313–325, Berlin, Heidelberg, 2008. Springer.
- [Sforza *et al.*, 2002] E. Sforza, S. Grandin, Christophe Jouny, T. Rochat, and V. Ibanez. Is waking electroencephalographic activity a predictor of daytime sleepiness in sleep-related breathing disorders? *European Respiratory Journal*, 19(4):645–652, 2002.
- [Shimura *et al.*, 2018] Akiyoshi Shimura, Hideo Sakai, Yoshikazu Takaesu, Ryota Nomura, Yoko Komada, and Takeshi Inoue. Comprehensive assessment of the impact of life habits on sleep disturbance, chronotype, and daytime sleepiness among high-school students. *Sleep Medicine*, 44:12 – 18, 2018.
- [Tataraidze *et al.*, 2015] Alexander Tataraidze, Lesya Anishchenko, Lyudmila Korostovtseva, Bert Jan Kooij, Mikhail Bochkarev, and Yurii Sviryaev. Sleep stage classification based on respiratory signal. In *EMBC 2015*, pages 358–361, Aug 2015.
- [Vapnik, 1998] Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, 1998.
- [Vicente *et al.*, 2016] José Vicente, Pablo Laguna, Ariadna Bartra, and Raquel Bailón. Drowsiness detection using heart rate variability. *Medical & Biological Engineering & Computing*, 54:927–937, 2016.
- [Yuan *et al.*, 2013] George Yuan, Nicole A. Drost, and R. Andrew McIvor. Respiratory rate and breathing pattern. *McMaster University Medical Journal*, 10(1):23–28, 2013.
- [Zhou, 2017] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5(1):44–53, 2017.