Apnea and Sleeping-state Recognition by Combination Use of Open-air/Contact Microphones

Abidah Alfi Maritsa[1,a)] Ayumi Ohnishi[1,b)] Tsutomu Terada[1,c)] Masahiko Tsukamoto[1,d)]

Abstract: The increasing importance of sleeping quality and the awareness of sleep related disorders have led to emerging research in the field of wearable sensor with the purpose to make the sleep sensing become more comfortable and accessible. Among various methods, the use of audio sensor is known to be one of the most direct approach. However, further research using audio sensor to detect sleeping-state and apnea should be done to explore various new contexts and possibilities. Thus, this study proposes a wearable system consisting two microphones in the form of open-air microphone and contact microphone to improve the number of recognized contexts and this system is able to detect breathing, heartbeat, swallowing, body movement, and oral sound for the further use of sleeping-state and apnea severity detection. Audio data combination methods of Aggregation Methods and Stacking Methods were evaluated to improve the accuracy of the context detection. The Stacking Method with Support Vector Machine Polynomial Kernel as both first and second level classification resulted the best performance of 85.1% accuracy and 18% average improvement.

1. Introduction

As one-third of human life, sleep is having a major role to maintain health and function as human[1]. Lacking in quality of sleep correlated with several body dysfunction and in long term it is also causing serious physical and mental disorders. Several complications associated by this situation are cardiac and cerebrovascular problems[2], traffic accidents, mood swings, and depression. Data also showed that this lack of sleep quality related with financial also social impact due to deficient attention and premature judgements in many vital areas[3]

Research in sleep sensing field is emerging due to the increasing prevalence and influence of sleep disorders in daily life[4]. These research are mainly focusing in finding alternatives of diagnosis modalities in sleep disorder and sleep quality detection, since currently polysomnography (PSG) is still being the standardized procedure as detection in hospital[5]. This procedure is known to be complex, less comfortable, and expensive also time consuming. Thus, the urge for a wearable detection system in sleep sensing especially sleep disorder diagnosis has encouraged studies in this field. Imtiaz et al. (2021) in their review study of Sensing Technologies for Wearable Sleep Staging presented the information that the number of publications in wearable sleep sensing has been increasing since 2014[3].

Among overall sleep disorders and problems, breathing-related sleep disorder is one of the most concerning problems[4]. Sleep apnea is a major breathing-related sleep disorder indicated by pauses of breathing during sleep[1] which affects 22% of male population and 17% of female population[6]. However, this number might be larger as the data shown that 80% of apnea patients are remain undiagnosed[7] due to the low awareness of the disease and the complexity of the detection using PSG in hospitals[8].

Apnea events during sleep are commonly detected by analyzing several bio-signals such as respiratory, heartbeat, movement, snoring, oxygen saturation, pharyngeal movement, and more. Several approaches used by wearable devices are based on few-channel electrocardiogram (ECG), accelerometer, electromyogram (EMG), electrooculogram (EOG), electroencephalogram (EEG), oronasal airflow, audio sensors, and plethysmograph[3], [8]. How-

1 Kobe University
a) abidah-alfi@stu.kobe-u.ac.jp
b) ohnishi@eedept.kobe-u.ac.jp
c) tsutomu@eedept.kobe-u.ac.jp
d) tuka@kobe-u.ac.jp

ever, detecting breathing pattern using airflow is usually considered as the most straightforward approach for detecting breathing disorders including sleep apnea[8].

Detecting breathing pattern using airflow sensor might be the most straightforward approach but at the same time it can be considered as not a comfortable and wearable solution as its placement can block or disturb airways. Considering this reason, we saw the use of audio sensors as a more wearable approach with similar purpose.

Wearable devices to detect apnea and sleeping state using popular modalities such as accelerometer, ECG, EEG, and plethysmograph can only detect limited type of contexts (2 or 3) simultaneously[3]. The popular examples are the use of accelerometer to only detect movement or ECG to detect heartbeat. While actually there are various contexts that can be used to detect apnea and sleeping state[3], [8]. Audio sensors are known for the ability to obtain a lot of signal information. In the field of bio-signals and bio-contexts recognition, the use of a common open-air microphone is known for it's capability to detect loud human sounds such as breathing[8], body movement, and oral sounds such as snoring[3]. While in order to detect intrinsic body sounds such as pharyngeal movement or swallowing and heartbeat pattern, a modified approach of contact microphone can be seen as an alternative. Thus, combining more than one type of audio sensors can expand the number of recognized contexts.

Among 131 studies reviewed by Imtiaz et al. in 2021, only three of them mentioned the use of audio sensor. Two of the mentioned studies used a single microphone while the other one used a microphone and accelerometer to detect respiratory and body movement. However, none of the study mentioned in the paper used the combination use of multiple audio sensors[3].

Thus, this research proposes a wearable device for detecting human contexts related with sleep-related disorders especially sleep apnea using the combination of two audio sensors in the form of contact microphone and open air microphone. The resulting system is aimed not to only as alternative for sleep disorder detection but to monitor and improve human sleep quality in general.

The contributions of this study are as follows:

- This study proposed novel and multiple contexts then outputted the number of each context, thus this value can be used not only to detect apnea or other sleeping disorders but it can also be used in daily life to monitor sleep quality.
- Device presented in this research uses two different audio sensors in the form of open-air and contact microphones to detect the contexts.
- This paper compares three different methods to combine data of the two audio sensors to improve the accuracy of the context recognition.

2. Related Research

2.1 Sleep Quality and Sleep-Related Disorders

Recently, concern in maintaining sleep quality and diseases associated with sleep disorders are growing. This is due to the increasing awareness of the indication that the lack of sleep quality also sleeps disorders can lead to growing health burden in modern societies[12]. Lack of quality of sleep causes serious physical conditions such as cardiovascular disease, cognitive impairment, and metabolic dysfunction[1], [12]. Sleep quality also proven to be highly associated with serious neurological disorder and mental conditions. Disturbance in sleep quality can lead to severe mood swings and depressions. This condition also recently associated with financial and economic loss due to the high cost of medications, decreased productivity, road accidents, and premature decisions which caused by the lack quality of sleep.

Sleep disorders are conditions that caused sleep impairments and decreasing quality of sleep. Several major sleep disorders are insomnia, circadian rhythm disorders, breathing related sleep disorder, hypersomnia/narcolepsy, parasomnia, and restless legs syndrome/periodic limb movement[1]. Compared to the other sleep disorders, the prevalence of sleep-related breathing disorders has raised the most attention[4]. And among the sleep-related breathing disorders, Apnea or Sleep Apnea Syndrome is introduced as the most common sleep-related breathing disorder with an increasing prevalence over the last two decades[4].

2.2 Apnea and the Detection Approaches

Apnea or Sleep Apnea Syndrome is a condition indicated by disturbed breathing during sleep[13]. Apnea events are usually followed by arousals and consciousness also fluctuating heartbeat pattern and often associated with loud and frequent snoring during sleep[2]. One apnea event is described as a stop of breathing for more than 10s period and the severity of apnea is assessed based on the Apnea-Hypopnea Index (AHI) which categorized as mild (5-15 apnea events/hour), moderate (15-30 events/hour), and severe (>30 events per/hour)[14].

Along with the other sleep disorders, sleep apnea is also diagnosed in a certified sleep center or hospital with polysomnography (PSG) as the standardized procedure. Polysomnography consists nine sensor leads; three electroencephalography (EEG) leads, two electrooculography (EOG) leads, and three electromyography (EMG) leads, and a single electrocardiography (ECG) lead along with nasal pressure sensor, thermistor, and two belts with respiratory inductive plethysmography to record the respiration activity[4].

This procedure is known to be very uncomfortable, expensive, and non-accessible. Thus, various studies are conducted to find alternatives for simpler recognition of bio-contexts related to the disease detection adapting the mentioned sensors in a standard polysomnography procedure. Nasal pressure sensor, thermistor, and belts respiratory inductive plethysmography are used for detecting respiratory events and patterns to estimate the apnea severity using the mentioned AHI score[15]. This method of detecting breathing is known to be the most direct approach since apnea is a breathing related sleep disorders[8], however the use of mentioned sensors might be the most direct, but it is often not the most comfortable to use during sleeping.

The next modality and also one of the most popular to be used in wearable devices to detect apnea is the use of pulse oximetry to measure O_2 saturation[8] since it may fall dramatically during an apneic event due to the stop of breathing. The stop of breathing and apnea event causes arousal as mentioned previously, thus EEG, EOG, EMG, also accelerometer are used for detecting brain activity and sleep stages[3], [15]. Apnea events also indicated by the changes of heartbeat pattern due to respiratory changes. Another well know indication of sleep apnea is snoring, thus audio sensors are also used to detect snoring and it can also used to detect breathing pattern[4], [8]. Recent research by Yagi et al. in 2014 and Bhutada et al. in 2020 mentioned that swallowing or pharyngeal impairments presence in apnea patients[16], [17]. Yagi et al. concluded that swallowing frequency during sleep increases with the increasing of apnea severity [16] while Bhutada et al. mentioned in their review study that 65% (n =11) of the paper they reviewed revealed the presence of pharyngeal swallowing impairments in patients with apnea[17]. However, we have not found any study on wearable devices using this novel swallowing context to detect apnea or sleeping state [3], [4], [8].

Thus, contexts such as respiratory or breathing pattern, heartbeat, body movements, snoring or speaking sounds, oxygen saturation[8], also swallowing or pharyngeal activity[16], [17] are known to be related and useful to detect apnea and potentially sleeping state or other sleep related problems.

2.3 Multiple Sensors and Audio Sensor for Sleep Sensing and Apnea Detection

Among modalities and context, the use of EEG, accelerometer, pulse oximetry and ECG are mentioned to be the most used sensors in wearable devices to detect apnea and sleeping state[3], [8]. However, the mentioned sensors can only detect limited number of 2-3 contexts, while as mentioned before that there are various type of signals can be used to detect the problems. Mentioned approaches also did not address the most direct context of breathing or respiratory pattern. In order to improve the number of detected contexts, a multi-sensors approach was investigated. Some wearable devices using this approach was mentioned by Imtiaz et al. and Mendonça et al. in their study such the use of audio sensor along with accelerometer[3], [8], [18]. Study by Kalkbrenner et. al in 2019 used audio sensor and accelerometer for automated sleep stage classification resulting in 86.9% accuracy for sleep/wake classification. However, the targeted contexts of mentioned system was limited to detect cardiorespiratory pattern and body movement[18].

The use of audio sensor is known as a wearable alternative for in sleep sensing and apnea detection device. Audio data is also well-known to contain extensive amount of data and information. The use of common or open-air microphone is also known to be able to detect respiratory pattern[8], while the use of contact microphone can detect intrinsic biological sounds. Configurations and combinations of microphones can obtain various contexts which are important in sleep sensing such as breathing pattern, snoring and other voice such as speaking, movement, swallowing, and others[4]. Thus, the use of multiple audio sensors as mentioned before have the possibility to improve the number of detected contexts.

In the construction of audio based wearable system for sleep sensing and apnea detection, several parameters should be considered such as the most suitable placement and attachment of the device to the body[19], [20], [22], window segmenting sizes, feature values to be extracted[21], [22], and also the most effective machine learning algorithm as detection modality[8].

Placement and attachment of sensor to the body is essential especially in the use of audio sensors since they are directly affect the quality of the data. Positions around neck such as near carotid artery and suprasternal notch[22] are considered the suitable position for wearable devices to detect bio-signals such as tracheal sounds, breathing, snoring[23], swallowing sound, and even heartbeat[19], [22].

Once the signals of the body sounds are obtained, various feature extraction methods can be used as preprocessing before inputting the data to any artificial intelligence algorithm. Time domain, statistical features, and also frequency domain features often used as features[8]. Speech recognition based approach also considered in several audio context recognition in the field of sleep sensing and apnea detection. Thus, Mel' s frequency cepstral coefficient (MFCC) is also preferred to detect oral sounds such as snoring and speaking during sleep and also to detect breathing events with the accuracy of 81%[8], [9].

Various research compares sleep sensing and apnea detection using multiple machine learning algorithms. However, the use of simple algorithms such as support vector machines and random forest achieved the best output[8] with the accuracy of 75.76%[10] and 86.3%[11] respectively. Both are also popular to be implemented in wearable devices[8].

2.4 Combination of Multiple Sensors

The use of multiple audio sensors in wearable device to detect human context can capture a broad range of activities and even small gestures and very sensitive signals. The combination of the sensors using the correct method can increase the accuracy and improve the results of the system[24].

One of the common approach of the combination of multiple sensors in the field of wearable devices is Aggregation Method. Aggregation method combines extracted features from sensors data to construct the training and testing dataset for the classification model[24]. Though this method is the most common approach, Garcia-Ceja (2019) in their work mentioned that this approach may also not optimal due to the different statistical properties of each type of sensor. The same research proposed what the author mentioned as multi-view stacking, that can be simply explained as a multi-level classification. The method trained first-level learner for each sensor and then the outputs are trained using stacked generalization. However, the mentioned research used accelerometer and sound sensor in the field of Human Activity Recognition (HAR) to detect activities such as standing, sitting, walking, and lying[24] and the used of these approaches for the combination of audio sensors in sleep sensing and apnea detection hasn' t been evaluated yet[3], [8]. On the other hand, research by Valipour et al. in 2017 used two audio sensors to detect vital signs of heartbeat and respiratory with the purpose of general health monitoring, however this research did not do any combination approach nor specified their purpose with more number of contexts[25].

Thus, this study would like to evaluate the use of multiple audio sensors as wearable device in recognizing contexts related with apnea and sleeping monitor. This research also inspected the use of several methods to combine data from the sensors and their contribution in improving the detection accuracy.

3. Proposed Method

This research constructed and evaluated a wearable device to detect human contexts related with apnea detection and sleep monitoring using multiple audio sensors. This research compared several parameters to build the most suitable design of the device and models for the system. And lastly, this study also adopted and compared several fusion methods to combine data from the sensors and evaluates the effect towards the accuracy of the system.

3.1 Preliminary Experiment and Prototype Construction

In the initial stage of the research, this study carried out a preliminary experiment to examine the most suitable positions to attach the device in terms of wear-ability and the quality of acquired data. Three positions were compared: a) position A=neck under the jaw area (around the *carotid artery*), b) position B= at the lower part of neck, between the clavicular bone (*suprasternal notch* area), and c) position C = the left chest area.

Audio data from the mentioned positions were collected using two different types of microphones: an open-air microphone and a contact microphone. Both of the microphones used in experiment were commercial KY-038 microphones, however one of the microphone is modified and connected to a chest piece of stethoscope to be functioning as the contact microphone.

A five minutes' length recording data were obtained from one subject. The data was separated into 4:1 training and testing data. Then the data was preprocessed using 3s segmenting with 0.25 step length with the sampling frequency of 44.1 kHz. Features extracted in this

図 1: Proposed Device

preliminary experiment was 10 FFT peaks, mean, variance, range, and energy value. Then a simple evaluation in Weka and MATLAB using SVM was conducted. During the preliminary experiment, three initial contexts were detected such as heart-beat, breathing, and swallowing context. For the heart-beat detection, the positions A, B, and C obtained the accuracy of 95.8%, 96.5%, and 95.9% respectively. While both breathing and swallowing context accuracy detection were 65.4%, 75.9%, and 36.9% respectively. Based on the mentioned result, Position B (*suprasternal notch*) was chosen as the most suitable position to attach the device.

A wearable device design was constructed as shown in Figure 1 following result of the preliminary experiment. 3D printed case was designed to maintain both of the microphones compact. The case was printed using acrylonitrile butadiene styrene (ABS) material and the printed case was then polished and smoothed so that it will be more comfortable for the user to wear. The device is connected to PC through two different audio channels for each microphone.

3.2 Experiment

Following the construction of the wearable device, experiments were conducted using the device. Data from seven healthy people in their 20s and 30s were collected during the experiment. The purpose of this experiment was to collect audio data for the construction of the prediction/classification model for the system to predict bio-contexts related to sleep sensing and apnea.

A 10 minutes' experiment was conducted for each subject. The participants were asked to lay down in order to mimic a sleeping position while the device was attached as shown in Figure 2 and recording their data. Supplementary audio and video data recorded using professional microphone and video camera also collected to construct ground-truth label.

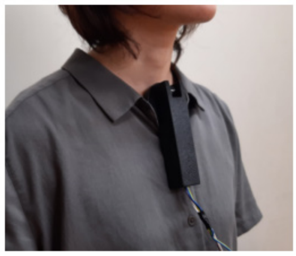

図 2: Device Attachment

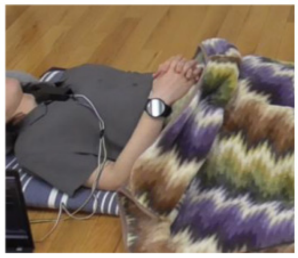

図 3: Experiment Setting

The experiment for each subject with the environment as shown in Figure 3 was divided into three parts which were: a) Minutes 0-8: participants were asked to perform natural breathing and periodic swallowing, b) Minutes 8-9: participants were asked to perform body movements such as rolling to the left and right, c) Minutes 9-10: participants were asked to perform oral sounds to mimic snoring and sleeping sound.

3.3 Data Analysis

Data acquired from the experiment was then processed and several aspects were compared to obtain the most suitable parameters for the system such as window sizes, feature values, and machine learning algorithms.

Initially, three window sizes were examined. The 10 minutes' data with frequency sampling of 44.1 kHz were segmented using three different length of window sizes of 1-,2-,3-s lengths with a step size of 0.25. Then, two different feature values were compared which were statistical-frequency domain features and MFCC features. The first type of feature consisted of 10 FFT peak values, 1 maximum amplitude, 1 range, 1 variance, and 1 energy value of each segment. While the MFCC features were 13 MFCC features and 1 log energy value. Both of types of the features were extracted in MATLAB programming environment.

The dataset was then separated into 4:1 training data and testing data to detect contexts labelled as breathing, others, swallowing, movement, and oral sound. In order to detect mentioned contexts, the prepared training data and testing data were evaluated in the WEKA program to compare two different machine learning algorithms which are the Support Vector Machine (SVM) and Random Forest algorithm. While the raw heartbeat signal was processed using peak detection to determine the number of heartbeat per minute also the status of heartbeat such as

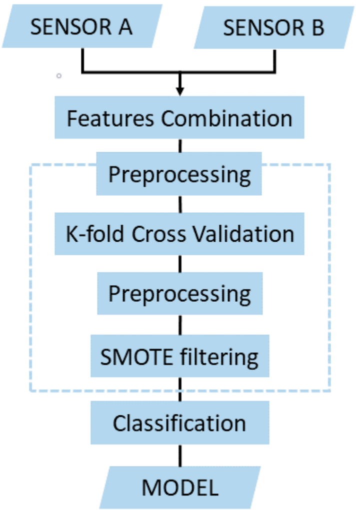

図 4: Aggregate Method Structure

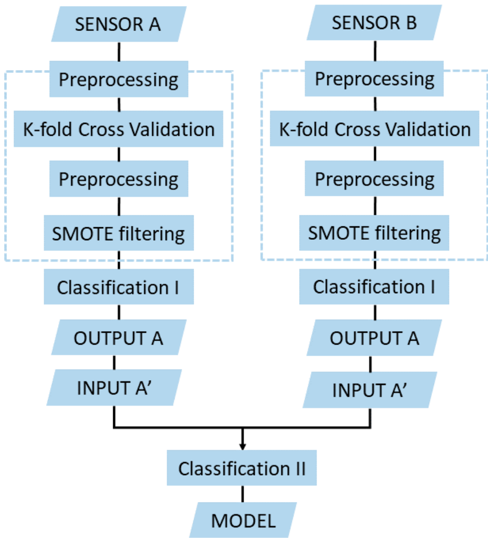

図 5: Stacking Method Structure

low, normal, or high.

3.4 Combination of Multiple Microphones Data

This study selected and modified two signal fusion methods based on the references to construct prediction based on the combination of data obtained from the two microphones used in the device, contact microphone and open air microphone. The methods evaluated in this study was classified into Aggregation Method and Stacking Methods. The second method is divided into Decision Tree-based Stacking Method and SVM-based Stacking Method. This process will be limited on evaluating the optimum parameters obtained in the previous data analysis.

3.4.1 Aggregation Method

Aggregation Method combined feature values of both Sensor A and Sensor B, open air microphone as Sensor A and contact microphone as Sensor B. The flow of this method can be seen in Figure 4. Following the combination of features from both sensors, the data is then preprocessed. In the preprocessing stage, this research used K-fold Cross Validation with k=2 followed by SMOTE filtering to prevent bias during the classification process. The preprocessed data was then classified using SVM-polynomial kernel. The preprocessing and classification processed were carried out in WEKA programming environment.

3.4.2 Stacking Method

Stacking Method was done by evaluating the data using two level classification process as seen in Figure 5. Data from each sensor was processed separately to output the first-level classification output. This classification result of each microphone was then combined and used as the input of the second-level classification. Before inputing the data into the first-level classification, the data was preprocessed using the same approach as the aggregation method which are K-fold Cross Validation with k=2 and SMOTE filtering.

This data was then inputted into the first-level of classification. The first-level classification algorithm was selected based on the optimum classifier obtained from the previous uncombined data analysis. The result of the first-level classification or *Output A* and *Output B* was then inputted into the second-level classification as *Input A'* and *Input B'*. Both preprocessing and classification method were performed in WEKA programming environment.

The proposed stacking method was then comparing models and results obtained from two the types of algorithm as the second-level classifier. Firstly, decision tree or rule based classification was selected as classification algorithm. J48 Decision Tree was used in WEKA to classify each training data. The second type selected algorithm was SVM-polynomial kernel.

Accuracy of the resulting data from both classifiers were then compared along with the comparison of all proposed combination methods to the non-combined results.

4. Results and Discussion

4.1 Optimum Parameters

Different parameters were compared and combined to obtain the optimum characteristics to be applied in the future system. Data from seven subjects which were ac-

表 1: Context Detection Accuracy

Index	Context	Open-air Mic [%]	Contact Mic [%]
1	Breath	71.9	59.2
2	Others	21.1	59.5
3	Swallowing	39.7	71.5
4	Movement	92.1	95.9
5	Oral Sound	92.3	86.9

quired in the experiment were analyzed. This analysis compared window sizes, feature values, classification algorithms, and also the performance of each type of microphones to detect the desired contexts. Processes held in MATLAB and WEKA programming environment resulted in the higher overall performance of contact microphone to detect the contexts when compared with the open-air microphone with the highest accuracy of 76.9% for the contact microphone while the open-air microphone resulted best accuracy of 65.4%.

The combination of data obtained with contact microphone, 1s window segmenting, extracted MFCC features, and SVM Polynomial Kernel resulted in better classification of overall data from all the seven subjects. While the combination of 3s window segmenting, extracted MFCC features, and SVM Polynomial Kernel performed better result to classify the open-air microphone data. Thus, based on the mentioned result it was assumed that mentioned combinations performed better to classify data from each microphone.

Each context accuracy evaluation were performed as shown in Table 1. Table 1 shows that despite the higher overall detection accuracy of the contact microphone data, the open-air microphone obtained better performance to detect breathing and oral sound contexts with both 71.9% and 92.3% accuracy. This result indicating that the performance and accuracy of the system might be improved by combining both microphones. Another phenomenon is that body movement sound noises affect the input of the open-air microphone, however it does not affect the contact microphone so much. Since some types of noises affect one of the microphones, combining the microphones can be be an effective alternative. Thus, in the next stage of the study this research evaluated different combination methods for the data from two microphones with aim to improve the context detection accuracy. The result of this optimum parameters analysis was used as the basis for the combination methods.

4.2 Combination Methods Result

This research evaluated two methods to combine audio data from two microphones with objective to improve the accuracy of the system. Methods used in this evaluation was based on the existing sensor combination approaches as mentioned in the previous related research explanations[24]. Data and parameter combinations used in this evaluation was based on the previous optimum parameters analysis. The previous evaluation confirmed following best parameters for open-air microphone data (Sensor A): 3s windowing size, 28 MFCC Features, and SVM Polynomial Kernel. While for contact microphone data (Sensor B): 3s windowing size, 28 MFCC Features, and SVM Polynomial Kernel.

4.2.1 Aggregation Method

In the evaluation using aggregation method, features from both sensors were combined resulting in the total of 56 features from both of the sensors data. In the dataset preparation, even though 1s windowing performed better for Sensor B data, the MFCC features extracted from the 3s segmented Sensor B data was still combined with the features obtained from the 3s segmented Sensor A data and vice versa.

The 2 fold cross validation process resulted in two different sets of training and test set. The training set of each data then processed with SMOTE filtering resulting in the same number of data for each classes. Training and testing data then classified using SVM Polynomial Kernel classifier due to the better performance of this classifier in the previous evaluation. Average of result obtained two different sets of training and test set for each trial were calculated.

Table 2 presented the result of the aggregation method. Model A and Model B were obtained from the 2-fold cross validation process and the average of these two models was calculated. This average score was then compared to the previous accuracy (Prev. Acc. in Table 2) which was obtained in the previous evaluation of uncombined data using the related parameters. The difference between the average accuracy using aggregation method with the previous accuracy was formulated in the delta score.

From the Table 2, it is shown that the method improved 10 out of the 14 trials performed in this evaluation with the average delta of 11.4%. Using the proposed method, 4 out of 14 accuracy of the data was decreased by the average of 15.3%. Average accuracy for each window size of 1s and 3s also experiencing improvement of 59.4% and 58.8% respectively, compared to the previous accuracy of

表 2: Aggregation Method Accuracy

Sub.	Win. Size	Model A	Model B	Avg. Models	Prev. Acc.	Delta
A	1	63.5	74.4	69.0	22.9	46.1
	3	58.8	60.6	59.7	26.7	33.0
B	1	61.9	75.0	68.5	66.2	2.3
	3	58.4	71.4	64.9	59.3	5.6
C	1	68.6	70.8	69.7	62.7	7.0
	3	61.4	73.0	67.2	60.9	6.3
D	1	22.8	20.3	21.6	52.2	-30.7
	3	32.8	63.5	48.2	58.2	-10.1
E	1	63.3	63.2	63.3	72.0	-8.8
	3	44.7	55.7	50.2	61.9	-11.7
F	1	66.5	71.9	69.2	67.9	1.3
	3	62.7	63.6	63.2	59.5	3.7
G	1	52.6	57.1	54.9	52.2	2.7
	3	57.1	59.5	58.3	52.7	5.6

表 3: Staking Method: Rule-based Classifier Accuracy

Sub.	Win. Size	Model A	Model B	Avg. Models	Prev Acc.	Delta
A	1	83.8	83.3	83.6	22.9	60.7
	3	75.1	74.7	74.9	26.7	48.2
B	1	70.2	60.8	65.5	66.2	-0.7
	3	79.1	73.8	76.5	59.3	17.2
C	1	71.1	79.1	75.1	62.7	12.4
	3	66.9	78.5	72.7	60.9	11.8
D	1	66.1	46.6	56.4	52.2	4.1
	3	66.1	69.9	68.0	58.2	9.8
E	1	54.2	72.1	63.2	72.0	-8.9
	3	74.7	66.2	70.5	61.9	8.6
F	1	73.1	68.8	71.0	67.9	3.0
	3	56.3	72.1	64.2	59.5	4.7
G	1	70.6	75.5	73.1	52.2	20.9
	3	70.7	70.1	70.4	52.7	17.7

56.5% and 54.2%.

4.2.2 Stacking Method: Rule-based Classification

Stacking method with rule-based classifier of J48 Decision Tree was performed as the second level classification process. Each sensor datasets with MFCC features were preprocressed with k-fold cross validation and SMOTE filtering and then inputted into the SVM algorithm as the first level classification. This process resulted in output in the format of predictions (1: breathing, 2: others, 3: swallowing, 4: movement, 5: oral sound) and the new dataset was prepared using five previous predictions from each sensors as features. Thus, the new dataset consisted of 10 features values: 5 features from Sensor A and 5 features from Sensor B. The data then inputted into the second level classifier resulting in two models for each trials. The same data analysis process as the previous aggregation method was performed for the result of this method.

Rule-based Classifier of J48 Decision Tree was used as the second classifier and result as presented in Table 3 was obtained. The proposed rule-based stacking method improved 12 out of 14 trials with the average improvement of 18.3% while the rest of the data experienced average of 4.8% loss. Both improvement and loss of this method were superior compared with the previous aggregation method. Average accuracy for each window size of 1s and 3s also experiencing improvement of 69.7% and 71% respectively which were better when compared with previous proposed method and the uncombined result.

表 4: Staking Method: SVM-based Classifier Accuracy

Sub.	Win. Size	Model A	Model B	Avg. Models	Prev. Acc.	Delta
A	1	83.7	85.1	84.4	22.3	62.1
	3	73.5	77.2	75.4	26.7	48.7
B	1	71.7	75.0	73.4	66.2	7.15
	3	76.5	77.7	77.1	59.3	17.8
C	1	70.0	77.9	74.0	62.7	11.3
	3	66.7	76.9	71.8	60.9	10.9
D	1	65.6	59.6	62.6	52.2	10.4
	3	66.7	69.5	68.1	58.2	9.9
E	1	74.3	68.6	71.5	72.0	-0.6
	3	54.8	66.7	60.8	61.9	-1.2
F	1	72.7	74.9	73.8	67.9	5.9
	3	69.5	71.8	70.7	59.5	11.2
G	1	69.5	74.4	72.0	52.2	19.8
	3	70.7	37.5	54.1	52.7	1.4

4.2.3 Stacking Method: SVM-based Classification

The same process as the previous rule-based stacking method were performed while in this proposed method, SVM polynomial kernel was used as the second-level classifier. This proposed method resulted the best accuracy compared with the previous evaluated methods. The highest average accuracy of 84.4% as shown in Table 4 and the average improvement of 18% were obtained. Both stacking methods have relatively similar average improvement of accuracy while the third method result in minimum loss of -0.85% which was the lowest out of all proposed methods. Average accuracy for each window size of 1s and 3s also experiencing improvement of 73.1% and 68.3% respectively. Table 2, Table 3, and Table 4 showed that data obtained from Subject A achieved the best accuracy and highest improvement using the proposed meth-

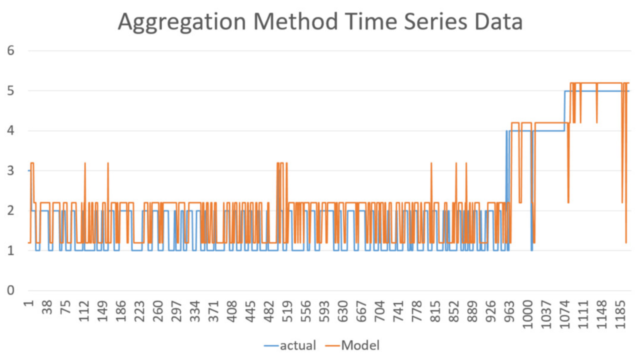

図 6: Ground Truth and Aggregation Method Time Series Result with 1:breathing, 2:others, 3:swallowing, 4:movement, and 5:oral sound

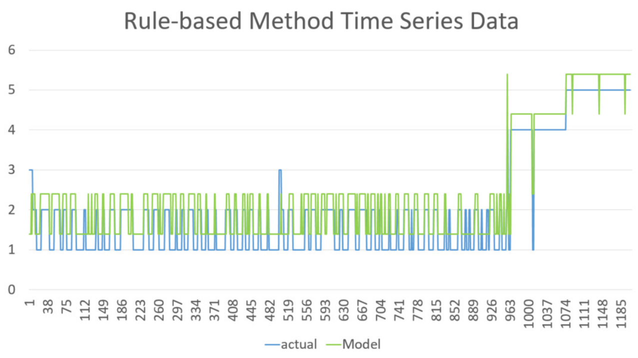

図 7: Ground Truth and Rule-based Method Time Series Result with 1:breathing, 2:others, 3:swallowing, 4:movement, and 5:oral sound

ods. Data from Subject A resulted in 74.4%, 83.8%, and 85.1% accuracy for all methods respectively with 46.6%-62.1% improvement while the previous accuracy show that the specific data have the worst value. This result shows that the proposed method can improve low accuracy data. In the opposite, data from the Subject E marked the best accuracy in the previous uncombined approach, however the specific data experienced loss using the proposed methods. Thus, it can be concluded that the combination approach does not support the detection improvement for this specific data.

Figure 6, 7, and 8 compared the time series result of each methods . From the figure, it can be seen that the SVM-based stacking method resulted the best prediction for most of the contexts.

5. Conclusion and Future Work

This research constructed a wearable device and evaluated methods to detect contexts for sleeping-state sensing and apnea detection. Breathing, others (non-breathing state), heartbeat, swallowing, movement, and oral sound

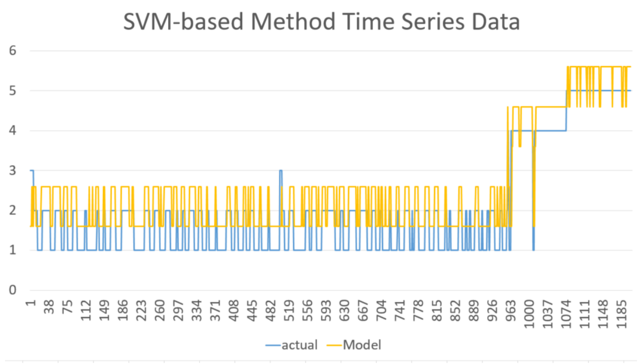

図 8: Ground Truth and SVM-based Method Time Series Result with 1:breathing, 2:others, 3:swallowing, 4:movement, and 5:oral sound

contexts were detected. The initial data analysis resulted in the optimum parameters of window sizes, feature values, and classification algorithms for each microphones. The contact microphone data resulted in better classification of 76.9% accuracy. However, there was a possibility of improving the system accuracy by combining both microphones data. Thus, this research proposed combination methods of aggregation method and stacking methods. The stacking methods have two types of approach which was rule-based classifier of J48 Decision Tree and SVM-based classifier as the second-level classification algorithm. The proposed methods resulted in the improvement of the overall accuracy of 11.4-18.3% with the last method of SVM-based stacking method marked the best accuracy of 85.1% and 62.1% best improvement.

Comparing this research result to the previous studies, this research was able to detect more number of contexts (six contexts) compared to the mentioned studies in both apnea and sleeping-state detection field [3], [8]. This study evaluated several aspects such as windowing sizes, feature values, and classification algorithms for this specific proposed system. Thus, even though previous related studies have evaluated such parameters with accuracy as mentioned in Section 2.3, it is hard to compare the result obtained in this study with previous research since the evaluated parameters were specific to the system constructed in this study. Then, this study also evaluated several combination methods of both microphone to investigate whether the mentioned methods can improve the accuracy of the contexts detection. This combination use of multiple audio sensors are novel [3], [8] with the result that the approach obtained more contexts and improved detection accuracy. A second prototype with a more comfortable and smaller design will be developed to improve the wear-ability of the device. Sleeping environment experiments will also be held to analyze the performance of the system.

謝辞 This work was supported by JST CREST Grant Number JPMJCR18A3, Japan.

参考文献

[1] M. K. Pavlova and V. Latreille: Sleep Disorders, *American Journal of Medicine*, Vol. 132, No. 3, pp. 292–299 (Mar. 2019).

[2] K. Zhu, M. Li, S. Akbarian, M. Hafezi, A. Yadollahi, and B. Taati: Vision-Based Heart and Respiratory Rate Monitoring during Sleep – a Validation Study for the Population at Risk of Sleep Apnea, *IEEE Journal of Translational Engineering in Health and Medicine*, Vol. 7, pp. 1–8 (Oct. 2019).

[3] S. Imtiaz: A Systematic Review of Sensing Technologies for Wearable Sleep Staging, *Sensors*, Vol. 21, No.5, pp. 1–21 (Feb. 2021).

[4] T. Penzel, C. Schöbel, and I. Fietze: New Technology to Assess Sleep Apnea: Wearables, Smartphones, and Accessories, *F1000Research*, Vol. 7, pp. 1–12 (Mar. 2018).

[5] H. Nakano, M. Hayashi, E. Ohshima, N. Nishikata, and T. Shinohara: Validation of a New System of Tracheal Sound Analysis for the Diagnosis of Sleep Apnea-hypopnea Syndrome, *Sleep*, Vol. 27, No. 5, pp. 951–957 (Aug. 2004).

[6] K. A. Franklin and E. Lindberg: Obstructive Sleep Apnea is A Common Disorder in The Population—A Review on the Epidemiology of Sleep Apnea, *Journal of Thoracic Disease*, Vol. 7, No. 8, pp. 1311–1322 (Aug. 2015).

[7] I. Fietze, N. Laharnar, A. Obst, R. Ewert, S. B. Felix, C. Garcia, S. Gläser, M. Glos, C. O. Schmidt, B. Stubbe, H. Völzke, S. Zimmermann, and T. Penzel: Prevalence and Association Analysis of Obstructive Sleep Apnea with Gender and Age Differences—Results of Ship-trend, *Journal of Sleep Research*, Vol. 28, No. 5, pp. 1–9 (Oct. 2018).

[8] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel: A Review of Obstructive Sleep Apnea Detection Approaches, *IEEE Journal of Biomedical and Health Informatics*, Vol. 23, No. 2, pp. 825–837 (Apr. 2018).

[9] R. F. Pozo, J. L. B. Murillo, L. H. Gomez, E. L. Gonzalo, J. A. Ramirez, and D. T. Toledano: Assessment of Severe Apnoea through Voice Analysis, Automatic Speech, and Speaker Recognition Techniques, *EURASIP Journal on Advances in Signal Processing*, Vol. 2009, No. 1, pp. 1–11 (June 2009).

[10] T. Praydas, B. Wongkittisuksa, and S. Tanthanuch: Obstructive Sleep Apnea Severity Multiclass Classification Using Analysis of Snoring Sounds, *The 2nd World Congress on Electrical Engineering and Computer Systems and Science (EECSS'16)*, pp. 16–20 (Aug. 2016).

[11] T. Rosenwein, E. Dafna, A. Tarasiuk, and Y. Zigel: Breath-by-breath Detection of Apneic Events for OSA Severity Estimation using Non-contact Audio Recordings, *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Vol. 2015, pp. 7688–7691 (Aug. 2015).

[12] N. Darchia, N. Oniani, and I. Sakhelashvili: Relationship between Sleep Disorders and Health Related Quality of Life—Results from the Georgia SOMNUS Study, *International Journal of Environmental Research and Public Health*, Vol. 15, No. 8, pp. 1–15 (July 2018).

[13] A. Ramachandran and A. Karuppiah: A Survey on Recent Advances in Machine Learning based Sleep Apnea Detection System, *Healthcare*, Vol. 9, No. 7, pp. 1–19 (July 2021).

[14] A. M. Osman, S. G. Carter, J. C. Carberry, and D. J. Eckert: Obstructive Sleep Apnea: Current Perspectives, *Nature and Science of Sleep*, Vol. 10, pp. 21–34 (Jan. 2018).

[15] M. Shokoueinejad, C. Fernandez, E. Carroll, F. Wang, J. Levin, S. Rusk, N. Glattard, A. Mulchrone, X. Zhang, A. Xie, M. Teodorescu, J. Dempsey, and J. Webster: Sleep Apnea: A Review of Diagnostic Sensors, Algorithms, and Therapies, *Physiological Measurement*, Vol. 38, No. 9, pp. R204–R252 (Aug. 2017).

[16] K. Yagi, A. A. Lowe, N. T. Ayas, J. A. Fleetham, and F. R. Almeida: Swallowing and Breathing Patterns during Sleep in Patients with Obstructive Sleep Apnea, *Sleep and Breathing*, Vol. 19, No. 1, pp. 377–384 (July 2014).

[17] A. M. Bhutada, W. A. Broughton, and K. L. Garand: Obstructive Sleep Apnea Syndrome (OSAS) and Swallowing Function—A Systematic Review, *Sleep and Breathing*, Vol. 24, No. 3, pp. 791–799 (Feb. 2020).

[18] C. Kalkbrenner, R. Brucher, T. Kesztyüs, M. Eichenlaub, W. Rottbauer, and D. Scharnbeck: Automated Sleep Stage Classification Based on Tracheal Body Sound and Actigraphy, *GMS German Medical Science*, Vol. 17, No. 2, pp. 1–12 (Feb. 2019).

[19] A. K. Sabil, M. Glos, A. Günther, C. Schöbel, C. Veauthier, I. Fietze, and T. Penzel: Comparison of Apnea Detection Using Oronasal Thermal Airflow Sensor, Nasal Pressure Transducer, Respiratory Inductance Plethysmography, and Tracheal Sound Sensor, *Journal of Clinical Sleep Medicine*, Vol. 15, No. 2, pp. 285–292 (Feb. 2019).

[20] S. Li, B. Lin, C. Tsai, C. Yang, and B. Lin: Design of Wearable Breathing Sound Monitoring System for Real-time Wheeze Detection, *Sensors*, Vol. 17, No. 1, pp. 1–15 (Jan. 2017).

[21] C. Avci and A. Akbaş: Sleep Apnea Classification based on Respiration Signals by Using Ensemble Method, *Bio-Medical Materials and Engineering*, Vol. 26, No. s1, pp. S1703–S1710 (Sep. 2015).

[22] C. Kalkbrenner, M. Eichenlaub, and R. Brucher: Development of a New Homecare Sleep Monitor using Body Sounds and Motion Tracking, *Current Directions in Biomedical Engineering*, Vol. 1, No. 1, pp. 30–33 (Sep. 2015).

[23] W. Shi, B. Xue, S. Guo, D. Y. T. Goh, and W. Ser: Obstructive Sleep Apnea Detection using Difference in Feature and Modified Minimum Distance Classifier, *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–4 (Oct. 2018).

[24] E. Garcia-Ceja, C. Galván-Tejada, and R. Brena: Multi-view Stacking for Activity Recognition with Sound and Accelerometer Data, *Information Fusion*, Vol. 40, pp. 45–56 (Mar. 2018).

[25] A. Valipour and R. Abbasi-Kesbi: A Heartbeat and Respiration Rate Sensor Based on Phonocardiogram for Healthcare Application, *25th Iranian Conference on Electrical Engineering (ICEE)*, pp. 45–48 (May 2017).