IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 28, 2021, accepted July 31, 2021, date of publication August 16, 2021, date of current version August 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3104805

Detecting Drowsy Learners at the Wheel of e-Learning Platforms With Multimodal Learning Analytics

RYOSUKE KAWAMURA[1], SHIZUKA SHIRAI[2], (Member, IEEE), NORIKO TAKEMURA[3], MEHRASA ALIZADEH[2], MUTLU CUKUROVA[4], HARUO TAKEMURA[2], (Member, IEEE), AND HAJIME NAGAHARA[3], (Member, IEEE)

[1]Fujitsu Ltd., Kawasaki, Kanagawa 211-8588, Japan
[2]Cybermedia Center, Osaka University, Osaka 560-0043, Japan
[3]Institute for Datability Science, Osaka University, Osaka 565-0871, Japan
[4]Institute of Education, University College London, London WC1H 0AL, U.K.

Corresponding author: Ryosuke Kawamura (k.ryosuke@fujitsu.com)

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) ''Innovation Platform for Society 5.0'' Program under Grant JPMXP0518071489.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Faculty Meeting of Cybermedia Center, Osaka University, under Application No. 2019-4.

ABSTRACT Learners are expected to stay wakeful and focused while interacting with e-learning platforms. Although wakefulness of learners strongly relates to educational outcomes, detecting drowsy learning behaviors only from log data is not an easy task. In this study, we describe the results of our research to model learners' wakefulness based on multimodal data generated from heart rate, seat pressure, and face recognition. We collected multimodal data from learners in a blended course of informatics and conducted two types of analysis on them. First, we clustered features based on learners' wakefulness labels as generated by human raters and ran a statistical analysis. This analysis helped us generate insights from multimodal data that can be used to inform learner and teacher feedback in multimodal learning analytics. Second, we trained machine learning models with multiclass-Support Vector Machine (SVM), Random Forest (RF) and CatBoost Classifier (CatBoost) algorithms to recognize learners' wakefulness states automatically. We achieved an average macro-F1 score of 0.82 in automated user-dependent models with CatBoost. We also showed that compared to unimodal data from each sensor, the multimodal sensor data can improve the accuracy of models predicting the wakefulness states of learners while they are interacting with e-learning platforms.

INDEX TERMS Drowsiness, online education, e-learning platforms, multimodal learning analytics, physical learning analytics.

I. INTRODUCTION

In recent years, e-learning has grown rapidly due to advances in information technologies. Online education is becoming mainstream worldwide thanks to the development of underlying ICT which has made it feasible technologically, economically, and operationally [1]. For instance, according to a report on distance education in American universities, from 2002 to 2012 both distance and overall enrollments in higher education institutions witnessed annual growth, yet since 2012 distance education has grown in demand steadily despite overall enrollment decline [2]. Moreover, in response to the COVID-19 pandemic, there is a burgeoning interest in e-learning and distance learning, and many schools and universities all over the world have adopted remote learning and teaching solutions [3], [4]. Despite the increasing demand for e-learning globally, there are still significant issues that lead to high dropout rates in online courses. Studies reveal that student dropout rates in e-learning are significantly higher than in traditional learning contexts [5]–[7].

The associate editor coordinating the review of this manuscript and approving it for publication was Chia-Wen Tsai.

High dropout rates in e-learning are caused by various factors. Lee and Choi [8] have reviewed studies conducted over 10 years on factors leading to dropouts in online courses and have classified them into three categories that affect learners' decision to drop out: *student factors*, *course/program factors*, and *environmental factors*. Student factors include psychological and behavioral attributes such as motivation and satisfaction during students' interaction with e-learning platforms. These measures and similar ones are difficult to identify in e-learning settings with traditional learning analytics approaches that are limited to the use of log data. For instance, learner wakefulness during e-learning is vital to effective engagement, which in turn impacts educational attainment [9]. However, mere clickstream data, widely deployed in learning analytics studies, is not sufficiently accurate to detect student wakefulness. On the other hand, the emerging area of multimodal learning analytics [10] is providing new methods and innovations to leverage and make sense of physical data to interpret constructs that are hard to measure and support with traditional learning analytics approaches.

In this paper, we present our approach to measuring learners' wakefulness as they interact with an e-learning platform using multimodal learning analytics. More specifically, we explore the following research questions.

- RQ1 What factors from multimodal data collected can be used to inform teacher and learner dashboards for feedback and reflection on learners' wakefulness?
- RQ2 To what extent can students' wakefulness be detected with the help of multimodal data from their heart rate, seat pressure, and facial expressions?

The current study contributes to the existing literature in the following ways: (1) The results of the first research question can have implications for the development of teacher and learner dashboards to help generate explainable insights into students' wakefulness states. (2) The findings of the second research question can have immediate implications for adaptive e-learning platforms to provide appropriate feedback and notifications to enhance students' learning experience. (3) The findings of both research questions have direct implications for student engagement and their learning outcomes in e-learning settings.

The paper is structured into 6 sections. Section II describes the motivation of the study and the literature on learner engagement and wakefulness detection. Section III describes the methodology we have adopted, and section IV presents the findings. The results are discussed in section V, which is followed by a conclusion and future work.

II. BACKGROUND AND RELATED WORK

In face-to-face learning contexts, teachers can often identify learners' internal states, such as engagement, boredom, concentration, and wakefulness, and are able to adapt their pedagogical approaches and teaching materials/methods accordingly. In e-learning settings, however, real-time adaptation is not easy due to physical distance and asynchrony. To address this problem, prediction methods are used to infer a variable based on other variables extracted from learning log data, such as response times to predict engagement [11] and conversational cues to predict affective states of boredom, confusion, flow, and frustration [12]. However, due to the complexities of learning processes and the non-linear relationships between observed and reported cognitive and emotional states, exclusively relying on log data to decipher complex internal states of learners is insufficient [13], [14]. To validate findings from log data, new technologies such as biometric sensors are used to unobtrusively gather physiological data from learners, but systems that use such data to monitor and predict learners' internal states as inferred from their physiological signals are still under development.

One aspect of learners' psychophysiological states that has garnered significant attention is engagement. Fredricks and colleagues have identified three forms of engagement, namely behavioral, emotional, and cognitive [15]. Behavioral engagement is associated with learners' participation and involvement in the learning process. Emotional engagement, as the name suggests, refers to learners' emotional attitudes towards teachers, peers, and learning, and cognitive engagement involves components that foster the learning process such as focused attention, memory, and creative thinking [16]. Wakefulness during online coursework can be considered an instance of behavioral engagement and thus essential to ensure improved learning outcomes. However, this type of behavioral engagement is largely under-researched in learning contexts.

There is very limited research specifically measuring students' wakefulness with multimodal learning analytics during their engagement with e-learning platforms. For example, in the context of writing tasks, [17] deployed a video-based estimation approach to investigate learners' engagement with the platform. In particular, they analyzed facial expression data as captured by Microsoft Kinect Face Tracker and heart rate measures from video-based sensing, using photoplethysmography. They were able to estimate engagement with a high level of accuracy as measured against concurrent and retrospective learner self-reports. Area under the ROC Curve (AUC) was used to evaluate classifier accuracy, and AUC = .758 for concurrent annotations and AUC = .733 for retrospective annotations were achieved. Despite the fact that fusion of multimodal data yielded overall best results, face tracking data alone was shown to be the best indicator. In another study, [18] examined students' drowsiness in online classes using a smart chair with a pressure sensor. There was no robust evaluation of the results; however, an average accuracy of 75.2% was reported for student engagement measure. In addition to these two studies, a past publication from our research group is one of the few previous studies we know of to date that explores learner wakefulness using facial expression and head pose analysis in a video-based online learning context [19]. Due to the limited

number of such studies, it can be argued that existing research on measuring wakefulness during students' engagement with e-learning platforms is at its early maturity level.

Due to the scarcity of research in this area within educational settings, we draw upon previous research in the field of driver drowsiness. In this area, drivers' wakefulness measures have been a significant point of focus with multiple research examples. As there can be fatal consequences of drowsiness while driving, there are a large number of technologically mature studies conducted to detect driver wakefulness, as reviewed by [20]. In general, three types of measures have been identified for monitoring driver drowsiness: (1) vehicle-based measures (2) behavioral measures, and (3) physiological measures. The first category of measures includes a number of metrics specific to the task of driving and not applicable to learning contexts. On the contrary, behavioral measures, such as eye closure and head pose, as well as physiological measures, such as EEG (electroencephalogram) and ECG (electrocardiogram) signals, can be collected from learners during e-learning using a range of devices from simple web cameras to EEG, heart rate, and eye tracking sensors. Those studies show the benefits of hybrid over single-mode measures and provide evidence for the potential of multimodal data to gain insights into detecting drowsy behaviors automatically. However, the motion patterns and psychological states of drivers may differ significantly from those of learners. Therefore, further research is needed to investigate the value and effectiveness of multimodal data for drowsiness detection in e-learners. Here, we use multimodal data from students' heart rate, seat pressure, and facial expressions and implement them in measuring their wakefulness during their interaction with an e-learning platform.

III. METHOD

In this section, we present information on the experimental design, multimodal data feature extraction, and the wakefulness estimation method based on multimodal data.

A. EXPERIMENTAL DESIGN

1) PARTICIPANTS AND LEARNING CONTEXT

Fifty-three first-year undergraduate students who were enrolled in an introductory computer science course agreed to participate in our study. The computer science course is a blended learning course consisting of an asynchronous e-learning session and a synchronous face-to-face session per week. Each e-learning session includes a number of video lectures of approximately 10 minutes of length each (Table 1). The videos are voice-over PowerPoint slideshows as displayed in Fig 1. Data for this study was collected as students were watching the video lectures. We asked the participants to come to an experiment room where they took the e-learning sessions. Therefore, a number of students were not able to join the experiment for some sessions. To maintain data balance, we discarded the data from all those students who took only one lecture in the experiment and had missing data. The remaining valid data was collected from 48 students (mean age of 18.4).

TABLE 1. The overview of the video lectures and the number of slides.

Topic	Video	Slide
Information Digitization	Computer Composition: Hardware	23
	Computer Composition: Software	17
	Information and Data	31
Information Networks & Information Security	Network Components	18
	Communication Protocol	23
	IP Address & Encryption Technology	24
Review Lesson	Review Lesson	40
Mechanism of Internet Services	How Internet works	17
	Server	19
	Web Page Structure	20
Information Technology used in Society	What is Information System	15
	Database and Data Model	24

FIGURE 1. Screenshot of a sample video slide.

2) DATA COLLECTION

Three main sources of data informed the findings of this study: (1) heart rate and other related parameters measured with wearable heart rate meters manufactured by Union Tool Co., (2) seat pressure measured with Smart Rubber Soft Vision seat pressure detection sensors from Sumitomo Riko Co. Ltd., and (3) facial expression videos recorded with Bandicam. Heart rate meters recorded R-R interval (RRI), body temperature, heart rate variability measures of low frequency/high frequency (LF/HF), and heart rate (HR). Seat pressure sensors yielded data on moving distance from center of gravity position, length of moving state, length/position/pressure of static state, etc.

3) PROCEDURE

The experiment was run on notebook computers (17.3-inch screen). While the participants took the e-learning session, we collected the multimodal data mentioned above and recorded their screens with recording software

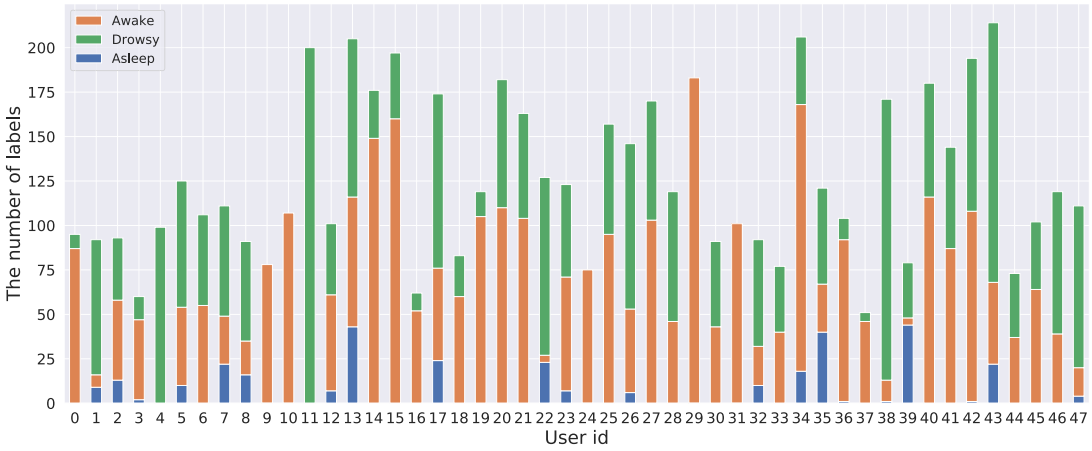

FIGURE 2. The number of slides covered by each learner as well as their wakefulness states.

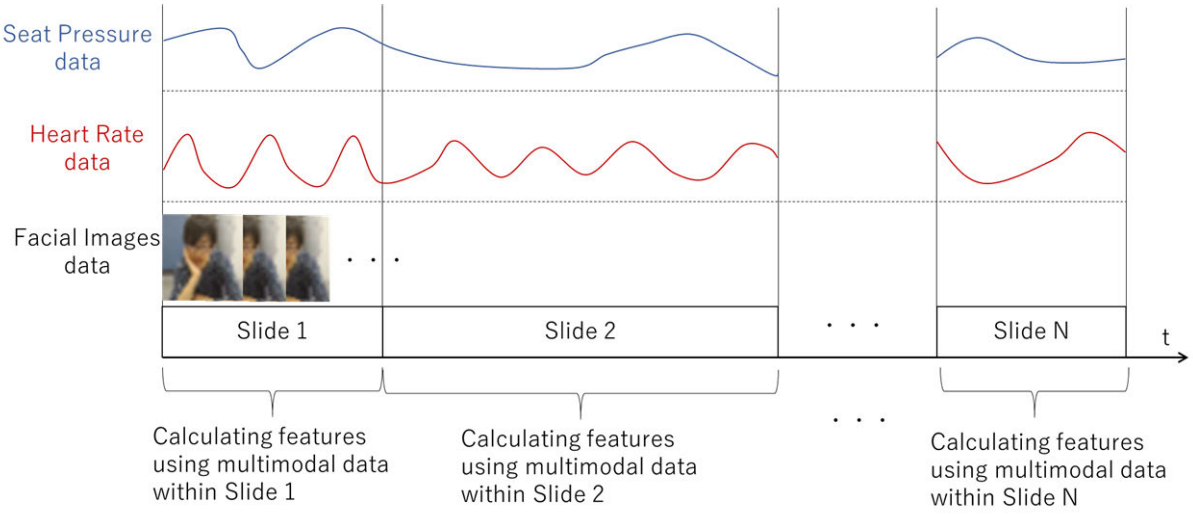

FIGURE 3. Generating multimodal features from every slide. Multimodal data is divided based on the time each slide is shown and features are calculated per slide.

(Bandicam, 30fps). After the e-learning session, each screen capture and its corresponding face recording were combined into one video. The participants watched this video and self-reported their wakefulness, understanding, and motivation per slide on a 4-point Likert scale. This study used only wakefulness data (1: Asleep, 2: Drowsy, 3: Awake, 4: Wide Awake). To minimize variability of self-assessments, we counted level 3 and 4 as "Awake." Figure 3 shows the number of slides covered by each learner as well as their wakefulness states from the observation data.

B. MULTIMODAL DATA FEATURE GENERATION

As presented in figure 3, the features are calculated for each slide since the self-reported data were collected per slide to reduce the burden of annotation work. Table 2 shows 40-dimensional features that were extracted from multimodal data.

1) HEART RATE FEATURES

The heart rate sensor we used in the experiment records RRI, LF/HF, and heart rate which are calculated using the data for the past 60 seconds, and body surface temperature. RRI stands for R-R interval and is an index of heart rate variability, showing cardiac beat-to-beat interval. LF and HF are the ratio of low frequency (0.04 Hz – 0.14 Hz) power and high frequency (0.14 Hz – 0.4 Hz) power in heart rate variability. They are frequently used as indices of stress and rest states. [21]. Two statistical values (mean, standard deviation) were extracted from RRI, LF/HF, Heart Rate and body surface temperature. In total 8 features were calculated as heart rate features.

2) SEAT PRESSURE FEATURES

Seat pressure can be a proxy of a learner's posture change. As we presented in section 2, in a recent study

TABLE 2. List of multimodal features.

Features	Metrics	Constructs Represented
Heart rate (HR)	Mean/ standard deviation of RRI, HR, LF/HF, Body surface temperature	Heart rate in general represents the activity of the autonomic nervous system. RRI is an index of heart rate variability. HR is the heart rate. Low frequency (LF) power and high frequency (HF) power represent stress and rest states. Body surface temperature is environmental temperature in clothing.
Seat Pressure (SP)	Mean pressure	Mean of each frame's total pressure and mean of pressure per second. They are used to estimate a learner's motions.
	Mean time of MS (moving state) and SS (static state)	Represents how long a learner moves or stays still.
	Ratio of MS (moving state)	Represents how often a learner changes posture.
	Mean of absolute pressure difference between pressure current and previous frame.	Represents how large and how often a learner changes posture along vertical axis.
Facial Expression (Face)	Mean/ standard deviation of AU 2, 15, 26, 45 (occurrence and intensity)	AU2: Outer Brow Raiser, AU15: Lip Corner Depressor, AU26: Jaw drop, AU45: Blink.
	Mean/ standard deviation of head rotation (yaw, pitch, roll)	Represents how large a learner's head rotation is.
	Mean/ standard deviation of head transition along x, y, z	Represents how large a learner's head transition is.

Nomura *et al.* (2018) used seat pressure data to estimate student engagement. In this paper, we use some of the features introduced in that study. For seat pressure features, the first step was to classify each frame into a moving state (MS) and a static state (SS) by using the distance of center gravity position between the current and previous frames. For this process, we used 0.1 as a threshold of MS and SS by checking the distribution of the distance as exemplified in previous research. We calculated the mean time of MS, SS, and the ratio of MS. These values represent how much a student's upper body moves. In addition, mean pressure and mean absolute pressure difference between current and previous pressure frames were also calculated as features. In total, 4 features were calculated from the seat pressure data for each learner.

3) FACIAL EXPRESSION FEATURES

To extract features from students' facial expressions, we used the facial action coding system (FACS) created by Ekman and Friesen [22]. FACS is commonly used to describe facial expressions in terms of Action Units (AU). AUs are fundamental motions of a facial muscle or a group of facial muscles. In learning contexts, head poses are also important indices of students' postures. Therefore, we used OpenFace [23] to extract AUs and head poses from facial images. OpenFace outputs were intensity and occurrence of 17 AUs and the 3-D transition and rotation of the student's head (yaw, pitch, and roll). We calculated the mean and standard deviation of AU 2, 15, 26, 45, and head pose. These AUs are Outer Brow Raiser, Lip Corner Depressor, Jaw drop, and Blink, respectively. More specifically, AU 26 is related to mouth opening and AU 2 and 15 are related to expressions when struggling to stay awake. All these AUs are chosen based on existing research on measuring wakefulness of drivers as referred to in Vural *et al.* [24]. In total, 28 features were extracted from facial images.

C. WAKEFULNESS ESTIMATION METHOD BASED ON MULTIMODAL DATA

In order to investigate the relationship between wakefulness levels and multimodal features described above to recognize learners' wakefulness, we conducted two types of analysis.

First, to find useful features to estimate learners' sleepiness level, we investigated differences in the features' means of each wakefulness level. As mentioned in section three, we classified all features based on learners' self-coding of three wakefulness levels. We examined the mean difference of three wakefulness levels (Asleep vs. Drowsy vs. Awake), as well as two levels (Asleep vs. Others and Awake vs. Others). Before the analyses, we conducted verification for normality of variances and used the non-parametric Friedman's test to investigate the difference in three levels. Wilcoxon signed-rank test was used to investigate the difference between levels. All statistical analyses were performed using SPSS 26.0.0.1.

Second, to investigate the potential of machine learning techniques utilizing multimodal data to predict learners' wakefulness automatically, we evaluated the classification accuracy of several models. We opted for supervised machine learning approaches. Existing research shows that supervised machine learning frequently outperforms unsupervised learning in real-world classification problems [15]. More

specifically, we have tested multi-class support vector machines (SVM), Random Forest Classifiers (RF) and CatBoost Classifiers [25] (CatBoost) to build wakefulness estimation models in user-independent and user-dependent settings for three types of classification; Awake vs. Others, Asleep vs. Others and three levels of wakefulness state (Asleep, Drowsy, Awake). Although in the user-independent setting we tested the potential of the model for unseen users, the influence of personalization in module building was evaluated in the user-dependent setting. To evaluate the accuracy of the models, we employed F1 scores as an index of three degrees of wakefulness recognition.

In essence, model construction and evaluation involved four main steps:

1) Splitting data into train and test datasets: We split data for cross validation of user-dependent and user-independent settings.
2) Over/undersampling to deal with imbalanced label distribution: To deal with imbalanced distribution of wakefulness state in train dataset, we use SMOTE [26] for oversampling and random sampling for undersampling.
3) Standardizing train and test datasets: Both train and test dataset are standardized to have zero-mean and unit variance by using mean and standard deviation of train dataset.
4) Building a model with train dataset and evaluating it with test dataset: SVM, RF and CatBoost are used to build the model.

To evaluate the model in the user-dependent setting, we employed leave one-group out cross validation. We divided 48 subjects into 6 groups, and every group was used as test dataset at a time.

Similarly, in the user-independent setting, we used leave one-group out cross validation. The data from one group is used as test dataset, and the data from other groups is used to train the model. These steps are repeated until every group's data is used as test. The output task was to recognize three degrees of wakefulness (Asleep, Drowsy and Awake), and two degrees of wakefulness (Asleep vs Others, Awake vs Others).

In the user-dependent setting, the first one tenth of data from the test group in each day's lecture is included in the train dataset. In this setting, train and test datasets include data from the same participants. This setting is not an unrealistic constraint. As a matter of fact, when a learner starts the e-learning task, the camera can use the data from the first few minutes as train data with labels (the person is expected to be awake).

IV. RESULTS

A. ANALYSIS OF MULTIMODAL FEATURES

Firstly, we classified all features based on three wakefulness levels (Asleep vs. Drowsy vs. Awake) and we ran Friedman's test, and post hoc tests (Scheffe's test). We discarded the data from those students who did not display all three patterns of wakefulness, and finally analyzed the data from the remaining 18 participants. Table 3 shows those features which yielded a significant difference. Post hoc comparisons show that the mean occurrence of AU26 (JawDrop) in Drowsy state is significantly higher than in other states ($p < .01$). The mean of intensity of AU26 in Asleep state is significantly lower than in other states. (Asleep vs. Drowsy: $p < .01$, Asleep vs. Awake: $p < .05$). The mean occurrence of AU45 (Blink) in Awake state is significantly lower than in other states (Awake vs. Drowsy: $p < .05$, Awake vs. Aspleep: $p < .01$). The mean of intensity of AU45 in Awake is significantly lower than in Asleep state ($p < .01$). Mean body surface temperature has a significant difference across all states ($p < .01$). The lower the participants' arousal levels, the higher their body surface temperature.

Secondly, we classified all features based on two wakefulness levels (Asleep vs. Others and Awake vs. Others), and we ran paired Wilcoxon signed rank tests. We present the features that show a significant difference for Asleep vs. Others in Table 4, and for Awake vs. Others in Table 5, respectively.

TABLE 3. Results of Friedman's test in three-level wakefulness classification.

Feature	χ^2	p	Ranks	
AU26 (Mean occurrence)	13.00	.002	Awake	1.83
			Drowsy	2.67
			Asleep	1.50
AU26 (Mean intensity)	7.44	.024	Awake	2.11
			Drowsy	2.39
			Asleep	1.50
AU45 (Mean occurrence)	7.00	.030	Awake	1.50
			Drowsy	2.17
			Asleep	2.33
AU45 (Mean intensity)	8.11	.017	Awake	1.56
			Drowsy	1.94
			Asleep	2.50
Body surface temperature (Mean)	27.97	.000	Awake	1.11
			Drowsy	2.03
			Asleep	2.86

TABLE 4. Results of Wilcoxon signed rank test in Asleep vs. others.

Feature	Z	p
AU02 (SD of occurrence)	-2.069	.039
AU02 (SD of intensity)	-2.286	.022
AU15 (SD of intensity)	-2.983	.003
AU26 (Mean occurrence)	-2.591	.010
AU26 (Mean intensity)	-2.025	.043
AU45 (Mean intensity)	-2.940	.003
Pitch (SD)	-1.982	.048
Roll (SD)	-1.982	.048
Head transition toward y-axis (Mean)	-1.982	.048
Body surface temperature (Mean)	-3.332	.001
Body surface temperature (SD)	-2.243	.025

TABLE 5. Results of Wilcoxon signed rank test in awake vs. others.

Feature	Z	p
AU02 (Mean of intensity)	-2.067	.039
AU02 (SD of intensity)	-2.274	.023
Body surface temperature (Mean)	-4.775	.000
Body surface temperature (SD)	-3.052	.002

TABLE 6. Average macro-F1 scores with features based on the results of ANOVA.

	user-independent	user-dependent
SVM	0.36	0.38
RF	0.36	0.38
CatBoost	0.36	0.38

The result of Asleep vs. Others comparison shows that the *SD* of occurrence and intensity of AU02 (Outer Brow Raiser), the *SD* of intensity of AU15 (Lip Corner Depressor), mean intensity of AU45, *SD* of Pitch and Roll, mean head transition toward y-axis, and mean body surface temperature in the Asleep state are higher than the other states. On the other hand, the mean occurrence and intensity of AU26 (Jaw Drop), and *SD* of body surface temperature are lower than in the other states.

The result of Awake vs. Others comparison shows that the mean and the *SD* intensity of AU02 and mean body surface temperature in the Awake state are lower than in other states. Additionally, the *SD* of body surface temperature in the Awake state is higher than in other states.

B. EVALUATION OF WAKEFULNESS ESTIMATION

In this section, we initially present the baseline recognition results of SVM, RF and Catboost classifiers and evaluate these models in user-independent and user-dependent settings for three types of classification; Awake vs. Others, Asleep vs. Others and three levels of wakefulness state.

1) AUTOMATED DETECTION OF LEARNERS' WAKEFULNESS STATE BASED ON THE RESULTS OF ANOVA AND T-TEST

Through ANOVA, we obtained a number of potential features which yielded statistically significant differences across wakefulness states (Table 3). These features are considered to be effective for more accurate automated detection of learners' wakefulness. Therefore, we evaluated the case of using these features to build a model and estimated learners' wakefulness state in user-dependent and independent settings.

Table 6 shows the average macro-F1 scores with features based on the results of ANOVA. In both user-independent and user-dependent settings, all three types of classifiers mark the same average F1-macro scores, 0.36 and 0.39. The scores in the user-dependent setting are higher than that of the user-independent.

TABLE 7. Average macro-F1 scores of Asleep vs Others features based on the results of t-test.

	user-independent	user-dependent
SVM	0.57	0.60
RF	0.57	0.59
CatBoost	0.62	0.66

TABLE 8. Average macro-F1 scores of awake vs others features based on the results of t-test.

	user-independent	user-dependent
SVM	0.56	0.60
RF	0.55	0.71
CatBoost	0.55	0.62

TABLE 9. Average macro-F1 scores of three degrees of wakefulness classification in user-independent setting.

	HeartRate	SeatPressure	Face	Multimodal
SVM	0.37	0.35	0.37	0.37
RF	0.37	0.34	0.36	0.37
CatBoost	0.37	0.34	0.37	0.37

TABLE 10. Average macro-F1 scores of Asleep vs Others in user-independent setting.

	HeartRate	SeatPressure	Face	Multimodal
SVM	0.54	0.53	0.59	0.61
RF	0.53	0.50	0.59	0.62
CatBoost	0.55	0.53	0.62	0.65

Table 7 and 8 show the average macro-F1 scores with features based on the results of t-test for recognizing two degrees of wakefulness state: Asleep vs Others, Awake vs Others. In Asleep vs Others, the scores of Catboost in both user-independent and user-dependent settings, 0.62 and 0.66, are higher than other classifiers. In Awake vs Others, RF in the user-dependent setting marks better performance than other classifiers. On the other hand, in the user-independent setting SVM yields the highest score, 0.56.

2) AUTOMATED DETECTION OF LEARNERS' WAKEFULNESS STATE WITH ALL MULTIMODAL FEATURES

For this task, we used unimodal (Face, SeatPressure, HeatRate) and multimodal features as input to the machine learning model. These features are explained in detail in section 3.3 and summarized in Table 2. Table 9, 10, and 11 show the average F1-macro scores of three types of classification task in SVM, RF, and Catboost with unimodal and multimodal features in each output task, respectively.

In the user-independent setting, multimodal features and HeartRate mark the highest scores, and there are no differences among the three classifiers. Although multimodal features outperformed other features in all classifiers in

TABLE 11. Average macro-F1 scores of Awake vs Others in user-independent setting.

	HeartRate	SeatPressure	Face	Multimodal
SVM	0.58	0.53	0.55	0.59
RF	0.56	0.54	0.54	0.55
CatBoost	0.57	0.54	0.55	0.55

TABLE 12. Average macro-F1 scores of three degrees of wakefulness classification in user-dependent setting.

	HeartRate	SeatPressure	Face	Multimodal
SVM	0.41	0.35	0.50	0.54
RF	0.42	0.35	0.46	0.47
CatBoost	0.43	0.37	0.48	0.53

TABLE 13. Average macro-F1 scores of Asleep vs Others in user-dependent setting.

	HeartRate	SeatPressure	Face	Multimodal
SVM	0.54	0.49	0.75	0.77
RF	0.50	0.48	0.58	0.71
CatBoost	0.55	0.51	0.73	0.82

TABLE 14. Average macro-F1 scores of Awake vs Others in user-dependent setting.

	HeartRate	SeatPressure	Face	Multimodal
SVM	0.65	0.60	0.72	0.75
RF	0.64	0.58	0.66	0.71
CatBoost	0.66	0.59	0.74	0.77

Asleep vs Others setting, HeartRate feature in RF and CatBoost marks a higher score than other features and only multimodal features mark the highest score in Awake vs Others setting. Compared to the results in Table 6, the score of three-level wakefulness state recognition in user-independent setting, 0.37 is lower than that in selected features based on ANOVA. On the other hand, the scores of two-level classifications (Asleep vs Others, Awake vs Others) are higher than three-level. These results indicate that effectiveness of feature value selection is limited.

Table 12, 13, and 14 show the average F1-macro scores of cross validation in SVM, RF, and Catboost with unimodal and multimodal features in three types of classification. As the table indicates, multimodal features outperformed unimodal features with regard to all classifiers. In addition, the scores of multimodal features, HeartRate and Face in the user-dependent setting is higher than the user-independent setting although there are minor differences in the scores of SeatPressure. In two-level classification, Awake vs. Others and Asleep vs. Others settings, multimodal features also outperformed unimodal features and Catboost marks the highest score in both Awake vs. Others and Asleep and Others. Compared to Asleep vs. Others, the scores in HR and SeatPressure are higher in case of Awake vs. Others. This result indicates that HR and SeatPressure data have the possibility to contain different kind information from facial images. It can be assumed that the features we extracted are affected by individual differences, and in the user-dependent setting, the effect of individual differences is decreased since data from the same person is included in the train and test datasets.

Compared to the results of features selected based on ANOVA and t-test in IV-B1, the results of all multimodal features yielded higher scores. This is because ANOVA does not take into account combination of features, and thus the model based on it might be less effective than a model which considers combined features.

V. DISCUSSION

Although the automated detection approach has good potential for adaptive instruction, in essence, our multimodal learning analytics approach aims to provide explicit and comprehensible ways of presenting information to learners and teachers in order to support them in making informed decisions [27]. Therefore, in our first research question, we investigated the factors from multimodal data that can be used to inform teacher and learner dashboards for feedback and reflections on learners' wakefulness. Our results showed that particularly the facial features, the head position, and body surface temperature were useful to identify characteristics of learners' drowsy behaviors. Before falling asleep learners tended to blink more frequently, turn down their lip at the corners, and move their heads more. Furthermore, similar to previous research showing that increasing skin temperature might affect sleep propensity [28], our investigation also obtained similar results.

Our second research question was: To what extent can students' wakefulness be detected with the help of multimodal data from their heart rate, seat pressure, and facial expressions? The results show that we can predict individual learners' asleep, drowsy, and awake states with a high accuracy and confidence with the help of multimodal data. The best performing machine learning model was built using CatBoost Classifier algorithm. These results are comparable to and outperform some of the state-of-the-art results in measuring learner engagement with e-learning platforms with multimodal data [17]. These kinds of automation approaches are particularly useful for the provision of personalized support to learners through intelligent tutoring systems and adaptive e-learning platforms which can be implemented at a scale.

In these investigations, one of our goals was to generate transparency in models that predict learners' drowsiness from multimodal data. These insights can be used to support teachers' and learners' interpretations of the machine learning decisions in predicting drowsy learner behaviors. Allowing opportunities for teachers and learners to interpret and scrutinize analytics suggestions generated can lead to better feedback and reflection opportunities. These insights generated from the models can be deployed in multimodal learning analytics tools that provide suggestions for interven-

tions to teachers and learners. As our answer to the second research question indicates, accurate detection of drowsy learner engagement with e-learning platforms is important. However, learners and teachers also need to know potential reasons for the analytics predictions of learner states. This increases human agency in transparent models which can also lead to better adoption in practice [29].

VI. CONCLUSION

In this paper, we focus on the multimodal sensor data to predict learners' three states of Asleep, Drowsy, and Awake during their engagement with an e-learning platform. First, we generated some insights into the multimodal characteristics of each state based on the results of the statistical analysis. Second, we showed the potential of the multimodal data and supervised machine learning for the automated predictions of learners' three engagement states, especially in user-dependent settings.

Our findings have significant implications for improving student engagement, and indirectly their learning outcomes, in e-learning contexts. Compared to traditional log data analysis and unimodal investigations, multimodal data provides significant improvements to the detection of learner engagement in e-learning platforms. In our future research, we plan to implement our prediction models in teacher and learner dashboards to evaluate their value in improving learning outcomes in e-learning contexts.

REFERENCES

[1] S. Palvia, P. Aeron, P. Gupta, D. Mahapatra, R. Parida, R. Rosner, and S. Sindhi, "Online education: Worldwide status, challenges, trends, and implications," *J. Global Inf. Technol. Manage.*, vol. 21, no. 4, pp. 233–241, Oct. 2018.

[2] J. E. Seaman, I. E. Allen, and J. Seaman. (2018). Grade increase: Tracking distance education in the United States. Babson Survey Research Group. [Online]. Available: https://files.eric.ed.gov/fulltext/ED580852.pdf

[3] J. Crawford, K. Butler-Henderson, J. Rudolph, B. Malkawi, M. Glowatz, R. Burton, P. Magni, and S. Lam, "COVID-19: 20 countries' higher education intra-period digital pedagogy responses," *J. Appl. Learn. Teach.*, vol. 3, no. 1, pp. 1–20, 2020.

[4] World Bank. (2020). *How Countries are Using Edtech (Including Online Learning, Radio, Television, Texting) to Support Access to Remote Learning During the COVID-19 Pandemic*. [Online]. Available: https://www.worldbank.org/en/topic/edutech/brief/how-countries-are-using-edtech-to-support-remote-learning-during-the-covid-19-pandemic

[5] W. Doherty, "An analysis of multiple factors affecting retention in web-based community college courses," *Internet Higher Educ.*, vol. 9, no. 4, pp. 245–255, 2006.

[6] Y. Levy, "Comparing dropouts and persistence in e-learning courses," *Comput. Educ.*, vol. 48, no. 2, pp. 185–204, Feb. 2007.

[7] S. F. Tello, "An analysis of student persistence in online education," *Int. J. Inf. Commun. Technol. Educ.*, vol. 3, no. 3, pp. 47–62, Jul. 2007.

[8] Y. Lee and J. Choi, "A review of online course dropout research: Implications for practice and future research," *Educ. Technol. Res. Develop.*, vol. 59, no. 5, pp. 593–618, Oct. 2011.

[9] R. M. Carini, G. D. Kuh, and S. P. Klein, "Student engagement and student learning: Testing the linkages," *Res. Higher Educ.*, vol. 47, no. 1, pp. 1–32, Feb. 2006.

[10] M. Cukurova, M. Giannakos, and R. Martinez-Maldonado, "The promise and challenges of multimodal learning analytics," *Brit. J. Educ. Technol.*, vol. 51, no. 5, pp. 1441–1449, Sep. 2020.

[11] J. E. Beck, "Engagement tracing: Using response times to model student disengagement," in *Proc. 12th Int. Conf. Artif. Intell. Educ. (AIED)*, C. Looi, G. I. McCalla, B. Bredeweg, and J. Breuker, Eds., Jul. 2005, pp. 88–95.

[12] S. K. D'Mello, S. D. Craig, A. Witherspoon, B. McDaniel, and A. Graesser, "Automatic detection of learner's affect from conversational cues," *User Model. User-Adapt. Interact.*, vol. 18, nos. 1–2, pp. 45–80, Feb. 2008.

[13] C. R. Henrie, R. Bodily, R. Larsen, and C. R. Graham, "Exploring the potential of LMS log data as a proxy measure of student engagement," *J. Comput. Higher Educ.*, vol. 30, no. 2, pp. 344–362, Aug. 2018.

[14] N. Bergdahl, J. Nouri, and U. Fors, "Disengagement, engagement and digital skills in technology-enhanced learning," *Educ. Inf. Technol.*, vol. 25, no. 2, pp. 957–983, Mar. 2020.

[15] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Rev. Educ. Res.*, vol. 74, no. 1, pp. 59–109, 2004.

[16] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagementfrom facial expressions," *IEEE Trans. Affect. Comput.*, vol. 5, no. 1, pp. 86–98, Jan. 2014.

[17] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Trans. Affective Comput.*, vol. 8, no. 1, pp. 15–28, Jan./Mar. 2017.

[18] K. Nomura, M. Iwata, O. Augereau, and K. Kise, "Estimation of student's engagement using a smart chair," in *Proc. ACM Int. Joint Conf. Int. Symp. Pervas. Ubiquitous Comput. Wearable Comput.*, Oct. 2018, pp. 186–189.

[19] S. Terai, S. Shirai, M. Alizadeh, R. Kawamura, N. Takemura, Y. Uranishi, H. Takemura, and H. Nagahara, "Detecting learner drowsiness based on facial expressions and head movements in online courses," in *Proc. 25th Int. Conf. Intell. User Interfaces Companion*, Mar. 2020, pp. 124–125.

[20] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 16937–16953, Dec. 2012.

[21] E. Smets, E. R. Velazquez, G. Schiavone, I. Chakroun, E. D'Hondt, W. De Raedt, J. Cornelis, O. Janssens, S. Van Hoecke, S. Claes, I. Van Diest, and C. Van Hoof, "Large-scale wearable data reveal digital phenotypes for daily-life stress detection," *npj Digit. Med.*, vol. 1, no. 1, pp. 1–10, Dec. 2018.

[22] P. Ekman and W. V. Friesen, *Manual for the Facial Action Coding System*. Palo Alto, CA, USA: Consulting Psychologists Press, 1978.

[23] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "OpenFace 2.0: Facial behavior analysis toolkit," in *Proc. 13th IEEE Int. Conf. Automat. Face Gesture Recognit. (FG)*, May 2018, pp. 59–66.

[24] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *Proc. Int. Workshop Hum.-Comput. Interact.* Berlin, Germany: Springer, 2007, pp. 6–18.

[25] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6638–6648.

[26] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[27] M. Cukurova, C. Kent, and R. Luckin, "Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring," *Brit. J. Educ. Technol.*, vol. 50, no. 6, pp. 3032–3046, Nov. 2019.

[28] E. J. W. Van Someren, "Mechanisms and functions of coupling between sleep and temperature rhythms," *Prog. Brain Res.*, vol. 153, no. 2, pp. 309–324, 2006.

[29] R. S. Baker, "Stupid tutoring systems, intelligent humans," *Int. J. Artif. Intell. Educ.*, vol. 26, no. 2, pp. 600–614, Jun. 2016.

RYOSUKE KAWAMURA received the B.S. and M.S. degrees in engineering from Osaka University. From 2017 to 2021, he was a Researcher at Fujitsu Laboratories Ltd. Since 2021, he has been a Researcher at the Research Unit, Fujitsu Ltd., Japan. His research interests include affective computing, pattern recognition, and human–computer interaction.

SHIZUKA SHIRAI (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in informatics and mediology from Mukogawa Women's University, Hyogo, Japan, in 2007, 2012, and 2015, respectively. She was an Assistant Professor with the School of Human Environmental Sciences, Mukogawa Women's University, from 2015 to 2018. She has been an Associate Professor at the Cybermedia Center, Osaka University, since 2018. Her research interests include human–computer-interaction, educational technology, learning analytics, and computer science education. She is a member of ACM and Information Processing Society of Japan (IPSJ).

NORIKO TAKEMURA received the B.S., M.E., and Ph.D. degrees in engineering from Osaka University, in 2006, 2007, and 2010, respectively. She is currently an Associate Professor with the Institute for Datability Science, Osaka University. Her research interests include gait recognition, ambient intelligence, and emotion estimation. She is a member of SICE, RSJ, and VRSJ.

MEHRASA ALIZADEH received the B.A. degree in English language and literature and the M.A. degree in English language teaching from Allameh Tabataba'i University, in 2009 and 2012, respectively, and the Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, in 2019. Her research interests include technology-enhanced learning and learning analytics.

MUTLU CUKUROVA received the B.S. and M.S. degrees in science and technology education and the Ph.D. degree in learning sciences was conferred from the University of York, in 2008, 2010, and 2014, respectively. He is currently an Associate Professor of learning technologies at University College London. His research interests include (multimodal) learning analytics, artificial intelligence in education, and learning sciences. He is a fellow of the Higher Education Academy, an Editor of the *British Journal of Educational Technology*, and an Associate Editor of the *International Journal of Child-Computer Interaction*.

HARUO TAKEMURA (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer science from Osaka University, Osaka, Japan, in 1982, 1984, and 1987, respectively. He was a Researcher at Advanced Telecommunication Research Institute, International (ATR-I), Kyoto, Japan, from 1987 to 1994, and an Associate Professor at Nara Institute of Science and Technology, Nara, Japan. He has been a Professor at the Infomedia Education Division, Cybermedia Center, Osaka University, since 2001. He is in charge of campus wide deployment of learning management system (LMS) and other IT systems for education. His research interests include interactive computer graphics, human–computer interaction, mixed reality, and their applications in education, including learning analytics.

HAJIME NAGAHARA (Member, IEEE) received the Ph.D. degree in system engineering from Osaka University, Japan, in 2001. He was a Research Associate of the Japan Society for the Promotion of Science, from 2001 to 2003. He was a Visiting Associate Professor at CREA, University of Picardie Jules Verne, France, in 2005. He was an Assistant Professor at the Graduate School of Engineering Science, Osaka University, from 2003 to 2010. He was an Associate Professor with the Faculty of Information Science and Electrical Engineering, Kyushu University, from 2010 to 2017. He was a Visiting Researcher at Columbia University, from 2007 to 2008 and from 2016 to 2017. He has been a Professor at the Institute for Datability Science, Osaka University, since 2017. His research interests include computational photography and computer vision. He received an ACM VRST2003 Honorable Mention Award, in 2003, IPSJ Nagao Special Researcher Award, in 2012, ICCP2016 Best Paper Runners-up, and SSII Takagi Award, in 2016. He served as the Program Chair for ICCP2019. He has been serving as an Associate Editor for IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING, since 2019.