Check for updates

Comprehensive evaluation of machine learning algorithms for predicting sleep–wake conditions and differentiating between the wake conditions before and after sleep during pregnancy based on heart rate variability

Xue Li[1], Chiaki Ono[2], Noriko Warita[3], Tomoka Shoji[1,3],
Takashi Nakagawa[1,2], Hitomi Usukura[4], Zhiqian Yu[4],
Yuta Takahashi[2], Kei Ichiji[5], Norihiro Sugita[6], Natsuko Kobayashi[2],
Saya Kikuchi[2], Ryoko Kimura[1], Yumiko Hamaie[2,4], Mizuki Hino[4],
Yasuto Kunii[2,4], Keiko Murakami[3], Mami Ishikuro[3], Taku Obara[3],
Tomohiro Nakamura[7], Fuji Nagami[8], Takako Takai[7],
Soichi Ogishima[7], Junichi Sugawara[9], Tetsuro Hoshiai[10],
Masatoshi Saito[10], Gen Tamiya[11], Nobuo Fuse[11], Susumu Fujii[12],
Masaharu Nakayama[12], Shinichi Kuriyama[3,13],
Masayuki Yamamoto[6,11], Nobuo Yaegashi[8,10], Noriyasu Homma[5]
and Hiroaki Tomita[1,2,3,4]*

[1]Department of Psychiatry, Tohoku University Graduate School of Medicine, Sendai, Japan,
[2]Department of Psychiatry, Tohoku University Hospital, Sendai, Japan, [3]Department of Preventive
Medicine and Epidemiology, Tohoku University Tohoku Medical Megabank Organization, Sendai, Japan,
[4]Department of Disaster Psychiatry, International Research Institute of Disaster Sciences, Tohoku
University, Sendai, Japan, [5]Department of Radiological Imaging and Informatics, Tohoku University
Graduate School of Medicine, Sendai, Japan, [6]Department of Management Science and Technology,
Graduate School of Engineering, Tohoku University, Sendai, Japan, [7]Department of Health Record
Informatics, Tohoku University Tohoku Medical Megabank Organization, Sendai, Japan, [8]Department of
Public Relations and Planning, Tohoku University Tohoku Medical Megabank Organization, Sendai,
Japan, [9]Department of Community Medical Supports, Tohoku University Tohoku Medical Megabank
Organization, Sendai, Japan, [10]Department of Obstetrics, Tohoku University Graduate School of
Medicine, Sendai, Japan, [11]Department of Integrative Genomics, Tohoku University Tohoku Medical
Megabank Organization, Sendai, Japan, [12]Department of Disaster Medical Informatics, International
Research Institute of Disaster Sciences, Tohoku University, Sendai, Japan, [13]Department of Disaster
Public Health, International Research Institute of Disaster Sciences, Tohoku University, Sendai, Japan

Introduction: Perinatal women tend to have difficulties with sleep along with autonomic characteristics. This study aimed to identify a machine learning algorithm capable of achieving high accuracy in predicting sleep–wake conditions and differentiating between the wake conditions before and after sleep during pregnancy based on heart rate variability (HRV).

Methods: Nine HRV indicators (features) and sleep–wake conditions of 154 pregnant women were measured for 1 week, from the 23rd to the 32nd weeks of pregnancy. Ten machine learning and three deep learning methods were applied to predict three types of sleep–wake conditions (wake, shallow sleep, and deep sleep). In addition, the prediction of four conditions, in which the wake conditions before and after sleep were differentiated—shallow sleep, deep sleep, and the two types of wake conditions—was also tested.

Results and Discussion: In the test for predicting three types of sleep−wake conditions, most of the algorithms, except for Naïve Bayes, showed higher areas under the curve (AUCs; 0.82−0.88) and accuracy (0.78−0.81). The test using four types of sleep−wake conditions with differentiation between the wake conditions before and after sleep also resulted in successful prediction by the gated recurrent unit with the highest AUC (0.86) and accuracy (0.79). Among the nine features, seven made major contributions to predicting sleep−wake conditions. Among the seven features, "the number of interval differences of successive RR intervals greater than 50ms (NN50)" and "the proportion dividing NN50 by the total number of RR intervals (pNN50)" were useful to predict sleep−wake conditions unique to pregnancy. These findings suggest alterations in the vagal tone system specific to pregnancy.

1. Introduction

In recent years, perinatal maternal care has become an increasingly important part of public health in society (1). Accelerometers have been widely used to ascertain sleep patterns through recording movement of the body throughout the night, which may be usable to monitor perinatal maternal health. Unlike traditional methods such as polysomnography (PSG), accelerometers allow estimation of sleep–wake conditions because the body tends to remain stationary when falling asleep. The motion amplitude during sleep becomes distinctively smaller than that in the waking state (2, 3). While the accuracy of accelerometer-based evaluations of sleep conditions remains limited, and accelerometers have not been listed by the American Academy of Sleep Medicine as one of the four types of devices that can evaluate sleep (4), numerous studies have investigated and summarized the validity of using accelerometers to measuring sleep objectively. Research to date has supported to some extent the validity of using accelerometers and actigraphy devices (e.g., fitbit, Jawbone, MovBand) for measuring sleep outside of the laboratory setting (5, 6). Various studies have examined the accuracy of such devices in patients with sleep problems and found no significant differences in sleep onset latency (SOL), wakefulness after sleep onset, and total sleep time between PSG and actigraphy measurements in subjects with or without sleep problems (7–10). Another study assessing the validity of the MovBand 3 against previously validated medical sleep monitors reported that it provided a valid and reliable assessment of sleep conditions, including the number of awakenings, deep sleep, light sleep, and physical activity (4). Therefore, ample evidence has been presented to confirm the stability and accuracy of such devices.

Regardless of the abovementioned limitations, accelerometers offer a significant advantage in terms of ease of use in daily life. For example, the sleep patterns of perinatal women can be monitored using accelerometers throughout the pregnancy and postpartum periods. Perinatal women tend to have difficulties sleeping. Many studies have reported that pregnant women experience significant sleep disruption and sleep disorders (11–13), and it is generally known that sleep disturbances occur during the third trimester of pregnancy (14–16). Sleep disorders can be related to mental health problems

during the perinatal period, including postpartum depression (PPD) (17). Pregnancy is marked by considerable physiological changes and a multitude of symptoms, many of which are likely to disrupt sleep. It is also accompanied by dramatic hormonal changes, which have a significant potential to impact sleep quality (18). Given the impact sleep has on physical and mental well-being, assessing sleep quality throughout pregnancy is crucial. Therefore, to recognize and evaluate the sleep quality of pregnant women and the risk of PPD, it is important to gain a better understanding of the characteristics of sleep–wake conditions during pregnancy.

There are several types of sleep in a sleep cycle: shallow sleep (stages 1 and 2) and deep sleep (stages 3 and 4), which are categorized as non-rapid eye movement (NREM) sleep, and rapid eye movement (REM) sleep (19–21). In the typical sleep process, the sleep cycle, including the wake condition, shallow sleep, deep sleep, and REM, is repeated several times during the night, and sometimes, each cycle is spanned by short-term wake conditions. There are several types of sleep disturbances, including difficulty falling asleep, maintaining sleep, and awakening in the early morning. To develop countermeasures to sleep problems, it is important to consider the sleep cycle and these types of sleep disturbances. Even though accelerometers cannot detect REM sleep precisely, they present the benefit of allowing us to grasp the overall sleep cycle patterns throughout sleep with only limited effort on the part of study participants.

It is noteworthy that the arousal levels in wake conditions before and after sleep can differ. The arousal level in the wake condition before sleep can be related to difficulty falling asleep, and that in the wake condition after sleep can be related to sleep quality. Therefore, it may be worthwhile to differentiate between the wake conditions before and after sleep.

Heart rate can be a useful marker for assessing sleep cycles because it is known to differ between wake and sleep conditions (22). Some investigators have reported correlations between sleep and autonomic nervous system (ANS) activity. Various studies have indicated that the parasympathetic nervous system (PNS) activity is principally influenced by the circadian system, whereas the sympathetic nervous system (SNS) activity is principally influenced by the sleep system (23). It is generally known that the cardiac PNS is

activated during NREM sleep (24–26). In contrast, the cardiac PNS is inactivated during REM sleep (27, 28). Heart rate variability (HRV) is the physiological phenomenon of variations in the time interval between heartbeats. HRV can reflect numerous physiological factors and is always used as a measure of ANS activity (29–31). Time domain features include the coefficient of variation R-R interval (CVRR), the standard deviation of all NN intervals (SDNN), the square root of the mean squared differences of successive NN intervals (RMSSD), the number of interval differences of successive RR intervals greater than 50 ms (NN50), and the proportion dividing NN50 by the total number of RR intervals (pNN50). Frequency domain features include low frequency (LF), high frequency (HF), and the ratio of LF to HF (LF/HF). Determining the influence of sleep on HRV is of considerable interest because of the relevance of sleep stages, HRV, and ANS. Many studies have reported a correlation between different sleep conditions and HRV. Some investigators (32–34) have used changes in HRV to analyze and consider sleep quality, sleep disorders, and mental diseases. Prediction of whether someone is asleep or awake based on HRV has also been attempted.

HRV may be related to not only the sleep–wake conditions, but also the physiological conditions underlying sleep disturbances, which are considered to be related to the dysregulation of the SNS and PNS. The perinatal period introduces a myriad of changes, such as sleep disturbances characterized by insomnia symptoms and poor sleep quality, which are highly prevalent during pregnancy and can increase depressive symptomatology and PPD (35). Regarding the correlation between the ANS, HRV, and sleep, prior studies have investigated whether accounting for sleep disturbances may explain some of the heterogeneity in the association between HRV and depression (36–39).

Machine learning algorithms are a widely used technology that have become one of the core technologies of artificial intelligence and data science. Regarding HRV and other types of clinical information, many scientists have also used machine learning to establish prediction models for different sleep conditions. In addition, several studies (40–42) have shown that the k-nearest neighbor (k-NN) can be used as a sleep condition classifier for different sleep conditions. Some studies (40, 43) have indicated that support vector machine (SVM) is an appropriate method for discriminating between the different sleep conditions. Various studies (41, 42, 44) have also reported that artificial neural network (ANN) is a suitable method for distinguishing the different sleep conditions. Other studies (41, 42, 44–48) have also shown that the random forest (RF) could solve the sleep condition recognition problem, which was considered arousal and valence prediction based on physiological signals. Some studies (49, 50) have confirmed that a deep learning algorithm, long short-term memory (LSTM), was an useful method for predicting different sleep conditions. Mendez (51) set HRV as an important feature and carried out a hidden Markov model (HMM) as a classifier. On this basis, the HMM was used to classify the different sleep stages (NREM and REM). However, since the publication of these previous studies, various new machine learning algorithms have been developed, including RF, gradient boosting trees, stochastic gradient descents, extreme gradient boosting, and ANNs. The application of these algorithms may be beneficial, allowing more efficient prediction of sleep conditions based on HRV.

This study aimed to indicate the extent to which the sleep–wake condition and differentiation between the wake conditions before and after sleep could be predicted in pregnant women based on heart rate-relevant information as indicators of ANS functioning using various algorithms, considering the difference in the arousal level during the wake conditions before and after sleep. Thirteen methods [k-NN, SVM, logistic regression (LR), RF, Naïve Bayes (NB), decision tree (DT), gradient boosting tree (GBT), stochastic gradient descent (SGD), extreme gradient boosting (XGBoost), ANN, convolutional neural network (CNN), LSTM, and gated recurrent unit (GRU)] were applied to predict three types of sleep–wake conditions in pregnant women ("wake," "shallow sleep," and "deep sleep") followed by prediction of four types of sleep–wake conditions ("wake condition before sleep," "wake condition after sleep," "shallow sleep," and "deep sleep"), and the accuracy of each prediction model was subsequently evaluated.

2. Materials and methods

2.1. Participants and procedures

The study participants were recruited from among women registered with the Tohoku Medical Megabank Project Birth and Three-Generation Cohort Study (TMM BirThree Cohort Study) (52–54). In the process of following up with the participants after delivery, a flyer was presented to notify women about the opportunity to participate in the present project after conceiving their next baby. From May 2018 to November 2019, 154 pregnant women (mean age, 32.1 ± 3.2 years) who were measured with an accelerometer for 1 week starting from the 23rd to 32nd weeks of pregnancy enrolled and completed the project. Among 154 participants, 78 gave birth to a baby girl and 76 to a baby boy. Three of the 154 mothers bottle-fed their babies with formula milk, 38 combined breastfeeding and bottle feeding, and 108 breastfed only. This study was approved by the Tohoku University Graduate School of Medicine Ethical Research Committee (2021-4-137, 2021-1-266). Written informed consent was obtained from all participants.

2.2. Measures

2.2.1. Sleep condition information

Sleep–wake conditions were evaluated based on data obtained from a wearable motion sensor installed on Health Care MovBand 3 Wristbands (WMB-03; Docomo, Tokyo, Japan) (55, 56). Every 5-min period during the observation period was classified as the "wake," "shallow sleep," or "deep sleep" condition based on body motions recorded by the device. The conditions of being awake for 30 min before sleep and being awake for 30 min after sleep were included in the "wake" condition. Although electroencephalographs (EEGs) may allow more accurate assessment of sleep–wake conditions, alternative methods based on wearable accelerometers are sometimes used to measure body motions during sleep because of their applicability to monitoring sleep conditions in daily life. Multiple studies have confirmed the validity of the accelerometer-based evaluation of sleep–wake conditions by simultaneous measurement with EEG to predict sleep–wake conditions (57, 58). In addition, the validity of the MovBand as a wrist-mounted accelerometer was confirmed in a previous study (59).

2.2.2. Heart rate variability

HRV was obtained using a wearable heart rate monitor (MyBeat; Union Tool, Tokyo, Japan) attached to the pregnant women's underwear (Toyobo, Osaka, Japan). The validity of the heart rate monitoring function of the MyBeat was confirmed in previous studies (60, 61). The HRVs (62) measured in the present study were CVRR, SDNN, RMSSD, NN50, pNN50, LF, HF, LF/HF, and LF/(LF + HF). HRV indicators were calculated for every 5-min segment over 7 days. Descriptive HRV information is summarized in Supplementary Table S1.

2.3. Statistical analyses

The HRV indicators were calculated separately for each sleep and wake pattern. Multiple sets of variables in the present study were consistent with homogeneity of variance (Brown-Forsythe test p-value < 0.05 and Bartlett's test p-value < 0.05). Therefore, analysis of variance was used to compare multigroup variables, and Tukey's multiple comparisons test was used to compare intergroup variables. A p-value < 0.05 was considered significant. Most statistical analyses were performed using Prism 8 (GraphPad Software, San Diego, CA) (63–65).

2.4. Machine learning and deep learning algorithms

Machine learning algorithms (k-NN, SVM, LR, NB, SGD, DT, RF, GBT, XGBoost, and ANN) and deep learning algorithms (CNN, LSTM, and GRU) (66–91) were applied to predict the different types of sleep–wake conditions and to differentiate between the wake conditions before and after sleep based on HRV data (78). Descriptions of the 10 machine learning algorithms (92) and three deep learning algorithms are provided in Supplementary Text 1. For the test set, we used the trained models to test and compare their prediction of sleep–wake conditions and differentiation between the wake conditions before and after sleep with actual data (93, 94). The accuracy, precision, sensitivity, specificity, F1 score, and the area under the receiver operating characteristic curve (AUC) have been described elsewhere (92). Multi-head attention was applied in the GRU prediction model. Weight initialization and an early stopping mechanism were applied in the model. The design of the present study is shown in Figure 1, and the structure of the deep learning prediction models is shown in Figure 2.

2.5. Evaluations of feature contributions

RF was used to evaluate feature contributions to predicting different sleep–wake conditions and differentiating between the wake conditions before and after sleep (92). Feature analysis evaluated all features and observed the important features significantly related to the different types of sleep–wake conditions and differentiation between the wake conditions before and after sleep based on feature contributions. Thus, as used in previous studies, RF is usually used as a classifier (95–97) and a method for evaluating the feature contribution method (98–100).

2.6. Validation of the analyses

2.6.1. Alternative applications for machine learning predictions

The Waikato Environment for Knowledge Analysis (WEKA) (101, 102) and JMP statistical software (SAS Institute, Cary, NC, United States) (103, 104) were used to analyze the same dataset to validate the abovementioned Python-based prediction models, as described elsewhere (92).

2.6.2. Alternative calculations of HRV indicators

We primarily used the HRV indicators calculated using the program installed in the MyBeat device. The source codes of the algorithms used to calculate the HRV indicators in the device are proprietary. To validate the HRV indicators provided by the device, we calculated HRV indicators in Python using open-source codes (92). The multiple formulas used to calculate the time domain features included CVRR, SDNN, RMSSD, NN50, and pNN50, and the frequency domain features included LF and HF. The formula (105–110) used to calculate the remaining HRV indicators was summarized in a previous paper (92). The HRV indicators given by the MyBeat were compared with those calculated using Python to ensure consistency (92).

2.7. Cross-validation of models for the hyper-parameter search

Three major approaches to automatic hyper-parameter tuning— GridSearch CrossValidation (GridSearchCV) (111–113), RandomizedSearch CrossValidation (RandomizedSearchCV) (114–116), and Bayesian optimization search (117–119)—were tested on the two datasets in a preliminary study. We selected RandomizedSearchCV as the most appropriate method because it provided the highest accuracy and required the least amount of time for calculation. The optimal parameters are listed in Supplementary Table S2.

2.8. Performance evaluation of the model effects

To evaluate the performance of each algorithm, we used k-fold cross-validation (KCV) (test size = 0, k = 5) and leave-one-out cross-validation (LOOCV). The AUC is widely used to evaluate the effects of different algorithms. Accuracy measures the prediction accuracy of a model at a specific threshold. Among the tested algorithms, the prediction model with the highest AUC and accuracy was determined.

3. Results

3.1. HRV indicators in different sleep–wake conditions and differentiation between the wake conditions before and after sleep

CVRR during the wake condition after sleep was significantly larger than that during shallow sleep. CVRR during shallow sleep was significantly larger than that during deep sleep. CVRR during the
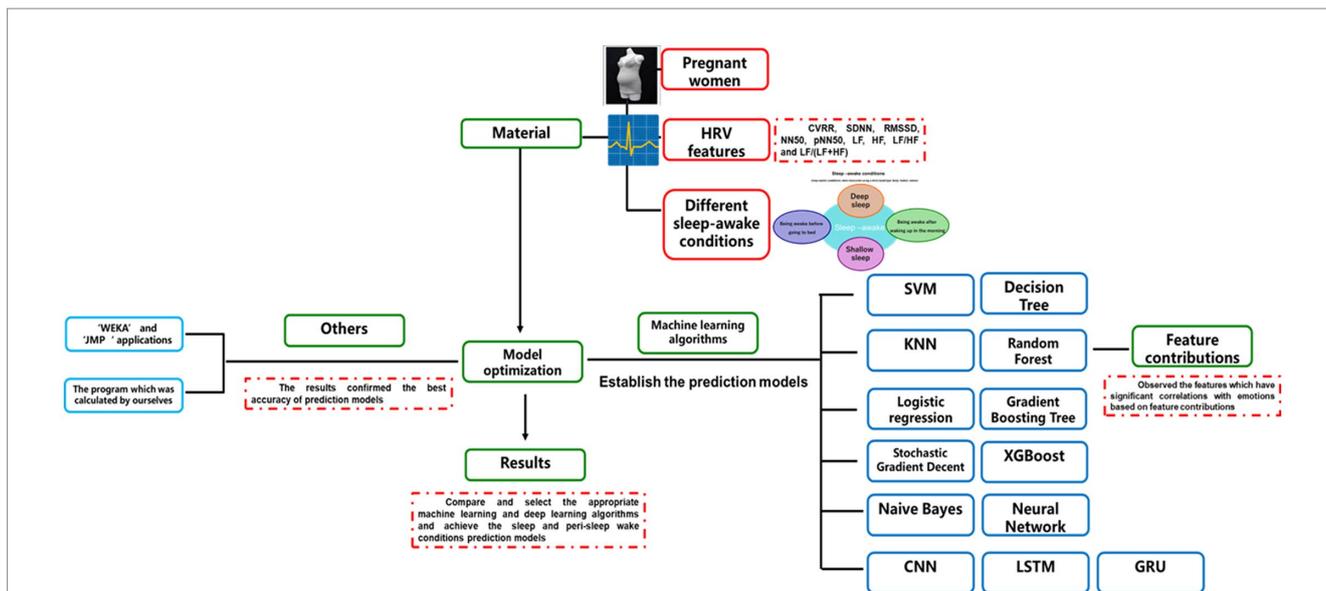
FIGURE 1
The design of this study. The figure indicates the types of information, including heart rate variability (HRV), sleep−wake condition, and the differentiated wake conditions of before and after sleep submitted for analysis, as well as the data analysis methods, including machine learning and deep learning algorithms.
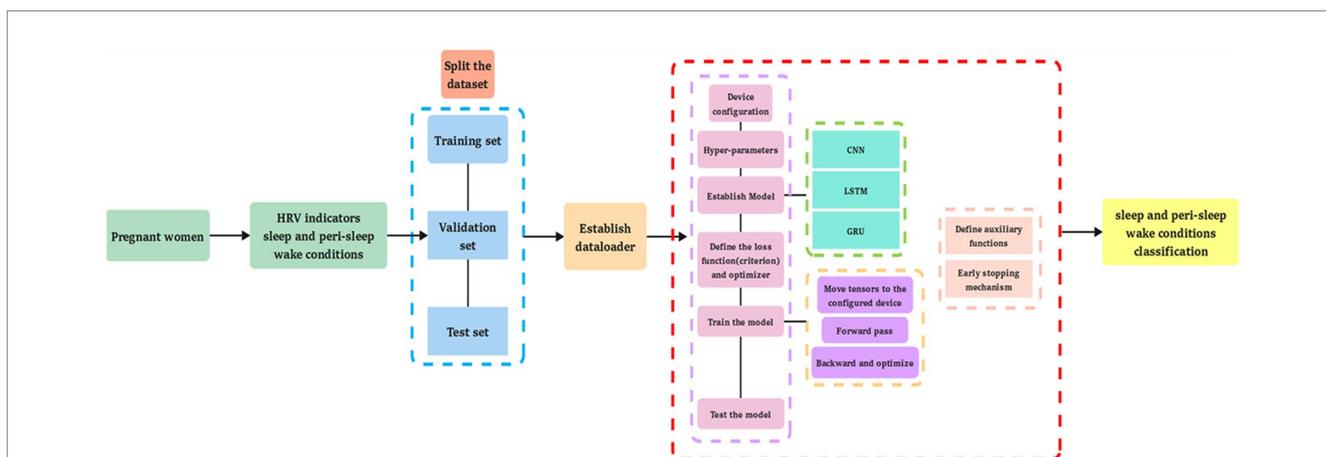

FIGURE 2
The structure of the deep learning prediction models.

wake condition before sleep was significantly larger than that during deep sleep. SDNN during shallow sleep was significantly larger than that during the wake condition after sleep. SDNN during the wake condition after sleep was significantly larger than that during the wake condition before sleep. SDNN during the wake condition before sleep was significantly smaller than that during the sleep condition. RMSSD and HF during deep sleep were significantly larger than those during shallow sleep, the wake condition before sleep, and the wake condition after sleep. NN50s during deep and shallow sleep were significantly larger than those during the wake conditions. pNN50 during shallow sleep was significantly larger than that during deep sleep. pNN50 during deep sleep was significantly larger than that during the wake condition after sleep. pNN50 during the wake condition after sleep was significantly larger than that during the wake condition before sleep. LF during the wake condition after sleep was significantly larger

than that during shallow sleep. LF during shallow sleep was significantly larger than that during deep sleep. LF during deep sleep was significantly larger than that during the wake condition before sleep. LF/HF and LF/(LF + HF) during the wake condition after sleep were significantly larger than those during the wake condition before sleep. LF/HF and LF/(LF + HF) during the wake condition before sleep were significantly larger than those during shallow sleep. LF/HF and LF/(LF + HF) during shallow sleep were significantly larger than those during deep sleep (Figure 3).

To test whether maternal HRV differed among gestational weeks, the subjects were divided into two groups based on gestational weeks at the time of recruitment (23–28 or 28–32 weeks) and then HRV indicators were compared between the two groups. None of the HRV indicators were significantly different between the two groups (Supplementary Figure S4).
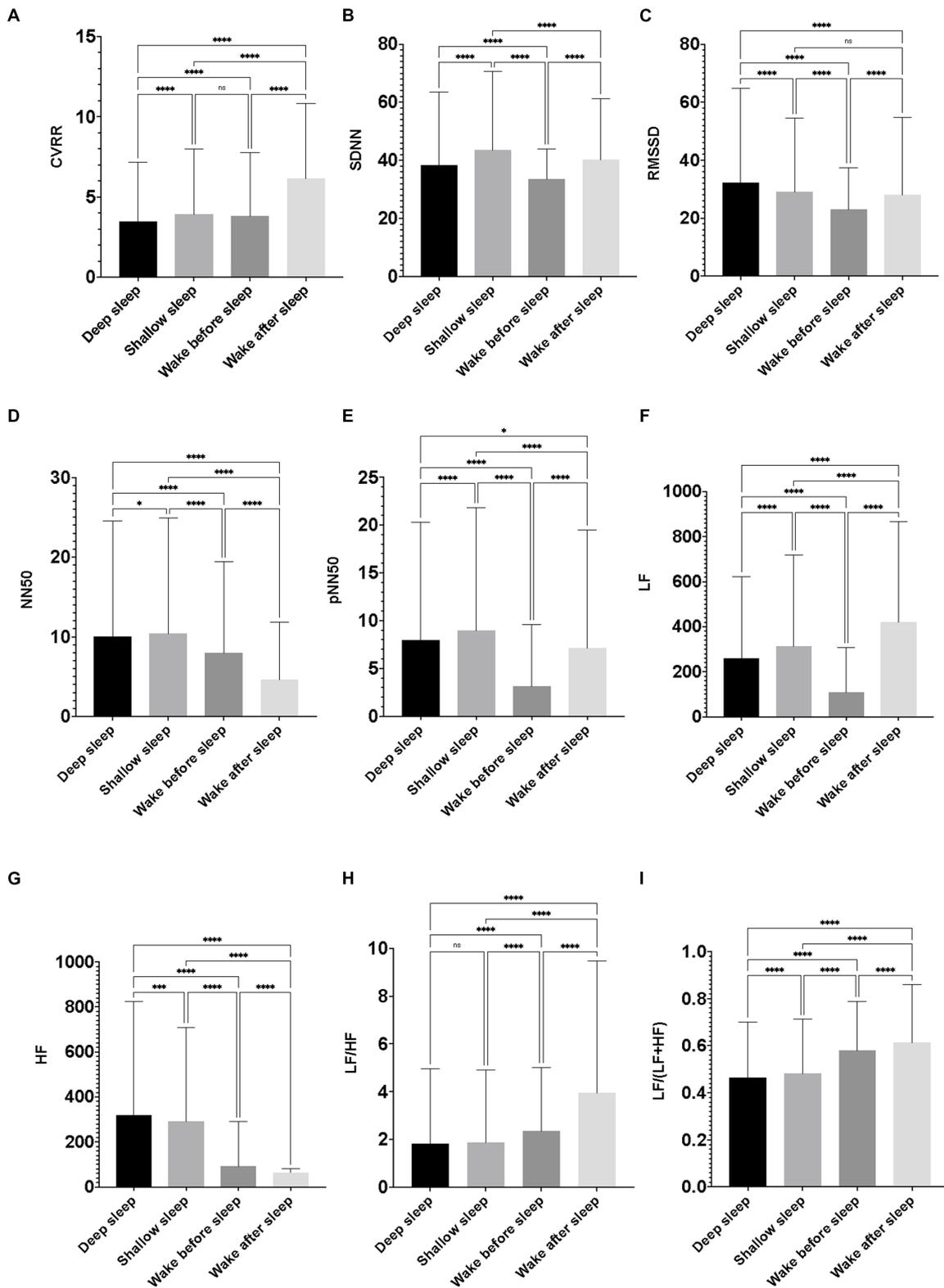
FIGURE 3

Differences in heart rate variability indicators among the sleep–wake conditions. The figures show differences in the nine heart rate variability (HRV) indicators among the four conditions (deep sleep, shallow sleep, wake before sleep, and wake after sleep). Featured HRV indicators were as follows: **(A)** coefficient of variation of R-R interval (CVRR), **(B)** the standard deviation of the time interval between successive normal heart beats (SDNN), **(C)** the square root of the mean of the sum of the squares of differences between adjacent RR intervals. Reflects high frequency (fast or parasympathetic) influences on HRV (RMSSD), **(D)** number of interval differences of successive RR intervals greater than 50ms (NN50), **(E)** the proportion dividing NN50 (the number of interval differences of successive RR intervals greater than 50ms) by the total number of RR intervals (pNN50), **(F)** low frequency from 0.04 to 0.15Hz (LF), **(G)** high frequency from 0.15 to 0.4Hz (HF), **(H)** the ratio of LF to HF (LF/HF), and **(I)** the ratio of LF to (LF+HF) [LF/(LF+HF)]. The

(Continued)

TABLE 1 Model evaluation indices of the 13 machine and deep learning algorithms in predicting the three sleep−wake conditions (wake, shallow sleep, and deep sleep).

Items	SVM	k-NN	SGD	LR	DT	NB	RF	GBT	XGBoost	ANN	CNN	LSTM	GRU
Accuracy	0.80	0.79	0.79	0.78	0.80	0.42	0.81	0.81	0.81	0.81	0.80	0.81	0.81
Precision	0.73	0.75	0.74	0.77	0.77	0.62	0.79	0.78	0.79	0.78	0.75	0.77	0.78
Sensitivity	0.80	0.79	0.79	0.78	0.80	0.42	0.81	0.81	0.81	0.81	0.80	0.81	0.81
F1 score	0.72	0.75	0.74	0.77	0.77	0.43	0.76	0.78	0.77	0.75	0.72	0.73	0.78
AUC	0.82	0.86	0.87	0.87	0.86	0.59	0.88	0.88	0.88	0.87	0.86	0.86	0.88

The table summarizes accuracies, precisions, sensitivities, F1 scores, and areas under the curve (AUC) of the 13 machine and deep learning predictions of the three sleep–wake conditions (wake, shallow sleep, and deep sleep). In these analyses, wake conditions before and after sleep were uniformly integrated into wake conditions. Ten machine learning algorithms, including support vector machine (SVM), k-nearest neighbor (k-NN), stochastic gradient descent (SGD), logistic regression (LR), decision tree (DT), naïve Bayes (NB), random forest (RF), gradient boosting trees (GBT), extreme gradient boosting (XGBoost), artificial neural network (ANN) and three deep learning algorithms including, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) were tested.

3.2. Prediction of sleep−wake conditions and differentiating between the wake conditions before and after sleep

Among the 13 machine and deep learning algorithms applied to predict three selected sleep–wake conditions (wake, shallow sleep, and deep sleep) based on HRV indicators, GRU, XGBoost, RF, and GBT showed AUCs and accuracy of 0.88 and 0.81, respectively, compared with ANN (0.87 and 0.81), LSTM (0.86 and 0.81), CNN, DT (0.86 and 0.80), SGD (0.87 and 0.79), LR (0.87 and 0.79), k-NN (0.86 and 0.79), SVM (0.82 and 0.80), and NB (0.59 and 0.42). According to the prediction model results of the three types of sleep–wake conditions with different methods, most tested algorithms, except for NB, were excellent with high AUC and accuracy. The accuracy, precision, sensitivity, F1 score, and AUCs of the algorithms are summarized in Table 1 and Supplementary Figures S5–S7.

Among the 13 machine and deep learning algorithms applied to predict four selected sleep–wake conditions and differentiate between the wake conditions before and after sleep (wake condition before sleep, wake condition after sleep, shallow sleep, and deep sleep) based on HRV indicators, GRU showed the highest AUC and accuracy of 0.86 and 0.79, respectively, followed by LSTM (0.78 and 0.73), ensemble learning methods (RF, GBT and XGBoost) (0.76 and 0.71), CNN (0.75 and 0.71), k-NN(0.75 and 0.67), ANN (0.74 and 0.70), DT (0.73 and 0.69), SVM (0.73 and 0.67), SGD (0.72 and 0.70), LR (0.72 and 0.67), and NB (0.61 and 0.42). According to the prediction model results for the four sleep–wake conditions, GRU was an appropriate method, achieving the highest AUC and accuracy. The accuracy, precision, sensitivity, F1 score, and AUCs of the algorithms are summarized in Table 2 and Supplementary Figures S8–S10. Confusion matrix for multi-classification is shown in Supplementary Figures S11, S12.

3.3. Evaluations of each feature

The importance scores of each feature for predicting sleep–wake condition and differentiating between the wake conditions before and after sleep based on the nine HRV indicators using RF revealed that pNN50, RMSSD, SDNN, CVRR, HF, and LF were important for the prediction of sleep–wake condition and differentiating between the wake conditions before and after sleep. In addition, pNN50, RMSSD, NN50, SDNN, CVRR, HF, and LF made major contributions to predictions for both three and four types of sleep–wake conditions and differentiating between the wake conditions before and after sleep. The importance scores of each feature in predicting sleep conditions based on the nine HRV indicators using RF are plotted in Figure 4. Cross-validation scores were plotted with the number of features used to predict sleep–wake conditions and differentiate between the wake conditions before and after sleep. The cross-validation scores increased as more features were included in the prediction; many of the above features are included in Figure 5. pNN50, RMSSD made major contributions to the stability of model predictions. Regarding the results of feature importance, the accuracy, precision, sensitivity, F1 score, and AUCs of the algorithms used for three and four types of sleep–wake conditions and differentiation between the wake conditions before and after sleep with important features are summarized in Supplementary Tables S3, S4.

3.4. Validation of analyses

Predictions of sleep–wake conditions and differentiation between the wake conditions before and after sleep using the Python-based open-source codes of algorithms to calculate HRV indicators (92) provided the same predictions using the HRV indicators produced by

TABLE 2 Model evaluation indices of the 13 machine and deep learning algorithms in predicting the four sleep−wake and differentiated wake conditions of before and after sleep (wake condition before sleep, wake condition after sleep, shallow sleep, and deep sleep).

Items	SVM	k-NN	SGD	LR	DT	NB	RF	GBT	XGBoost	ANN	CNN	LSTM	GRU
Accuracy	0.67	0.67	0.70	0.67	0.69	0.42	0.71	0.71	0.71	0.70	0.71	0.73	0.79
Precision	0.65	0.66	0.58	0.55	0.64	0.42	0.67	0.67	0.67	0.65	0.70	0.71	0.76
Sensitivity	0.68	0.67	0.70	0.67	0.69	0.42	0.71	0.71	0.71	0.70	0.71	0.72	0.78
F1 score	0.63	0.66	0.60	0.56	0.63	0.39	0.64	0.66	0.63	0.61	0.70	0.72	0.76
AUC	0.73	0.75	0.72	0.72	0.73	0.61	0.76	0.76	0.76	0.74	0.75	0.78	0.86

The table summarizes accuracies, precisions, sensitivities, F1 scores, and areas under the curve (AUC) of the 13 machine and deep learning predictions of the four sleep−wake condition and differentiating wake conditions before and after sleep (wake condition before sleep, wake condition after sleep, shallow sleep, and deep sleep). Ten machine learning algorithms, including support vector machine (SVM), k-nearest neighbor (k-NN), stochastic gradient descent (SGD), logistic regression (LR), decision tree (DT), naïve Bayes (NB), random forest (RF), gradient boosting trees (GBT), extreme gradient boosting (XGBoost) and artificial neural network (ANN), and three deep learning algorithms including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) were tested.
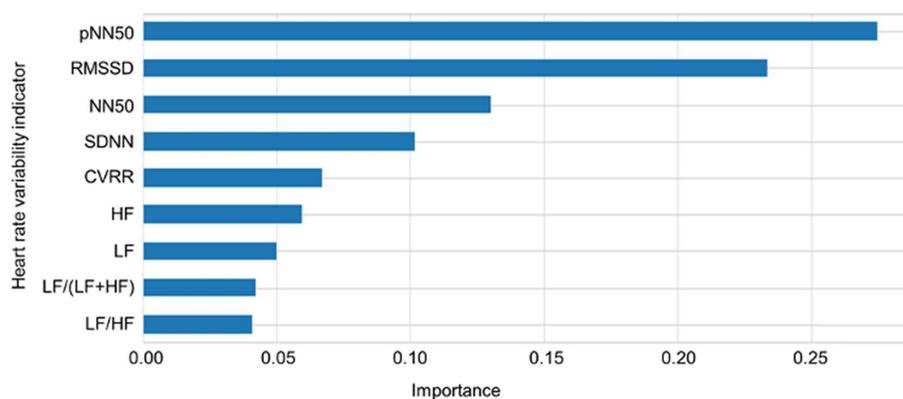

FIGURE 4
Importance of each heart rate variability indicator. The importance scores of each feature in predicting sleep conditions based on the nine heart rate variability indicators using random forest are plotted. CVRR: coefficient of variation RR intervals, SDNN: standard deviation of all NN intervals, RMSSD: square root of the mean squared differences of successive NN intervals, NN50: number of interval differences of successive RR intervals greater than 50ms, pNN50: the proportion derived by dividing NN50 by the total number of RR intervals; LF: frequency domain features including low frequency; HF: high frequency; LF/HF: the ratio of low frequency to high frequency.
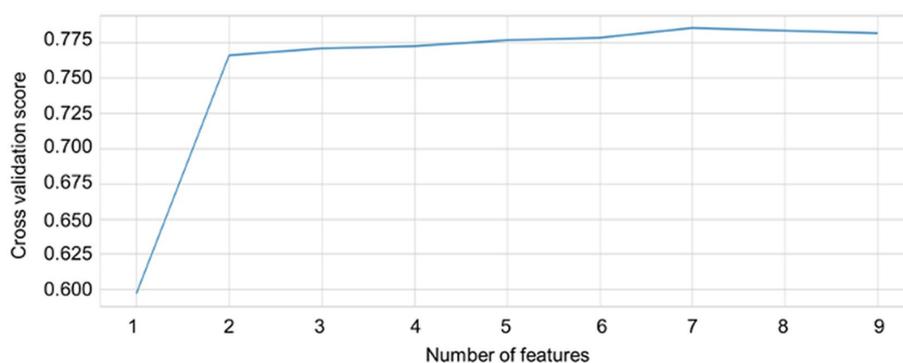
FIGURE 5
Numbers of features and cross-validation scores of random forest-based predictions of sleep−wake and differentiated wake conditions of before and after sleep. Cross-validation scores for each number of features used in the prediction of sleep conditions are plotted. pNN50, RMSSD made major contributions to the stability of model predictions. As more features are included in the prediction, cross-validation scores increase. A plateau is reached when the features are included.

the program installed in the MyBeat device. In addition, WEKA and JMP analyses of the same dataset produced the same results regarding the AUC of the predictions using the 10 algorithms. The results of other applications are provided in Supplementary Tables S5, S6.

With respect to validating the machine learning algorithms for building prediction models of the different sleep−wake conditions and differentiating between the wake conditions before and after sleep, RandomizedSearchCV achieved the highest accuracy as well as the

fastest calculation time. Regarding the performance of each algorithm, we performed KCV (test size = 0, k = 5) to evaluate training performance. The features were separated into five folds: four were used as training data and the remaining one as a validation dataset. The results showed the performance on accuracy, precision, sensitivity, F1 score, and AUCs from the test dataset for all iterations. The optimal parameters are listed in Supplementary Tables S7, S8. The AUCs of the algorithms with LOOCV are summarized in Supplementary Table S9.

4. Discussion

In this study, significant differences among sleep–wake conditions were found for all nine HRV indicators [CVRR, SDNN, RMSSD, NN50, pNN50, LF, HF, LF/HF, and LF/(LF + HF)]. *Post hoc* analysis indicated that CVRR, SDNN, NN50, pNN50, LF, LF/HF, and LF/(LF + HF) were significantly lower and RMSSD and HF were significantly higher during deep than during shallow sleep. These findings support previous studies that reported significant differences in CVRR, SDNN, RMSSD, LF, HF, and LF/HF among different sleep–wake conditions (120, 121). It is noteworthy that many previous studies, with several exceptions, ignored NN50 and pNN50. Furthermore, several studies investigating NN50 and pNN50 reported finding no significant correlations between these two HRV indicators and sleep–wake conditions (122–124). Contrary to these previous findings, the present study found that NN50 and pNN50 significantly differed among sleep–wake conditions. The major difference between the present and previous studies was the target participants: pregnant women in the present study, compared with non-pregnant women and men in previous studies.

NN50 is the number of interval differences of successive RR intervals greater than 50 ms, and pNN50 is the proportion dividing NN50 by the total number of RR intervals, both of which reflect vagal tone (125–128). Previous studies have revealed that vagal tone is altered during pregnancy. In general, vagal tone and reactivity are reported to decrease as pregnancy progresses (129–131). While CVRR, SDNN, LF, LF/HF, and LF/(LF + HF) are useful SNS markers for differentiating sleep–wake conditions, NN50 and pNN50 may not be distinguishable between different sleep–wake conditions among subjects without pregnancy. As pregnancy conditions alter the vagal tone, NN50 and pNN50 may constitute useful factors that vary according to and allow differentiation between sleep–wake conditions (significantly lower NN50 and pNN50 during wake conditions compared with sleep conditions).

In the present study, we investigated whether there were differences in HRV indicators between the wake conditions before and after sleep as an unprecedented trial. Interestingly, the same HRV indicators [CVRR, SDNN, NN50, pNN50, LF, LF/HF, and LF/(LF + HF)] showed significantly lower values during the wake condition before sleep than during that after sleep. In addition, HF was significantly higher during the wake condition before sleep than during that after sleep. These results suggest that there may be apparent differences in arousal levels and ANS conditions between the wake conditions before and after sleep, similar to the differences among the sleep–wake conditions.

Unlike previous studies, which applied a limited number of different algorithms to predict sleep–wake conditions based on HRV indicators, the present study conducted comprehensive evaluations of

widely used machine and deep learning algorithms. Among the 13 Python-based machine and deep learning algorithms, 12, except NB, provided high AUCs (0.82–0.88) and accuracy (0.78–0.81) for predicting sleep–wake conditions based on HRV indicators. As for the predictions of four conditions (deep sleep, shallow sleep, wake condition before sleep, and wake condition after sleep), different algorithms showed a wider range of AUCs and accuracy than did those in the prediction of three conditions (deep sleep, shallow sleep, and wake conditions). GRU showed the highest AUC and accuracy (0.86 and 0.79), respectively, followed by LSTM (0.78 and 0.73), RF (0.76 and 0.71), GBT (0.76 and 0.71), XGBoost (0.76 and 0.71), CNN (0.75 and 0.71), k-NN (0.75 and 0.67), ANN (0.74 and 0.70), DT (0.73 and 0.69), SVM (0.73 and 0.67), SGD (0.72 and 0.70), LR (0.72 and 0.67), and NB (0.61 and 0.42) in predicting the four conditions. These results suggested that the differences in ANS conditions among the three types of sleep–wake conditions were clear enough to be equally predictable by most algorithms.

In contrast, the differences in ANS conditions among the four types of sleep–wake conditions were not as evident as those among the three types of sleep–wake conditions. Therefore, a specific algorithm was needed to obtain high AUCs and accuracy for predicting the four conditions. Deep learning (GRU and LSTM) and ensemble learning methods in machine learning (RF, GBT, and XGBoost) may be suitable for differentiating between the wake conditions before and after sleep in predictions for the four conditions.

In previous studies conducted to predict sleep–wake conditions based on HRV indicators, one or several algorithms were tested in each study, as summarized in Table 3. In contrast, all previous studies testing deep learning algorithms used only one algorithm. Among the studies testing multiple algorithms, those carried out to test machine learning algorithms, including RF, consistently indicated that RF showed the highest prediction accuracy. The studies testing machine learning algorithms, including NB, indicated that NB showed lower prediction accuracy compared with the others. It is reasonable that the accuracy of NB (132–134) would be the lowest among these methods because it is a simple algorithm that has the advantage of predicting a binary classification. Therefore, NB is not suitable for predicting multiple classifications, as conducted in the present research for prediction of three types of sleep–wake conditions using nine HRV indicators.

GRU showed the highest AUC and accuracy of 0.86 and 0.79, respectively, in predicting the four conditions. As mentioned in the Methods section, we used GRU with a multi-head attention prediction model instead of GRU only. GRU with a multi-head attention prediction model can effectively improve the task processing effect. It calculates the probability weight of different data, improves the quality of hidden layer feature extraction, and reduces the problem of information loss in feature extraction. A multi-head attention mechanism is widely used in various sequence prediction tasks. A number of previous studies (135–137) reported that a model with an attention mechanism achieved high accuracy. We conducted a preliminary test using GRU without a multi-head attention mechanism, which resulted in a lower AUC and accuracy compared with GRU with a multi-head attention mechanism and LSTM (data not shown). GRU can automatically classify different sleep–wake conditions based on HRV indicators, and thus outperforms conventional methods because it considers the complex and cyclic characteristics of sleep and wakefulness.

TABLE 3 Previous machine and deep learning studies predicting sleep conditions based on heart rate variability indicators.

First author	Year	Training dataset (*n*)	Test dataset (*n*)	Total sample (*n*)	Machine learning algorithms	Algorithms to show higher accuracy
Bozkurt (40)	2018	50%	50%	8,452 epochs	SVM, k-NN	SVM, k-NN
Adnane (43)	2012	20%	80%	16	SVM	SVM
Piotrowski (41)	2017	k-fold cross-validation		4	k-NN, NB, ANN	ANN
Sharma (42)	2017	10-folds cross-validation		15,136 epochs	NB, k-NN, ANN, DT, RF	RF
Taran (46)	2020	10-folds cross-validation		8 (30 epochs)	DT, k-NN, Ensemble learning	Ensemble learning
Xiao (48)	2013	20%	80%	45	RF	RF
Mendez (51)	2010	10,000 leave-one-out cross-validation		20,000 epochs	HMM	HMM
Ebrahimi (44)	2008	N. A.	N. A.	7 (30 epochs)	ANN	ANN
Li (49)	2018	k-fold cross-validation		5,793 epochs	CNN	CNN
Radha (50)	2019	4-fold cross-validation		292 (541,214 epochs)	LSTM	LSTM
Erdenebayar (72)	2020	89 (80,316 epochs)	23 (20,079 epochs)	112 (100,395 epochs)	GRU	GRU

The table summarizes previous studies of first author, year of the publish, validation of training and test set, sample size, tested machine learning algorithm(s), and the algorithm(s) to show high accuracy to predict sleep conditions based on heart rate variability indicators. The previous machine learning and deep learning algorithms included support vector machine (SVM), k-nearest neighbor (k-NN), decision tree (DT), random forest (RF), gradient boosting trees (GBT), extreme gradient boosting (XGBoost), artificial neural network (ANN), Hidden Markov Model (HMM), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU).

RF-based evaluation of important features in the prediction indicated that pNN50 was the most important feature for predicting sleep–wake conditions and differentiating between the wake conditions before and after sleep, followed in order by RMSSD, NN50, SDNN, CVRR, HF, and LF. pNN50, RMSSD, NN50, SDNN, CVRR, HF, and LF may be sensitive SNS-based biomarkers for predicting sleep–wake conditions and differentiating between the wake conditions before and after sleep. Previous studies have indicated that RMSSD, SDNN (138), CVRR (139), HF (122, 139–141), and LF (139, 140) are useful features for predicting sleep stages, which supports our results. Although some previous studies have examined pNN50 and NN50, these features did not help differentiate between sleep stages (123, 124, 142). The present study noted that pNN50 and NN50 were important features for predicting sleep stages among pregnant women, suggesting that these features may be related to the ANS characteristics during pregnancy conditions. pNN50 and NN50 were identified as important features for predicting sleep–wake conditions and allow differentiation between the wake conditions before and after sleep among pregnant women probably because these features differed significantly among the sleep–wake conditions and the wake conditions before and after sleep.

The three cross-validation methods with automatic hyper-parameter tuning, including GridSearchCV, RandomizedSearchCV, and Bayesian optimization searches (117, 143), were also tested. Among these, RandomizedSearchCV achieved the highest accuracy with the shortest time consumption. GridSearchCV can ensure the accuracy of parameters within the specified parameter range by traversing all possible parameter combinations. However, this is very time-consuming and provides low accuracy in the case of large datasets and multiple parameters. Regarding hyper-parameter optimization, RandomizedSearchCV search is more effective than GridSearchCV. Some previous studies have indicated that Bayesian optimization searching provides the best accuracy (117, 144), whereas others (145) have indicated that the limitations of Bayesian

optimization sometimes makes its search effectiveness unstable and not significantly better than RandomizedSearchCV. Compared with RandomizedSearchCV, the drawback of Bayesian optimization is the greater consumption of computational resources, which may result in its taking longer to escape the local optimum, and its deployment in a distributed system (146). Memory consumption, training time, power consumption, and parallelism are essential for deep learning, while the feature of borrowing ideas from previous results prevents the Bayesian method from direct parallelization. Although many recent developments have solved this problem, it is still not as natural as RandomizedSearchCV, which is easy to combine with early stopping strategies. Such a combination could be expected to vastly improve the efficiency of narrowing down the search space (145). Regarding the performance of the model, while KCV and LOOCV showed almost equivalent performance, KCV was selected because LOOCV was time-consuming, taking longer to fit a dataset compared with KCV.

A verification study with alternative usages of JMP and WEKA, as well as Welch's method on Python to extract the HRV based on RR intervals, assured the validity of the present findings, which showed replicated prediction accuracies with slight differences due to the variability in parameter regulations.

The major finding of the present study in the clinical context is that vagal tone appears to be an important factor for differentiating between sleep–wake conditions, specifically among pregnant women. The findings also suggest that pregnancy conditions alter vagus nerve conditions in different ways between sleep and wake conditions. Another major finding in the clinical context is that HRV indicators can be useful for differentiating between not only sleep–wake conditions, but also the wake conditions before and after sleep. The wake condition before sleep may reflect a drowsy state, whereas that after sleep may reflect an alert state. By observing HRV indicators, drowsiness and readiness to falling asleep can be objectively evaluated. This information may be useful in interventions to improve sleep health in pregnant women.

5. Limitations

This research has several limitations. First, the sample size was relatively small ($N = 154$). In the future, the accuracy of the model should be verified with a greater quantity of data, and the most suitable algorithms should be selected. Second, sleep stages were defined based on the information collected through body motion sensors without recording EEGs; therefore, there was a limitation in terms of the accuracy of sleep stage determination. Third, REM sleep was not taken into account in the present study because it is not detectable by an accelerometer. Fourth, the types of machine learning and deep learning were limited. Transformer was not applicable because of the small number of sleep condition observations. After collecting more data, an advanced method such as transformer may be applicable in order to optimize prediction models and parameters in the future. Fifth, concerning wake conditions, among the two types of wake conditions, we arbitrarily extracted the 30 min of wakefulness before sleep and that immediately after waking up in the morning (i.e., the wake condition after sleep). In the future, different durations of observations, such as 15, 45, or 60 min (both before sleep and immediately after waking) could alternatively be applied to more completely investigate the nature of wake conditions. Finally, as this study focused on the heart rate of pregnant women, the possibility that the fetal heart rate might have interfered with the observed maternal heart rate remains. However, unless an ultrasound transducer is placed correctly over the fetal heart, the fetal heart rate signal cannot be reliably acquired. The disposable electrode placed on the maternal heart in this study rarely detects the fetal heart rate. Therefore, the observed heart rate should accurately reflect the maternal heart rate without considerable interference from the fetal heart rate.

6. Conclusion and future research

This research tested 10 machine and three deep learning algorithms to predict sleep–wake conditions based on HRV indicators. Most of the tested algorithms except NB could provide a suitably accurate prediction of three types of sleep–wake conditions (wake, shallow sleep, and deep sleep). GRU was the most accurate method for predicting four sleep conditions (including differentiation between two wake conditions): the wake conditions before and after sleep, shallow sleep, and deep sleep. Moreover, pNN50, RMSSD, NN50, SDNN, CVRR, HF, and LF were important features for predicting sleep–wake conditions and differentiating between the wake conditions before and after sleep.

Statement of significance

The present study is the first trial aiming to predict the sleep–wake condition and differentiate between the wake conditions before and after sleep in pregnant women. We tested 13 algorithms to predict four conditions (deep sleep, shallow sleep, and wake conditions before and after sleep) considering the difference in the arousal level of the two wake conditions. We successfully predicted the conditions by the gated recurrent unit with the highest area under the receiver operating characteristic curve (0.86) and accuracy (0.79). In addition, we demonstrated the usability of "the number of interval differences

of successive RR intervals greater than 50 ms (NN50)" and "the proportion dividing NN50 by the total number of RR intervals (pNN50)" to predict the sleep–wake condition and differentiate between the wake conditions before and after sleep unique to pregnancy. These findings suggest the existence of alterations in the vagal tone system during pregnancy.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the ethics committee of Graduate School of Medicine, Tohoku University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XL took major role in the data analysis. CO, ZY, YT, KI, NS, NH, and HT also contributed to the data analysis. CO, NW, TS, TaN, HU, KM, MI, TO, FN, NF, JS, ShK, MY, NY, and HT contributed to the acquisition of data. CO, ToN, TT, SO, and GT contributed to the data management. XL, CO, KI, NS, NK, SaK, RK, YH, MH, YK, JS, TH, MS, SF, MN, ShK, NH, and HT contributed to the interpretation of the data. XL and HT drafted the manuscript. KI, NS, NK, SaK, SF, NH, and HT critically revised the manuscript for important scientific content. XL, CO, NW, FN, ShK, and HT made substantial contributions to the conception and design of the study. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1104222/full#supplementary-material

References

1. Li X, Lu Y, Fu X, Qi Y. Building the internet of things platform for smart maternal healthcare services with wearable devices and cloud computing. *Futur Gener Comput Syst*. (2021) 118:282–96. doi: 10.1016/j.future.2021.01.016

2. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep*. (1992) 15:461–9. doi: 10.1093/sleep/15.5.461

3. Paquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. *Sleep*. (2007) 30:1362–9. doi: 10.1093/sleep/30.10.1362

4. Liang Z, Chapa Martell MA. Validity of consumer activity wristbands and wearable EEG for measuring overall sleep parameters and sleep structure in free-living conditions. *J Healthc Inform Res*. (2018) 2:152–78. doi: 10.1007/s41666-018-0013-1

5. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act*. (2015) 12:159. doi: 10.1186/s12966-015-0314-1

6. Van de Water AT, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography – a systematic review. *J Sleep Res*. (2011) 20:183–200. doi: 10.1111/j.1365-2869.2009.00814.x

7. McCall C, McCall WV. Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs. *J Sleep Res*. (2012) 21:122–7. doi: 10.1111/j.1365-2869.2011.00917.x

8. Mundt JM, Crew EC, Krietsch K, Roth AJ, Vatthauer K, Robinson ME, et al. Measuring treatment outcomes in comorbid insomnia and fibromyalgia: concordance of subjective and objective assessments. *J Clin Sleep Med*. (2016) 12:215–23. doi: 10.5664/jcsm.5488

9. Choi SJ, Kang M, Sung MJ, Joo EY. Discordant sleep parameters among actigraphy, polysomnography, and perceived sleep in patients with sleep-disordered breathing in comparison with patients with chronic insomnia disorder. *Sleep Breath*. (2017) 21:837–43. doi: 10.1007/s11325-017-1514-5

10. Scott H, Lack L, Lovato N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Med Rev*. (2020) 49:101227. doi: 10.1016/j.smrv.2019.101227

11. Hedman C, Pohjasvaara T, Tolonen U, Suhonen-Malm AS, Myllylä VV. Effects of pregnancy on mothers' sleep. *Sleep Med*. (2002) 3:37–42. doi: 10.1016/s1389-9457(01)00130-7

12. Mindell JA, Cook RA, Nikolovski J. Sleep patterns and sleep disturbances across pregnancy. *Sleep Med*. (2015) 16:483–8. doi: 10.1016/j.sleep.2014.12.006

13. Pien GW, Fife D, Pack AI, Nkwuo JE, Schwab RJ. Changes in symptoms of sleep-disordered breathing during pregnancy. *Sleep*. (2005) 28:1299–305. doi: 10.1093/sleep/28.10.1299

14. Facco FL, Kramer J, Ho KH, Zee PC, Grobman WA. Sleep disturbances in pregnancy. *Obstet Gynecol*. (2010) 115:77–83. doi: 10.1097/AOG.0b013e3181c4f8ec

15. Hutchison BL, Stone PR, McCowan LM, Stewart AW, Thompson JM, Mitchell EA. A postal survey of maternal sleep in late pregnancy. *BMC Pregnancy Childbirth*. (2012) 12:144. doi: 10.1186/1471-2393-12-144

16. Neau JP, Texier B, Ingrand P. Sleep and vigilance disorders in pregnancy. *Eur Neurol*. (2009) 62:23–9. doi: 10.1159/000215877

17. Christian LM, Carroll JE, Teti DM, Hall MH. Maternal sleep in pregnancy and postpartum part I: mental, physical, and interpersonal consequences. *Curr Psychiatry Rep*. (2019) 21:20. doi: 10.1007/s11920-019-0999-y

18. Wilson DL, Barnes M, Ellett L, Permezel M, Jackson M, Crowe SF. Decreased sleep efficiency, increased wake after sleep onset and increased cortical arousals in late pregnancy. *Aust N Z J Obstet Gynaecol*. (2011) 51:38–46. doi: 10.1111/j.1479-828X.2010.01252.x

19. Antrobus J. REM and NREM sleep reports: comparison of word frequencies by cognitive classes. *Psychophysiology*. (1983) 20:562–8. doi: 10.1111/j.1469-8986.1983.tb03015.x

20. McCarley RW. Neurobiology of REM and NREM sleep. *Sleep Med*. (2007) 8:302–30. doi: 10.1016/j.sleep.2007.03.005

21. Nielsen TA. A review of mentation in REM and NREM sleep: "covert" REM sleep as a possible reconciliation of two opposing models. *Behav Brain Sci*. (2000) 23:851-66; discussion 904-1121–66. doi: 10.1017/s0140525x0000399x

22. Tobaldini E, Costantino G, Solbiati M, Cogliati C, Kara T, Nobili L, et al. Sleep, sleep deprivation, autonomic nervous system and cardiovascular diseases. *Neurosci Biobehav Rev*. (2017) 74:321–9. doi: 10.1016/j.neubiorev.2016.07.004

23. Burgess HJ, Trinder J, Kim Y, Luke D. Sleep and circadian influences on cardiac autonomic nervous system activity. *Am J Phys*. (1997) 273:H1761–8. doi: 10.1152/ajpheart.1997.273.4.H1761

24. Takeuchi S, Iwase S, Mano T, Okada H, Sugiyama Y, Watanabe T. Sleep-related changes in human muscle and skin sympathetic nerve activities. *J Auton Nerv Syst*. (1994) 47:121–9. doi: 10.1016/0165-1838(94)90073-6

25. Varoneckas G, Plauška K, Kauk J. Components of the heart rhythm power spectrum in wakefulness and individual sleep stages. *Int J Psychophysiol*. (1986) 4:129–41. doi: 10.1016/0167-8760(86)90006-1

26. Vaughn BV, Quint SR, Messenheimer JA, Robertson KR. Heart period variability in sleep. *Electroencephalogr Clin Neurophysiol*. (1995) 94:155–62. doi: 10.1016/0013-4694(94)00270-u

27. Negoescu RM, Csiki IE. Autonomic control of the heart in some vagal maneuvers and normal sleep. *Physiologie*. (1989) 26:39–49.

28. Orr W, Lin B, Adamson P, Vanoli E. Autonomic control of heart rate variability during sleep. *Sleep Res*. (1993) 22:26.

29. Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. *Med Biol Eng Comput*. (2006) 44:1031–51. doi: 10.1007/s11517-006-0119-0

30. Sztajzel J. Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Med Wkly*. (2004) 134:514–22. doi: 10.4414/smw.2004.10321

31. van Ravenswaaij-Arts CM, Kollée LA, Hopman JC, Stoelinga GB, van Geijn HP. Heart rate variability. *Ann Intern Med*. (1993) 118:436–47. doi: 10.7326/0003-4819-118-6-199303150-00008

32. Gil E, Mendez M, Vergara JM, Cerutti S, Bianchi AM, Laguna P. Discrimination of sleep-apnea-related decreases in the amplitude fluctuations of PPG signal in children by HRV analysis. *IEEE Trans Biomed Eng*. (2009) 56:1005–14. doi: 10.1109/tbme.2008.2009340

33. Penzel T, Kantelhardt JW, Grote L, Peter JH, Bunde A. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. *IEEE Trans Biomed Eng*. (2003) 50:1143–51. doi: 10.1109/tbme.2003.817636

34. Stein PK, Pu Y. Heart rate variability, sleep and sleep disorders. *Sleep Med Rev*. (2012) 16:47–66. doi: 10.1016/j.smrv.2011.02.005

35. Okun ML. Disturbed sleep and postpartum depression. *Curr Psychiatry Rep*. (2016) 18:66. doi: 10.1007/s11920-016-0705-2

36. Bylsma LM, Salomon K, Taylor-Clift A, Morris BH, Rottenberg J. Respiratory sinus arrhythmia reactivity in current and remitted major depressive disorder. *Psychosom Med*. (2014) 76:66–73. doi: 10.1097/psy.0000000000000019

37. da Estrela C, McGrath J, Booij L, Gouin JP. Heart rate variability, sleep quality, and depression in the context of chronic stress. *Ann Behav Med*. (2021) 55:155–64. doi: 10.1093/abm/kaaa039

38. Keller PS, Kouros CD, Erath SA, Dahl RE, El-Sheikh M. Longitudinal relations between maternal depressive symptoms and child sleep problems: the role of parasympathetic nervous system reactivity. *J Child Psychol Psychiatry*. (2014) 55:172–9. doi: 10.1111/jcpp.12151

39. Werner GG, Ford BQ, Mauss IB, Schabus M, Blechert J, Wilhelm FH. Cardiac vagal control and depressive symptoms: the moderating role of sleep quality. *Behav Sleep Med*. (2017) 15:451–65. doi: 10.1080/15402002.2016.1150280

40. Bozkurt MR, Uçar MK, Bozkurt F, Bilgin C. In obstructive sleep apnea patients, automatic determination of respiratory arrests by photoplethysmography signal and heart rate variability. *Australas Phys Eng Sci Med*. (2019) 42:959–79. doi: 10.1007/s13246-019-00796-9

41. Piotrowski Z, Szypulska M. Classification of falling asleep states using HRV analysis. *Biocybern Biomed Eng*. (2017) 37:290–301. doi: 10.1016/j.bbe.2017.02.003

42. Sharma R, Pachori RB, Upadhyay A. Automatic sleep stages classification based on iterative filtering of electroencephalogram signals. *Neural Comput Applic*. (2017) 28:2959–78. doi: 10.1007/s00521-017-2919-6

43. Adnane M, Jiang Z, Yan Z. Sleep–wake stages classification and sleep efficiency estimation using single-lead electrocardiogram. *Expert Syst Appl*. (2012) 39:1401–13. doi: 10.1016/j.eswa.2011.08.022

44. Ebrahimi F, Mikaeili M, Estrada E, Nazeran H. Automatic sleep stage classification based on EEG signals by using neural networks and wavelet packet coefficients. *Annu Int Conf IEEE Eng Med Biol Soc*. (2008) 2008:1151–4. doi: 10.1109/iembs.2008.4649365

45. Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhaus H. Automated sleep stage identification system based on time-frequency analysis of a single EEG Channel and random forest classifier. *Comput Methods Prog Biomed*. (2012) 108:10–9. doi: 10.1016/j.cmpb.2011.11.005

46. Taran S, Sharma PC, Bajaj V. Automatic sleep stages classification using optimize flexible analytic wavelet transform. *Knowl-Based Syst*. (2020) 192:105367. doi: 10.1016/j.knosys.2019.105367

47. Wang Y, Wang D, Zhang L, Liu C, Li J, Hou F, et al. Automatic identification of rapid eye movement sleep based on random Forest using heart rate variability. *Phys A Stat Mech Appl*. (2019) 527:121421. doi: 10.1016/j.physa.2019.121421

48. Xiao M, Yan H, Song J, Yang Y, Yang X. Sleep stages classification based on heart rate variability and random forest. *Biomed Signal Process Control*. (2013) 8:624–33. doi: 10.1016/j.bspc.2013.06.001

49. Li Q, Li Q, Liu C, Shashikumar SP, Nemati S, Clifford GD. Deep learning in the cross-time frequency domain for sleep staging from a single-lead electrocardiogram. *Physiol Meas*. (2018) 39:124005. doi: 10.1088/1361-6579/aaf339

50. Radha M, Fonseca P, Moreau A, Ross M, Cerny A, Anderer P, et al. Sleep stage classification from heart-rate variability using long short-term memory neural networks. *Sci Rep*. (2019) 9:14149. doi: 10.1038/s41598-019-49703-y

51. Mendez MO, Matteucci M, Castronovo V, Ferini-Strambi L, Cerutti S, Bianchi A. Sleep staging from heart rate variability: time-varying spectral features and hidden Markov models. *Int J Biomed Eng Technol*. (2010) 3:246–63. doi: 10.1504/IJBET.2010.032695

52. Kikuchi S, Murakami K, Obara T, Ishikuro M, Ueno F, Noda A, et al. One-year trajectories of postpartum depressive symptoms and associated psychosocial factors: findings from the Tohoku Medical Megabank Project birth and three-generation cohort Study. *J Affect Disord*. (2021) 295:632–8. doi: 10.1016/j.jad.2021.08.118

53. Kuriyama S, Metoki H, Kikuya M, Obara T, Ishikuro M, Yamanaka C, et al. Cohort profile: Tohoku Medical Megabank Project birth and three-generation cohort Study (TMM BirThree Cohort Study): rationale, progress and perspective. *Int J Epidemiol*. (2020) 49:18–9m. doi: 10.1093/ije/dyz169

54. Kuriyama S, Yaegashi N, Nagami F, Arai T, Kawaguchi Y, Osumi N, et al. The Tohoku Medical Megabank Project: design and mission. *J Epidemiol*. (2016) 26:493–511. doi: 10.2188/jea.JE20150268

55. Kodama M. E-healthcare service innovations: in depth case studies in Japan In: *Collaborative dynamic capabilities for service innovation: creating a new healthcare ecosystem*. Palgrave Macmillan, Cham (2018). 91–111.

56. Yamauchi T, Hiyama S. (2016) Research laboratories Daisuke Ochi 2016: 23–29.

57. Fonseca P, van Gilst MM, Radha M, Ross M, Moreau A, Cerny A, et al. Automatic sleep staging using heart rate variability, body movements, and recurrent neural networks in a sleep disordered population. *Sleep*. (2020) 43:zsaa048. doi: 10.1093/sleep/zsaa048

58. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Deep neural network sleep scoring using combined motion and heart rate variability data. *Sensors (Basel)*. (2020) 21:25. doi: 10.3390/s21010025

59. Newton AT. *Validity of a commercially-available, low-cost, wrist-mounted accelerometer in a laboratory and free-living environment*. Kent, OH: Kent State University (2016).

60. Hashimoto K, Harada N, Kasamaki Y. Can a patch electrocardiographic device be a leading actor for detecting atrial fibrillation? Diversifying electrocardiographic monitoring devices. *Circ J*. (2022) 86:189–91. doi: 10.1253/circj.CJ-21-0644

61. Okubo Y, Tokuyama T, Okamura S, Ikeuchi Y, Miyauchi S, Nakano Y. Evaluation of the feasibility and efficacy of a novel device for screening silent atrial fibrillation (Mybeat trial). *Circ J*. (2022) 86:182–8. doi: 10.1253/circj.CJ-20-1061

62. Camm AJ, Malik M, Bigger JT, Breithardt G, Cerutti S, Cohen RJ, et al. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*. (1996) 17:354–81. doi: 10.1093/oxfordjournals.eurheartj.a014868

63. Scheaffer SM, Lee D, Whitener B, Ying B, Wu K, Jani H, et al. Bivalent Sars-Cov-2 Mrna vaccines increase breadth of neutralization and protect against the Ba. 5 omicron variant. *Nat Med*. (2023) 29:247–57. doi: 10.1038/s41591-022-02092-8

64. Wang H, Fan M, An Y, He D. Molecular mechanism of Long noncoding RNA SNHG 14 in osteogenic differentiation of bone marrow-derived mesenchymal stem cells through the NEDD 4l/FOXA2/PCP 4 axis. *Stem Cells Int*. (2023) 2023:7545635. doi: 10.1155/2023/7545635

65. Zhang X, Scheijen JL, Stehouwer CDA, Wouters K, Schalkwijk C. Increased methylglyoxal formation in plasma and tissues during a glucose tolerance test is derived from exogenous glucose. *Clin Sci (Lond)*. (2023) 137:697–706. doi: 10.1042/cs20220753

66. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adam M, Gertych A, et al. A deep convolutional neural network model to classify heartbeats. *Comput Biol Med*. (2017) 89:389–96. doi: 10.1016/j.compbiomed.2017.08.022

67. Agarap AF. (2017). An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. *arXiv*. doi: 10.48550/arXiv.1712.03541. [Epub ahead of preprint].

68. Banerjee I, Ling Y, Chen MC, Hasan SA, Langlotz CP, Moradzadeh N, et al. Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. *Artif Intell Med*. (2019) 97:79–88. doi: 10.1016/j.artmed.2018.11.004

69. Bottou L, Bousquet O. *The tradeoffs of large scale learning, advances in neural information processing systems, 20*. Cambridge, MA: MIT Press (2008).

70. Chueh T-H, Chen T-B, Lu HH-S, Ju S-S, Tao T-H, Shaw J-H. Statistical prediction of emotional states by physiological signals with manova and machine learning. *Int J Pattern Recognit Artif Intell*. (2012) 26:1250008. doi: 10.1142/s0218001412500085

71. Arbib MA. *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press (2002).

72. Erdenebayar U, Kim Y, Park J-U, Lee S, Lee K-J. Automatic classification of sleep stage from an ECG signal using a gated-recurrent unit. *Int J Fuzzy Logic Intell Syst*. (2020) 20:181–7. doi: 10.5391/ijfis.2020.20.3.181

73. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw*. (2005) 18:602–10. doi: 10.1016/j.neunet.2005.06.042

74. Jang E-H, Park B-J, Kim S-H, Chung M-A, Park M-S, Sohn J-H, editors. (2014). Emotion classification based on bio-signals emotion recognition using machine learning algorithms. 2014 International conference on information science, electronics and electrical engineering; IEEE. Available at: https://doi.org/10.1109/infoseee.2014.6946144

75. Joo S, Choi K-J, Huh S-J. Prediction of spontaneous ventricular tachyarrhythmia by an artificial neural network using parameters gleaned from short-term heart rate variability. *Expert Syst Appl*. (2012) 39:3862–6. doi: 10.1016/j.eswa.2011.09.097

76. Karim F, Majumdar S, Darabi H, Harford S. Multivariate LSTM-FCNS for time series classification. *Neural Netw*. (2019) 116:237–45. doi: 10.1016/j.neunet.2019.04.014

77. Kelwade J, Salankar S. Prediction of cardiac arrhythmia using artificial neural network. *Int J Comput Appl*. (2015) 115:30–5. doi: 10.5120/20270-2679

78. Kido S, Hirano Y, Hashimoto N, editors. (2018). Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN). 2018 International workshop on advanced image technology (IWAIT); IEEE. Available at: https://doi.org/10.1109/iwait.2018.8369798

79. Kumar S, Hussain L, Banerjee S, Reza M, editors. (2018). Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster. 2018 Fifth international conference on emerging applications of information technology (EAIT); IEEE. Available at: https://doi.org/10.1109/eait.2018.8470406

80. Lee HG, Noh KY, Ryu KH. (2007). Mining biosignal data: coronary artery disease diagnosis using linear and nonlinear features of HRV. Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer 2007: 218–228.

81. Li Y, Hao Z, Lei H. Survey of convolutional neural network. *J Comput Appl*. (2016) 36:2508. doi: 10.11772/j.issn.1001-9081.2016.09.2508

82. Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing*. (2019) 337:325–38. doi: 10.1016/j.neucom.2019.01.078

83. Natarajan S, Prabhakar A, Ramanan N, Bagilone A, Siek K, Connelly K, editors. (2017). Boosting for postpartum depression prediction. 2017 IEEE/ACM international conference on connected health: applications, systems and engineering technologies (CHASE); IEEE. Available at: https://doi.org/10.1109/chase.2017.82

84. Plewa L, Student CO. Istress: stress classification from heart rate variability. *Cal Poly: CSC*. (2015) 520:96.

85. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324

86. Tao Q, Liu F, Li Y, Sidorov D. Air pollution forecasting using a deep learning model based on 1D Convnets and bidirectional GRU. *IEEE Access*. (2019) 7:76690–8. doi: 10.1109/access.2019.2921578

87. Urtnasan E, Park J-U, Lee S, Lee K-J. Optimal classifier for detection of obstructive sleep apnea using a heartbeat signal. *Int J Fuzzy Logic Intell Syst*. (2017) 17:76–81. doi: 10.5391/ijfis.2017.17.2.76

88. Wang S, Pathak J, Zhang Y. Using electronic health records and machine learning to predict postpartum depression. *Stud Health Technol Inform*. (2019) 264:888–92. doi: 10.3233/shti190351

89. Yoo SK, Lee CK, Park YJ, Kim NH, Lee BC, Jeong KS, editors. (2005). Neural network based emotion estimation using heart rate variability and skin resistance. Advances in natural computation: first international conference, ICNC 2005, Changsha, China, August 27-29, 2005, Proceedings, Part I 1; Springer.

90. Zhang D, Kabuka MR. Combining weather condition data to predict traffic flow: a GRU-based deep learning approach. *IET Intell Transp Syst*. (2018) 12:578–85. doi: 10.1049/iet-its.2017.0313

91. Zhang W, Liu H, Silenzio VMB, Qiu P, Gong W. Machine learning models for the prediction of postpartum depression: application and comparison based on a cohort Study. *JMIR Med Inform*. (2020) 8:e15516. doi: 10.2196/15516

92. Li X, Ono C, Warita N, Shoji T, Nakagawa T, Usukura H, et al. Heart rate information-based machine learning prediction of emotions among pregnant women. *Front Psych*. (2021) 12:799029. doi: 10.3389/fpsyt.2021.799029

93. Bowes D, Hall T, Gray D, editors. (2012). Comparing the performance of fault prediction models which report multiple performance measures: recomputing the confusion matrix. Proceedings of the 8th international conference on predictive models in software engineering.

94. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag*. (2009) 45:427–37. doi: 10.1016/j.ipm.2009.03.002

95. Belgiu M, Drăguţ L. Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens*. (2016) 114:24–31. doi: 10.1016/j.isprsjprs.2016.01.011

96. Mursalin M, Zhang Y, Chen Y, Chawla NV. Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. *Neurocomputing*. (2017) 241:204–14. doi: 10.1016/j.neucom.2017.02.053

97. Paul D, Su R, Romain M, Sébastien V, Pierre V, Isabelle G. Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier. *Comput Med Imaging Graph*. (2017) 60:42–9. doi: 10.1016/j.compmedimag.2016.12.002

98. Palczewska A, Palczewski J, Marchese Robinson R, Neagu D. Interpreting random Forest classification models using a feature contribution method. *Integr Reusable Syst*. (2014) 263:193–218. doi: 10.1016/j.compmedimag.2016.12.002

99. Palczewska A, Palczewski J, Robinson RM, Neagu D, editors. (2013). Interpreting random Forest models using a feature contribution method. 2013 IEEE 14th international conference on Information Reuse & Integration (IRI); IEEE. Available at: https://doi.org/10.1109/iri.2013.6642461

100. Whitmore LS, George A, Hudson CM. (2018). Explicating feature contribution using random Forest proximity distances. *arXiv*. doi: 10.48550/arXiv.1807.06572. [Epub ahead of preprint].

101. Frank E, Hall M, Holmes G, Kirkby R, Pfahringer B, Witten IH, et al. Weka – a machine learning workbench for data mining In: *Data mining and knowledge discovery handbook*. Boston, MA: Springer (2010). 1269–77.

102. Markov Z, Russell I. An introduction to the Weka data mining system. *ACM SIGCSE Bull*. (2006) 38:367–8. doi: 10.1145/1140123.1140127

103. Lehman A, O'Rourke N, Hatcher L, Stepanski E. *JMP for basic univariate and multivariate statistics: methods for researchers and social scientists*. Cary, NC: SAS Institute (2013).

104. Sall J, Stephens ML, Lehman A, Loring S. *JMP start statistics: a guide to statistics and data analysis using JMP*. Cary, NC: SAS Institute (2017).

105. Doret M, Spilka J, Chudáček V, Gonçalves P, Abry P. Fractal analysis and hurst parameter for intrapartum fetal heart rate variability analysis: a versatile alternative to frequency bands and LF/HF ratio. *PLoS One*. (2015) 10:e0136661. doi: 10.1371/journal.pone.0136661

106. Glos M, Romberg D, Fietze I, Rottig J, Knobe M, Witt C, editors. (1999). Analysis of heart rate and blood pressure variability during nasal continuous positive airway pressure therapy in patients with obstructive sleep apnea. Computers in Cardiology 1999 Vol 26 (Cat No 99CH37004); IEEE. Available at: https://doi.org/10.1109/cic.1999.826043

107. Posada-Quintero HF, Florian JP, Orjuela-Cañón AD, Aljama-Corrales T, Charleston-Villalobos S, Chon KH. Power spectral density analysis of electrodermal activity for sympathetic function assessment. *Ann Biomed Eng*. (2016) 44:3124–35. doi: 10.1007/s10439-016-1606-6

108. Schaffer T, Hensel B, Weigand C, Schüttler J, Jeleazcov C. Evaluation of techniques for estimating the power spectral density of RR-intervals under paced respiration conditions. *J Clin Monit Comput*. (2014) 28:481–6. doi: 10.1007/s10877-013-9447-4

109. Verma A, Cabrera S, Mayorga A, Nazeran H, editors. (2013). A robust algorithm for derivation of heart rate variability spectra from ECG and PPG signals. 2013 29th southern biomedical engineering conference; IEEE. Available at: https://doi.org/10.1109/sbec.2013.26

110. Wang H-M, Huang S-C. SDNN/RMSSD as a surrogate for LF/HF: a revised investigation. *Modell Simul Eng*. (2012) 2012:1–8. doi: 10.1155/2012/931943

111. Ahmed MR, Ahammed MS, Niu S, Zhang Y, editors. (2020). Deep learning approached features for ASD classification using SVM. 2020 IEEE international conference on artificial intelligence and information systems (ICAIIS); IEEE. Available at: https://doi.org/10.1109/icaiis49377.2020.9194791

112. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. (2013). API design for machine learning software: experiences from the Scikit-learn project. *arXiv*. doi: 10.48550/arXiv.1309.0238. [Epub ahead of preprint].

113. Ranjan G, Verma AK, Radhika S, editors. (2019). K-nearest neighbors and grid search CV based real time fault monitoring system for industries. 2019 IEEE 5th international conference for convergence in technology (I2CT); IEEE. Available at: https://doi.org/10.1109/i2ct45611.2019.9033691

114. Bisong E. *More supervised machine learning techniques with Scikit-learn. Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners*. (2019): 287–308. Berkeley, CA: Apress

115. Kirori Z. Hyper-parameter optimization: toward practical sentiment analysis using a convolutional neural network. *Res J Comput Inf*. (2019) 7:1–5.

116. Ng Y-L, Lo C-K, Lee K-H, Xie X, Kwong TN, Ip M, et al. Supporting information to" development of an open-access and explainable machine learning prediction system to assess the mortality and recurrence risk factors of clostridioides difficile infection patients: model training and hyperparameter optimization with cross-validation". *Authorea Preprints*. (2020) 3:2000188. doi: 10.1002/aisy.202000188

117. Ke H, Chen D, Shi B, Zhang J, Liu X, Zhang X, et al. Improving brain E-health services via high-performance EEG classification with grouping Bayesian optimization. *IEEE Trans Serv Comput*. (2019) 13:696–708. doi: 10.1109/tsc.2019.2962673

118. Snoek J, Larochelle H, Adams RP. (2012). Practical Bayesian optimization of machine learning algorithms. Advances in Neural Information Processing Systems 25. Available at: https://doi.org/10.48550/arXiv.1206.2944

119. Wu J, Chen X-Y, Zhang H, Xiong L-D, Lei H, Deng S-H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J Electron Sci Technol*. (2019) 17:26–40. doi: 10.11989/JEST.1674-862X.80904120

120. Ebrahimi F, Setarehdan SK, Ayala-Moyeda J, Nazeran H. Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals. *Comput Methods Prog Biomed*. (2013) 112:47–57. doi: 10.1016/j.cmpb.2013.06.007

121. Versace F, Mozzato M, De Min TG, Cavallero C, Stegagno L. Heart rate variability during sleep as a function of the sleep cycle. *Biol Psychol*. (2003) 63:149–62. doi: 10.1016/s0301-0511(03)00052-8

122. Bonnet MH, Arand DL. Heart rate variability: sleep stage, time of night, and arousal influences. *Electroencephalogr Clin Neurophysiol*. (1997) 102:390–6. doi: 10.1016/s0921-884x(96)96070-1

123. Ferri R, Parrino L, Smerieri A, Terzano MG, Elia M, Musumeci SA, et al. Cyclic alternating pattern and spectral analysis of heart rate variability during normal sleep. *J Sleep Res*. (2000) 9:13–8. doi: 10.1046/j.1365-2869.2000.00190.x

124. Hall M, Vasko R, Buysse D, Ombao H, Chen Q, Cashmere JD, et al. Acute stress affects heart rate variability during sleep. *Psychosom Med*. (2004) 66:56–62. doi: 10.1097/01.psy.0000106884.58744.09

125. Braeken MA, Jones A, Otte RA, Widjaja D, Van Huffel S, Monsieur GJ, et al. Anxious women do not show the expected decrease in cardiovascular stress responsiveness as pregnancy advances. *Biol Psychol*. (2015) 111:83–9. doi: 10.1016/j.biopsycho.2015.08.007

126. Chrysostomakis SI, Simantirakis EN, Schiza SE, Karalis IK, Klapsinos NC, Siafakas NM, et al. Continuous positive airway pressure therapy lowers vagal tone in patients with obstructive sleep Apnoea-Hypopnoea syndrome. *Hell J Cardiol* (2006) 47(1): 13–20. Available at: https://hellenicjcardiol.org/archive/full_text/2006/1/2006_1_13.pdf

127. Lee JH, Huh IY, Lee JM, Lee HK, Han IS, Kang HJ. Relation of various parameters used to estimate cardiac vagal activity and validity of Pnn50 in anesthetized humans. *Kosin Med J*. (2018) 33:369–79. doi: 10.7180/kmj.2018.33.3.369

128. Sztajzel J, Jung M, Sievert K, Bayes De Luna A. Cardiac autonomic profile in different sports disciplines during all-day activity. *J Sports Med Phys Fitness*. (2008) 48:495–501.

129. DiPietro JA, Costigan KA, Gurewitsch ED. Maternal psychophysiological change during the second half of gestation. *Biol Psychol*. (2005) 69:23–38. doi: 10.1016/j.biopsycho.2004.11.003

130. Fantozzi MPT, Artoni F, Faraguna U. Heart rate variability at bedtime predicts subsequent sleep features. *Annu Int Conf IEEE Eng Med Biol Soc*. (2019) 2019:6784–8. doi: 10.1109/embc.2019.8857844

131. Walther T, Wessel N, Baumert M, Stepan H, Voss A, Faber R. Longitudinal analysis of heart rate variability in chronic hypertensive pregnancy. *Hypertens Res*. (2005) 28:113–8. doi: 10.1291/hypres.28.113

132. Diab S. (2019). Optimizing stochastic gradient descent in text classification based on fine-tuning hyper-parameters approach. A case study on automatic classification of global terrorist attacks. *arXiv*. doi: 10.48550/arXiv.1902.06542. [Epub ahead of preprint].

133. Lin Y, Lv F, Zhu S, Yang M, Cour T, Yu K, et al., editors. (2011). Large-scale image classification: fast feature extraction and SVM training. CVPR 2011; IEEE. Available at: https://doi.org/10.1109/cvpr.2011.5995477

134. Shirouzu S, Seno Y, Tobioka K, Yagi T, Takahashi T, Sasaki M et al., editors. (2016). For children's sleep assessment: can we trace the change of sleep depth based on ECG data measured at their respective home with a wearable device? 2016 IEEE-EMBS international conference on biomedical and health informatics (BHI); IEEE. Available at: https://doi.org/10.1109/bhi.2016.7455871

135. Chen J, Jiang D, Zhang Y. A hierarchical bidirectional GRU model with attention for EEG-based emotion classification. *IEEE Access*. (2019) 7:118530–40. doi: 10.1109/access.2019.2936817

136. Khan R, Islam MS, Kanwal K, Iqbal M, Hossain M, Ye Z. (2022). A deep neural framework for image caption generation using GRU-based attention mechanism. *arXiv*. doi: 10.48550/arXiv.2203.01594. [Epub ahead of preprint].

137. Zhang B, Xiong D, Xie J, Su J. Neural machine translation with GRU-gated attention model. *IEEE Trans Neural Netw Learn Syst*. (2020) 31:4688–98. doi: 10.1109/tnnls.2019.2957276

138. Otzenberger H, Gronfier C, Simon C, Charloux A, Ehrhart J, Piquard F, et al. Dynamic heart rate variability: a tool for exploring sympathovagal balance continuously during sleep in men. *Am J Phys*. (1998) 275:H946–50. doi: 10.1152/ajpheart.1998.275.3.H946

139. Toscani L, Gangemi PF, Parigi A, Silipo R, Ragghianti P, Sirabella E, et al. Human heart rate variability and sleep stages. *Ital J Neurol Sci*. (1996) 17:437–9. doi: 10.1007/bf01997720

140. Herzig D, Eser P, Omlin X, Riener R, Wilhelm M, Achermann P. Reproducibility of heart rate variability is parameter and sleep stage dependent. *Front Physiol*. (2017) 8:1100. doi: 10.3389/fphys.2017.01100

141. Villa MP, Calcagnini G, Pagani J, Paggi B, Massa F, Ronchetti R. Effects of sleep stage and age on short-term heart rate variability during sleep in healthy infants and children. *Chest*. (2000) 117:460–6. doi: 10.1378/chest.117.2.460

142. Wennerblom B, Lurje L, Tygesen H, Vahisalo R, Hjalmarson A. Patients with uncomplicated coronary artery disease have reduced heart rate variability mainly affecting vagal tone. *Heart*. (2000) 83:290–4. doi: 10.1136/heart.83.3.290

143. Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N. Taking the human out of the loop: a review of Bayesian optimization. *Proc IEEE*. (2015) 104:148–75. doi: 10.1109/jproc.2015.2494218

144. Klein A, Falkner S, Bartels S, Hennig P, Hutter F. Fast Bayesian Hyperparameter optimization on large datasets. *Electron J Statist*. (2017) 11:4945–68. doi: 10.1214/17-ejs1335si

145. Yu T, Zhu H. (2020). Hyper-parameter optimization: a review of algorithms and applications. *arXiv*. doi: 10.48550/arXiv.2003.05689. [Epub ahead of preprint].

146. Tong Y, Yu B. Research on hyper-parameter optimization of activity recognition algorithm based on improved cuckoo search. *Entropy (Basel)*. (2022) 24:845. doi: 10.3390/e24060845