scientific reports

Check for updates

OPEN

Machine learning-based prediction of restless legs syndrome using digital phenotypes from wearables and smartphone data

Jingyeong Jeong[1,7], Yoonseo Jeon[1,7], Hyungju Kim[2,7], Ji Won Yeom[3], Yu-Bin Shin[3], Sujin Kim[3], Seung Pil Pack[4], Heon-Jeong Lee[3], Taesu Cheong[2] & Chul-Hyun Cho[3,5,6]✉

Restless legs syndrome (RLS) is a relatively common neurosensory disorder that causes an irresistible urge for leg movement. RLS causes sleep disturbances and reduced quality of life, but accurate diagnosis remains challenging owing to the reliance on subjective reporting. This study aimed to propose a predictive machine learning model based on digital phenotypes for RLS diagnosis. Self-reported lifestyle data were integrated via a smartphone application with objective biometric data from wearable devices to obtain 85 features processed based on circadian rhythms. Prediction models used these features to distinguish between the non-RLS (International Restless Legs Study Group Severity Rating Scale [IRLS] score ≤ 10) and RLS symptom groups (10 < IRLS ≤ 20) and between the non-RLS and severe RLS symptom groups (IRLS > 20). The RF model showed the highest performance in predicting the RLS symptom group and XGB model in the severe RLS symptom group. For the RLS symptom group, when using only wearable device data, the AUC, accuracy, precision, recall, and F1 scores were 0.78, 0.70, 0.66, 0.84, and 0.74, respectively, while these scores combining wearable device and application data were 0.86, 0.76, 0.68, 1.00, and 0.81, respectively. For the severe RLS symptom group, when using only wearable device data, XGB achieved AUC, accuracy, precision, recall, and F1 scores of 0.66, 0.84, 0.89, 0.93, and 0.91, respectively, while these scores combining wearable device and application data were 0.70, 0.80, 0.88, 0.90, and 0.89, respectively. Diverse digital phenotypes clinically associated with RLS were processed based on circadian rhythms to demonstrate the potential of digital phenotyping for RLS prediction. Thus, our study establishes early detection and personalized management of RLS.

Trial Registration: Clinical Research Information Service (CRIS) KCT0009175 (Registration data: Feb-15-2024) (https://cris.nih.go.kr/cris/search/detailSearch.do?search_lang=E&focus=reset_12&search_page=M&pageSize=10&page=undefined&seq=26133&status=5&seq_group=26133).

Keywords Restless legs syndrome, Digital phenotypes, Circadian rhythm, Machine learning, Prediction

Abbreviations

RLS	Restless Legs Syndrome
IRLS	International Restless Leg Study Group Severity Assessment Scale
ISI	Insomnia Severity Index
RF	Random Forest
XGB	Extreme Gradient Boost
LGBM	Light Gradient Boosting Method
ROC	Receiver Operating Characteristic

[1]Korea University College of Medicine, Seoul, Republic of Korea. [2]School of Industrial Management Engineering, Korea University, Seoul, Republic of Korea. [3]Department of Psychiatry, Korea University College of Medicine, Seoul, Republic of Korea. [4]Department of Biotechnology and Bioinformatics, Korea University, Sejong, Republic of Korea. [5]Department of Biomedical Informatics, Korea University College of Medicine, Seoul, Republic of Korea. [6]Department of Psychiatry and Biomedical Informatics, Korea University College of Medicine, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Republic of Korea. [7]Jingyeong Jeong, Yoonseo Jeon and Hyungju Kim contributed equally. ✉email: david0203@korea.ac.kr; david0203@gmail.com

AUC	Area Under the Receiver Operating Characteristic
SMOTE	Synthetic Minority Oversampling Technique
SHAP	Shapley Additive exPlanations

Restless legs syndrome (RLS) is a neurosensory disorder characterized by an irresistible urge to move the legs. RLS is diagnosed based on five criteria established by International Restless Legs Syndrome Study Group (IRLSSG): an urge to move the legs, symptoms present during rest, relieved by movement, a circadian pattern that worsens at night, and symptoms not being secondary to another medical condition[1]. RLS is a relatively common disease affecting approximately 3.9–14.35% of the general population[2]. Clinically, RLS often causes sleep disturbances and a reduced quality of life[3]. Several studies have revealed that RLS is related to metabolic dysregulation and cardiovascular diseases[3,4]. Fortunately, RLS can be relatively well managed with treatments like iron supplementation, anti-seizure medications, and dopamine agonists[5]. However, despite its high prevalence and effective treatment options, only a small number of patients receive proper diagnosis and treatment[6]. This is partly because RLS is a symptom-based syndrome, and diagnosis relies not on specific biological tests, but rather on physicians asking about RLS symptoms in patients who seek help for sleep problems. Moreover, many patients may not recognize their symptoms as a medical condition requiring treatment, leading them to avoid seeking medical help[6]. Even when they go see a doctor, the variation in symptoms such as restless arm, diverse ways in which patients describe their experiences, and the presence of other conditions with overlapping symptoms, such as cramps, make diagnosis challenging[7–9]. Furthermore, sleep-related questionnaires are largely based on subjective recall, which limits their reliability[10]. Consequently, it often takes decades for patients to receive an accurate diagnosis after symptoms first appear, emphasizing the need for new diagnostic tools[11].

The growth and evolution of digital technology has enabled the moment-by-moment collection of an individual's state using data from personal digital devices[12]. Advances in sensor technology along with the introduction of artificial intelligence, including machine learning, have enabled the accurate detection of social and behavioral signs of disease in the daily life of an individual[13]. In particular, evolving wearable device technologies in recent years have expanded the potential of the digital phenotype by providing accurate physiological data of clinical relevance[14].

The digital phenotype has attracted attention in psychiatry for overcoming the subjectivity and recall bias inherent in traditional interview-based diagnoses by offering a precise, continuous, and multidimensional assessment[15]. Numerous studies have reported the advantages of digital phenotypes in understanding various psychiatric conditions such as schizophrenia, mood disorders, anxiety disorders, and attention deficit hyperactivity disorder[16–19]. Recent studies revealed the potential of digital phenotypes in sleep disorders by disclosing sleep patterns derived from wearable devices or smartphones[20].

In the current research on digital phenotypes in the psychiatric field, we discovered the potential for applying digital phenotyping in the context of RLS. In addition to the relationship with sleep quality, the association of RLS with the cardiovascular system and sympathetic tone motivates the further application of digital phenotypes in RLS through heart rate analysis[21,22]. Several studies have revealed a clear circadian rhythm of RLS symptoms, which worsens immediately after midnight and alleviates in late-morning hours[23]. Real-time evaluation of lifestyle and vital signs using wearable devices can enable circadian rhythm analysis, thus enhancing the clinical evidence for prediction. In addition, lifestyle risk factors, such as activity level, alcohol consumption, and caffeine intake, can be effectively assessed using a digital phenotypic methodology[24,25]. Therefore, digital phenotypes can help address the issue of delayed and missed diagnoses of RLS, by screening the general population and alerting individuals who may require medical attention. Also, since the current diagnostic system relies on patient's subjective report, we can support diagnosis by predicting RLS through objective data, and further help manage the prognosis by identifying factors that may exacerbate symptoms.

This study aimed to propose a predictive machine learning model based on digital phenotypes for RLS diagnosis. Through passive or active data collection on the activities, sleep, heart rate, and daily lifestyles of participants, we established a comprehensive set of digital phenotypes based on circadian rhythms. Predicting the risk of RLS using this methodology can provide insights into the use of digital phenotypes in disease screening and diagnosis, potentially complementing current subjective and ambiguous diagnostic processes. In addition, the factor-specific contributing features of predictive algorithms provide valuable clinical insights into RLS, enabling the discovery of possible disease-related factors.

Methods

Study population

Between January 2023 and July 2024, 338 participants, consisting of 119 people with insomnia and 89 normal sleepers, were recruited from Korea University and Datamaker (Daejeon, South Korea) to collect clinical data and digital phenotypes for 4 weeks. Participants aged 19–70 years were recruited from the local community through advertisements on various internet communities and campus noticeboards. The insomnia group included individuals experiencing subjective insomnia symptoms at least three times per week in the past 3 months with an insomnia severity index (ISI) score of ≥ 15[26]. Exclusion criteria for the insomnia group included evidence of intellectual disability or organic brain damage, diagnosis of schizophrenia spectrum disorder, ongoing treatment for sleep disorders, and absence of a smartphone. The normal sleep group included individuals with subjective insomnia symptoms occurring three or less times per month, an ISI score of ≤ 8, and an average sleep duration of at least 6-h over the past 3 months. In the normal sleep group, patients diagnosed with mood disorders, anxiety disorders, and current sleep medication users were additionally excluded compared to the insomnia group.

At the beginning of the study, eligible participants provided informed consent and completed clinical report forms including demographic information, family history, and current illnesses. They were provided with wearable devices (Fitbit Inspire 1 or 2, Fitbit Inc., USA) to collect data related to step count, heart rate, and sleep.

Participants downloaded a smartphone application called "SOMDAY (Lumanlab Inc, Seoul, Korea)" and were instructed to log their daily habits using the app. Seven participants were excluded because of missing data, and finally, 338 individuals were included for analysis (Fig. 1A).

Ethical considerations

Ethical considerations were strictly adhered to throughout the study, and approval was obtained from the Institutional Review Board of Korea University Anam Hospital (No. 2022AN0587). This study was conducted according to the Declaration of Helsinki. Written informed consent was obtained from all participants after a comprehensive explanation of the study. To protect participants' privacy, their personal information was anonymized. Participants were compensated about $80 for their time and participation in the study.

Measures

Clinical assessment

Demographic information, including age, sex, height, weight, neck circumference, work pattern, life pattern, and sleep environment, were collected. Additionally, the International Restless Legs Scale (IRLS) was used to assess RLS symptoms, providing a standardized diagnostic method for grading the severity of core symptoms through self-reported assessments. Developed by the International Restless Legs Syndrome Study Group (IRLSSG), this 10-question survey is each assigned with 5 response options graded from 0 to 5 for a total score of 0–40. Each question evaluates the severity of symptoms and their impact on daily life in the previous week. The IRLS validity analyses showed good criterion validity when compared to the Clinical Global Impression (0.73–0.74), strong concurrent validity with Patient Global Impression (0.78–0.82), and significant discriminant validity ($F = 577.0$, d.f. = 2, $P < 0.01$)[27]. As a clinical standard for evaluating RLS symptoms and their consequences, the IRLS has also been used as an endpoint in nearly all academic researches[28].

Participants were categorized into groups based on their IRLS scores, with the RLS symptom group defined as IRLS > 10 ($n = 96$), and the non-RLS group ($n = 242$), which we designated as the control group defined as IRLS ≤ 10. Further classification was made within the RLS symptom group, with 10 < IRLS ≤ 20 indicating the moderate RLS symptom group ($n = 62$), and IRLS > 20 indicating the severe RLS symptom group ($n = 34$) (Fig. 1).

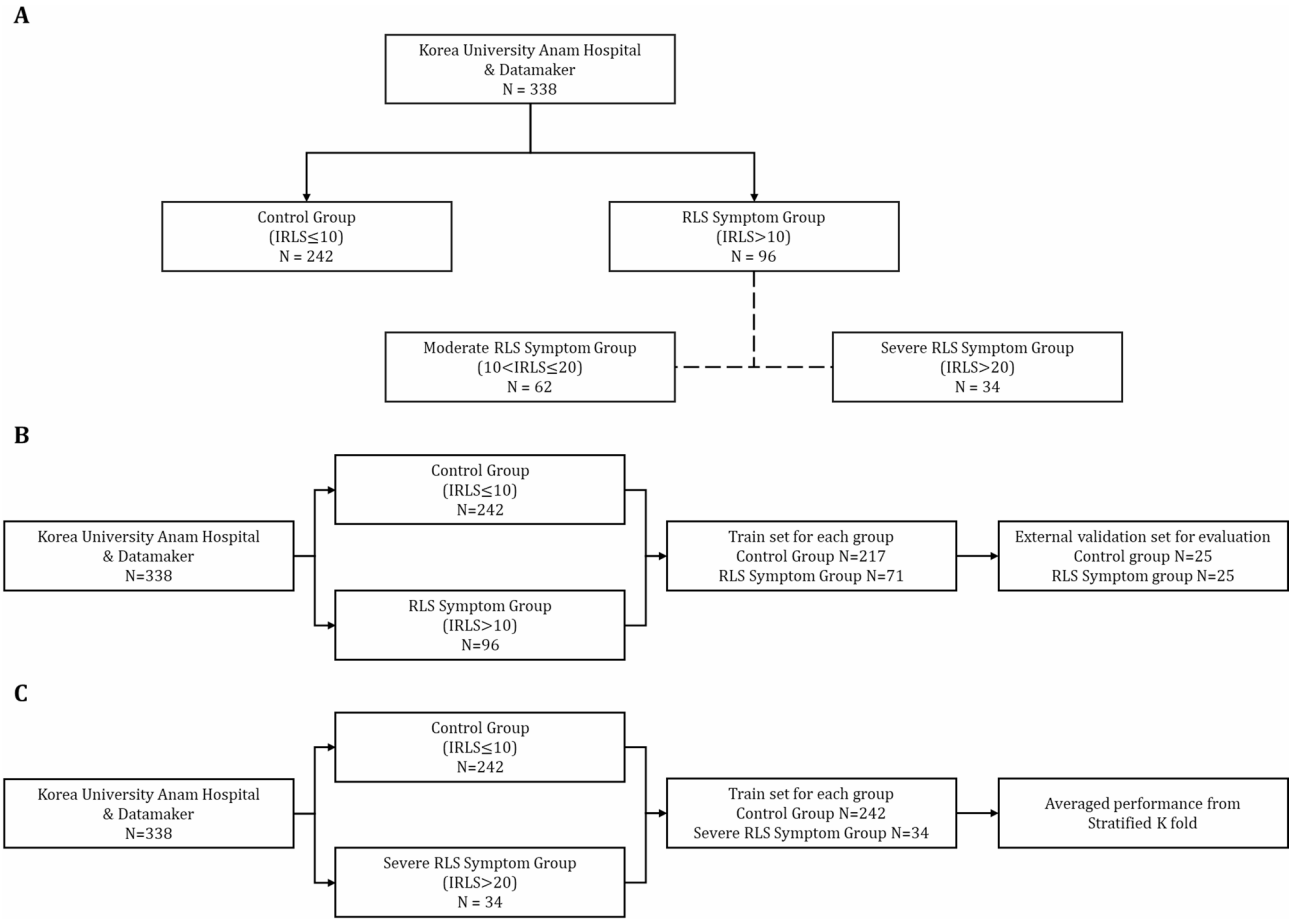


Fig. 1. Flow chart of study population and prediction. *RLS* Restless Legs Syndrome, *IRLS* International Restless Legs Syndrome Study Group rating scale. (**A**) Flow chart of study population. (**B**) Flow chart of prediction for RLS symptom group. (**C**) Flow chart of prediction for severe RLS symptom group.

However, it is important to note that this classification is based solely on IRLS cutoff scores[27]. The IRLS is a validated severity scale, not a diagnostic tool, and participants in the RLS symptom group were not clinically diagnosed with RLS. Therefore, the "RLS symptom group" in this study should be interpreted as a group at risk of RLS rather than confirmed clinical cases.

Digital phenotypes

Our team developed the SOMDAY smartphone application, a comprehensive tool for collecting information on the daily habits and lifestyle choices of participants. The application is compatible with both the Android OS (Google) and iOS (Apple) platforms and is accessible to a wide range of users. Specifically designed for collecting digital phenotypic data, the app prompts participants to record their daily entries every night at 9 PM, including information on alcohol consumption, caffeine intake, napping habits, stress levels, self-reported sleep duration, and frequency of waking up during the night.

In addition, the Fitbit device automatically captured various features, including sleep duration, number of awake periods, sleep quality, exercise time, heart rate, step count, and walking distance[29,30]. Integrating SOMDAY and Fitbit allows data synchronization, enabling participants to access their Fitbit data through the SOMDAY application. This information was seamlessly connected to an online database, providing accessibility for the research team for further analysis.

Datasets

Following the collection of data from 338 participants using SOMDAY and Fitbit, we preprocessed the digital phenotypic information. Based on the periodic nature of the RLS, this process considers the maximum, minimum, and average values aligned with daily circadian rhythms, weekdays, and holidays. Additionally, continuous variables, such as heart rate and step count, underwent cosinor analysis (see Supplementary Information 1) within 72-h intervals to reflect circadian rhythms[31,32]. We ultimately obtained 85 features from 338 individuals over 4 weeks. Missing values were replaced by the mean value of the samples in each IRLS.

Prediction model construction

Machine learning modeling

Tree-based machine learning models have significant performance in healthcare data[33]. We used XGboost (XGB), LightGBM (LGBM), and random forest (RF) classification models to predict the RLS and severe RLS symptom groups compared with the control group[34–36]. We tuned the hyperparameters when predicting RLS symptom group and did not tune them when predicting the severe RLS symptom group because the data was too imbalanced. We performed parameter tuning with cross-validation and evaluated the performance of the tuning using the scikit-learn cross validation score. To determine the degree to which different variables affect the classification model, we calculated the Shapley Additive exPlanations (SHAP) value to obtain variable importance[37].

Data training and validation

Model prediction used Python scikit-learn version 1.4.1, XGboost version 2.0.3, LightGBM version 4.3.0, shap version 0.46.0, imbalanced_learn version 0.12.2.

For predicting RLS symptom and controls, we separated the data into a training set and an independent validation set. We extracted 25 samples from each of the RLS symptom and control groups and used the rest as train set (Fig. 1B). For model performance, we used Bayesian optimization which is a hyperparameter tuning method finding the best parameters to lead best performance during training, and also used SMOTE and stratified K-fold with K = 5 for overfitting and imbalanced data[38].

For the prediction of severe RLS symptom and control groups, we did not separate the data into trainset and independent validation set, because the sample size of the severe RLS symptom group was too small and imbalanced. Instead, we used stratified K-fold cross-validation with K = 5. This means that the data was divided into 5 folds, 4 (80%) as training set and 1 (20%) as validation set, and evaluated the average performance after 5 iterations (Fig. 1C). SMOTE was also used.

The performance of the machine learning models was compared by using five metrics: accuracy, precision, recall, F1 scores and area under the curve (AUC). The performance was evaluated in 3 cases: (1) using only wearable data, (2) combining wearable data with smartphone app data, and (3) using selected features based on mutual information. Mutual information was used as a feature selection method before model training (see Supplementary Information 2). Additionally, SHAP values were used separately for feature importance analysis after training to interpret the model and gain insights. The top 20 features of the SHAP value on the model were used to verify computational and clinical efficiency.

Results

Baseline characteristics of participants

The RLS symptom group displayed a significantly lower most active 10-h period (M10) step count on weekdays (271.6 vs. 307.9, $t = 2.48$, $P = 0.01$), and a lower relative amplitude (RA) on holidays (0.4 vs. 0.5, $t = 3.38$, $P < 0.01$). In sleep-related features, the RLS symptom group had longer total sleep (375.1 vs. 362.3, $t = -2.35$, $P = 0.02$), longer REM sleep (211.6 vs. 200.2, $t = -2.26$, $P = 0.03$), and longer deep sleep (54.7 vs. 51.4, $t = -2.08$, $P = 0.04$). The RLS symptom group showed lower maximum nighttime heart rate on weekdays (105.0 vs. 109.4, $t = 3.40$, $P < 0.01$). Additionally, the RLS symptom group displayed higher stress level (29.1 vs. 17.9, $t = -2.18$, $P = 0.03$) and lower alcohol consumption (13.1 vs. 26.6, $t = 3.06$, $P < 0.01$).

In comparing the severe RLS symptom group to the control group, significant baseline differences were observed in step count cosine analyzed features. The severe RLS symptom group had lower MESOR (midline

estimating statistics of rhythm) (172.7 vs. 210.2, t = 2.33, $P = 0.02$), and lower amplitude (142.2 vs. 189.6, t = 2.13, $P = 0.04$). Also, the severe RLS symptom group exhibited lower least active 5-h period (L5) on both weekdays (106.9 vs. 158.7, t = 2.69, $P = 0.01$) and holidays (101.6 vs. 155.3, t = 2.05, $P = 0.04$). Most active 10-h period (M10) was also lower on both weekdays (252.9 vs. 307.9, t = 2.53, $P = 0.01$) and holidays (227.9 vs. 288.5, t = 2.19, $P = 0.03$). A lower relative amplitude (RA) was observed on holidays (0.4 vs. 0.5, t = 2.59, $P = 0.01$). The severe RLS symptom group exhibited lower physical activity level at nighttime on weekdays with lower mean exercise time (8.6 vs. 13.0, t = 2.25, $P = 0.03$), lower walking distance (0.2 vs. 0.4, t = 2.70, $P = 0.01$), and lower step count (283.2 vs. 508.6, t = 2.64, $P = 0.01$). In nighttime heart rate, the severe RLS symptom group exhibited lower maximum (102.3 vs. 109.4, t = 3.85, $P < 0.01$) and mean (69.1 vs. 71.7, t = 2.29, $P = 0.03$) heart rate on weekdays, and lower mean heart rate on holidays (69.3 vs. 72.1, t = 2.38, $P = 0.02$). Additionally, alcohol consumption was significantly lower in the severe RLS symptom group (8.8 vs. 26.6, t = 3.88, $P < 0.01$).

Performance evaluation of the RLS prediction model

See Table 1.

Performance evaluation of the RLS symptom group prediction model

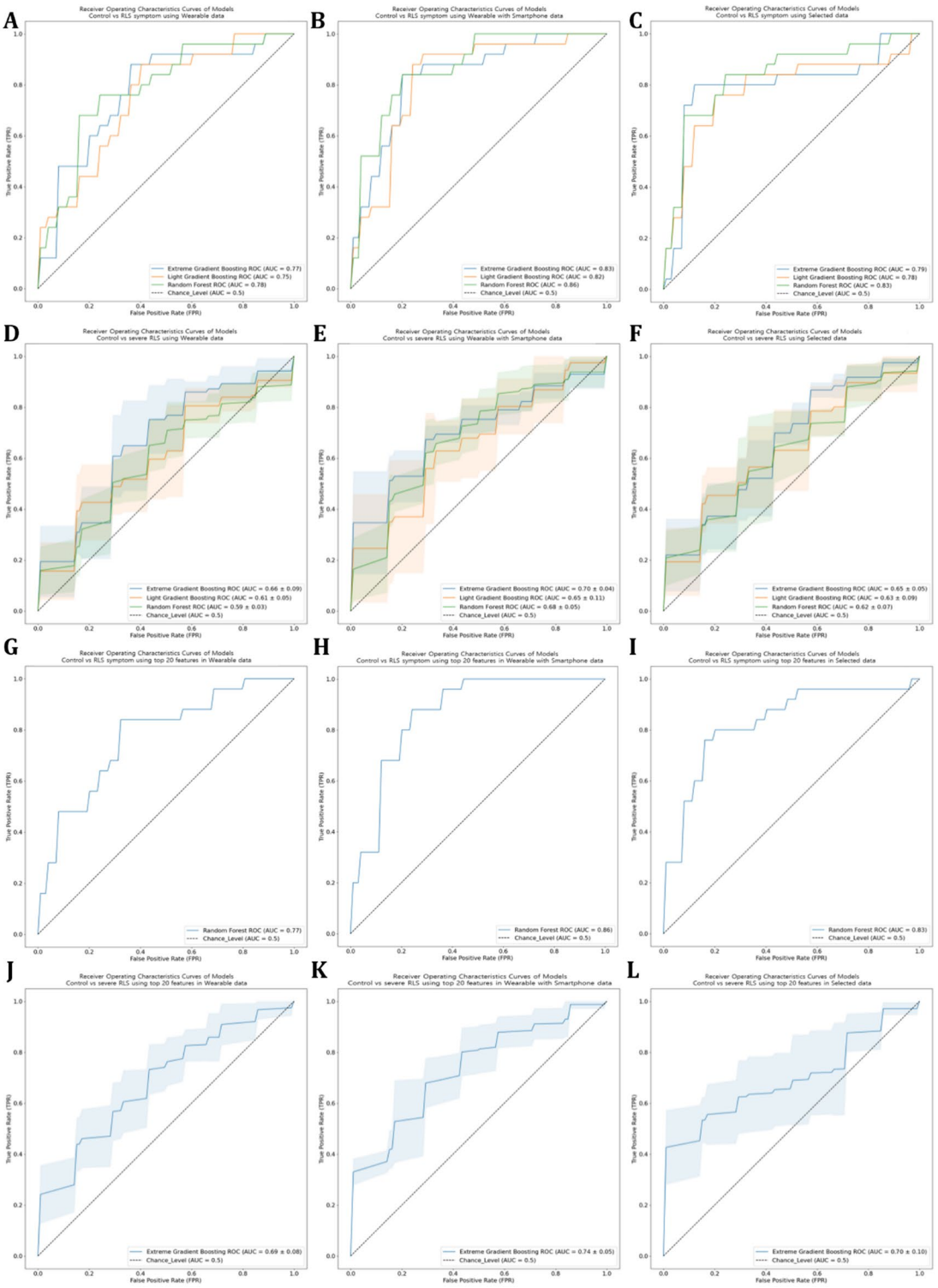
Figure 2A–C and Table 1 display the classification prediction accuracy of the three machine learning model types using ROC Curves and five parameters for RLS symptom group prediction models. When using wearable device data alone, the AUC was 0.77, 0.75, and 0.78 for XGB, LGBM, and RF, respectively. When wearable device data were combined with application data, an improved performance was achieved with AUC of 0.83, 0.82, and 0.86 for XGB, LGBM, and RF, respectively. For data selected were selected by feature selection method, an improved performance was achieved with AUC of 0.79, 0.78, and 0.83 for XGB, LGBM, and RF, respectively. The RF model showed the highest performance in both cases. When using only wearable device data, the accuracy, precision, recall, and F1 scores were 0.70, 0.66, 0.84, and 0.74, respectively, while these scores for the XGB model combining wearable device and application data were 0.76, 0.68, 1.00, and 0.81, respectively. For selected data were 0.80, 0.78, 0.84, and 0.81, respectively.

Figure 3A shows the top 20 contributing features of RF using a wearable device and the application data measured by the SHAP analysis.

Stress and step count features such as most active 10-h period on holidays and on weekdays (M10_holiday, M10_week), and minimum step count at day on holiday (min_steps_holiday_day) had the highest impact,

			Accuracy score	Precision score	Recall score	F1 score	AUC
Control vs. RLS symptom group	Wearable device data	XGB	0.72	0.66	0.92	0.77	0.77
		LGBM	0.72	0.68	0.84	0.75	0.75
		RF	0.70	0.66	0.84	0.74	0.78
		RF using top 20 features of Shapley value	0.66	0.61	0.88	0.72	0.77
	Wearable device + Smartphone App data	XGB	0.72	0.67	0.88	0.76	0.83
		LGBM	0.74	0.68	0.92	0.78	0.82
		RF	0.76	0.68	1.00	0.81	0.86
		RF using top 20 features of Shapley value	0.78	0.71	0.96	0.81	0.86
	Selected data by feature selection	XGB	0.74	0.71	0.80	0.75	0.79
		LGBM	0.74	0.71	0.80	0.75	0.78
		RF	0.80	0.78	0.84	0.81	0.83
		RF using top 20 features of Shapley value	0.72	0.69	0.80	0.74	0.83
Control vs. Severe RLS symptom group	Wearable device data	XGB	0.84	0.89	0.93	0.91	0.66
		LGBM	0.80	0.88	0.90	0.89	0.61
		RF	0.82	0.87	0.93	0.90	0.59
		XGB using top 20 features of Shapley value	0.83	0.89	0.92	0.90	0.69
	Wearable device + Smartphone App data	XGB	0.80	0.88	0.90	0.89	0.70
		LGBM	0.84	0.90	0.93	0.91	0.65
		RF	0.84	0.87	0.95	0.91	0.68
		XGB using top 20 features of Shapley value	0.85	0.90	0.93	0.91	0.74
	Selected data by feature selection	XGB	0.83	0.91	0.90	0.90	0.65
		LGBM	0.82	0.89	0.90	0.90	0.63
		RF	0.82	0.89	0.91	0.90	0.62
		XGB using top 20 features of Shapley value	0.84	0.89	0.93	0.91	0.70

Table 1. Performance evaluation of RLS prediction models in different settings. Performance evaluation of Extreme Gradient Boosting (XGB), light-gradient boosting model (LGBM), Random Forest (RF), and XGB using the top 20 features of the Shapley values in different subgroups. *AUC* area under curve. Selected data were selected by mutual information, which is feature selection method, the top 25% of the most relevant features.



followed by relative amplitude of step count on holidays (RA_holiday), alcohol, and interdaily stability of step count on weekdays (IS_week).

Performance evaluation of the severe RLS symptom group prediction model

As shown in Fig. 2D–F, and Table 1, when using only wearable device data, XGB achieved an AUC of 0.66, with accuracy, precision, recall, and F1 scores of 0.84, 0.89, 0.93, and 0.91, respectively. The LGBM and RF analysis yielded AUC values of 0.61 and 0.59, respectively. When combining wearable device data with application data, XGB maintained a superior performance, with an AUC of 0.70 and accuracy, precision, recall, and F1 scores of 0.80, 0.88, 0.90, and 0.89, respectively. In this setting, LGBM and RF exhibited AUC values of 0.65 and 0.68, respectively. When using selected data, XGB is still superior with an AUC of 0.65 and accuracy, precision,

◀**Fig. 2**. Receiver operating characteristics (ROC) curves of prediction models. The dashed line describes what would be expected by chance, while the three curves describe each machine learning model's differentiating sensitivity and specificity. Each 95% confidence interval (CI) was calculated based on the list of AUC values from machine learning models. The figure below is the ROC curve of models predicting the RLS symptom group and severe RLS symptom group, respectively. (**A**–**F**) display the ROC curves of three machine learning models, while (**G**–**L**) shows the ROC curves of XGB based on the top 20 features of the Shapley value, utilizing only wearable device data or both wearable device data and application data. (**A**) Prediction for RLS symptom group using only wearable device data. (**B**) Prediction for RLS symptom group using wearable device and application data together. (**C**) Prediction for RLS symptom group using selected data. (**D**) Prediction for severe RLS symptom group using wearable device data. (**E**) Prediction for severe RLS symptom group using wearable device and application data together. (**F**) Prediction for severe RLS symptom group selected data. (**G**) Prediction for RLS symptom group using only wearable device data. (**H**) Prediction for RLS symptom group using wearable device and application data together. (**I**) Prediction for RLS symptom group using selected data. (**J**) Prediction for severe RLS symptom group using wearable device data. (**K**) Prediction for severe RLS symptom group using wearable device and application data together. (**L**) Prediction for severe RLS symptom group using selected data.

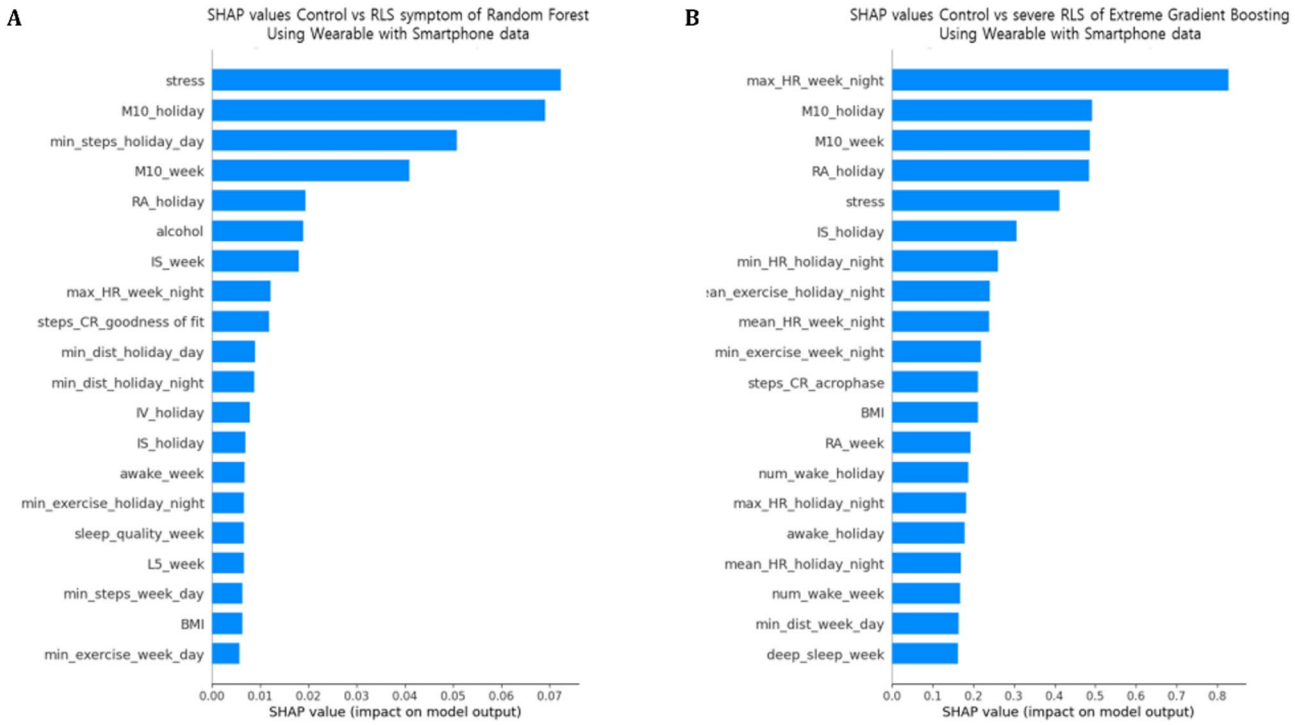


Fig. 3. Summary of Shapley additive explanations importance by prediction models. Each figure presents SHAP of RF, XGB Models response to prediction for RLS symptom, severe RLS symptom group each. (**A**) Prediction for RLS symptom group using wearable device and application data together. (**B**) Prediction for severe RLS symptom group using wearable device and application data together.

recall, and F1 scores of 0.83, 0.91, 0.90 and 0.90 while LGBM, and RF exhibited AUC values of 0.63 and 0.62, respectively.

Figure 3B shows the top 20 SHAP values of the XGB model using wearable device and application data, highlighting the high ranking of maximum heartrate at night on weekdays (max_HR_week_night), step count features such as most active 10-h period on holidays and on weekdays (M10_holiday, M10_week), and relative amplitude of step count on holidays (RA_holiday). Stress and interdaily stability of step count on holidays (IS_holiday) were also important.

Performance evaluation of the RLS symptom group and the severe RLS symptom group prediction model using top 20 contributing features

A refined RF, XGB machine learning model was created using the 20 features with the highest SHAP values from the previous analysis for each RLS group, which showed improved performance (Fig. 2G–L; Table 1.) When distinguishing the RLS symptom group, the AUC of RF improved 0.77, 0.86, 0.83, when using only wearable data and combining wearable data with application data and selected data, respectively. Similarly, for distinguishing the severe RLS symptom group, the AUC of XGB increased from 0.61 to 0.77, from 0.69 to 0.79 and from 0.60

to 0.80 when using only wearable device data, wearable device data combined with application data and selected data respectively.

Discussion

Principal findings

This study proposed a machine learning prediction model for RLS that uses data from wearable devices and self-reported application records. To the best of our knowledge, this is the first attempt at applying digital phenotyping to RLS. Satisfactory prediction accuracy was achieved across various feature settings.

For the prediction of the RLS symptom group, the RF model using a combination of self-reported application data and wearable device data achieved the highest performance with an AUC of 0.86. For predicting the severe RLS symptom group, the XGB model achieved the highest AUC of 0.70 using a combination of self-reported application data and wearable device data. When predicting the severe RLS symptom group and control group, the data was highly imbalanced ($n=34$). So, oversampling and cross validation were used to overcome this. Nevertheless, the performance decreased compared to the RLS symptom group. It can be seen as a failure to sufficiently compensate for this high imbalance. RF was the model with the best performance in predicting the control group and the RLS symptom group. However, when only wearable device data was used, XGB and LGBM were better than RF in some performance indicators. Since the ranking was based on AUC, RF was the best of the three models. The data of the control group ($n=242$) and the RLS symptom group ($n=96$) are imbalanced. In classification tasks with imbalanced data, commonly used metrics such as accuracy, precision, recall, and F1 score can be misleading. For example, a model that simply predicts the majority class may achieve high accuracy while completely failing to identify minority cases, which are often of greatest clinical importance. Precision and recall are also sensitive to threshold selection and class distribution, which limits their generalizability across datasets. F1 score, though intended to balance precision and recall, can still obscure poor performance on the minority class when class imbalance is severe. As a result, these metrics may not reflect the true discriminative ability of the model in such settings. To address this, we used AUC as a threshold-independent metric that better captures class separation regardless of imbalance. This was also reflected in the performance results of the severe RLS symptom group, which were predicted using a combination of self-reported application data and wearable device data. In addition, using the top 20 contributing features increased overall performance. With data extracted from the wearable device alone, the RF model achieved an AUC of 0.77 for predicting the RLS symptom group and XGB achieved an AUC of 0.69 for the severe RLS symptom group. By selecting top 25% of the most relevant to IRLS from a combination of application data and wearable device data using a feature selection method, RF models achieved an AUC of 0.83 for predicting RLS symptoms and XGB achieved an AUC of 0.70 for severe RLS symptoms. Mutual information, a feature selection method, selects features that are related to the target value before training. Performance was improved by using the top 25% of features while excluding features with little correlation, and many of the selected features were also found among the top 20 SHAP values.

Although the average performance of the models was acceptable, achieving a higher accuracy was limited owing to the restricted information of the digital phenotypes. Digital phenotypes only utilize daily life data, rather than variables directly associated with RLS pathophysiology or symptoms. Moreover, the complex etiology of RLS, which involves various genetic, molecular, and lifestyle-related factors, and neurotransmitters, further complicates the prediction of RLS solely based on the digital phenotype[24,39].

To overcome this limitation, we used a dataset that is both theoretically and clinically relevant to RLS, including variables known to be related to the condition. These include psychological factors (e.g., stress), lifestyle factors (e.g., alcohol, caffeine), physical factors (e.g., activity level, heart rate, sleep), and circadian rhythm factors (e.g., data processing method based on circadian rhythm)[24,25].

RLS is also closely linked with insomnia, particularly affecting sleep fragmentation, reduced REM sleep, and lightening of sleep[40,41]. Therefore, we included sleep-related features such as the number of awakenings and the durations of REM, deep, and light sleep. Regarding circadian rhythms, studies have shown that diurnal variability in multiple biological factors may serve as a mechanism. Circadian rhythms of iron and dopamine levels associated with fluctuations in biomarkers, such as melanin, melanocortin, and TSH, may exacerbate symptoms at night[42,43]. Several studies have also shown that fluctuations in spinal cord excitability or diurnal disturbance of the default mode network are related to the periodic change of RLS symptoms[44,45].

Thus, our use of circadian rhythms in digital phenotyping is both appropriate and effective, given the established links between RLS and biological rhythms.

Baseline characteristic analysis highlighted significantly lower most active 10-h period (M10) and lower relative amplitude (RA) across both the RLS and severe RLS symptom groups (Table 2). SHAP analysis also showed these variables were key contributors to RLS prediction (Fig. 3). A lower M10 reflects decreased activity level during wake periods, while a lower RA indicates a less pronounced circadian rhythm, both suggesting relatively disturbed circadian rhythm. Although there is no direct study on circadian rhythm disturbance influencing RLS, previous study investigated night shift workers exhibited higher prevalence of RLS[46]. Furthermore, a previous study found that night shift workers have a disrupted cosine regression curve of heart-rate rhythm with a lower amplitude and delayed acrophase, which support probable association between circadian rhythm disturbance and RLS[47]. Moreover, the severe RLS symptom group demonstrated a lower MESOR (midline estimating of statistics rhythm) for step count, measure reflecting the rhythm-adjusted mean activity level, which may be a more accurate indicator than simple mean levels[48]. This finding aligns with previous study the lower physical activity tends to increase the RLS prevalence[24]. Stress levels were high in both the RLS and severe RLS symptom groups, contributing high SHAP value (Table 2; Fig. 3). This finding supports a previous study showing that patients with RLS have more psychological distress and reduced quality of life, although the causality remains unclear[4]. Moreover, number of wake or duration of awake ranked high in SHAP values, particularly in predicting

Category	Variables	Mean (SD)			Control vs. RLS symptom group			Control vs. Severe RLS symptom group		
		Control ($n=242$)	RLS symptom ($n=96$)	Severe RLS symptom ($n=34$)	t-value	df	p-value	t-value	df	p-value
Demographic information										
	Sex, male (%)	99 (40.9%)	20 (20.8%)	6 (17.6%)			0.98			0.13
	Age	29.1 (9.1)	31.6 (10.6)	31.9 (10.7)	− 2.03	153	0.04*	− 1.45	40	0.15
	BMI	22.4 (3.4)	22.2 (3.5)	21.4 (3.1)	0.65	173	0.52	1.76	45	0.08
Features based on wearable device data										
Sleep	Weekdays									
	Duration of									
	Total sleep	362.3 (48.0)	375.1 (43.9)	373.7 (45.2)	− 2.35	189	0.02*	− 1.37	44	0.18
	Awake	61.9 (15.3)	63.4 (13.8)	60.5 (15.7)	− 0.88	193	0.38	0.48	42	0.64
	REM	200.2 (43.1)	211.6 (40.9)	209.1 (51.4)	− 2.26	183	0.03*	− 0.96	40	0.35
	Light sleep	72.6 (20.2)	75.9 (17.9)	69.6 (19.8)	− 1.51	195	0.13	0.81	43	0.42
	Deep sleep	51.4 (12.6)	54.7 (13.1)	53.6 (14.2)	− 2.08	169	0.04*	− 0.82	41	0.42
	Number of awake	25.1 (6.4)	25.7 (6.8)	24.8 (8.1)	− 0.78	166	0.44	0.16	39	0.87
	Quality of sleep	0.830 (0.036)	0.832 (0.031)	0.840 (0.036)	− 0.47	203	0.64	− 1.46	43	0.15
	Holidays[a]									
	Duration of									
	Total sleep	396.7 (63.0)	410.2 (66.0)	405.5 (71.6)	− 1.72	168	0.09	− 0.68	41	0.50
	Awake	67.9 (18.8)	68.4 (17.8)	66.7 (20.6)	− 0.24	183	0.81	0.32	41	0.75
	REM	219.1 (48.6)	228.9 (51.1)	226.4 (56.4)	− 1.61	167	0.11	− 0.72	40	0.48
	Light sleep	81.2 (24.9)	82.4 (22.7)	77.4 (22.5)	− 0.42	190	0.68	0.90	45	0.37
	Deep sleep	56.5 (14.2)	59.6 (14.5)	59.2 (17.5)	− 1.81	171	0.07	− 0.87	39	0.39
	Number of awake	27.5 (7.4)	27.5 (7.5)	26.9 (9.0)	0.03	173	0.97	0.37	39	0.71
	Quality of sleep	0.830 (0.039)	0.833 (0.037)	0.835 (0.040)	− 0.79	185	0.43	− 0.77	42	0.44
Exercise time	Weekdays									
	Daytime[b]									
	Minimum	5.0 (9.7)	4.4 (7.4)	3.2 (4.2)	0.64	227	0.53	1.93	94	0.06
	Maximum	26.4 (19.1)	25.4 (18.4)	25.7 (23.5)	0.45	180	0.66	0.15	39	0.88
	Mean	11.2 (11.5)	10.4 (8.8)	9.6 (6.7)	0.65	227	0.52	1.19	64	0.24
	Nighttime[b]									
	Minimum	6.1 (16.3)	6.9 (18.6)	2.9 (7.6)	− 0.36	157	0.72	1.91	85	0.06
	Maximum	30.6 (28.1)	34.5 (37.5)	27.6 (24.3)	− 0.93	139	0.35	0.65	46	0.52
	Mean	13.0 (19.4)	14.0 (23.7)	8.6 (8.8)	− 0.38	148	0.71	2.25	88	0.03*
	Holidays[a]									
	Daytime[b]									
	Minimum	4.8 (17.3)	4.2 (7.0)	2.8 (4.5)	0.49	336	0.62	1.47	194	0.14
	Maximum	25.2 (25.8)	25.1 (23.7)	22.2 (28.9)	0.05	189	0.96	0.57	41	0.57
	Mean	10.9 (18.6)	10.1 (9.2)	8.1 (6.1)	0.56	320	0.58	1.76	142	0.08
	Nighttime[b]									
	Minimum	8.3 (19.7)	9.9 (25.6)	7.0 (14.6)	− 0.56	142	0.58	0.48	52	0.64
	Maximum	31.8 (40.6)	34.0 (42.8)	27.5 (21.7)	− 0.43	167	0.67	0.95	71	0.35
	Mean	15.2 (26.0)	16.5 (29.3)	11.8 (14.7)	− 0.39	157	0.69	1.11	66	0.27

Continued

Category	Variables	Mean (SD)			Control vs. RLS symptom group			Control vs. Severe RLS symptom group		
		Control ($n=242$)	RLS symptom ($n=96$)	Severe RLS symptom ($n=34$)	t-value	df	p-value	t-value	df	p-value
Walking distance	Weekdays									
	Daytime[b]									
	Minimum	0.1 (0.3)	0.1 (0.2)	0.1 (0.1)	1.28	285	0.20	1.46	77	0.15
	Maximum	0.9 (0.5)	0.8 (0.5)	0.8 (0.6)	0.92	177	0.36	0.48	40	0.64
	Mean	0.3 (0.3)	0.3 (0.2)	0.2 (0.2)	1.39	251	0.16	1.39	60	0.17
	Nighttime[b]									
	Minimum	0.2 (0.6)	0.1 (0.5)	0.1 (0.2)	0.11	203	0.91	1.92	116	0.06
	Maximum	1.0 (0.9)	1.0 (1.0)	0.8 (0.7)	0.35	161	0.73	1.85	54	0.07
	Mean	0.4 (0.7)	0.3 (0.6)	0.2 (0.2)	0.39	189	0.69	2.70	121	0.01*
	Holidays[a]									
	Daytime[b]									
	Minimum	0.1 (0.4)	0.1 (0.2)	0.1 (0.2)	0.29	318	0.77	0.64	80	0.53
	Maximum	0.7 (0.7)	0.7 (0.7)	0.6 (0.8)	0.41	167	0.68	0.82	40	0.42
	Mean	0.2 (0.4)	0.2 (0.3)	0.2 (0.2)	0.66	264	0.51	1.38	80	0.17
	Nighttime[b]									
	Minimum	0.2 (0.6)	0.2 (0.9)	0.1 (0.3)	−0.44	135	0.66	0.75	68	0.46
	Maximum	0.9 (1.2)	0.9 (1.3)	0.7 (0.6)	0.02	160	0.99	1.76	80	0.08
	Mean	0.4 (0.8)	0.4 (0.9)	0.3 (0.4)	−0.14	157	0.89	1.61	96	0.11
Step count	Mesor[c]	210.2 (141.8)	188.1 (122.4)	172.7 (77.7)	1.43	201	0.15	2.33	69	0.02*
	Amplitude[c]	189.6 (163.7)	182.2 (152.1)	142.2 (98.9)	0.40	187	0.69	2.13	62	0.04*
	Acrophase[c]	4.6 (31.5)	−6.0 (25.1)	6.9 (17.8)	−0.41	218	0.68	−0.63	67	0.53
	Goodness of fit[c]	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	−0.45	191	0.65	0.09	51	0.93
	Weekdays									
	Daytime[b]									
	Minimum	155.6 (379.9)	119.3 (236.8)	99.7 (193.8)	1.06	275	0.29	1.35	75	0.18
	Maximum	1225.3 (763.9)	1170.4 (770.2)	1193.6 (923.5)	0.59	173	0.55	0.19	40	0.85
	Mean	417.1 (444.3)	370.9 (319.7)	350.1 (281.5)	1.07	241	0.29	1.19	59	0.24
	Nighttime[b]									
	Minimum	220.8 (822.5)	219.9 (723.6)	77.9 (302.2)	0.01	197	0.99	1.93	120	0.06
	Maximum	1435.9 (1317.0)	1424.3 (1481.7)	1146.6 (967.4)	0.07	158	0.95	1.55	52	0.13
	Mean	508.6 (960.2)	484.8 (915.9)	283.2 (345.0)	0.21	182	0.83	2.64	124	0.01*
	L5	158.7 (192.8)	135.2 (169.1)	106.9 (85.6)	1.11	198	0.27	2.69	90	0.01*
	M10	307.9 (153.9)	271.6 (106.2)	252.9 (113.1)	2.48	251	0.01*	2.53	52	0.01*
	RA	0.4 (0.3)	0.4 (0.3)	0.4 (0.2)	−0.06	174	0.95	0.20	52	0.84
	IV	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	−0.44	183	0.66	0.86	50	0.40
	IS	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	−0.15	171	0.88	0.31	43	0.76
	Holidays[a]									
	Daytime									
	Minimum	123.0 (530.0)	115.0 (272.4)	91.6 (259.2)	0.18	315	0.86	0.56	79	0.58
	Maximum	1070.7 (985.1)	1043.4 (1070.8)	932.0 (1185.9)	0.22	162	0.83	0.65	40	0.52
	Mean	361.0 (606.9)	335.9 (410.6)	283.4 (302.0)	0.44	255	0.66	1.20	78	0.24
	Nighttime									
	Minimum	293.5 (943.2)	355.5 (1282.3)	211.1 (508.9)	−0.43	138	0.67	0.78	70	0.44
	Maximum	1361.7 (1778.7)	410.6 (1282.3)	1068.22 (870.6)	−0.06	165	0.95	1.56	79	0.12
	Mean	560.0 (1234.3)	468.5 (1891.3)	373.7 (515.2)	−0.17	160	0.86	1.57	99	0.12
	L5	155.3 (333.0)	114.2 (121.5)	101.6 (87.9)	1.66	334	0.10	2.05	193	0.04*
	M10	288.5 (172.4)	249.4 (174.3)	227.9 (148.1)	1.87	173	0.06	2.19	47	0.03*
	RA	0.5 (0.2)	0.4 (0.3)	0.4 (0.3)	3.38	143	< 0.01*	2.59	39	0.01*
	IV	0.1 (0.1)	0.1 (0.1)	0.1 (0.1)	−0.49	173	0.63	1.14	64	0.26
	IS	0.2 (0.1)	0.3 (0.1)	0.3 (0.2)	−0.89	174	0.37	−0.51	42	0.54

Continued

Category	Variables	Mean (SD)			Control vs. RLS symptom group			Control vs. Severe RLS symptom group		
		Control ($n=242$)	RLS symptom ($n=96$)	Severe RLS symptom ($n=34$)	t-value	df	p-value	t-value	df	p-value
Heart rate	Mesor[c]	75.2 (9.2)	75.1 (6.3)	73.7 (6.9)	0.20	252	0.84	1.16	51	0.25
	Amplitude[c]	12.0 (6.1)	11.2 (4.1)	11.9 (4.9)	1.37	254	0.17	0.11	48	0.91
	Acrophase[c]	6.9.4 (3.8)	10.3 (4.8)	10.3 (3.9)	−1.66	144	0.10	−1.22	42	0.23
	Goodness of fit[c]	0.4 (0.1)	0.4 (0.1)	0.4 (0.1)	−0.53	160	0.60	−1.32	41	0.19
	Weekdays									
	Daytime[b]									
	Minimum	62.4 (7.4)	63.8 (7.6)	616.6 (5.3)	−1.62	171	0.11	0.74	53	0.47
	Maximum	112.7 (12.8)	112.2 (13.5)	110.5 (13.2)	0.32	166	0.75	0.94	42	0.35
	Mean	80.2 (8.1)	81.0 (7.7)	79.6 (5.8)	−0.82	181	0.41	0.50	53	0.62
	Nighttime[b]									
	Minimum	56.0 (6.5)	56.6 (5.1)	54.7 (4.5)	−0.83	222	0.41	1.43	54	0.16
	Maximum	109.4 (12.4)	105.0 (10.2)	102.3 (9.7)	3.40	221	<0.01*	3.85	49	<0.01*
	Mean	71.7 (7.9)	71.1 (6.1)	69.1 (5.9)	0.74	226	0.58	2.29	52	0.03*
	Holidays[a]									
	Daytime[b]									
	Minimum	59.5 (7.3)	60.7 (6.5)	59.3 (5.8)	−1.40	194	0.16	0.24	49	0.81
	Maximum	109.1 (14.7)	107.3 (13.8)	105.0 (11.9)	1.11	185	0.27	1.83	48	0.07
	Mean	77.4 (8.5)	77.7 (7.2)	76.5 (6.8)	−0.32	204	0.75	0.67	49	0.50
	Nighttime[b]									
	Minimum	56.8 (7.4)	57.2 (5.7)	55.1 (5.0)	−0.56	204	0.75	1.72	56	0.09
	Maximum	106.1 (13.3)	103.4 (11.6)	102.1 (11.5)	1.87	198	0.06	1.86	46	0.07
	Mean	72.1 (8.8)	71.2 (6.7)	69.3 (5.9)	1.00	227	0.32	2.38	56	0.02*
Features based on application data										
	Caffeine	16.9 (19.9)	19.5 (25.9)	21.7 (28.6)	−0.89	142	0.38	−0.96	38	0.35
	Nap time	5.6 (7.3)	6.9 (10.3)	8.6 (14.9)	−10.7	135	0.28	−1.15	35	0.26
	Stress	17.9 (33.3)	29.1 (46.0)	38.1 (60.4)	−2.18	136	0.03*	−1.91	36	0.06
	Alcohol	26.6 (55.4)	13.1 (25.7)	8.8 (16.7)	3.06	328	<0.01*	3.88	161	<0.01*

Table 2. Baseline characteristics of demographic information and data set. Significance between continuous variables was determined using independent *t*-tests. Significance between categorical variables was determined using the Chi-square test. *Significance at $p<0.05$.. [a]Holidays include weekends (Saturdays and Sundays) and Korean national holidays. [b]Daytime is defined as 8 AM to 6 PM, and nighttime as 6 PM to 8 AM. [c]These parameters were generated by cosinor analysis to reflect the circadian rhythm. Detailed descriptions are provided in Supplementary Methods. *REM* rapid eye movement, *IS* interdaily stability, *IV* interdaily variability, *L5* least active 5-h period, *M10* most active 10-h period, *MESOR* midline estimating statistics of rhythm, *RA* relative amplitude.

severe RLS symptoms (Fig. 3). This suggests that sleep fragmentation is one of the key sleep feature in RLS patients, as supported by previous polysomnography studies[40,41].

In addition, through the SHAP analysis, we extracted the top 20 contributing features to construct a more focused models and confirmed that they demonstrated similar or superior scores across all performance metrics. By reducing the amount of data used, we can lower computational costs, which increase the possibility of actual clinical application. We chose RF and XGB models to construct this focused model because of the superior performance compared to other models.

Comparison to prior work

While there are several attempts to explore use of digital phenotyping of daily life to predict mental disorders such as schizophrenia, mood disorder, and anxiety disorder, its application to sleep disorders remains limited[14]. For example, one study clustered insomnia subtypes based on sleep patterns, daily activities, and personal demographics, revealing significant differences in sleep and behavioral characteristics across clusters[49]. Another study utilized digital phenotype of wearable device to predict sleep problems in children, achieving an accuracy of 0.737[18]. Additionally, single-lead ECG, which can also be incorporated into smartwatches, has been utilized to predict periodic limb movement disorder, with F1 scores of 0.92[50]. In the context of RLS, digital phenotypes have typically focused on measures such as heart rate variability (HRV) and physical activity in attempts to reveal pathophysiology of RLS[51,52]. However, comprehensive studies using a wider range of digital phenotypes for RLS prediction and management are still lacking.

Compared to these studies, our study is the first study to utilize a wide range of daily life metrics reflecting life patterns and cosinor variation to predict RLS, offering a more general approach to understanding and managing the disorder.

Strengths and limitations

The strength of this study lies in the novel use of digital phenotyping and circadian rhythm-based feature generation while achieving meaningful accuracy. A diverse range of daily life parameters that were clinically acknowledged to be associated with RLS were included. Additionally, we employed circadian rhythm-based feature generation, considering the various mechanisms suggested to explain the circadian rhythm of RLS symptoms[42,43]. Moreover, while several studies have attempted to analyze features reflecting lifestyle patterns, our investigation proposed a more thorough approach to feature categorizing. Specifically, we divided them into daytime and nighttime and distinguished between weekdays and holidays for analysis, which reflects detailed life patterns and suggests guidance for future digital phenotype research. Subsequently, we stratified participants into subgroups based on symptom severity. We diversified the data settings used in machine learning to compare the analysis accuracy according to the subgroups or types of datasets.

Despite its strengths, our study has several limitations. First, the skewed age distribution (mean [SD] age: 29.8 [9.6] years) necessitates broader validation across diverse age groups, given that RLS prevalence tends to increase with age[2]. However, no difference was found in the age distribution between the groups. Second, the scarcity of the total dataset produced an inconsistent performance across the five performance metrics. Additionally, the dataset was highly imbalanced toward control group, necessitating the use of oversampling techniques. In particular, the severe RLS symptom subgroup ($n = 34$) had an even more limited sample size. Due to this imbalance, we used a small external validation dataset for the RLS symptom group and did not create a separate validation set for the severe RLS symptom subgroup, meaning that complete external validation was not performed. Although stratified K-fold cross-validation and SMOTE oversampling were employed to address class imbalance, the small sample size restricts the generalizability of our conclusions for severe RLS specifically, and the findings related to this subgroup should be interpreted with caution. Future studies should aim to include larger and more balanced samples, particularly for moderate-to-severe RLS populations. Moreover, medical conditions, such as iron deficiency or blood glucose levels, were not adjusted for, despite their potential to cause secondary RLS or being crucial for RLS management[39,53]. However, considering that the secondary form of RLS has a low prevalence in the younger generation, we believe that this did not significantly impact our research findings[54].

Finally, the prevalence of RLS in our study group (28.4%) is notably higher than the prevalence in general population (3.9–14.3%). This discrepancy is largely due to the characteristics of our study sample. Our research focused on sleep disorders, resulting in a large insomnia group ($n = 119$) within the study sample. Numerous studies have reported that insomnia patients exhibit a higher prevalence of RLS (15.6–45%), and have identified insomnia as a main predictor of RLS (odds ratios: 1.71–3.16)[55–57]. Consequently, the overrepresentation of insomnia patients may contribute to the higher RLS prevalence observed in our study. Additionally, demographic factors played a role in our findings. Our study included a higher proportion of female participants (219/338, 64.7%), who are known to have twice the prevalence of RLS compared to males[54]. This sex imbalance likely contributed to the elevated RLS prevalence in our sample. Also, the average age of participants was relatively young (mean [SD] age: 29.8 [9.6]). While younger age is typically associated with a lower prevalence of RLS, this factor does not seem to have significantly influenced our findings[54].

This study focused on a sample enriched with insomnia participants, who are more likely to exhibit RLS, and thus does not provide an epidemiological estimate of RLS prevalence. Therefore, generalization of these findings to the general population, particularly the elderly, should be done carefully. Moreover, the RLS symptom group in our study was not diagnosed by clinicians, nor defined by IRLSSG diagnostic criteria, which may have caused slight overestimation of RLS symptom group. However, as the IRLS questionnaire is regarded as a global standard for RLS severity assessment, we believe this limitation had minimal impact on the overall results.

To further explore this topic, future studies could focus on using clinician diagnoses and ensuring a more balanced representation of age groups and sex.

Implications and future research

The accumulation of digital phenotyping data will enable disease screening using daily life data from personal digital platforms. This can enable early intervention by identifying at-risk groups in the general population, addressing delayed and missed diagnosis problem of RLS. Moreover, along with current attempts to phenotype RLS, the accumulation of digital phenotyping data could help future research classify RLS subtypes, paving the way for personalized treatment options. Future research in this area will benefit from larger sample sizes and the inclusion of a broader range of clinical features, such as HRV, considering its association with autonomic disturbances in RLS[22]. The evolving wearable device technology with longer battery lives and wider range of data collection along with advanced strategies for handling outliers and missing data will ensure the attainment of more accurate and comprehensive set of data[58,59]. Moreover, enhanced methods for analyzing patterns from the data such as human activity recognition will further help uncovering meaningful digital phenotypes[60].

Conclusions

In conclusion, we developed an RLS prediction model that used novel digital phenotyping techniques. Digital phenotyping has increasingly been applied to identify at-risk groups and guide treatment strategies for sleep disorders, but comprehensive studies using a wider range of digital phenotype for RLS prediction and management are lacking. By incorporating features known for their clinical relevance and conducting a circadian rhythm-based analysis, the model demonstrated an enhanced capability for the early detection of

RLS. Our findings represent an initial step towards integrating digital phenotyping in RLS management, laying a foundation for future research aimed at validating these approaches in clinical settings. Our findings also highlight the potential of digital health technologies for managing neuropsychiatric disorders and contributing to the progress of personalized medicine.

Data availability

Data supporting the findings of this study are available upon request from the corresponding authors. These data are not publicly available because of privacy concerns.

Received: 15 January 2025; Accepted: 5 May 2025
Published online: 10 May 2025

References

1. Allen, R. P. et al. Restless legs syndrome/Willis–Ekbom disease diagnostic criteria: Updated international restless legs syndrome study group (IRLSSG) consensus criteria–history, rationale, description, and significance. *Sleep. Med.* **15** (8), 860–873 (2014).
2. Ohayon, M. M., O'Hara, R. & Vitiello, M. V. Epidemiology of restless legs syndrome: A synthesis of the literature. *Sleep Med. Rev.* **16** (4), 283–295 (2012).
3. Earley, C. J. & Silber, M. H. Restless legs syndrome: Understanding its consequences and the need for better treatment. *Sleep Med.* **11** (9), 807–815 (2010).
4. Innes, K. E., Selfe, T. K. & Agarwal, P. Restless legs syndrome and conditions associated with metabolic dysregulation, sympathoadrenal dysfunction, and cardiovascular disease risk: A systematic review. *Sleep Med. Rev.* **16** (4), 309–339 (2012).
5. Clavadetscher, S. C., Gugger, M. & Bassetti, C. L. Restless legs syndrome: Clinical experience with long-term treatment. *Sleep Med.* **5** (5), 495–500 (2004).
6. Allen, R. P. et al. Restless legs syndrome prevalence and impact: REST general population study. *Arch. Intern. Med.* **165** (11), 1286–1292 (2005).
7. Haghshenas, M., Veisani, Y. & Sahebi, A. Restless legs syndrome variants: A systematic review. *Heliyon* (2024).
8. Ruppert, E. Restless arms syndrome: Prevalence, impact, and management strategies. *Neuropsychiatr. Dis. Treat.*, 1737–1750 (2019).
9. Hening, W. A., Allen, R. P., Washburn, M., Lesage, S. R. & Earley, C. J. The four diagnostic criteria for restless legs syndrome are unable to exclude confounding conditions (mimics). *Sleep Med.* **10** (9), 976–981 (2009).
10. Ibáñez, V., Silva, J. & Cauli, O. A survey on sleep assessment methods. *PeerJ* **6**, e4849 (2018).
11. Jaarsma, J. & Sevborn, S. Diagnosing restless legs syndrome: The patient experience. *Sleep Med.* **14**, e156–e157 (2013).
12. Jain, S. H., Powers, B. W., Hawkins, J. B. & Brownstein, J. S. The digital phenotype. *Nat. Biotechnol.* **33** (5), 462–463 (2015).
13. Galatzer-Levy, I. R. & Onnela, J-P. Machine learning and the digital measurement of psychological health. *Annu. Rev. Clin. Psychol.* **19**, 133–154 (2023).
14. De Boer, C. et al. A call to expand the scope of digital phenotyping. *J. Med. Internet Res.* **25**, e39546 (2023).
15. Oudin, A. et al. Digital phenotyping: Data-driven psychiatry to redefine mental health. *J. Med. Internet Res.* **25**, e44502 (2023).
16. Bufano, P., Laurino, M., Said, S., Tognetti, A. & Menicucci, D. Digital phenotyping for monitoring mental disorders: Systematic review. *J. Med. Internet Res.* **25**, e46778 (2023).
17. Lee, H-J. et al. Prediction of impending mood episode recurrence using real-time digital phenotypes in major depression and bipolar disorders in South Korea: A prospective nationwide cohort study. *Psychol. Med.* **53** (12), 5636–5644 (2023).
18. Kim, W-P. et al. Machine learning–based prediction of attention-deficit/hyperactivity disorder and sleep problems with wearable data in children. *JAMA Netw. Open.* **6** (3), e233502 (2023).
19. Chun, J. Y. et al. Prediction of specific anxiety symptoms and virtual reality sickness using in situ autonomic physiological signals during virtual reality treatment in patients with social anxiety disorder: Mixed methods study. *JMIR Serious Games* **10** (3), e38284 (2022).
20. Cho, C-H. Revolutionizing sleep health: The promise and challenges of digital phenotyping. *Chronobiol. Med.* **5** (3), 95–96 (2023).
21. Ferini-Strambi, L., Walters, A. S. & Sica, D. The relationship among restless legs syndrome (Willis–Ekbom Disease), hypertension, cardiovascular disease, and cerebrovascular disease. *J. Neurol.* **261**, 1051–1068 (2014).
22. Barone, D. A., Ebben, M. R., DeGrazia, M., Mortara, D. & Krieger, A. C. Heart rate variability in restless legs syndrome and periodic limb movements of sleep. *Sleep Sci.* **10** (2), 80 (2017).
23. Trenkwalder, C. et al. Circadian rhythm of periodic limb movements and sensory symptoms of restless legs syndrome. *Mov. Disord.* **14** (1), 102–110 (1999).
24. Batool-Anwar, S. et al. Lifestyle factors and risk of restless legs syndrome: Prospective cohort study. *J. Clin. Sleep Med.* **12** (2), 187–194 (2016).
25. Lutz, E. G. Restless legs, anxiety and caffeinism. *J. Clin. Psychiatry* **39** (9), 693–698 (1978).
26. Bastien, C. H., Vallières, A. & Morin, C. M. Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med.* **2** (4), 297–307 (2001).
27. IRLSS Group. Validation of the international restless legs syndrome study group rating scale for restless legs syndrome. *Sleep Med.* **4** (2), 121–132 (2003).
28. Walters, A. S. et al. Review of severity rating scales for restless legs syndrome: Critique and recommendations. *Mov. Disord Clin. Pract.* **1** (4), 317–324 (2014).
29. Lee, H-A. et al. Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults. *Psychiatry Invest.* **14** (2), 179 (2017).
30. Kang, Y-W. et al. Design and methods of a prospective smartphone app-based study for digital phenotyping of mood and anxiety symptoms mixed with centralized and decentralized research form: The search your Mind (SYM, 'ث') project. *Psychiatry Investig.* **19** (7), 588 (2022).
31. Cornelissen, G. Cosinor-based rhythmometry. *Theor. Biol. Med. Model.* **11** (1), 1–24 (2014).
32. Jeong, S. et al. Circadian rhythm of heart rate assessed by wearable devices tends to correlate with the circadian rhythm of salivary cortisol concentration in healthy young adults. *Chronobiol. Med.* **2** (3), 109–114 (2020).
33. Hu, L. & Li, L. Using tree-based machine learning for health studies: Literature review and case series. *Int. J. Environ. Res. Public Health* **19** (23), 16080 (2022).
34. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
35. Chen, T. et al. Xgboost: Extreme gradient boosting. *R Package Version 04-2* **1** (4), 1–4 (2015).
36. Zhang, J., Mucs, D., Norinder, U. & Svensson, F. LightGBM: An effective and scalable algorithm for prediction of chemical toxicity–application to the Tox21 and mutagenicity data sets. *J. Chem. Inf. Model.* **59** (10), 4150–4158 (2019).
37. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30** (2017).
38. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).

39. Gossard, T. R., Trotti, L. M., Videnovic, A. & St Louis, E. K. Restless legs syndrome: Contemporary diagnosis and treatment. *Neurotherapeutics* **18** (1), 140–155 (2021).
40. Hornyak, M., Feige, B., Voderholzer, U., Philipsen, A. & Riemann, D. Polysomnography findings in patients with restless legs syndrome and in healthy controls: A comparative observational study. *Sleep* **30** (7), 861–865 (2007).
41. Zhang, H. et al. Polysomnographic features of idiopathic restless legs syndrome: A systematic review and meta-analysis of 13 sleep parameters and 23 leg movement parameters. *J. Clin. Sleep Med.* **18** (11), 2561–2575 (2022).
42. Tang, M. et al. Circadian rhythm in restless legs syndrome. *Front. Neurol.* **14**, 1105463 (2023).
43. Earley, C. J. et al. Altered brain iron homeostasis and dopaminergic function in restless legs syndrome (Willis–Ekbom disease). *Sleep Med.* **15** (11), 1288–1301 (2014).
44. Dafkin, C., Green, A., Olivier, B., Mckinon, W. & Kerr, S. Circadian variation of flexor withdrawal and crossed extensor reflexes in patients with restless legs syndrome. *J. Sleep Res.* **27** (5), e12645 (2018).
45. Ku, J. et al. Diurnal variation of default mode network in patients with restless legs syndrome. *Sleep Med.* **41**, 1–8 (2018).
46. Sharifian, A. et al. Restless legs syndrome in shift workers: A cross sectional study on male assembly workers. *J. Circ. Rhythms.* **7**, 1–7 (2009).
47. Tan, C., Mineyama, K. & Shiotani, H. Influence of night shift work on circadian heart-rate rhythm in nurses: Using a Holter electrocardiogram that can be continuously measured for two weeks. *Ind. Health* **2024**, 2023–2157 (2024).
48. Refinetti, R., Cornélissen, G. & Halberg, F. Procedures for numerical analysis of circadian rhythms. *Biol. Rhythm Res.* **38** (4), 275–325 (2007).
49. Park, S., Lee, S. W., Han, S. & Cha, M. Clustering insomnia patterns by data from wearable devices: Algorithm development and validation study. *JMIR mHealth uHealth*. **7** (12), e14473 (2019).
50. Urtnasan, E., Park, J-U., Lee, J-H., Koh, S-B. & Lee, K-J. Deep learning for automatic detection of periodic limb movement disorder based on electrocardiogram signals. *Diagnostics* **12** (9), 2149 (2022).
51. Yıldız, A., Yıldız, C. & Karakurt, A. Assessment of cardiac autonomic functions by heart rate variability in patients with restless leg syndrome. *Arch. Turk. Soc. Cardiol./Türk Kardiyoloji Derneği Arşivi* **46** (3) (2018).
52. Reimers, A., Heidenreich, V., Bittermann, H., Knapp, G. & Reimers, C. Accelerometer-measured physical activity and its impact on sleep quality in patients suffering from restless legs syndrome. *BMC Neurol.* **21**, 1–10 (2021).
53. Bosco, D. et al. Role of the oral glucose tolerance test (OGTT) in the idiopathic restless legs syndrome. *J. Neurol. Sci.* **287** (1–2), 60–63 (2009).
54. Garcia-Borreguero, D., Egatz, R., Winkelmann, J. & Berger, K. Epidemiology of restless legs syndrome: The current status. *Sleep Med. Rev.* **10** (3), 153–167 (2006).
55. Brown, L. K., Dedrick, D. L., Doggett, J. W. & Guido, P. S. Antidepressant medication use and restless legs syndrome in patients presenting with insomnia. *Sleep Med.* **6** (5), 443–450 (2005).
56. Bonakis, A., Androutsou, A., Koloutsou, M-E. & Vagiakis, E. Restless legs syndrome masquerades as chronic insomnia. *Sleep Med.* **75**, 106–111 (2020).
57. Bjorvatn, B. et al. Prevalence, severity and risk factors of restless legs syndrome in the general adult population in two Scandinavian countries. *Sleep Med.* **6** (4), 307–312 (2005).
58. Currey, D. & Torous, J. Increasing the value of digital phenotyping through reducing missingness: A retrospective review and analysis of prior studies. *BMJ Ment. Health* **26** (1) (2023).
59. Dlima, S. D., Shevade, S., Menezes, S. R. & Ganju, A. Digital phenotyping in health using machine learning approaches: Scoping review. *JMIR Bioinform. Biotechnol.* **3** (1), e39618 (2022).
60. Perez-Pozuelo, I., Spathis, D., Clifton, E. A. & Mascolo, C. Elsevier,. Wearables, smartphones, and artificial intelligence for digital phenotyping and health. In *Digital Health* 33–54 (2021).

Acknowledgements
We would like to thank Editage (www.editage.co.kr) for editing the English language.

Author contributions
SPP, HJL, CHC conceived and designed the study. JJ, YJ, HK, YBS, TC, and CHC performed analyses. JJ, YJ, HK, and CHC wrote the first draft of the manuscript. JJ, YJ, HK, YBS, and CHC collected data. JJ, YJ, HK, YBS, JWY, SK, SPP, HJL, TC, and CHC edited the manuscript. All authors were involved in interpreting the results and have read, commented on, and approved the final version of the manuscript.

Funding
This work was supported by National Research Foundation (NRF) of Korea grants funded by the Ministry of Science and Information and Communications Technology (MSIT), Government of Korea (NRF-2021R1A5A8032895), Information and Communications Technology and Future Planning for Convergent Research in the Development Program for R&D Convergence over Science and Technology Liberal Arts (NRF-2022M3C1B6080866), and Institute of Information and communications Technology Planning and Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2023-00224823).

Declarations

Competing interests
The authors declare no competing interests.

Human ethics and consent to participate
All study procedures were reviewed and approved by the Institutional Review Board (IRB) of Korea University Anam Hospital (IRB No. 2022AN0587). Written informed consent was obtained from all the participants.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-01215-8.

Correspondence and requests for materials should be addressed to C.-H.C.