npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-01999-z

Actigraphy-based detection of isolated REM sleep behavior disorder: multicenter validation across devices and populations

Li Zhou[1,2,3,10], Andreas Brink-Kjaer[4,10], Katarina Gunter[5], Salonee Marwaha[6], Ambra Stefani[7], Birgit Högl[7], Michele T. Hu[5], Emmanuel Mignot[8], Ankit Parekh[1], Qi Tang[7], Merve Aktan-Süzgün[7], Bei Huang[2,3,9], Shi Tang[2,3], Siyi Gong[2,3], Yuhua Yang[2,3], Xie Chen[2,3], Jianzhang Ni[2,3], Ningning Li[2,3], Zhixuan He[2,3], Yun Kwok Wing[2,3,9,11] & Emmanuel During[1,6,11] ✉

Scalable home-based detection of isolated REM sleep behavior disorder (iRBD) is essential for early care, prevention trials, and identifying candidates for neuroprotective interventions against synucleinopathies. We previously showed that high-resolution wrist actigraphy (Axivity AX6) could identify iRBD based on abnormal sleep (AUC of 0.916) and rest-activity-rhythms (RAR, AUC of 0.856) using machine learning. Here, we aimed to assess generalizability across: (1) other actigraphs using lower resolutions; (2) different populations. We tested the analysis pipeline directly in cohorts from the International RBD Study Group using: Axivity AX6 (50–100 Hz), Philips Actiwatch (60-second epoch), and MicroMini-Motionlogger (30-second epoch). The cohorts included a total of 352 iRBD and 258 non-RBD participants from 4 centers (Mount Sinai, Oxford, Hong Kong, and Innsbruck). Two conversion pipelines were created to map activity counts from Actiwatch and MicroMini-Motionlogger to AX6 from 14 volunteers co-wearing two devices. In addition to the actigraphy analysis, four synucleinopathy prodromes—RBD symptoms, hyposmia, constipation, orthostatic hypotension—were tested in a two-stage screening approach. The sleep model achieved AUCs of 0.838–0.865 across centers, and the RAR model 0.520–0.818. Screening based on prodromes followed by actigraphy achieved sensitivities, specificities, and positive predictive values (PPVs) of 59.4–78.3%, 84.1–98.2%, and 56.0–98.6% (RBD symptoms), 46.5%, 99.0%, and 98.9% (hyposmia), 25.8–43.3%, 95.5–98.8%, 96.3–98.0% (constipation), and 11.6–36.8%, 96.0–100%, and 96.2-100% (orthostatic hypotension), respectively. Resolution (high versus low) did not affect the performance. After adjusting for a real-world iRBD prevalence of 1.5%, the corresponding PPVs would range from 6.3% to 100.0% depending on the prodromes. This multicenter study shows that the original actigraphy-based detection model of iRBD using sleep features but not RAR features generalizes well across independent cohorts and devices. Combined with key prodromes of synucleinopathies, it could enable precise, scalable population-level screening.

Isolated rapid eye movement (REM) sleep behavior disorder (iRBD) is a parasomnia characterized by the loss of normal muscle atonia during REM sleep due to the disinhibition of motor neurons, leading to repeated motoric behaviors potentially causing injuries to both patients and their bed partners[1–4]. Although its prevalence is 1–1.5% in the general population, affecting 30–50 million individuals worldwide, only a small proportion is diagnosed, to a large extent due to the cost and limited availability of the gold standard diagnostic test, the in-lab polysomnography[5–8]. In most cases, iRBD is an early manifestation of synucleinopathies, including Parkinson's disease, Dementia with Lewy bodies, and less commonly, multiple system atrophy[1,2]. Early diagnosis of iRBD is needed for timely clinical intervention, ensuring safety and symptom control, as well as enabling prevention trials

A full list of affiliations appears at the end of the paper. ✉e-mail: emmanuel.during@mssm.edu

and, eventually, the selection of suitable candidates for neuroprotective interventions. The traditional diagnostic method and gold-standard, video-polysomnography (vPSG), is resource-intensive and not always feasible for widespread use. Despite the development of several questionnaires to screen for and identify iRBD, there remain limitations of moderate specificity in the context of low prevalence rate of iRBD in the community, resulting in excessively low positive predictive value (PPV)[9–12]. Therefore, accurate yet scalable detection methods are needed to enable broader and earlier diagnosis of these individuals.

Our previous work demonstrated that wrist actigraphy and machine learning techniques could effectively identify iRBD through the automated analysis of abnormal movement patterns in sleep based on patients-filled sleep diaries, achieving very high accuracy[13]. A follow up study intended to test a fully automated approach that would spare the need for sleep diaries, and maintain the model's accuracy by adding 24-h rest-activity rhythms (RAR) features[14]. In this study, sleep and RAR features alone achieved areas under the curve (AUC) of 0.916 and 0.856, respectively, and a higher AUC of 0.954 when combined[14]. However, the study had several limitations. First, the iRBD sample was relatively small ($n = 42$) and drawn from a single center, which may limit generalizability. Additionally, the control group was derived from a different dataset (UK Biobank), collected a decade earlier and from a different geographic population (United Kingdom vs. United States), which might introduce bias into the RAR analyses. Second, it is not clear whether the device used (AX3 or AX6, Axivity Ltd, Newcastle) including its high-resolution capability (25–100 Hz) is required to achieve similar performance. In other words, the generalizability of the fully automated model across different wearable devices with varying resolutions remained unknown.

Therefore, this study aimed to assess the generalizability of our original fully automated actigraphy-based model for detecting iRBD by: (1) assessing the model's performance across diverse clinical cohorts with varying demographics; (2) testing its replicability across different wearable devices with lower sampling resolutions.

Methods

Study participants

This is a multi-center study involving 4 independent cohorts from the International RBD Study Group (IRBDSG), including Mount Sinai, Oxford, Hong Kong, and Innsbruck (Fig. 1: Schematic diagram of this study). All centers received approval from respective Institutional Review Boards. The inclusion criteria of iRBD patients were: 1) diagnosis of RBD confirmed by vPSG; 2) absence of overt neurodegenerative disease; 3) absence of other apparent causes such as narcolepsy. The control subjects were free of RBD diagnosis and any neurodegenerative disorders.

Synucleinopathy prodromal features measures

RBD symptoms. RBD was screened using the Innsbruck RBD Inventory summary-question (RBD-I-1Q) in the Mount Sinai cohort[15], Innsbruck RBD Inventory 5-item Version (RBD-I-5, total score ≥0.25) in the Oxford and Innsbruck cohorts[15], and RBD questionnaire-Hong Kong (RBDQ-HK, total score ≥19) in the Hong Kong cohort[16].

Constipation. At Mount Sinai, constipation was assessed by one-item question *("Do you experience constipation [difficulty having a bowel movement every day or having to strain hard to pass stools] OR need to use laxatives or other treatments to improve your bowel transit")*[13]. In Oxford, constipation was measured by MDS-UPDRS-part I[17]. In Hong Kong, constipation was defined as use of laxatives ≥1 times per week or bowel frequency ≤2 times per week, as measured by the SCOPA-AUT questionnaire[18].

Hyposmia. At Mount Sinai, was evaluated by a subjective one-item question ("*Have you been told, or have you noticed that your sense of smell or taste is reduced compared to others or to what it used to be?*")[13], while olfactory function was measured by objective olfactory tests in Oxford (Sniffin 16-item odor identification test, different cutoffs based on age) and Hong Kong cohorts (Olfactory Identification test, total score <3)[19,20].

Orthostatic hypotension. Orthostatic hypotension is defined across cohorts as a drop in systolic blood pressure of ≥ 20 mmHg or diastolic blood pressure of ≥ 10 mmHg after 2 or 3 min of standing, compared to blood pressure measured in the supine position.

Actigraphy data

Actigraphy data were collected across four centers using three different devices (Supplementary Table 1).

- *Axivity AX6* (Mount Sinai and Oxford). Participants wore an AX6 device (Axivity, Ltd, Newcastle, UK) on one hand for at least 2 weeks. The device was set at a resolution of 50 or 100 Hz and worn on the dominant hand or non-dominant hand in the Mount Sinai and Oxford cohorts, respectively.
- *Philips Actiwatch* (Hong Kong). Participants wore an Actiwatch Spectrum Plus (Philips Actiwatch Spectrum PRO, Philips Respironics, UK) for 1 week on their non-dominant hand. The device was set at a resolution of 60 s per epoch.
- *MicroMini-Motionlogger* (Innsbruck). Participants wore a MicroMini-Motionlogger (Ambulatory Monitoring, NY) on their dominant hand for one week. The device was set at a resolution of 30 s per epoch. Proportional Integrating Measure data was used for further analysis and non-wear time was defined as more than 60 min of consecutive zeros[21,22].

Sleep and RAR features

The details of sleep and RAR features analysis were provided in the previous work[13,14]. In brief, sleep features were extracted based on a previously developed sleep analysis pipeline that automatically detects sleep periods, extracts sleep activity features, and outputs an iRBD score based on these features. The pipeline was developed to analyze activity counts summarized in epoch sizes of 1-, 30-, or 60 s. Following automatic sleep detection, extracted features (total $n = 119$) describing various activity features across both the full sleep period and within time windows of interest (e.g., past first hour of sleep for increasing the probability of REM sleep)[8]. The features were fed to our machine learning model using boosted decision trees, which provides for each subject a per-night prediction score, which is then averaged across all available nights.

For RAR features, the raw accelerometer data from AX6 were processed using the biobankAccelerometerAnalysis software package, then cosinor and nonparametric analysis performed with "cosinor" and "nparACT" R packages[23,24]. The parameters, including mesor, amplitude, acrophase, interdaily variability, intradaily variability, L5 start time and count, and M10 start time and count, were generated for RAR model validation[23–26]. For the Actiwatch and MicroMini-Motionlogger devices, the 24-h physical activity level were directly used to generate the RAR features with the "cosinor" and "nparACT" R packages. Non-wear time was excluded from the analyses. Data were excluded at the daily level if invalid or non-wear data exceeded 20%.

Conversion pipeline for sleep model

Our analysis pipeline was developed to summarize raw accelerometer data into activity counts at 1-, 30-, and 60-second epochs (details were provided in the previous work)[13,14]. However, different actigraphy devices compute activity counts using proprietary algorithms, such as those used in the Actiwatch and MicroMini-Motionlogger. To enable cross-device application of our iRBD detection algorithm—which was originally developed using AX6 data—we sought to harmonize activity counts across devices via conversion models.

To do so, we recruited eight healthy volunteers (mean age ± SD: 28.4 ± 4.0 years, 37.5% male) who wore both the Actiwatch and AX6 concurrently on the same wrist over a 3-day period. Similarly, we collected data from another six healthy volunteers (29.8 ± 10.7 years, 12.5% male) co-

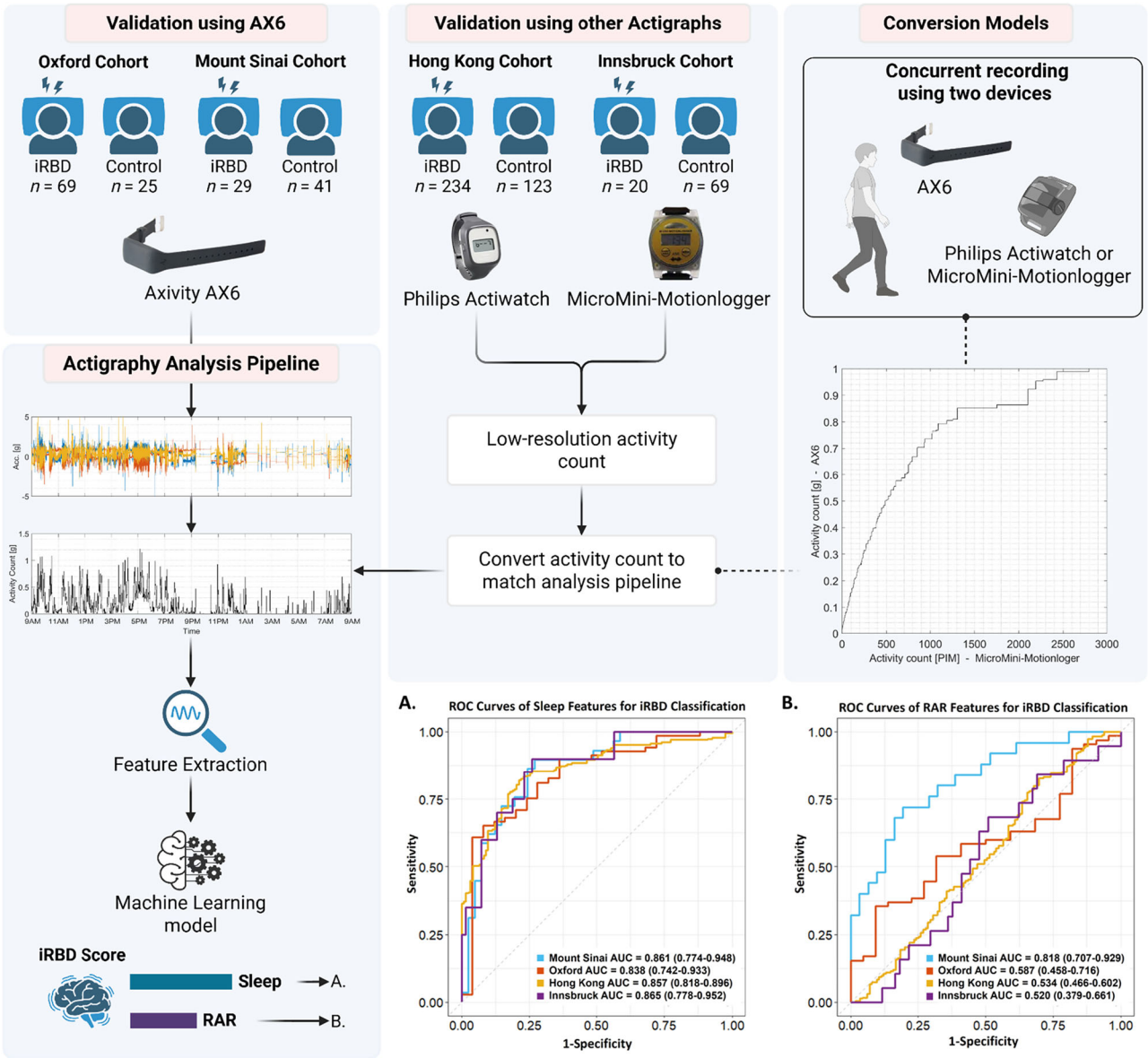

Fig. 1 | Schematic diagram of this study. AUC area under the curve, iRBD isolated rapid eye movement sleep behavior disorder, RAR rest-activity rhythm.

wearing the MicroMini-Motionlogger and AX6 over a 6-day period (Supplementary Table 2).

Then we fitted separate conversion models to map activity counts from each low-resolution device (MicroMini-Motionlogger and Actiwatch) to those of the AX6 in 30- and 60-second epochs. These models were trained using data from co-worn devices, ensuring aligned conditions across recordings. Prior to model fitting, we temporally aligned the activity count signals using cross-correlation within a ± 5-min window to correct for potential time shifts between devices.

The conversion models were developed using a consistent methodology in a leave-one-subject-out cross-validation setup. We evaluated two types of models: isotonic regression and polynomial regression of degrees 1, 3, and 5 (without a constant term). For polynomial models, we considered two input variants: (i) using the current activity count alone, and (ii) including the current and adjacent counts to account for local temporal context. To ensure a physiologically plausible, monotonically increasing mapping, the 3rd- and 5th-degree models were constrained to have positive derivatives, optimized using an interior-point algorithm. For the MicroMini-Motionlogger, activity counts were divided by a factor of 10,000 to provide a more similar range for fitting the polynomial models.

Following conversion, the harmonized activity counts were passed through our standard iRBD detection pipeline, allowing us to evaluate cross-device generalizability of the classifier without retraining.

Statistical analyses

Continuous variables were summarized as means and SD, while categorical variables were presented as number and percentages. Conversion model performance was assessed using two metrics: mean absolute error (MAE) and Spearman correlation. These were calculated both for the aligned activity counts and for the resulting iRBD score output by our classification model when applied to the converted activity counts. Conversion model evaluation metrics were computed using Matlab R2024b (The MathWorks, Natick, Massachusetts, USA). Finally, the best conversion model was chosen for each device based on the output iRBD score MAE. In case of ties, the simplest conversion model was selected.

Performance was evaluated separately for the classifiers using sleep features alone, RAR features alone, or combined sleep and RAR features, based on areas under the curve (AUC), sensitivity, specificity, accuracy, and PPV within each cohort. The 95% confidence intervals were estimated using Wald method. We also evaluated the performance of actigraphy-based

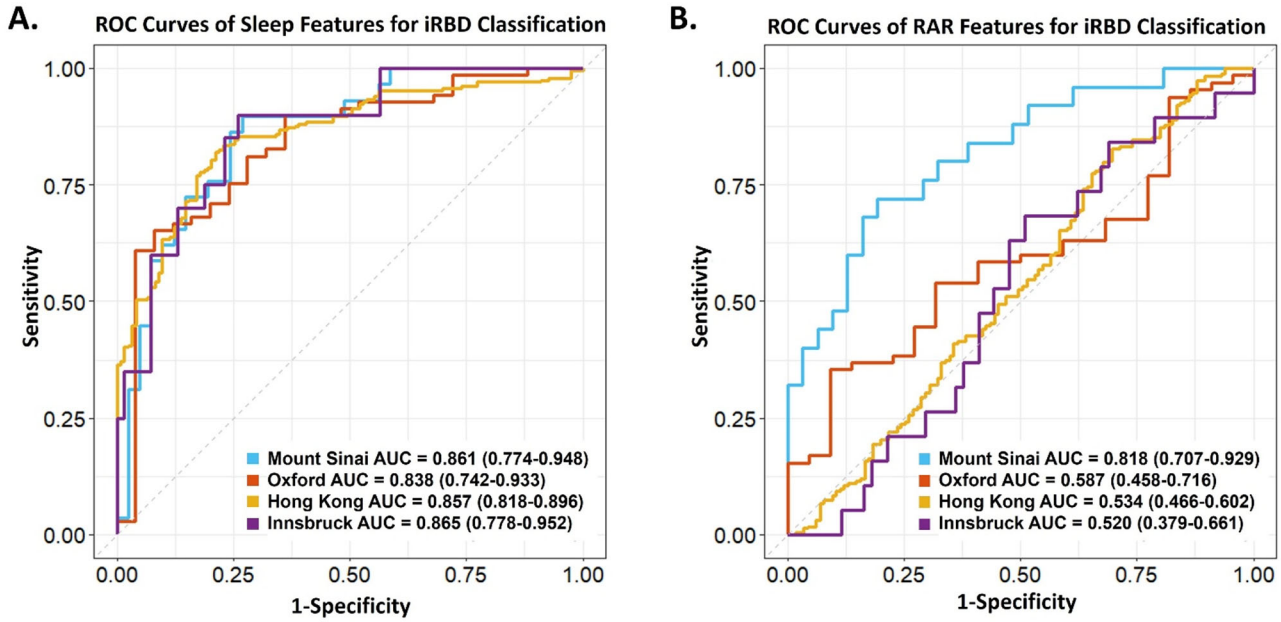

Fig. 2 | Sleep and RAR features predictive performance across cohorts. AUC area under the curve, iRBD isolated rapid eye movement sleep behavior disorder, RAR rest-activity rhythm.

classifiers combined with synucleinopathy prodromal features (two-stage model), in which subjects were stratified based on both prodromal symptoms and actigraphy classifier.

Additionally, we calculated projected real-world PPVs for all models. The adjusted PPVs were calculated using reported sensitivity and specificity estimates along with an assumed population prevalence of 1.5% based on the latest population-based study in which all subjects underwent PSG assessment[5]. The PPV was adjusted according to the following formula:

$$PPV = (Sensitivity \times Prevalence)/(Sensitivity \times Prevalence + (1 - Specificity) \times (1 - Prevalence))$$

Confidence intervals for the adjusted PPV were computed by applying this formula to the lower and upper bounds of the sensitivity and specificity 95% confidence intervals, providing a range of plausible PPV values given the uncertainty in the test parameters. These statistical analyses were performed using SPSS version 27.0 (IBM, Armonk, NY) and R software (version 4.4.0).

Results

A total of 352 vPSG-confirmed iRBD patients and 258 controls across sites––Mount Sinai (29 iRBD and 41 controls), Oxford (69 iRBD and 25 controls), Hong Kong (234 iRBD and 123 controls), and Innsbruck (20 iRBD and 69 controls)––were included in the analysis (Table 1).

Validation performance in cohorts using AX6 (Mount Sinai and Oxford)

In the Mount Sinai cohort, sleep and RAR models achieved AUC values of 0.861 (0.774–0.948) (Fig. 2a) and 0.818 (0.707–0.929) (Fig. 2b, respectively. The sensitivity, specificity, and accuracy for the sleep model were 72.4%, 80.5%, and 77.1%, while for the RAR model, the sensitivity, specificity, and accuracy are 80.0%, 61.3%, 69.6% (Supplementary Table 4). When combining the sleep and RAR features, the model achieved a sensitivity of 72.0%, specificity of 83.9%, and an accuracy of 78.6%. The PPVs were 72.4% for the sleep model, 62.5% for the RAR model, and 78.3% for the combined model. After adjusting for a population prevalence of 1.5%, the PPVs were 5.4%, 3.1%, and 6.4% for the sleep, RAR, and combined models, respectively.

In the Oxford cohort, our sleep and RAR models achieved AUC values of 0.838 (0.742–0.933) (Fig. 2a) and 0.587 (0.458-0.716) (Fig. 2b), respectively. The performance metrics are summarized in Table 2. The sensitivity, specificity, accuracy, and PPV for the sleep model were 84.1%, 64.0%, 78.7%, and 86.6% respectively, 6.2%, 100.0%, 29.9%, and 100.0% for RAR model, and 39.1%, 91.3%, 52.9%, and 92.6% for the combined model. After adjusting for a population prevalence of 1.5%, the PPVs were 3.43%, 100%, and 6.4% for the sleep, RAR, and combined models, respectively.

Supplementary Table 3 shows the comparisons of RAR features among Mount Sinai and Oxford cohorts. In the Mount Sinai cohort, iRBD patients showed significant differences from controls in mesor, amplitude, acrophase, relative amplitude, and M10. However, within the Oxford cohort, only relative amplitude and L5 significantly differed between groups.

Validation performance in cohorts using other devices

Conversion models. All conversion models were fitted and evaluated in a leave-one-subject-out cross validation scheme. The comparison of activity counts and resulting iRBD prediction score can be seen in Supplementary Tables 5–8.

For the Actiwatch, the MAE of activity counts was lowest at 0.014 ± 0.004 g for the 5th degree polynomials with and without neighboring context. However, for the iRBD prediction score, the lowest MAE at 1.41 was found using isotonic regression and a 1st degree polynomial without neighboring context. Therefore, the simple conversion with a 1st degree polynomial as $\hat{a}_{AX6} = 0.0014 \cdot a_{Actiwatch}$, where $a_{Actiwatch}$ is measured activity count from the Actiwatch and $\hat{a}_{AX6}$ is the estimated AX6 activity count.

For the MicroMini-Motionlogger, the MAE of the activity counts was similarly lowest at 0.009 ± 0.004 g for the 5th degree polynomials with and without neighboring context. For the iRBD prediction score, the isotonic regression provided a better fit with a MAE of 2.24. Therefore, it was chosen as the conversion model for this device with the curve as displayed in Supplementary Fig. 1.

Validation performance in Actiwatch (Hong Kong)

After converting Actiwatch data, our sleep model achieved an AUC value of 0.857 (0.818-0.933) (Fig. 2a) with a sensitivity of 63.2% and specificity of 90.2% in the Hong Kong cohort (Table 3). Due to the significant differences in the magnitudes of the mesor, amplitude, L5 count, and M10 count features between the Actiwatch and AX6 devices, we averaged the ratios from data collected from eight healthy subjects. These averaged ratios were

Table 1 | Demographics and neurodegenerative symptoms across four cohorts

	Mount Sinai cohort (*N* = 70)		Oxford cohort (*N* = 94)		Hong Kong cohort (*N* = 357)		Innsbruck cohort (*N* = 89)	
	iRBD (*N* = 29)	Control (*N* = 41)	iRBD (*N* = 69)	Control (*N* = 25)	iRBD (*N* = 234)	Control (*N* = 123)	iRBD (*N* = 20)	Control[a] (*N* = 69)
Age, year	69.7 ± 10.2	57.3 ± 12.5	70.7 ± 7.2	70.2 ± 9.9	67.3 ± 7.9	62.2 ± 11.4	70.9 ± 8.4	47.0 ± 13.9
Male	20 (69.0)	18 (43.9)	6 (8.7)	13 (52.0)	167 (71.4)	68 (55.3)	17 (85.0)	50 (72.5)
BMI	26.5 ± 5.3	25.2 ± 3.8 N = 28	26.4 ± 4.2	27.2 ± 7.2	24.4 ± 3.6	25.4 ± 3.7	26.6 ± 3.5	26.8 ± 5.4
Ethnicity								
Asian	-	-	1 (1.4)	0 (0)	234 (100)	123 (100)	0 (0)	3 (4.3)
White	20 (69.0)	34 (82.9)	67 (97.1)	25 (100)	-	-	20 (100)	66 (95.7)
Others	9 (31.0)	7 (17.1)	1 (1.4)	0 (0)	-	-	-	-
RBD symptoms[b]	16 (64.0) *N* = 25	2 (22.2) *N* = 9	56 (93.3) *N* = 60	3 (13.6) *N* = 22	214 (91.5) *N* = 234	16 (14.2) *N* = 113	16 (80.0)	21 (30.4)
Hyposmia[c]	10 (45.5) *N* = 22	2 (20.0) *N* = 10	36 (55.4) *N* = 65	0 (0) *N* = 3	143 (71.5) *N* = 200	17 (16.2) *N* = 105	-	-
Constipation	11 (45.8) *N* = 25	1 (10.0) *N* = 10	29 (48.3) *N* = 60	5 (22.7) *N* = 22	76 (39.2) *N* = 194	7 (8.1) *N* = 86	-	-
Orthostatic hypotension[d]	5 (21.7) *N* = 23	-	26 (38.2) *N* = 68	3 (12.0)	30 (14.5) *N* = 207	3 (3.6) *N* = 83	-	-

Data was shown as mean ± SD or *n* (%).
iRBD isolated rapid eye movement sleep behavior disorder, *BMI* body mass index.
[a]Control includes 19 healthy controls, 20 RLS patients, 10 patients with RLS and sleep apnea, 20 sleep apnea patients.
[b]RBD symptoms were measured by RBD single-question in Mount Sinai cohort, RBD screening questionnaire (total score ≥5) in Oxford and Innsbruck cohorts, and RBD questionnaire-Hong Kong (total score ≥19) in Hong Kong cohort.
[c]Hyposmia was evaluated by subjective one-item question at Mount Sinai, while olfactory function was measured by objective olfactory tests in Oxford (Sniff test, different cutoffs based on age) and Hong Kong cohorts (Olfactory Identification test, total score <3).
[d]Orthostatic hypotension is defined as a drop in systolic blood pressure of ≥ 20 mmHg or diastolic blood pressure of ≥ 10 mmHg after 2 min of standing, compared to blood pressure measured in the supine position.

then used to convert the raw RAR features for the RAR model. The converted RAR features showed comparable magnitude to those obtained from the AX6 device, with significant differences between iRBD patients and controls in mesor, amplitude, Interdaily stability, L5 start time and count, and M10 start time and count (Supplementary Table 3). Leveraging the converted RAR features, the RAR model achieved an AUC value of 0.534 (0.466-0.602) (Fig. 2b) with a sensitivity of 37.9%, specificity of 58.3%, and an accuracy of 44.7% (Table 3). The combined model achieved a sensitivity of 49.3%, specificity of 73.0%, and an accuracy of 57.3%. The PPVs and adjusted PPVs were 92.5% and 8.9% for sleep model, 64.2% and 1.37% for the RAR model, and 78.3% and 2.7% for the combined model, respectively.

Validation performance in MicroMini-Motionlogger (Innsbruck)

Our sleep model achieved an AUC of 0.865 (0.778–0.948) (Fig. 2a) with a sensitivity of 90.0%, specificity of 65.2%, accuracy of 70.8%, PPV of 42.9%, and an adjusted PPV of 3.8% in the Innsbruck cohort (Supplementary Table 9). Similar to the Actiwatch, significant differences were also observed in the magnitudes of the mesor, amplitude, L5 count, and M10 count features between the MicroMini-Motionlogger and AX6 devices. Thus, we applied the same conversion of RAR features for Motionlogger using data from six healthy subjects. Supplementary Table 3 shows comparable magnitude for the converted mesor, amplitude, L5 count, and M10 count features between the MicroMini-Motionlogger and AX6 devices. No significant differences were found among RAR features between iRBD patients and controls. Leveraging the converted RAR features, the RAR model alone achieved an AUC value of 0.520 (0.379–0.661) (Fig. 2b), with a sensitivity of 31.6%, specificity of 55.7%, accuracy of 50.0%, PPV of 18.2%, and an adjusted PPV of 1.08% (Supplementary Table 9). The model combining sleep and RAR achieved a sensitivity of 57.9%, specificity of 59.0%, accuracy of 58.8%, PPV of 30.6%, with an adjusted PPV of 2.1%.

Performance of a two-stage screening approach based on prodromal synucleinopathy features and wearable data

In participants endorsing RBD symptoms, the two-stage screening method resulted in a raise in specificities up to 95.5% in the Oxford cohort, 98.2% in the Hong Kong cohort, and 84.1% in the Innsbruck cohort, with corresponding PPVs of 97.9%, 98.6%, and 56.0%, respectively. After adjusting for general population prevalence, the PPVs ranged from 6.3% to 33.4% across the three cohorts, with sensitivities ranging from 59.4% to 78.3%.

In those with constipation, the specificity, PPV, and adjusted PPV increased to 95.5%, 96.3%, and 12.8% in the Oxford cohort, and to 98.8%, 98.0%, and 24.7% in the Hong Kong cohort, with corresponding sensitivities of 43.3% and 25.8%, respectively.

In participants with orthostatic hypotension, the specificity, PPV, and adjusted PPV reached to 100% in the Hong Kong cohort, though the sensitivity reduced to 11.6%. In the Oxford cohort, the specificity, PPV, and adjusted PPV were 96.0%, 96.2%, and 12.3%, respectively, with a sensitivity of 36.8%.

In terms of hyposmia, data were available only in the Hong Kong cohort, showing a specificity of 99.0%, PPV of 98.9%, and adjusted PPV of 41.5%, with a sensitivity of 46.5%. Due to the limited sample size for synucleinopathy prodromal features in the Mount Sinai cohort, we could not perform a two-stage analysis.

Discussion

This study is the first to demonstrate generalizability of a fully automated algorithm for detecting iRBD across clinical cohorts with different ethnicities, geographical backgrounds, and wrist-worn accelerometer devices. Our model using sleep movement data achieved good performance with AUCs ranging between 0.838 and 0.865—only modestly reduced from its initial performance (AUC: 0.916), independent of high (50−100 Hz) versus low (30–60 s) resolution. This highlights the model's robustness not only across cohorts with varying demographic and clinical characteristics, but

Table 2 | Performance of iRBD classifier in Oxford cohort

	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	PPV (95% CI)	Adjusted PPV (95% CI)
Actigraphy sleep features only	84.1 (75.4, 92.7)	64.0 (45.2, 82.8)	78.7 (70.4, 82.8)	86.6 (78.4, 94.7)	3.43 (2.05, 7.58)
Actigraphy sleep features + symptoms					
RBD symptoms	78.3 (67.9, 88.8)	95.5 (86.8, 100.0)	82.9 (74.8, 91.1)	97.9 (93.9, 100.0)	20.9 (7.3, 100)
Constipation	43.3 (30.8, 55.9)	95.5 (86.8, 100.0)	57.3 (46.6, 68.0)	96.3 (89.2, 100.0)	12.8 (3.4, 100)
Orthostatic hypotension	36.8 (25.3, 48.2)	96.0 (88.3, 100.0)	52.7 (42.5, 62.8)	96.2 (88.8, 100.0)	12.3 (3.2, 100)
Actigraphy RAR features only	6.2 (0.3, 12.0)	100.0 (100.0, 100. 0)	29.9 (20.3, 39.5)	100.0 (100.0, 100.0)	100.0 (100.0, 100.0)
Actigraphy sleep + RAR features	39.1 (27.1, 51.0)	91.3 (79.8, 100)	52.9 (42.4, 63.4)	92.6 (82.7, 100)	6.4 (2.0, 100)

iRBD isolated rapid eye movement sleep behavior disorder, *RAR* rest activity rhythm, *CI* confidence interval, *PPV* positive predictive value.

Table 3 | Performance of iRBD classifier in Hong Kong cohort

	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	PPV (95% CI)	Adjusted PPV (95% CI)
Actigraphy sleep features only	63.2 (57.1, 69.4)	90.2 (85.0, 95.5)	72.5 (67.9, 77.2)	92.5 (88.4, 96.6)	8.9 (5.5, 19.0)
Actigraphy sleep features + symptoms					
RBD symptoms	59.4 (53.1, 65.7)	98.2 (95.8, 100)	72.0 (67.3, 76.8)	98.6 (96.6, 100)	33.4 (16.1, 100)
Constipation	25.8 (19.6, 31.9)	98.8 (96.6, 100)	48.2 (42.4, 54.1)	98.0 (94.2, 100)	24.7 (8.1, 100)
Orthostatic hypotension	11.6 (7.2, 16.0)	100 (100, 100)	36.9 (31.3, 42.5)	100 (100, 100)	100 (100, 100)
Hyposmia	46.5 (39.6, 53.4)	99.0 (97.2, 100)	64.6 (59.2, 70.0)	98.9 (96.9, 100)	41.5 (17.7, 100)
Actigraphy RAR features only	37.9 (31.6, 44.2)	58.3 (49.2, 67.3)	44.7 (39.5, 50.0)	64.2 (56.1, 72.3)	1.37 (0.94, 2.02)
Actigraphy sleep + RAR features	49.3 (42.8, 55.8)	73.0 (64.9, 81.2)	57.3 (52.1, 62.6)	78.3 (71.6, 85.1)	2.7 (1.8, 4.3)

iRBD isolated rapid eye movement sleep behavior disorder, *RAR* rest-activity rhythm, *CI* confidence interval, *PPV* positive predictive value.

also across device brands, using different resolution and proprietary methods for measuring activity.

Our results show that strong model performance can be achieved with as little as 7 nights of data, even when using low-resolution devices with 60-second sampling rates. Notably, the highest performance was observed in the Innsbruck cohort using a different device (MicroMini-Motionlogger), while the lowest was in the Oxford cohort using the same device (AX6, as used to train the original model). Although customized conversion models for activity counts may be required, these findings support the feasibility of applying the model across diverse research-grade and possibly consumer devices—potentially enabling scalable RBD detection in broader populations.

When testing a two-stage screening considering synucleinopathy prodromal features, as expected, sensitivities generally reduced; however, specificity and resulting PPVs increased. The sleep-based classifier yielded sensitivities and specificities as high as 78.3% and 98.2%, respectively, in individuals reporting RBD symptoms, and even higher specificity in those with hyposmia (99.0%), constipation (up to 98.8%), and orthostatic hypotension (up to 100%). Because RBD questionnaires lack specificity in community settings, efforts to improve their predictive value have included follow-up structured phone interviews or clinical evaluations by sleep specialists. In a recent study, a sequential screening protocol—starting with a single-item screen, followed by a phone interview, and then clinical evaluation in selected cases—achieved a final PPV of 33.3%[27]. In the largest Hong Kong cohort, combining actigraphy with RBD questionnaire yielded a comparable projected real-world PPV of 33.4%. These findings support the feasibility and rationale of a staged screening strategy that leverages both wrist movement data and synucleinopathy prodromes for cost-effective, scalable early detection of high-risk individuals.

In comparison to the sleep-based detection model, the RAR model showed significantly lower performance, with AUCs ranging from 0.520 to 0.818, and poor generalization. This could be due to heterogeneity in individual rest-activity patterns, the influence of comorbid conditions, or variability in daily routines that obscure disease-specific alterations[28,29]. Our findings also revealed inconsistent differences in RAR features between iRBD patients and controls across cohorts. In contrast to relatively stable sleep-based features, which are assumably related to the neurophysiologic hallmark of RBD (loss of the REM sleep atonia), rest-activity patterns can be influenced by multiple other factors (socioeconomical, cultural, seasonality, the mere influence of weather) with a significant global disparity limiting their utility for individual-level iRBD detection[30].

In summary, our findings support the use of wearable devices for a scalable, non-invasive screening in the community or at-risk populations, and offers greater accuracy than RBD questionnaires alone. Going forward, actigraphy could be integrated in population screening to enhance recruitment pipelines in neuroprotective trials and accelerate drug discovery[31]. Further, earlier diagnosis of individuals with RBD can facilitate access to symptomatic care and reduce the risk of injuries[4]. In the clinical setting, individuals endorsing RBD symptoms and screening positively on actigraphy could be prioritized for vPSG assessments, potentially minimizing diagnostic delays[3,20,32].

Several limitations need to be acknowledged. First, this is a retrospective study involving sites and cohorts that differed in research protocols related to wearable data collection, clinical assessments and questionnaire administration. As a result, there was significant data heterogeneity and incomplete availability of certain prodromal features across some cohorts. Second, although the sleep model generalized well, its performance still varies across cohorts with different cultural or environmental contexts, with sensitivities ranging from 63.2% to 90.0% and specificities from 64.0% to 90.2%. This variability underscores the need for threshold optimization tailored to the specific application and the desired balance between sensitivity and specificity. Third, it remains unclear whether our models can effectively capture the prodromal stage of iRBD, which is defined as the phase in which symptoms and signs of evolving RBD are present but do not yet meet the behavioral or polysomnography diagnostic criteria for iRBD[5,33–35]. Future studies focusing on prodromal RBD are needed. Finally, although we adjusted PPVs for estimated general population prevalence, these values may not fully generalize to real-world settings where disease

prevalence varies, especially in populations with different demographic or clinical characteristics. This highlights the need for validation of predictive models in diverse cohorts and implementation in the community setting to ensure reliable estimation of PPV.

Conclusion

Our study supports the utility and generalizability of an actigraphy-based machine learning model for the early and scalable detection of iRBD, particularly when combined with prodromal features of synucleinopathy (two-stage model). Sleep-based features demonstrated robust generalizability with only 7 nights of data, whereas the original RAR-based model showed lower and more variable performance, highlighting the need for further refinement and alternative approaches. Continued model development leveraging large and diverse datasets is warranted, as are studies evaluating how combinations of prodromal features followed by actigraphy can be operationalized in a staged screening strategy for RBD and early neurodegeneration.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Received: 9 August 2025; Accepted: 8 September 2025;
Published online: 29 October 2025

References

1. Dauvilliers, Y. et al. REM sleep behaviour disorder. *Nat. Rev. Dis. Prim.* **4**, 19 (2018).
2. Wing, Y. K. et al. Prospective outcome of rapid eye movement sleep behaviour disorder: psychiatric disorders as a potential early marker of Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **83**, 470–472 (2012).
3. Fernández-Arcos, A., Iranzo, A., Serradell, M., Gaig, C. & Santamaria, J. The clinical phenotype of idiopathic rapid eye movement sleep behavior disorder at presentation: a study in 203 consecutive patients. *Sleep* **39**, 121–132 (2016).
4. Yang, Y. et al. Sleep related injury and its correlates in isolated rapid eye movement sleep behavior disorder. *Sleep. Med* **126**, 9–18 (2025).
5. Lee, W. J. et al. REM sleep behavior disorder and its possible prodromes in general population: prevalence, polysomnography findings, and associated factors. *Neurology* **101**, e2364–e2375 (2023).
6. Cicero, C. E. et al. Prevalence of idiopathic REM behavior disorder: a systematic review and meta-analysis. *Sleep* **44**, zsaa294 (2021).
7. Sasai-Sakuma, T., Takeuchi, N., Asai, Y., Inoue, Y. & Inoue, Y. Prevalence and clinical characteristics of REM sleep behavior disorder in Japanese elderly people. *Sleep* **43**, zsaa024 (2020).
8. Haba-Rubio, J. et al. Prevalence and determinants of rapid eye movement sleep behavior disorder in the general population. *Sleep* **41**, zsx197 (2018).
9. Stiasny-Kolster, K. et al. The REM sleep behavior disorder screening questionnaire–a new diagnostic instrument. *Mov. Disord.* **22**, 2386–2393 (2007).
10. Stefani, A. et al. Low Specificity of Rapid Eye Movement Sleep Behavior Disorder Questionnaires: Need for Better Screening Methods. *Mov. Disord.* **38**, 1000–1007 (2023).
11. Postuma, R. B. et al. A single-question screen for rapid eye movement sleep behavior disorder: a multicenter validation study. *Mov. Disord.* **27**, 913–916 (2012).
12. Cesari, M. et al. Isolated REM sleep behaviour disorder-is screening possible? *J Sleep Res.* e70109 (2025).
13. Brink-Kjaer, A. et al. Ambulatory detection of isolated rapid-eye-movement sleep behavior disorder combining actigraphy and questionnaire. *Mov. Disord.* **38**, 82–91 (2023).
14. Brink-Kjaer, A. et al. Fully automated detection of isolated rapid-eye-movement sleep behavior disorder using actigraphy. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.* **2023**, 1–5 (2023).
15. Frauscher, B. et al. Validation of the Innsbruck REM sleep behavior disorder inventory. *Mov. Disord.* **27**, 1673–1678 (2012).
16. Li, S. X. et al. Validation of a new REM sleep behavior disorder questionnaire (RBDQ-HK). *Sleep. Med* **11**, 43–48 (2010).
17. Gallagher, D. A., Goetz, C. G., Stebbins, G., Lees, A. J. & Schrag, A. Validation of the MDS-UPDRS Part I for nonmotor symptoms in Parkinson's disease. *Mov. Disord.* **27**, 79–83 (2012).
18. Visser, M., Marinus, J., Stiggelbout, A. M. & van Hilten, J. J. Assessment of autonomic dysfunction in Parkinson's disease: The SCOPA-AUT. *Mov. Disord.* **19**, 1306–1312 (2004).
19. Chan, A., Tam, J., Murphy, C., Chiu, H. & Lam, L. Utility of olfactory identification test for diagnosing Chinese patients with Alzheimer's disease. *J. Clin. Exp. Neuropsychol.* **24**, 251–259 (2002).
20. Zhou, L. et al. Early- and late-onset of isolated rapid eye movement sleep behavior disorder: A retrospective cohort study. *Sleep. Med* **105**, 1–8 (2023).
21. Bucko, A. G., Armstrong, B., McIver, K. L., McLain, A. C. & Pate, R. R. Longitudinal associations between sleep and weight status in infants and toddlers. *Pediatr. Obes.* **18**, e13056 (2023).
22. Troiano, R. P. et al. Physical activity in the United States measured by accelerometer. *Med. Sci. sports Exerc.* **40**, 181–188 (2008).
23. Sachs, M. cosinor: Tools for Estimating and Predicting the Cosinor Model. R Package Version 1.1, Available at: https://CRAN.Rproject.org/package=cosinor Accessed October 1, 2019 (2014).
24. Blume, C., Santhi, N. & Schabus, M. 'nparACT' package for R: A free software tool for the non-parametric analysis of actigraphy data. *MethodsX* **3**, 430–435 (2016).
25. Feng, H. et al. Rest-activity pattern alterations in idiopathic REM sleep behavior disorder. *Ann. Neurol.* **88**, 817–829 (2020).
26. Minaeva, O. et al. Individual-specific change points in circadian rest-activity rhythm and sleep in individuals tapering their antidepressant medication: an actigraphy study. *Sci. Rep.* **14**, 855 (2024).
27. Cicero, C. E. et al. Prevalence of isolated RBD in the city of Catania, Italy: a population-based study. *J. Clin. Sleep. Med.* **17**, 2241–2248 (2021).
28. Makarem, N. et al. Rest-activity rhythms are associated with prevalent cardiovascular disease, hypertension, obesity, and central adiposity in a nationally representative sample of US adults. *J. Am. Heart Assoc.* **13**, e032073 (2024).
29. Brito, L. C. & Thosar, S. S. Rest–activity rhythm as a clinical biomarker: what are the next steps?. *Lancet Healthy Longev.* **4**, e179–e180 (2023).
30. Lim, D. C. et al. The need to promote sleep health in public health agendas across the globe. *Lancet Public Health* **8**, e820–e826 (2023).
31. Postuma, R. B. Neuroprotective trials in REM sleep behavior disorder: the way forward becomes clearer. *Neurology* **99**, 19–25 (2022).
32. White, C., Hill, E. A., Morrison, I. & Riha, R. L. Diagnostic delay in REM sleep behavior disorder (RBD). *J. Clin. Sleep. Med* **8**, 133–136 (2012).
33. Cesari, M. et al. Video-polysomnography procedures for diagnosis of rapid eye movement sleep behavior disorder (RBD) and the identification of its prodromal stages: guidelines from the International RBD Study Group. *Sleep* **45**, zsab257 (2021).
34. Liu, Y. et al. Evolution of prodromal REM sleep behavior disorder to neurodegeneration: a retrospective longitudinal case-control study. *Neurology* **99**, e627–e637 (2022).
35. Huang, B. et al. Isolated dream-enactment behaviours as a prodromal hallmark of rapid eye movement sleep behaviour disorder. *J Sleep Res.* e13791 (2022).

Author contributions

L.Z., A.B.K., Y.K.W., and E.D. contributed to the conception and design of the study. L.Z., A.B.K., and E.D wrote the main paper. L.Z., A.B.K., K.G., A.P., and E.D handled data analysis and interpretation. L.Z., K.G., S.M., A.S., B.G., M.H., E.M., Q.T., M.A.S., B.H., S.T., S.Y.G., Y.H.Y., X.C., J.Z.N., Z.X.H., Y.K.W., E.D. contributed to data collection and preparation. All authors have reviewed and approved the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-025-01999-z.

Correspondence and requests for materials should be addressed to Emmanuel During.

[1]Department of Medicine, Division of Pulmonary, Critical Care and Sleep Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2]Department of Psychiatry, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China. [3]Li Chiu Kong Family Sleep Assessment Unit, Department of Psychiatry, Faculty of Medicine, The Chinese University of Hong Kong, Ma Liu Shui, Hong Kong SAR, China. [4]Department of Health Technology, Denmark Technical University, Kongens Lyngby, Denmark. [5]Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, Oxford, UK. [6]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York City, NY, USA. [7]Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria. [8]Department of Psychiatry and Behavioral Sciences, Stanford Center for Sleep Sciences and Medicine, Stanford University, Redwood City, CA, USA. [9]Li Ka Shing Institute of Health Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong, China. [10]These authors contributed equally: Li Zhou, Andreas Brink-Kjaer. [11]These authors jointly supervised this work: Yun Kwok Wing, Emmanuel During. emmanuel.during@mssm.edu