Sleep Medicine Reviews 88 (2026) 102298

Contents lists available at ScienceDirect

Sleep Medicine Reviews

journal homepage: www.elsevier.com/locate/smrv

Clinical Review

Machine learning to diagnose, classify and predict phenoconversion in isolated REM sleep behavior disorder

Kausar Raheel [a,1], Laurent Sheybani [b,1], Nazanin Biabani [a], See Qi Rui [a], Alessio Delogu [c], Jan Rosenzweig [d], Zoran Cvetkovic [d], Robert Leech [a,e], Nir Grossman [f,g], Masoud Tahmasian [h,i,j], Peter J. Goadsby [k,l], K. Ray Chaudhuri [m], Carlos H. Schenck [n], Ivana Rosenzweig [a,o,*]

[a] Sleep and Brain Plasticity Centre, Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, UK
[b] Department of Clinical and Experimental Epilepsy, University College London, London, UK
[c] Basic and Clinical Neuroscience, IoPPN, King's College London, London, WC2R 2LS, UK
[d] Department of Engineering, King's College London, UK
[e] Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, UK
[f] Department of Brain Sciences, Imperial College London, London, UK
[g] UK Dementia Research Institute, Imperial College London, London, UK
[h] Institute of Neurosciences and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany
[i] Institute of Systems Neuroscience, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[j] Department of Nuclear Medicine, University Hospital and Medical Faculty, University of Cologne, Cologne, Germany
[k] NIHR-Wellcome Trust King's Clinical Research Facility, King's College London, London, WC2R 2LS, UK
[l] Division of Biomedical Sciences, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[m] Kings College Hospital London, Dubai, United Arab Emirates
[n] Minnesota Regional Sleep Disorders Center, and Departments of Psychiatry, Hennepin County Medical Center, and University of Minnesota Medical School, Minneapolis, MN, USA
[o] Sleep Disorders Centre, Guy's and St Thomas' NHS Foundation Trust, London, UK

ARTICLE INFO

Keywords:
REM sleep behaviour disorder
Machine learning
α-synucleinopathies
Parkinson's disease
Dementia with Lewy bodies
Multiple system atrophy

ABSTRACT

Rapid eye movement (REM) sleep behaviour disorder (RBD), particularly its idiopathic/isolated form (iRBD), is a prodromal marker for α-synucleinopathies, including Parkinson's disease, dementia with Lewy bodies and multiple system atrophy. Machine learning (ML) offers opportunities to improve diagnosis and risk stratification in this high-risk group. We conducted a systematic review of PubMed, Embase (Ovid) and Medline (Ovid) from 2014 to September 2025, following PRISMA guidelines. From 335 records identified, 202 remained after duplicate removal and 75 studies on adult humans with clinically diagnosed RBD or iRBD that applied and validated an ML model were included. Fifty-eight studies addressed diagnosis, four studied RBD phenotypes, and thirteen evaluated prediction of phenoconversion to overt α-synucleinopathy. Across diagnostic studies, reported accuracies ranged from ~63% to ~99.7%, with median values around 90%, using polysomnography, EEG, neuroimaging, molecular and behavioural markers. Phenoconversion models (often using dopaminergic imaging or multimodal features) achieved AUCs up to ~0.94, but frequently relied on small, single-centre cohorts with heterogeneous definitions of phenoconversion and limited external validation. A wide variety of ML algorithms was used (n ~ 30), most commonly support vector machines, random forests and logistic regression. Overall, ML approaches show promise for scalable diagnosis and risk stratification in iRBD, but progress is constrained by methodological bias, inconsistent endpoints, data imbalance and a lack of explainable, externally validated models. We outline methodological priorities to make future ML tools clinically interpretable and translatable.

* Corresponding author. Sleep and Brain Plasticity Centre, Department of Neuroimaging, Box 089, Institute of Psychiatry, Psychology and Neuroscience, De Crespigny Park, London, SE5 8AF, UK.
E-mail address: ivana.1.rosenzweig@kcl.ac.uk (I. Rosenzweig).
[1] Joint authorship.

https://doi.org/10.1016/j.smrv.2026.102298
Received 13 August 2025; Received in revised form 30 March 2026; Accepted 20 April 2026
Available online 29 April 2026

1. Introduction

Neurodegenerative diseases are prevalent, detrimental neurological conditions with a high burden on quality of life and health cost. Among them, Parkinson's disease (PD) is the second most frequent neurodegenerative disease after Alzheimer disease and affects ~1% of the population >60 years old [1]. By 2040, the prevalence of PD could reach 12-17 million worldwide, and it is expected that the economic burden will reach almost $80 billion in the US [2]. In 2016, PD contributed to 3.2 million disability-adjusted life-years worldwide [3]. Despite this considerable burden on healthcare systems and quality of life, we still lack disease-modifying drugs against PD and other α-synucleopathies (Multiple System Atrophy, MSA and Dementia with Lewy Body, DLB), although ongoing research has made some promising findings in the last few years [4]' [5]. For example, certain antidiabetic medications have been associated with better outcome in people with PD [6–9]. This should motivate early diagnosis and interventions to slow down progression. For that purpose, the identification of reliable biomarkers of the prodromal phases of neurodegenerative disease is a priority for optimising neuroprotective treatments [9,10].

Subtle symptoms or syndromes can occur years to decades before overt α-synucleopathies, including idiopathic/isolated REM Sleep Behaviour Disorder (iRBD). RBD is a REM sleep parasomnia characterized by loss of the normal muscle atonia of REM sleep, a phenomenon known as REM sleep without atonia (RSWA) [11–14]. Diagnosis of RBD requires video-polysomnography to document REM sleep without atonia (RSWA), combining electroencephalography (EEG) and electro-oculography (EOG) to identify REM sleep and electromyography (EMG) to demonstrate loss of muscle atonia [11–15]. RSWA can be quantified using validated scoring approaches (e.g., AASM or SINBAR-derived metrics), and quantitative phasic/tonic cut-offs have been proposed to support diagnosis, although thresholds depend on montage, comorbidities and scoring method [16,17]. Although vPSG is the diagnostic gold standard, the sensitivity of a single-night study can be limited by night-to-night variability in RSWA and dream-enactment behaviours; therefore, repeat recordings may be required in clinically suspicious cases with equivocal initial findings. Behaviourally, individuals with RBD may exhibit simple or complex motor behaviours such as limb jerks, talking, or more vigorous enactment of dreams; while serious violent episodes can occur, large video-PSG series show that the majority of events are brief and relatively simple, with injury-causing behaviours representing a minority [18]. These dream-enactment episodes may result in injury to themselves or their bed partners [11–13].

Meta-analytic data suggest that the majority of patients with iRBD eventually develop an α-synucleinopathy [19]. Longitudinal cohorts report phenoconversion rates of 33.5% at 5 years, 82.4% at 10.5 years and 96.6% at 14 years [19], highlighting iRBD as one of the most specific prodromal markers for Parkinson's disease and related α-synucleinopathies [10,20]. RBD is thus a critical candidate biomarker to identify and eventually treat people in the very early stages of a neurodegenerative disease. However, diagnosis of RBD requires a video polysomnography (vPSG), an expensive medical procedure usually requiring a night in a medical setting. This is an obstacle to diagnosis. Questionnaire-based estimates of probable RBD in the general population are higher (around 5–6%) than vPSG-confirmed RBD (approximately 1.2–1.3%) [20–22]. However, multicentre validation work has shown that questionnaires have limited specificity for RBD and tend to overestimate true prevalence [23]. Therefore, we interpret these figures as showing a discrepancy between screening tools and definitive vPSG-based diagnosis, rather than as evidence that 'true' RBD prevalence is 5.5%. Alternative means to identify RBD patients are thus needed.

To this end, machine learning (ML) could be used to aid in disease detection and clinical decision in RBD [24–28]. ML is a branch of AI that focuses on development of algorithms that learn, make inferences and identify patterns in data to eventually generalize the learnt model on unseen dataset [29]. Regressions and Support Vector Machine (SVM) are examples of relatively simple and complex ML models respectively. An ML algorithm typically involves selecting and extracting specific features from raw data, acquired using, e.g., neuroimaging or neurophysiological techniques, and training the algorithm to classify subjects based on a variable of interest to eventually validate the accuracy of the algorithm on unseen data (see Fig. 1) [27].

Recent works have reported the use of various ML algorithms to differentiate RBD, other sleep disorders and healthy individuals using different biomarkers [26,30,31]. ML algorithms offer promising avenues not only because they can identify subjects at risk in a resource efficient manner, but also because they can potentially reveal features, such as electrophysiological biomarkers, that were not previously identified by the human eye using classical diagnostic tools.

In this systematic review, we present studies from 2014 to September 2025 that focused on diagnosis and phenotype of RBD, as well as prediction of phenoconversion to an overt α-synucleopathy using ML approaches. We then focus on the ML algorithms themselves, and a quality assessment of the different methods. Most included cohorts comprised idiopathic/isolated RBD (iRBD). For clarity, we use 'RBD' throughout, but our synthesis and conclusions primarily refer to iRBD. When studies included secondary RBD or RBD in the context of established α-synucleinopathies, this is specified.

2. Methods

2.1. Search strategy and selection criteria

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [32]. Search criteria (see Supplementary Table 1) were applied on PubMed, Embase (Ovid) and Medline (Ovid) databases and candidate studies were identified by two reviewers (KR and SQR). The ML-related keywords in Supplementary Table 1 were deliberately broad, spanning linear models, tree-based methods, kernel methods, neural networks and umbrella terms such as 'supervised learning', 'unsupervised learning' and 'deep learning'. This strategy was chosen to minimise the risk of missing less common model families while keeping the search reproducible, recognising that no keyword list can be perfectly exhaustive. PICOS statement is available in Supplementary Table 2. Using the PICOS framework (Supplementary Table 2), we addressed three questions: (1) Which ML approaches have been used to diagnose or screen iRBD/RBD against controls or other sleep/neurological disorders? (2) How have ML models been applied to classify clinical or biomarker-defined phenotypes within RBD? (3) Which ML models predict phenoconversion from iRBD/RBD to overt α-synucleinopathies, and with what performance? Eligible papers were extracted from 2014 to September 2025 (Tables 1–3). We chose 2014 as the starting year because modern ML applications to RBD and related sleep data (including support vector machines, random forests and deep learning) begin to appear around this time and become progressively more frequent thereafter, whereas earlier work was either absent or limited to classical statistical models. The references of the selected articles were also examined to retrieve documents missed by the literature search. Across the different studies, we counted a total of 38 different sources or databases (Supplementary Table 3).

Each article was first considered by title and abstract. This systematic review included: 1) Original research articles; 2) only papers written in English; 3) observational, descriptive, longitudinal, retrospective, cross-sectional, and cohort studies; 4) studies that used ML techniques, validated and tested their model performance to classify, assess, or detect RBD (including prediction of phenoconversion or classification of RBD phenotypes) and 5) human studies. We also required that the applied ML model was appropriate for the nature of the input data, with an architecture suited to the biomarker modality, a described feature representation and training procedure, and reported performance metrics;

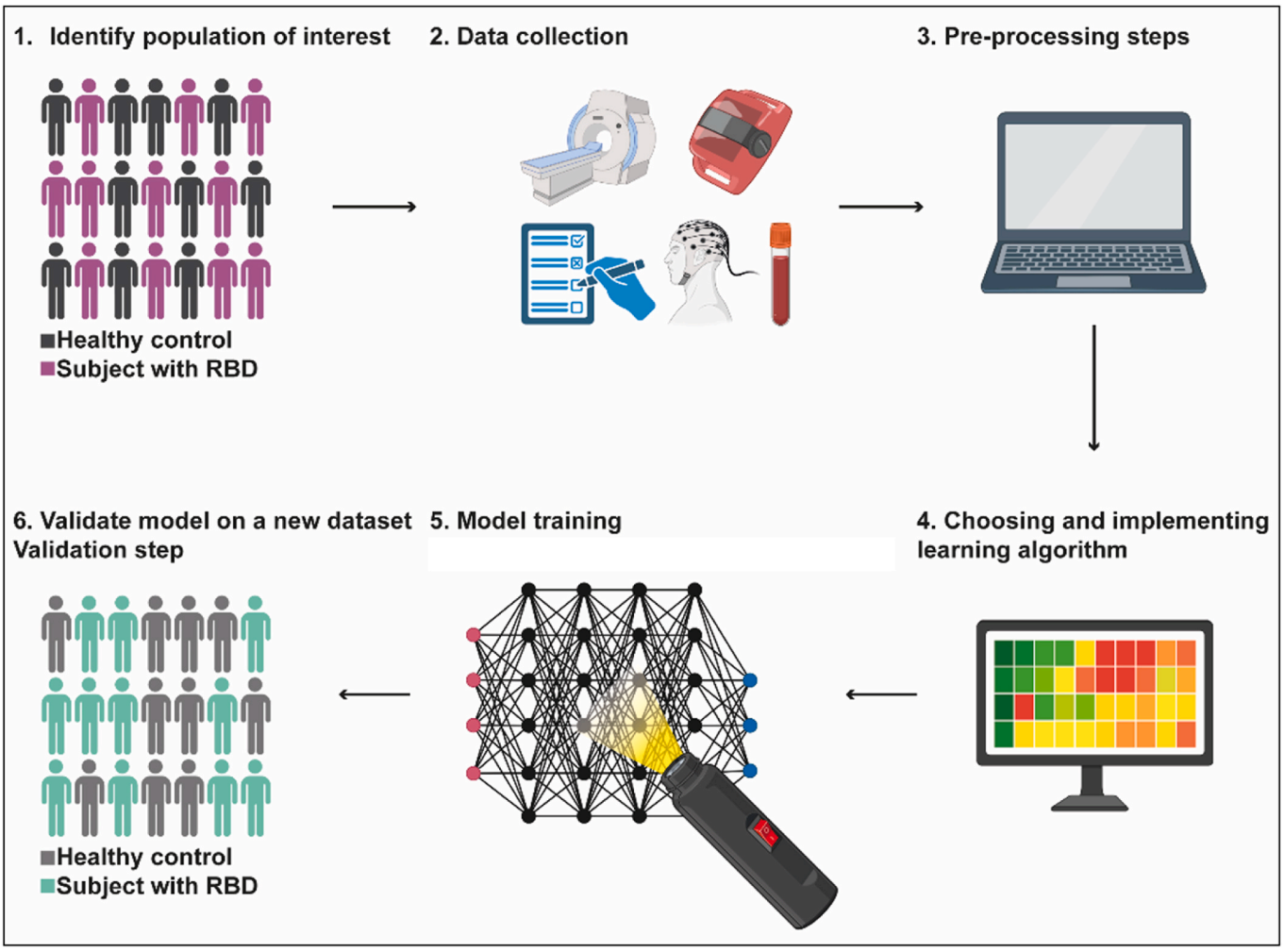

Fig. 1. Schematic diagram of a machine learning framework. In the first step (1), a population of interest is defined and the prediction task is specified (e.g., diagnosis of iRBD vs controls or other sleep disorders, classification of clinical or biomarker-defined RBD phenotypes, or prediction of phenoconversion to overt α-synucleinopathy). (2) Data from participants with and without the clinical condition or outcome of interest (e.g., iRBD vs other sleep disorders, converters vs non-converters) are then collected. (3) Pre-processing steps include data cleaning and curation, specific to each acquisition method. (4) The ML algorithm is selected and implemented, including pre-defined feature engineering and, for classical models, explicit feature selection and hyperparameter optimisation. In deep learning approaches, feature learning is typically embedded within the network architecture rather than performed explicitly. (5) The model is trained on the original dataset and learns decision boundaries or latent representations that best separate the predefined classes. (6) The trained model is then tested on a new population to evaluate generalisability through internal and/or external validation. Created in BioRender. Rosenzweig, I. (2026) https://BioRender.com/3d5z4ry.

studies that simply named an algorithm without such detail were excluded. Exclusion criteria included manuscript characteristics (e.g., animal studies, guidelines), patients' diagnosis (e.g., no use of a diagnostic tool), biomarker measurement (i.e., unclear method of measurement), machine learning feature (e.g., no validation step). A comprehensive list of exclusion criteria is available in Supplementary Table 4. Two reviewers (KR and SQR) independently screened each eligible study, and disagreements were resolved through discussion after retrieving full text.

2.2. Data extraction

For each article, two reviewers (KR and SQR) independently extracted the following data: study name and year, country, type of study, study aim, biomarker used, sample size, age of patients, methods used, main findings and critical evaluation of the study. Then, articles were classified and grouped according to the subtype of biomarkers used (e.g., molecular, physiological, imaging, or clinical assessments). For imaging biomarkers, we only included MRI, PET or DaT-SPECT measures acquired using established sequences and processing protocols that had previously been validated in clinical or methodological studies.

2.3. Study objectives

To maximise generalisability and eventual scalability of identified studies, we categorised the included studies in two ways: (1) according to the nature of the biomarker used in the ML model (e.g., neuroimaging, molecular) and (2) according to each study main and primary objective (although there could be overlaps) which we divided into three main categories: (a) diagnosis of RBD, (b) RBD phenotypes and (c) predicting the phenoconversion of RBD to overt α-synucleinopathies. We use the term 'RBD phenotypes' to refer to clinical or biomarker-based subtypes within RBD, for example, RBD with vs without mild cognitive impairment, distinct MRI atrophy patterns, or differing motor and cognitive profiles, rather than to etiological categories such as idiopathic vs secondary or drug-induced RBD.

2.4. Model evaluation

Classification accuracy was used to compare the performance of ML

Table 1
Performance of ML algorithms for diagnosis.

Author/Study	RBD Database	RBD Sample Size	Modality	Classifier	Validation Approach	Best performing model performance
Polysomnography (PSG) [Including PSG-derived physiological features (i.e., ECG, EOG, EMG)]: ALL or combination (i.e. EMG + EOG)						
1. Rechichi et al. (2022)	1. CAP 2. TURIN Database (TuSDi)	32	PSG	KNN SVM	5-fold CV	Classifying RSWA versus non-RSWA (K-NN): Accuracy: 87% Sensitivity: 93% Precision: 87.5%
2.Brink-Kjaer et al. (2022)	1. DCSM 2. STNF	82	PSG	Ensemble model using LR	5-fold CV	Accuracy: 89.0% Sensitivity: 91.4% Specificity: 86.3%
3.Zhuang et al. (2022	1. CAP	22	PSG	CNN RF SVM	5-fold CV	Model DL-R: Sensitivity:94.8%. Specificity: 98.4%.
4.Cooray et al. (2019)	1. CAP 2. JR	53	PSG: ECG + EOG	RF	10-fold CV	Features that incorporate sleep architecture: Accuracy: 96% Sensitivity: 98% Specificity: 94% *Critical biomarker: EMG - Atonia Index (REM)
5.Andreaotti et al. (2018)	1. CAP 2. JR	43	PSG	CNN	Leave-One-Out CV 5-fold CV	Healthy: Cohen kappa: 0.75 Disease (different sleep disorders): Cohen kappa: 0.64 *Critical biomarker: EEG + EOG
6.Christensen et al. (2014)	1. DCSM	31	PSG: EEG + EOG	Lasso-regularized regression model	8-fold CV	Sensitivity: 91.4% Specificity: 68.8%
7.Sharma et al. (2022)	1. CAP	19	PSG: EMG + EOG	EBTC SVM Boosted Trees	10-fold CV	EBTC: EMG-based features Accuracy: 96.6% EOG-based features Accuracy: 96% Both: 98.3%
8.Cooray et al. (2021)	1. CAP 2. JR	50	PSG	RF	10-fold CV	EMG + ECG: Accuracy: 93% Sensitivity: 94% Specificity: 92%
9.Wallis et al. (2020)	1. Private (Veteran Affairs Hospital, Portland)	692	PSG	FFNN LSTM ID CNN	-	1D CNN (MC + res): Balanced Accuracy: 91%
10. Gunter et al. (2023)	1. DCSM 2. STFD	96	PSG	Simple CNN Dilated CNN SViT	-	(EEG + EOG) SViT: F1 score: 93%
11. Brink-Kjaer et al. (2021)	1. DeNoPa	114	PSG	Multiple Regression	N/A (multiple linear regression; no classifier validation)	N/A (no diagnostic performance metric reported)
12. Markov et al. (2025)	1. CAP	18	PPG	Random Forest	20-fold CV	ADASYN-balanced dataset: Accuracy: 77.8% (RBD)
13. Feuerstein et al. (2024)	1. Private (Medical University of Innsbruck)	86	PSG	ResNet Deep Learning	5-fold CV	The best performing model per subject basis reached a mean accuracy of 0.784 and a mean macro F1 score of 0.783.
Polysomnography (PSG): ECG only						
14.Sharma et al. (2023)	1. CAP	22	PSG: ECG	KNN EbagT SVM EboostT	10-fold CV	EbagT: Accuracy for RBD: 99.65%
15.Urtnasan et al. (2021)	1. CAP	7	PSG: ECG	CNN	-	F1 score for RBD only: 91% Precision for RBD only: 91% Recall for RBD only: 100%
16.Widasari et al. (2020)	1. CAP	22	PSG: ECG	EBT SVM Decision Tree	5-fold CV	EBT: Sensitivity: 90.91% Specificity: 82.76% Accuracy: 86.27% Cohen's Kappa: 0.73
17.Salsone et al. (2022)	1. Private (Not specified)	20	PSG: ECG	RF XGBoost LR	Leave-One-Out CV	RF model: Accuracy: 94% Sensitivity: 95% Specificity: 92%
Polysomnography (PSG): EMG only						
18.Cesari et al. (2019)	1. DCSM	29	PSG: EMG	SVM with linear kernel SVM with radial basis function RF	5-fold CV	Both REM and NREM features used (3 classification mixture): Overall validation accuracy: 70.76% Sensitivity: 69.43% Specificity: 85.64%

(*continued on next page*)

Table 1 (*continued*)

Author/Study	RBD Database	RBD Sample Size	Modality	Classifier	Validation Approach	Best performing model performance
Polysomnography (PSG) [Including PSG-derived physiological features (i.e., ECG, EOG, EMG)]: ALL or combination (i.e. EMG + EOG)						
19.Kempfer et al. (2014)	1. DCSM	16	PSG: EMG	One-Class SVM	Leave-One-Out CV	Average AUC of 0.993
20. Cooray et al. (2018)	1. CAP 2. JR	44	PSG: EMG	RF	Leave-One-Out CV	Atonia index ratios + Sleep Architecture + EMG metrics: Accuracy: 92% Sensitivity: 93% Specificity: 91%,
21. Choi et al. (2026)	1. Private (5 tertiary hospitals, South Korea))	227	PSG: EMG	Deep Learning (EEGNet)	Stratified 5-fold cross-validation	Manual sleep staging: ROC AUC 0.812 (95% CI 0.744–0.871), PR AUC 0.866 (95% CI 0.796–0.909), sensitivity 87.3%, specificity 56.5%, F1 score 0.824, accuracy 73.6%
Polysomnography (PSG): EEG only (including EEG sleep events such as CAP, Sleep Spindles)						
22.Buettner et al. (2020)	1. CAP	22	PSG: EEG	RF	10-fold CV	Accuracy: 90.12%.
23.Kazemi et al. (2022)	1. CAP	20	PSG: EEG	SCDS KNN RF	Leave-One-Out CV	Fpz-Cz & Pz-Oz using RF: Accuracy: 83.3%
24.Srivastava (2022)	1. CAP	19	PSG: EEG	CNN	-	Accuracy: 97.92% Accuracy increased dramatically with increased data generated from FFT and interpolation (62.63% to 97.92%).
25.Rechichi et al. (2022)	1. CAP 2. TURIN Database	32	PSG	SVM K-NN NB DT	5-fold CV	SWS features using SVM: Accuracy: 86% Sensitivity: 91% Specificity: 83% Precision: 77% F1: 83% AUC: 0.94
26.Dimitriadis et al. (2021)	1. CAP	22	PSG: EEG	RF	10-fold CV	P4-O2 EEG: Accuracy: 74% Macro-average precision: 83.7% Macro-average recall: 59.1% Macro-average F-score: 64.9%
27.Wadichar et al. (2023)	1. CAP	22	PSG: CAP	CNN LSTM	Hold-out validation and 10-fold CV	LSTM-CNN: Healthy-Unhealthy: Accuracy of 91.45% Disease classification: Accuracy of 90.55%
28.Murarka et al. (2022)	1. CAP	22	PSG: CAP	CNN	5-fold CV	Accuracy: 79.48% Sensitivity: 72.30% Specificity: 86.84% Precision: 84.92% F1: 78.10%
29.Sharma et al. (2021)	1. CAP	22	PSG: CAP	SVM K-NN EBagT EBoosT	10-fold CV	Combining both C4-A1 and F4-C4 channels using EBagT: F1: 0.71 Kappa: 0.44 AUC: 0.80
30.Christensen et al. (2014)	1. DCSM	15	PSG: Sleep Spindles	MP SVM	Leave-one-subject-out CV	Sleep spindle detector: Sensitivity: 84.7% Specificity: 84.5%
31. Sharma et al. (2021)	1. CAP	22	EEG	EBagT SVM KNN EBooT	10-fold CV	Using both C4-A1 & F4-C4 using Eboot: Accuracy: 98.98% Kappa: 0.97 AUC: 1.00
32. Sharma et al. (2021)	1. CAP	22	EEG	Decision Trees LR Naïve Bayes SVM KNN EBT Classifier and Discriminant	10-fold CV	EBT classifier: Overall classification accuracy: 71.9% Kappa: 0.6372 ± 0.0039. After balancing the dataset, the classification accuracy increased from 71.9% to 81.0%.
Neuroimaging: Electroencephalogram (EEG) (as a standalone)						
33. Kim et al. (2023	1. Private: SNU	16	EEG: ERP	CNN	10-fold CV	Training set Accuracy: 100% Validation set Accuracy: 90.86%
34. Kim et al. (2023)	1. Private: SNU	49	EEG: ERP	CNN	-	3dCNN: Mean test accuracy: 99.81% Precision: 99.77% Recall: 99.85% AUC: 99.49%
Neuroimaging: Magnetic Resonance Imaging (MRI)						
35. Lee et al. (2022)	1. Private (Haeundae Paik Hospital)	20	DTI	SVM	-	Conventional DTI measures: Accuracy: 87.5%

(*continued on next page*)

Table 1 (*continued*)

Polysomnography (PSG) [Including PSG-derived physiological features (i.e., ECG, EOG, EMG)]: ALL or combination (i.e. EMG + EOG)						
Author/Study	RBD Database	RBD Sample Size	Modality	Classifier	Validation Approach	Best performing model performance
						Sensitivity: 80.0% Specificity: 100% *Critical biomarker: Conventional DTI measures
36. Matsushima et al. (2024)	1. J-PPMI cohort	55	fMRI: FC	LR SVM	10-fold CV	Logistic regression: Accuracy: 64.9% Precision: 65.9% F1 score: 64.5%
37. Zhi et al. (2024)	1. Private (Second Affiliated Hospital of Soochow University)	15	MRI	AutoGluon: Ensemble approach that combines outputs of multiple models	6-fold CV	AUC: 0.8533
38. Yang et al. (2021)	1. Private: (Movement Disorders Clinic of the Xuanwu Hospital of Capital Medical University)	35	MRI	LR RF	Two-level nested CV	MDMR with Random Forest: Accuracy: 93.3%
39. Gaurav et al. (2022)	1. ICEBERG study	47	MRI	CNN	-	AUC: Volume: 0.78 Correct Volume: 0.79 Signal-to-Noise Ratio: 0.56 Contrast-to-noise Ratio: 0.63
Molecular: Serum, mRNA, gut microbiome						
40. Li et al. (2024)	1. Private (Ruijin Hospital, Shanghai)	56	Molecular	SVM	5-fold CV	Accuracy: 91.49% Sensitivity: 96% Specificity: 86.36%
41. Soto et al. (222)	1. Private (Hospital Clinic de Barcelona)	51	Molecular	GBM	Leave-one-out CV	RBD detection: Accuracy: 94% Sensitivity: 98% Specificity: 86% Discriminative ability of miRNAs: AUC = 0.98
42. Huang et al. (2023)	1. Private (Li Chiu Kong Family Sleep Assessment Unit, the Chinese University of Hong Kong)	170	Molecular	RF	10-fold CV	RBD vs Controls Mean Accuracy: 0.68 AUC: 0.75 RBD vs RBD-First Degree relatives Mean accuracy: 0.67 AUC: 0.72
43. Hallqvist et al. (2024)	1. DeNoPa 2. Kassel data of Healthy Brain Ageing	72	Molecular	Linear SVM OPLS-DA	5-fold CV	PD vs HC classification: Precision: 0.87 ± 0.09, Recall: 0.87 ± 0.08, F1 score: 0.86 ± 0.09 Balanced accuracy score: 0.82 ± 0.12,
Behavioural: Non-Motor & Motor Symptoms						
44. Lo et al. (2021)	1. Oxford Discovery Cohort study 2. Tracking cohort study 3.Department of Neurology, University of Marburg, Germany	278	Behavioural	RF	Leave-One-Subject-Out CV	Using olfactory function: Sensitivity: 65% Specificity: 100% Positive predictive value: 100% Negative predictive value: 30%
45. Arora et al. (2021)	1. Oxford Discovery Cohort	112	Behavioural	RF	Leave-One-Subject-Out CV	Sensitivity: 60.7% Specificity: 69.6%
46. Jeancolas et al. (2022)	1. ICEBERG cohort	41	Behavioural	SVM	10-fold CV	Balanced accuracy: 63% (Increased up to 70% when the analysis was restricted to the 22 iRBD participants with mild motor symptoms)
47. Arora et al. (2018)	1. Oxford Parkinson's Disease Centre (OPDC) Discovery Study	104	Behavioural	RF	10-fold CV Leave-One-Subject-Out CV Leave-one-recording-out	Controls vs. iRBD: Sensitivity: 91.9% Specificity: 90.0% iRBD vs. PD: Sensitivity: 87.5% Specificity: 90.1%
48. Brink-Kjaer et al. (2023)	1. Private (Stanford Sleep center)	42	Behavioural	SVM LR	Leave-one-out CV	Actigraphy classifier: Sensitivity: 95.2% Precision: 90.9% Questionnaire classifier: Accuracy: 90.6%

(*continued on next page*)

Table 1 (*continued*)

Polysomnography (PSG) [Including PSG-derived physiological features (i.e., ECG, EOG, EMG)]: ALL or combination (i.e. EMG + EOG)						
Author/Study	RBD Database	RBD Sample Size	Modality	Classifier	Validation Approach	Best performing model performance
						Precision: 92.7% Concordant predictions between actigraphy and questionnaire: Specificity: 100% Precision: 100% Sensitivity: 88.1%
49. Cochen et al. (2022)	1. Private (Regional University of Hospital of Nimes, France)	21	Behavioural	Lasso sparse LR	-	Accuracy: 95% Sensitivity: 100% Specificity: 90%
50. Cesari et al. (2023)	1. Private (Medical University of Innsbruck)	53	Behavioural	LR	10-fold CV	Using short movements in all the four ROIs: Accuracy: 86.6% F1 score: 78.3%
51. Rusz et al. (2021)	1. Private (General University Hospital in Prague, Czech Republic)	16	Behavioural	SVM	20-fold CV	Combination of four distinctive speech dimensions, including aperiodicity, irregular alternating motion rates, articulatory decay, dysfluency: Sensitivity: 96% Specificity: 79%
52. Rusz et al. (2021)	1. Private (7 Multicenters)	150	Behavioural	Linear Regression	Cross-validation (leave-1-subject-out cross-validation Follow-up analysis at 12 months	AUC: 0.65
53. Rusz et al. (2018)	1. Private (General University Hospital, Czech Republic)	50	Behavioural	LR	Leave-one-out CV	AUC for differentiating groups: PD vs. Controls: 0.85 (Sensitivity: 75.0%, Specificity: 78.6%) PD vs. RBD: 0.78 (Sensitivity: 66.7%, Specificity: 71.0%) RBD vs. Controls: 0.69 (Sensitivity: 69.8%, Specificity: 64.7%)
54. Benba et al. (2019)	1. Private (General University Hospital, Czech Republic)	50	Behavioural	SVM	Leave-one-out CV	MLP kernel of SVM classifier: Accuracy: 85%
55. Barber et al. (2017)	1. Private: Three centres: (1) JR Hospital, (2) Papworth Hospital, (3) Sheffield Teaching Hospital 2. Oxford Discovery Cohort study	171	Behavioural	LR	-	74% of RBD participants met the MDS criteria for probable prodromal Parkinson's disease (PD) The odds ratio (OR) of hyposmia for high-risk RBD patients compared to controls was 45.5 (95% CI 21.1–98.0)
56. Simonet et al. (2024)	1. Private (Sleep Clinic at Guy's Hospital)	33	Behavioural	LR	Leave-one-out CV	BRAIN test: 72.7% sensitivity, 62.1% specificity DFT test: 81.8% sensitivity, 69.0% specificity Combination of motor markers: 90.3% sensitivity, 89.3% specificity (AUC: 0.94)
57. Guarin et al. (2025)	1.Private	16	Behavioural	LR SVM Naïve Bayes Decision Tree RF	-	RF model provided the best performance in distinguishing iRBD and healthy controls, with a balanced accuracy of 79.8 ± 0.2%, sensitivity of 84.0 ± 0.0% and specificity of 76.2 ± 0.4%.
58. Abdelfattah et al. (2025)	1.Stanford Sleep Center	81	Behavioural (2D camera from vPSG recordings)	Non-linear MLP	10-fold CV	Accuracies for detecting iRBD ranged from 84.9% (with 2 features) to 87.2% (with 5 features). Combining all 5 features but only analysing short movements achieved the highest accuracy of 91.9%.

Also refer to Fig. 3.

models (Supplementary Table 5), although other performance metrics such as sensitivity, F1 scores and specificity are also reported, especially when accuracy is not available. For each biomarker category, we summarized the best performing features of each category, together with their model performance metrics.

2.5. Quality assessment and risk-of-bias analysis

Two reviewers (KR and SQR) independently evaluated the quality of studies that were included using the Quality Assessment Tool for Quantitative Studies, developed by the Effective Public Health Practice Project (EPHPP; www.ephpp.ca/tool.html) for observational, descriptive, longitudinal, cross-sectional, or cohort studies from original research articles [33]. Any disagreements were resolved by discussion or by consulting with a senior reviewer. For the EPHPP scale, each study was rated on a scale of strong, moderate, weak according to the following criteria: selection bias, study design, blinding, data collection methods, confounders, and withdrawals/attrition (if any).

Table 2
Performance of ML algorithms for phenotype.

Author/Study	RBD Database	RBD Sample Size	Modality Used	Phenotype/subtype classified	Classifier Algorithm	Validation Approach	Model Performance
Polysomnography (PSG): ECG only							
1. Salsone et al. (2024)	1. Private (Not specified)	42	PSG: ECG	iRBD with periodic limb movements vs iRBD without periodic limb movements	SVM RF XGBoost	Leave-One-Out CV	RF: Accuracy: 86% Sensitivity: 96% Specificity: 74%
Polysomnography (PSG): EMG only							
2. Cesari et al. (2019) (1)	1. DeNoPa	14	PSG: EMG	PD(+) vs PD(−) RBD cohort	Linear Regression	5-fold CV	German RBD cohort, including PD(+) and PD (−): Accuracy: 84.17% Sensitivity: 59.02% Specificity: 92.73%
Neuroimaging: Positron Emission Tomography (PET)							
3. Ryoo et al. (2023)	1. Private (SNU)	50	PET	RBD with mild cognitive impairment vs RBD without mild cognitive impairment	CNN	-	RBD-MCI versus RBD-nonMCI: AUC: 0.70
Neuroimaging: Magnetic Resonance Imaging (MRI)							
4. Joza et al. (2025)	1. PPMI	451	MRI	Cortical-first vs subcortical-first atrophy subtype	SuStain (unsupervised ML algorithm)	10-fold CV	SuStain classifier identified two brain atrophy subtypes: 1. Cortical-first (CVS = 0.90, f = 0.57) 2. Subcortical-first (CVS = 0.91, f = 0.43)

Phenotypes refer to within-RBD clinical or biomarker-defined subgroups rather than to aetiological categories such as idiopathic versus secondary RBD. Also refer to Fig. 3.

Subsequently, these ratings were compiled to form a global rating: studies were rated as strong if they had no weak ratings, moderate if they had one weak rating, and weak if they received two or more weak ratings. We chose the EPHPP Quality Assessment Tool because it is widely used for observational and cohort designs in health research; our aim was to evaluate the robustness of the underlying clinical study designs and datasets rather than the sophistication of the algorithms alone [33]. Also, although many included studies used retrospective data, blinding remains relevant because manual PSG scoring, feature selection, or clinical labelling can be influenced by knowledge of diagnosis or outcome. We therefore retained blinding as a quality domain, following the EPHPP framework.

Risk of bias of selected studies was also analysed. The aspects considered in the analysis of risk bias were based on the Cochrane Guidelines for Systematic Reviews [34], but the exact criteria were adapted for the purposes of the systematic review (e.g., methodology and goals of studies, creating/validating a classification model, see Supplementary Table 6). In the 'database' domain, we focused on internal methodological consistency (e.g. standardised acquisition and annotation) rather than generalisability. Open or widely used datasets such as CAP were therefore rated as lower risk in this domain because of their standardised protocols, even though their iRBD sample size is limited. Single-centre non-public cohorts such as DCSM were rated as medium risk; for example, Christensen et al. (2014) received Database = 1 on this basis [35]. We explicitly acknowledge that external validity and generalisability depend on sample size and diversity, which we discuss separately.

For the 'mathematical development/architecture' domain, we did not require full mathematical derivations of standard algorithms. A study was rated as 'low' risk when the model architecture, input features and training procedure were described in enough detail to allow reproduction or were clearly referenced to prior methodological work. Studies that only named the algorithm without describing inputs, architecture, or training were rated as 'high' risk.

3. Results

3.1. Studies characteristics

Between 2014 and 2025, 335 records were identified; 202 remained after duplicate removal, and 75 studies were retained after screening and full-text review (Fig. 2). Fifty-one studies used neuroimaging and physiological features as inputs, 5 used molecular biomarkers, and 19 used behavioural markers (Tables 1–3; Supplementary Tables 5, 7 and 8).

More specifically, the included studies spanned DaT-SPECT, MRI-derived measures, PET, EEG, PSG-derived EMG/EOG/ECG features, molecular biomarkers and behavioural assessments (Tables 1–3). On top of this, in PSG studies, we identified 14 studies using either all or a combination of PSG measures (i.e., EMG + EOG) [35–48], 5 studies using only EMG measures [40,49–51], 6 studies using electrocardiography (ECG) measures only [31,52–55] and 11 studies using EEG only [26,30,56–64]. Altogether, 19 behavioural studies were identified using a range of non-motor and motor symptoms assessments such as cognition, gait parameters, and speech dysfunction [65–82].

In terms of ML models used, Support Vector Machine (SVM) was the most popular in RBD studies with 32% of studies using it, followed by Random Forest (RF) with a percentage of 27% and Logistic Regression with a percentage of 21%.

Across modalities and study designs, higher reported performance most often coincided with (i) smaller, single-centre datasets, (ii) extensive feature engineering or deep learning applied to high-dimensional signals, and (iii) internal validation only. Lower or more variable performance was typically seen when models were tested on heterogeneous cohorts, when endpoints were less well defined (particularly for phenoconversion), or when external validation was attempted. This pattern suggests that differences in cohort size, case definition, and validation strategy may explain more of the performance variability than the choice of algorithm alone.

3.2. RBD diagnosis

Altogether, 58 studies focused on RBD diagnosis. Reported

Table 3
| Performance of ML algorithms for phenoconversion.

Author/Study	RBD Database	RBD Sample Size	Modality Used	Classifier	Validation Approach	Follow-up duration/ prediction horizon	Best performing model performance
Neuroimaging: Electroencephalogram (EEG) (as a standalone)							
1. Ruffini et al. (2019)	1. Private: Hospital du Sacre-Coeur de Montreal	121	EEG: Resting State	DCNN RNN	Leave-pair-out CV	Cross-sectional comparison; no prospective follow-up reported	DCNN: Healthy Control (HC) vs PD: AUC = 87% HC + RBD vs PD + DLB: AUC = 78%
2. Jeong et al. (2024)	1. Private: SNU	236	EEG: Resting State	XGBoost RF LR KNN	5-fold CV	Internal mean 3.5 y (0.9-8.6); external mean 2.17 ± 1.53 y	Random survival forest model was the best model for phenoconversion time prediction: Integrated Brier score: 0.114 Concordance index: 0.775. KNN model was the best model for phenoconversion subtype prediction: AUC: 0.901
Neuroimaging: Polysomnography (PSG)							
3. Cesari et al. (2024)	1. Private: Medical University of Innsbruck	66	PSG: EEG & EMG	Random Survival Forest	4-fold CV	Converted patients phenoconverted within 2.7 ± 1.0 y	Best test performance was obtained when considering EEG features in REM sleep only (Harrell's C-index: 0.723 ± 0.113; Uno's C-index: 0.741 ± 0.11; integrated Brier score: 0.174 ± 0.06).
Neuroimaging: Positron Emission Tomography (PET)							
4. van Veen et al. (2022)	1. Private (University Medical Centre of Groningen, Netherlands, and Philipps-Universität Marburg, Germany)	20	PET	SSM/PCA GMLVQ	10-fold CV	Annualised longitudinal change; exact duration NR	Distance travelled by iRBD subjects through GMLVQ space per year (i.e., velocity) was correlated with the change in motor symptoms per year (Spearman's rho = 0.62, p = 0.004).
5. Feng et al. (2023)	1. Private (Huashan Hospital, Shanghai, China)	33	PET	SVM	Leave-one-out CV	0-2, 2-4, 4-6 and >6 y prediction windows	Predicting conversion in 0-2 years: AUC: 0.879 Sensitivity: 90% Specificity: 88.3% Predicting conversion in 2-4 years: AUC: 0.807 Sensitivity: 72.7% Specificity: 83.3% Predicting conversion in 4-6 years: AUC: 0.940 Sensitivity: 100% Specificity: 84.6% Predicting conversion over 6 years: AUC: 0.879 Sensitivity: 100 Specificity: 80.7%
6. Mattioli et al. (2023)	1. Private: Two Italian Centres (Genoa and Rome Tor Vergata)	67	PET	LR	Leave-One-Subject-Out CV	Converters: 21 ± 14 mo; non-converters: 33 ± 19 mo	AUC: 0.85 Sensitivity: 87% Specificity: 72%
Neuroimaging: Presynaptic Dopaminergic Imaging							
7. Arnaldi et al. (2024)	1. International Multicenter Study	405	DaT	Decision Trees SVM KNN	10-fold CV	NR	RBD phenoconversion: DaT-SPECT in combination with clinical data: Sensitivity: 77% Specificity: 85% RBD detection: Decision tree: Sensitivity = 0.89, Specificity = 0.84, Error rate = 0.14 Support vector machine: Sensitivity = 0.86, Specificity = 0.89, Error rate = 0.12 KNN: Sensitivity = 0.88, Specificity = 0.88, Error rate = 0.12
8. Arnaldi et al. (2021)	1. International Multicenter Study	344	DaT	LR	Bootstrap (500x)	Mean 2 y	RBD phenoconversion: Generalised logistic regression was highly significant in predicting conversion (P < 0.000001) using DAT-SPECT data RBD detection: AUC varied from 0.51 to 0.62
Molecular: Serum, mRNA, gut microbiome							
9. Laguna et al. (2021)	1. Private: Neurology Service from the Hospital Clinic of Barcelona, Spain	33	Molecular	Elastic Net	10-fold CV	NR	To distinguish between controls and pre-DLB patients including area glycB: Corrected AUC: 0.765 Sensitivity: 53.33% Specificity: 96.55% To distinguish between pre-DLB and

(*continued on next page*)

Table 3 (*continued*)

Author/Study	RBD Database	RBD Sample Size	Modality Used	Classifier	Validation Approach	Follow-up duration/ prediction horizon	Best performing model performance
							pre-PD patients including small HDL-P and HDL-Z: Corrected AUC: 0.759 Sensitivity: 72.22% Specificity: 86.67%
Behavioural: Non-Motor & Motor Symptoms							
10. Fereshtehnejad et al. (2019)	1. Private (Hôpital du Sacre-Coeur de Montreal, Canada)	154	Behavioural	Linear Regression	-	Up to 6 y before diagnosis	Olfaction was most consistent with its lack of progression over the prodromal period. Overall accuracy dropped only modestly between Years 0 and -6 (0.889 to 0.872). Sensitivity remained at 67% even 6 years before diagnosis Specificity: 95.4%
11. Zhang et al. (2023)	1. Private (Tianjin Medical General Hospital)	45	Behavioural	LR	-	NR	Optimal combined features in GLM: Differentiating iRBD converters and non-converters: Sensitivity = 95.0% Specificity = 75.0% Differentiating iRBD converters with motor phenotype and non-converters: Sensitivity = 100% Specificity = 66.7% Differentiating iRBD converters with cognitive phenotype and non-converters: Sensitivity = 83.3% Specificity = 91.7%
12. Shin et al. (2025)	1. Private	178	Behavioural	Extreme Gradient Boosting	10-fold CV	Median follow-up 3.6 y	For phenoconversion time prediction, the extreme-gradient-boosting survival-embeddings Kaplan-neighbours model showed the best performance (concordance index: 0.823; integrated Brier score: 0.123). For subtype classification, the random-forest classifier achieved the highest performance (Matthews correlation coefficient: 0.697).
13. Fereshtehnejad et al. (2025)	1. Montreal RBD cohort	95	Behavioural	Random Survival Forest	CV	NR	The RSF model achieved a mean training accuracy of 0.7002.

Follow-up duration or prediction horizon is shown where it could be extracted reliably from the cited study; NR indicates that it was not reported explicitly or could not be recovered with confidence. Also refer to Fig. 3. For the EEGNet REM-EMG diagnosis study previously cited as Jung et al. (2025), we replaced the conference report with the subsequent full peer-reviewed publication (Choi et al., 2026), which provides the complete multicentre cohort, validation design and performance metrics. This replacement does not change the number of included studies or the overall conclusions.

classification accuracy ranges from 63% to 99.7% (Fig. 3A). Several ECG-only studies reported very high accuracies (up to 99.7%) [44], but these findings come largely from small datasets and predominantly internal cross-validation, which increases the risk of optimistic performance estimates and reduced generalisability. ECG may nonetheless perform well in some cohorts because REM-related autonomic and cardiopulmonary alterations are captured in the signal, but apparent gains may also reflect simpler case-control contrasts, cohort selection and overfitting in small datasets. By contrast, multichannel PSG-based approaches are more consistently validated against the diagnostic gold standard, and when sample size and validation strategy are considered, there is no clear evidence that ECG outperforms PSG overall. Methodological concerns that can challenge generalisability include mixing idiopathic and secondary RBD, male-dominant samples, age mismatch between groups, and variability in acquisition and processing across laboratories and sleep centres (see Methodological Quality Assessment Results). There was no systematic trend for any modality to outperform others in terms of accuracy ($p = 0.74$, Kruskal-Wallis test, Fig. 3A).

Actigraphy is a classical tool in sleep medicine. Using an ensemble of decision trees, home actigraphy detected iRBD with 95.2% sensitivity and 90.5% specificity [70]. Furthermore, using automated 3D video-based motor quantification during REM sleep across four regions of interest (head, hands, upper body, lower body), logistic regression achieved an accuracy of 86.6% in distinguishing iRBD from non-RBD participants [72]. Hence, the lack of information provided by EEG and EOG, which, as previously said, are formally necessary for scoring REM sleep, does not, however, make actigraphy futile.

Particular emphasis must be given to olfactory loss, which is widely regarded as an early symptom of PD, particularly in the context of iRBD [65,83]. Indeed, in iRBD, olfactory deficits were found in 72% of patients, compared with only 15% of controls [65]. Hence, assessing olfaction in patients with diagnosed iRBD could provide a unique window of opportunity to identify patients at risk of PD.

3.3. Classification of RBD phenotypes

Altogether, four studies focused on within-RBD phenotypic subtypes (Table 2). Reported classification accuracy ranges from 84.2% to 86% (Fig. 3B). In these studies, ‘phenotype’ refers to clinically or biomarker-defined subgroups within iRBD/RBD (e.g., cognitive status, imaging-based atrophy patterns), and therefore depends on the assessment modality rather than on aetiological categories such as idiopathic versus secondary RBD. Identifying such subtypes is clinically relevant because certain profiles, including olfactory deficit and mild cognitive

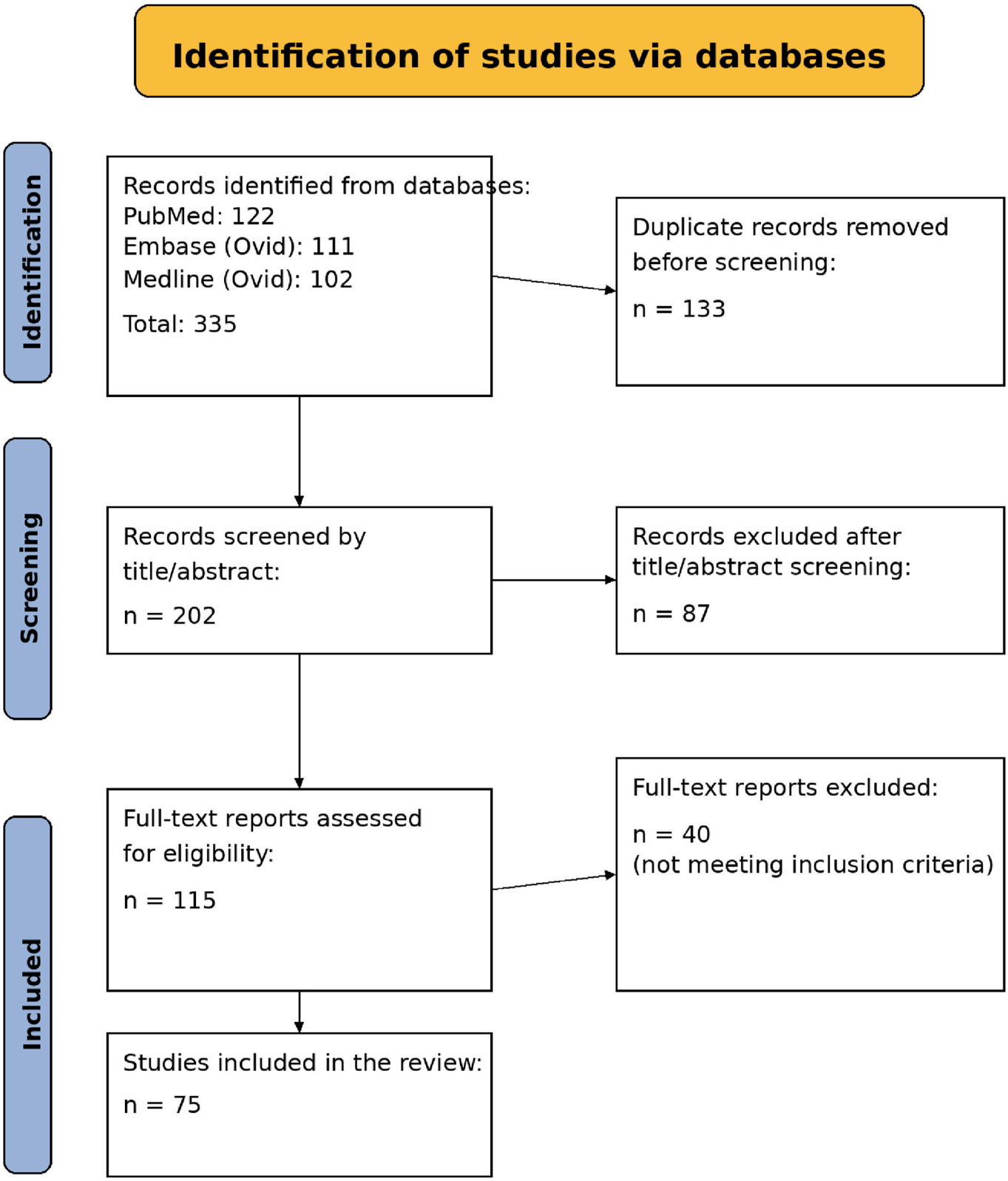

Fig. 2. PRISMA flowchart for study selection
A total of 335 records were identified in database searches. After removal of 133 duplicates, 202 records were screened by title and abstract. Of 115 full-text reports assessed for eligibility, 40 were excluded because they were outside the scope of the current review or did not meet the inclusion criteria, leaving 75 studies in the final analysis.

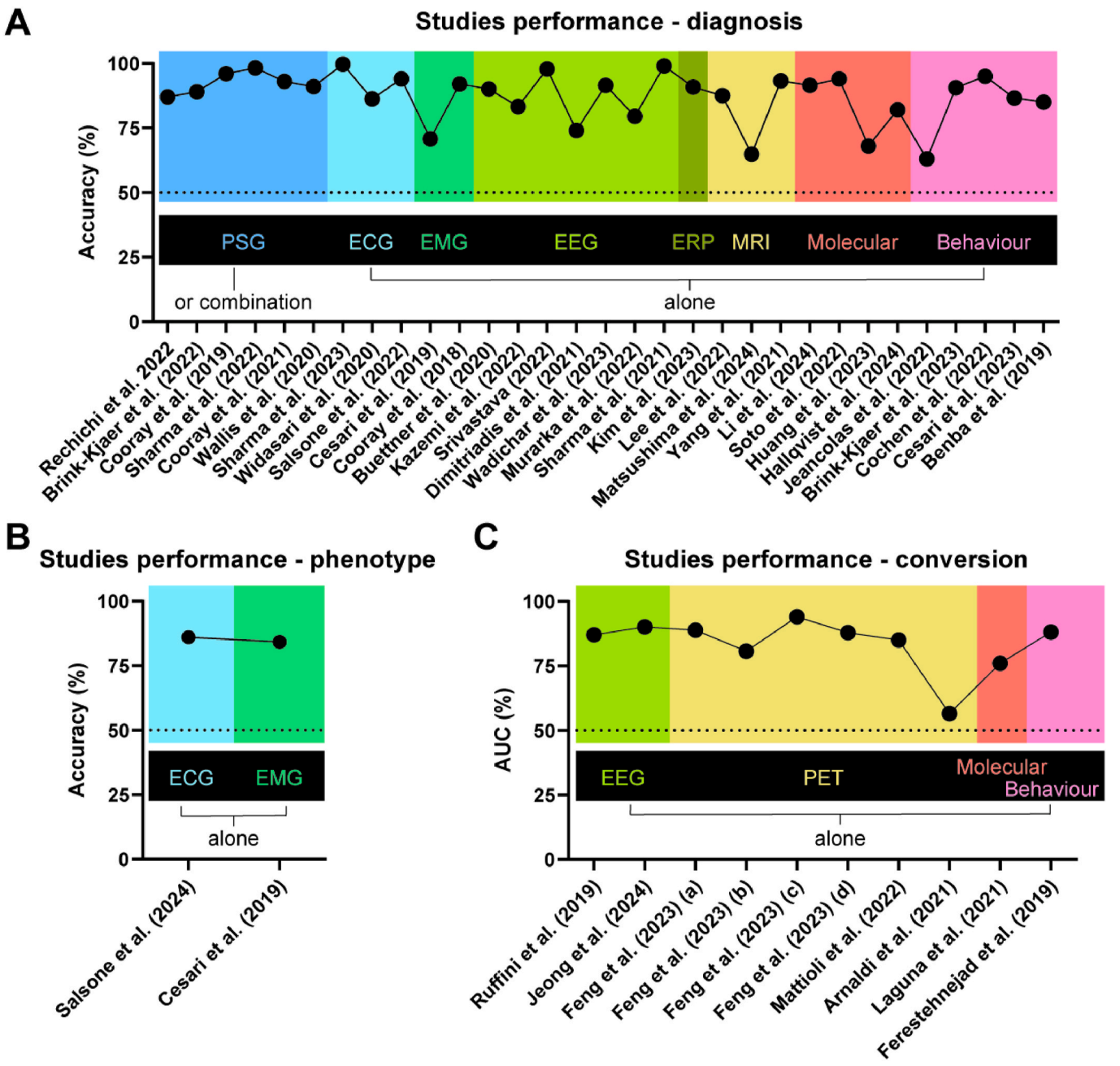

Fig. 3. Performance of identified studies
(A) Accuracy of the ML algorithm across diagnostic studies that reported it. Several ECG-only studies reported very high values, but these generally came from small datasets with internal validation only and should be interpreted cautiously. (B) Accuracy across phenotype-classification studies that reported accuracy; see Table 2 for phenotype definitions and for studies reporting alternative metrics such as AUC or subtype-separation measures. (C) AUC across phenoconversion studies that reported it; see Table 3 for follow-up duration or prediction horizon where available.

impairment, have been associated with a higher risk of future phenoconversion to overt α-synucleinopathy [84].

3.4. Predicting phenoconversion of RBD

Altogether, 13 studies evaluated prediction of phenoconversion to overt α-synucleinopathy (Table 3). Across studies that reported AUC, values ranged from 0.56 to 0.94 (Fig. 3C), with prediction horizons and follow-up duration varying substantially across cohorts. Models using dopaminergic imaging, particularly DaT-SPECT and dopaminergic PET, generally showed stronger prognostic performance than single-modality clinical measures, consistent with early nigrostriatal involvement in prodromal synucleinopathy [85,86].

3.5. Machine learning methods

Across the 75 included studies, we identified 32 distinct algorithm families (see Fig. 4 and Supplementary Table 5). In total, 128 model uses were reported across studies (mean 1.7 algorithm types per study), reflecting the fact that individual papers often evaluated more than one algorithm. Overall, Support Vector Machine (SVM) was the most commonly used algorithm family (24/75 studies, 32%), followed by random forests (20/75, 26.7%) and logistic regression (16/75, 21.3%). When stratified by study objective, SVM accounted for the largest share

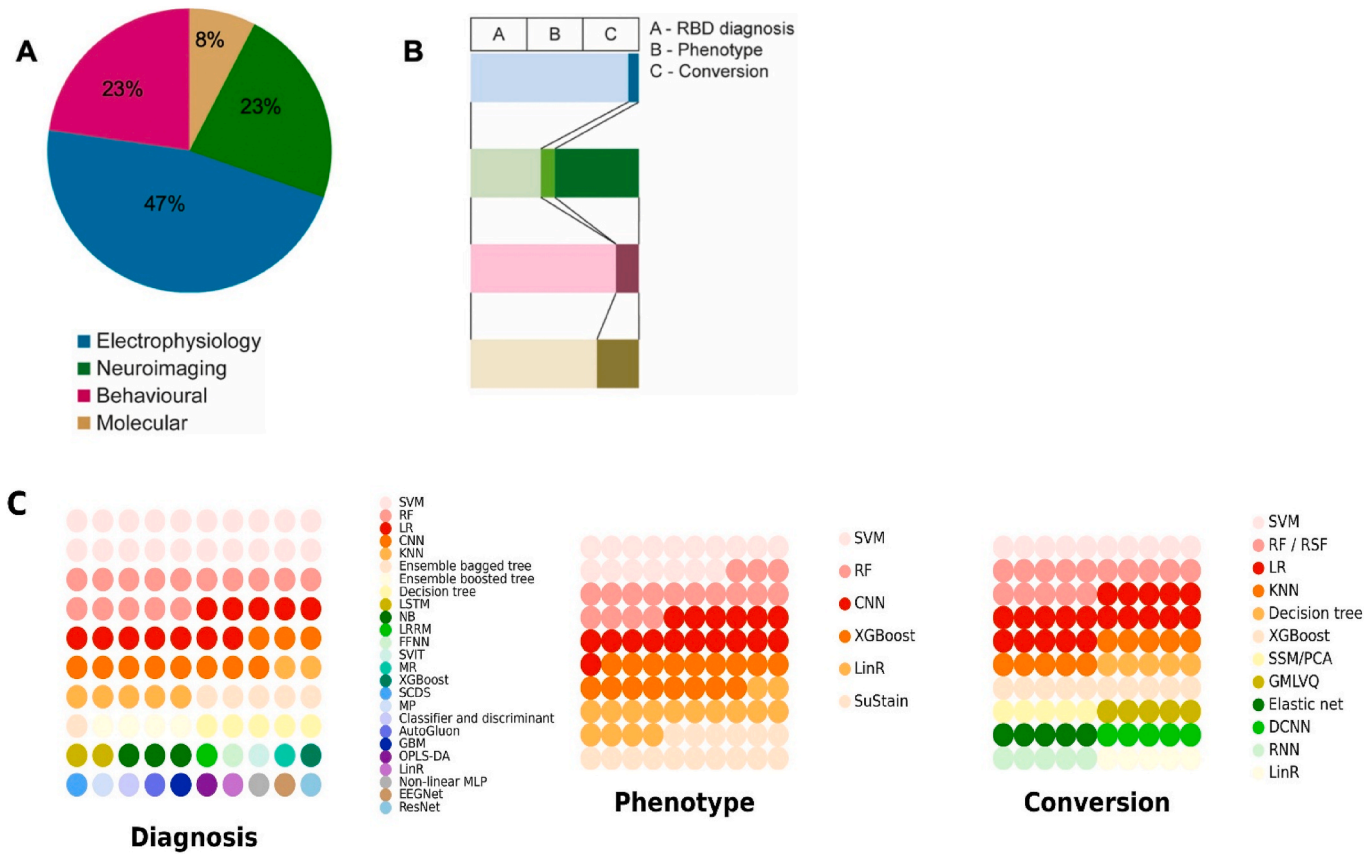

Fig. 4. Characteristics of ML and datasets used across the literature
(A) Percentage of studies that used electrophysiology, neuroimaging, behavioural and molecular data across the literature. (B) Across studies that used electrophysiology (blue), neuroimaging (green), behaviour (pink) and molecular (brown) information, the respective proportion of studies that used ML for diagnosis, phenotype classification and phenoconversion to an α-synucleinopathy. (C) Type of ML algorithm used for diagnosis (left), phenotype classification (middle) and phenoconversion (right). The number of dots is normalised to the total number of algorithm implementations within each objective. See Supplementary Table 5 for the underlying counts. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

of model uses in diagnosis studies (21/102 model uses, 20.6%), whereas logistic regression was most common in phenoconversion studies (4/20, 20.0%). Phenotype-classification studies applied a small and heterogeneous set of approaches: SVM, random forests, XGBoost, linear regression, convolutional neural networks (CNN) and SuStain each appeared once (1/6 model uses; 16.7% each).

3.6. *Methodological Quality Assessment Results*

Comprehensive tables of the methodological quality assessment of studies using the EPHPP Scale [33] are displayed in Supplementary Tables 6–8 and Supplementary Fig. 1. Almost all studies (92%) exhibited moderate selection bias, with no study having a weak selection bias. In contrast, blinding was a weak bias in 88% of studies and strong in only 3% of them. Overall, bias was weak in 43% of studies, moderate in 49% and strong in 8% according to global rating. Importantly, there was no discernible association between global study quality assessment ratings and the positivity of findings (Kruskal-Wallis test on accuracy between weak, moderate and strong bias: $p = 0.08$, see Supplementary Figs. 1 and 2 and Supplementary Table 7). The few studies classified as high risk (8%, n = 6/75) were due to the absence of validation and the lack of detailed algorithmic development.

4. Discussion

This systematic review identified 335 records, screened 202 after duplicate removal, and ultimately included 75 studies (diagnosis n = 58, phenotype n = 4, phenoconversion n = 13) (Fig. 2; Tables 1–3). Across diagnostic studies, reported performance was generally high (median accuracy ~91%), with PSG-derived signals (including EMG-based RSWA metrics) and EEG-based features most consistently supporting robust classification, albeit often without external validation (Table 1; Fig. 3A). Phenotype and phenoconversion studies were fewer and more heterogeneous, but dopaminergic imaging and multimodal approaches achieved the strongest prognostic performance (AUC up to ~0.94) (Tables 2 and 3; Fig. 3C). However, across all objectives, performance estimates were frequently limited by class imbalance, small single-centre cohorts, heterogeneous endpoints and limited external validation, making direct modality-to-modality comparisons inherently uncertain.

Hence, the vast majority of studies focused on RBD diagnosis, followed by phenoconversion (Tables 1–3). This trend is unsurprising, given the clinical risks associated with RBD and the availability of symptomatic treatment for dream-enactment behaviours, whereas no treatment can currently slow the progression from RBD to an α-synucleinopathy [87]. Since various disease-modifying treatments are being developed against these neurodegenerative diseases and will be tested in the near future, priority might be given to the development of algorithms able to predict them [4]. In fact, the Ki-Young Jung group in Seoul has conducted “explainable” ML research in RBD, including an important clinical application using EEG-based ML models for predicting α-synucleinopathy phenoconversion timing and subtype [88]. In 236 patients with iRBD who were followed for a mean 3.5 years (up to 8 years), 31 patients phenoconverted (n = 16, PD; n = 9, DLB; n = 6 MSA) [88]. The best model for survival prediction was the random forest model, and the best model for subtype prediction was the K-nearest

neighbour model [88]. EEG slowing contributed substantially to both models [88].

Furthermore, from a clinical perspective, high predictive performance is not sufficient on its own [89]. In many jurisdictions, including the EU, regulatory and professional frameworks now emphasise the need for explainable ML in medicine [90]. Models should therefore provide transparent feature importance or post-hoc explanations that clinicians can understand and scrutinise, rather than functioning as opaque black boxes [89]. On a fundamental perspective, identifying which features of RBD are associated with a future α-synucleopathy could shed light on potential mechanisms linking these 2 entities.

4.1. Methodological caveats for ML studies

There are specific methodological caveats to consider for MRI studies. Most are single-centre, cross-sectional protocols (see Table 1; Supplement). Because MRI acquisition is costly, labour-intensive and technically demanding, external validation cohorts are harder to obtain. Consequently, the number of participants in MRI studies is usually small. Performance and generalisability would probably improve with better standardisation of screening protocols, imaging acquisition and data preprocessing.

Another concern in some sleep-ML studies is the calculation of classification performance on epochs of data acquisition rather than on subjects [89]. This matters because subject-specific signal characteristics may be more discriminating than the disease itself. If epochs, rather than individuals, are used for classification, all epochs from a given subject can be assigned to a specific class simply because they belong to that subject and not necessarily because they belong to the corresponding diagnostic class. Other caveats include the need to normalise data sourced from different domains. When combining different metrics in one algorithm, those with larger numeric ranges can overshadow those with lower values; z-scoring is one common way to mitigate this.

In terms of methodological considerations for translational use in clinical settings, several trends emerged. First, no acquisition modality (e.g., EEG, MRI, behaviour) showed systematically better performance for diagnosing RBD (Fig. 3A–Table 1). Despite the small number of studies on phenotype classification and phenoconversion, no consistent modality-level trend emerged in those categories either (Tables 2 and 3, Fig. 3B and C). This observation needs confirmation in larger datasets, but it nonetheless encourages continued work on cost-effective systems for diagnosis, phenotyping and phenoconversion. At the ML level, only limited validation studies have tested clinical reliability [89]. However, encouraging results show that phenoconversion, including subtype prediction, can be modelled, and larger studies are needed to improve currently available models. Future work will need to extract general and reproducible rules using prospective, international, multi-cohort patient groups. It is imperative that these approaches are validated in clinical settings to ensure generalisability and clinical interpretability [89]. The availability of large pre-existing datasets (see Supplementary Table 3), international collaborations such as ENIGMA-PD and ENIGMA-Sleep, pre-registration, and code sharing should help control bias and improve reproducibility.

There are several challenges to consider in measuring RBD progression using ML models. An important one is the heterogeneous definition of "phenoconversion" (also see Table 3). There is still a dire need to define it reproducibly before any attempt to predict phenoconversion using ML models. It is also worth noting that different studies followed patients at different intervals, which inevitably affects model accuracy (Table 3). Critically, progression changes in clinical manifestations may not follow a linear trajectory, and they may differ substantially for PD, DLB and MSA. A perennial concern when devising an ML model for phenoconversion is how to define clinically meaningful endpoints that measure stages across disease progression. The ideal progression marker needs to be accessible, quantitative, easy to use, and applicable at different disease stages. The lack of attention to these issues challenges the effectiveness and clinical utility of the ML models employed. Overall, time-to-event analysis is still underused in this field, despite recent adoption of random survival forests and survival embeddings in a few studies [e.g. Refs. [81,82,88]; Tables 2 and 3]. Framing phenoconversion as a survival problem, rather than a simple binary endpoint at arbitrary follow-up, would better capture censoring and unequal follow-up and aligns more naturally with how clinicians think about risk over time [89].

From an ecological standpoint, studies using digital forms of testing offer a cost-effective approach that enables clinicians and researchers to capture real-world data and detect fluctuations over time, i.e., across days, weeks and months [72,76]. However, there are also challenges in digital technology. Data acquired outside laboratory and clinical settings may be more variable in quality and signal fidelity, which can affect the performance and generalisability of machine learning models. Furthermore, home-based recordings may capture less comprehensive signal information than laboratory polysomnography, although they offer important advantages in scalability and ecological validity (see Refs. [70,72,79]).

5. Conclusion

Taken together, our findings suggest several priorities for the field. First, large, prospective, multi-centre cohorts with harmonised PSG, imaging and behavioural protocols are needed to allow robust external validation of ML models and to reduce site-specific biases. Second, future work should converge on reproducible definitions of phenoconversion and clinically meaningful milestones, coupled with appropriate time-to-event modelling. Third, explainable ML methods, with clear reporting of input features and decision rules, are essential for clinical adoption and regulatory acceptance. Fourth, scalable and lower-cost biomarkers such as ECG, actigraphy and smartphone-based motor and speech testing should be systematically benchmarked against vPSG and advanced imaging to assess whether they can genuinely reduce dependence on expensive investigations. In conclusion, while current studies are encouraging, there are methodological considerations that limit their reproducibility. Methodological guidelines could contribute to standardize the methodological approaches and ultimately favour the comparability across studies [89]. Subsequently, future studies should focus on prospective multi-centre protocols for a more integrated, optimised and refined RBD ML model that could be implemented in a clinical setting. Without these steps, high reported performance will remain difficult to translate into clinically reliable tools.

Contributors

KR selected, reviewed, and assessed all the eligible studies and wrote the first draft of the manuscript.

LS wrote the manuscript and drew the figures with the help of all other authors.

NB reviewed all the eligible studies and participated in reviewing and drafting of the manuscript.

SQ reviewed and assessed all the eligible studies and participated in reviewing and drafting of the manuscript.

AD participated in reviewing and drafting of the manuscript.

JR participated in reviewing and drafting of the manuscript.

ZC participated in reviewing and drafting of the manuscript.

RL participated in reviewing and drafting of the manuscript.

NG participated in reviewing and drafting of the manuscript.

MT participated in reviewing and drafting of the manuscript.

PG participated in reviewing and drafting of the manuscript.

KRC participated in reviewing and drafting of the manuscript.

CS participated in reviewing and drafting of the manuscript.

IR supervised the study, participated in reviewing and drafting of the manuscript.

Data sharing

No data was collected for this review.

Author contributions

KR selected, reviewed, and KR, NB and SQR assessed all the eligible studies. All authors were involved in reviewing and drafting the manuscript.

Declaration of interests

KR declares no competing interest.
LS declares no competing interest.
NB declares no competing interest.
SQ declares no competing interest.
AD declares no competing interest.
JR declares no competing interest.
ZC declares no competing interest.
RL declares no competing interest.
NG declares no competing interest.
MT declares no competing interest.
PG declares no competing interest.
KRC declares no competing interest.
CS declares no competing interest.
IR declares no competing interest.

Acknowledgments

This research was funded in whole, or in part, by the Wellcome Trust [103952/Z/14/Z]. Other fundings include a Guarantors of Brain fellowship (LS). For the purpose of open access, the author IR has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. This paper represents independent research in part funded by the NIHR Maudsley Biomedical Research Centre in SouthLondon, the Maudsley NHS Foundation Trust, and King's College London. The views expressed are those of the author (s) and not necessarily those of the NIHR or Department of Health and Social Care.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.smrv.2026.102298.

References

[1] Erkkinen MG, Kim M-O, Geschwind MD. Clinical neurology and epidemiology of the major neurodegenerative diseases. Cold Spring Harbor Perspect Biol 2018;10:a033118.
[2] Yang W, Hamilton JL, Kopil C, Beck JC, Tanner CM, Albin RL, et al. Current and projected future economic burden of Parkinson's disease in the US. npj Parkinson's Dis 2020;6:15.
[3] Dorsey ER, Elbaz A, Nichols E, Abbasi N, Abd-Allah F, Abdelalim A, et al. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol 2018;17. 939-53.
[4] McFarthing K, Rafaloff G, Baptista M, Mursaleen L, Fuest R, Wyse RK, et al. Parkinson's disease drug therapies in the clinical trial pipeline: 2022 update. J Parkinsons Dis 2022;12:1073–82.
[5] McFarthing K, Larson D, Simuni T. Clinical trial Highlights–GLP-1 agonists. J Parkinsons Dis 2020;10:355.
[6] Brauer R, Wei L, Ma T, Athauda D, Girges C, Vijiaratnam N, et al. Diabetes medications and risk of Parkinson's disease: a cohort study of patients with diabetes. Brain 2020;143:3067–76.
[7] Aviles-Olmos I, Limousin P, Lees A, Foltynie T. Parkinson's disease, insulin resistance and novel agents of neuroprotection. Brain 2013;136:374–84.
[8] Athauda D, Maclagan K, Skene SS, Bajwa-Joseph M, Letchford D, Chowdhury K, et al. Exenatide once weekly versus placebo in Parkinson's disease: a randomised, double-blind, placebo-controlled trial. Lancet 2017;390:1664–75.
[9] Poewe W, Seppi K, Marini K, Mahlknecht P. New hopes for disease modification in Parkinson's Disease. Neuropharmacology 2020;171:108085.
[10] Mahlknecht P, Seppi K, Poewe W. The concept of prodromal Parkinson's disease. J Parkinsons Dis 2015;5:681–97.
[11] Roguski A, Rayment D, Whone AL, Jones MW, Rolinski M. A neurologist's guide to REM sleep behavior disorder. Front Neurol 2020;11:610.
[12] Schenck CH. REM sleep behaviour disorder (RBD): personal perspectives and research priorities. J Sleep Res 2025;34:e14228.
[13] Dauvilliers Y, Schenck CH, Postuma RB, Iranzo A, Luppi PH, Plazzi G, et al. REM sleep behaviour disorder. Nat Rev Dis Primers 2018;4:19.
[14] Ray Chaudhuri K, Leta V, Bannister K, Brooks DJ, Svenningsson P. The noradrenergic subtype of Parkinson disease: from animal models to clinical practice. Nat Rev Neurol 2023;19:333–45.
[15] Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. Chest 2014;146:1387–94.
[16] McCarter SJ, St Louis EK, Duwell EJ, Timm PC, Sandness DJ, Boeve BF, et al. Diagnostic thresholds for quantitative REM sleep phasic burst duration, phasic and tonic muscle activity, and REM atonia index in REM sleep behavior disorder with and without comorbid obstructive sleep apnea. Sleep 2014;37:1649–62.
[17] Leclair-Visonneau L, Feemster JC, Bibi N, Gossard TR, Jagielski JT, Strainis EP, et al. Contemporary diagnostic visual and automated polysomnographic REM sleep without atonia thresholds in isolated REM sleep behavior disorder. J Clin Sleep Med : JCSM : official publication of the American Academy of Sleep Medicine 2024;20:279–91.
[18] Mariño N, Serradell M, Gaig C, Mayà G, Montini A, Matos N, et al. Audiovisual analysis of the diagnostic video polysomnography in patients with isolated REM sleep behavior disorder. Journal of neurology 2025;272:146.
[19] Galbiati A, Verga L, Giora E, Zucconi M, Ferini-Strambi L. The risk of neurodegeneration in REM sleep behavior disorder: a systematic review and meta-analysis of longitudinal studies. Sleep Med Rev 2019;43:37–46.
[20] Postuma RB, Gagnon JF, Bertrand JA, Génier Marchand D, Montplaisir JY. Parkinson risk in idiopathic REM sleep behavior disorder: preparing for neuroprotective trials. Neurology 2015;84:1104–13.
[21] Lee WJ, Baek SH, Im HJ, Lee SK, Yoon JE, Thomas RJ, et al. REM sleep behavior disorder and its possible prodromes in general population: prevalence, polysomnography findings, and associated factors. Neurology 2023;101:e2364–75.
[22] Mahlknecht P, Seppi K, Frauscher B, Kiechl S, Willeit J, Stockner H, et al. Probable RBD and association with neurodegenerative disease markers: a population-based study. Mov Disord : official journal of the Movement Disorder Society 2015;30:1417–21.
[23] Stefani A, Serradell M, Holzknecht E, Gaig C, Ibrahim A, Marrero P, et al. Low specificity of rapid eye movement sleep behavior disorder questionnaires: need for better screening methods. Mov Disord : official journal of the Movement Disorder Society 2023;38:1000–7.
[24] Dinov ID. Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. GigaScience 2016;5:s13742. 016-0117-6.
[25] Salari N, Kazeminia M, Sagha H, Daneshkhah A, Ahmadi A, Mohammadi M. The performance of various machine learning methods for Parkinson's disease recognition: a systematic review. Curr Psychol 2023;42:16637–60.
[26] Sharma M, Tiwari J, Patel V, Acharya UR. Automated identification of sleep disorder types using triplet half-band filter and ensemble machine learning techniques with EEG signals. Electronics 2021;10:1531.
[27] Staartjes VE, Kernbach JM. Foundations of machine learning-based clinical prediction modeling: part III—Model evaluation and other points of significance. Machine learning in clinical neuroscience: foundations and applications. Springer; 2022. p. 23–31.
[28] Badrulhisham F, Pogatzki-Zahn E, Segelcke D, Spisak T, Vollert J. Machine learning and artificial intelligence in neuroscience: a primer for researchers. Brain Behav Immun 2024;115:470–9.
[29] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8–17.
[30] Sharma M, Tiwari J, Acharya UR. Automatic sleep-stage scoring in healthy and sleep disorder patients using optimal wavelet filter bank technique with EEG signals. Int J Environ Res Publ Health 2021;18:3087.
[31] Salsone M, Quattrone A, Vescio B, Ferini-Strambi L, Quattrone A. A machine learning approach for detecting idiopathic REM sleep behavior disorder. Diagnostics 2022;12:2689.
[32] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group* t. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine 2009;151:264–9.
[33] Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evidence-Based Nurs 2004;1:176–84.
[34] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. second ed. ed. Wiley; 2019.
[35] Christensen JA, Zoetmulder M, Koch H, Frandsen R, Arvastson L, Christensen SR, et al. Data-driven modeling of sleep EEG and EOG reveals characteristics indicative of pre-parkinson's and Parkinson's disease. J Neurosci Methods 2014;235:262–76.
[36] Rechichi I, Iadarola A, Zibetti M, Cicolin A, Olmo G. Assessing rem sleep behaviour disorder: from machine learning classification to the definition of a continuous dissociation index. Int J Environ Res Publ Health 2021;19:248.
[37] Brink-Kjaer A, Gunter KM, Mignot E, During E, Jennum P, Sorensen HB. End-to-end deep learning of polysomnograms for classification of REM sleep behavior disorder. 2022 44th annual international conference of the IEEE engineering in medicine & Biology Society (EMBC). IEEE; 2022. p. 2941–4.
[38] Zhuang D, Rao I, Ibrahim AK. A machine learning approach to automatic classification of eight sleep disorders. arXiv preprint arXiv:220406997. 2022.

[39] Cooray N, Andreotti F, Lo C, Symmonds M, Hu MT, De Vos M. Detection of REM sleep behaviour disorder by automated polysomnography analysis. Clin Neurophysiol 2019;130:505–14.
[40] Andreotti F, Phan H, Cooray N, Lo C, Hu MT, De Vos M. Multichannel sleep stage classification and transfer learning using convolutional neural networks. 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2018. p. 171–4.
[41] Sharma M, Darji J, Thakrar M, Acharya UR. Automated identification of sleep disorders using wavelet-based features extracted from electrooculogram and electromyogram signals. Comput Biol Med 2022;143:105224.
[42] Cooray N, Andreotti F, Lo C, Symmonds M, Hu MT, De Vos M. Proof of concept: screening for REM sleep behaviour disorder with a minimal set of sensors. Clin Neurophysiol 2021;132:904–13.
[43] Wallis P, Yaeger D, Kain A, Song X, Lim M. Automatic event detection of rem sleep without atonia from polysomnography signals using deep neural networks. ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2020. p. 4112–6.
[44] Gunter KM, Brink-Kjær A, Mignot E, Sørensen HB, During E, Jennum P. SViT: a spectral vision transformer for the detection of REM sleep behavior disorder. IEEE J Biomed Health Inform 2023;27(9):4285–92.
[45] Brink-Kjær A, Cesari M, Sixel-Döring F, Mollenhauer B, Trenkwalder C, Mignot E, et al. Arousal characteristics in patients with Parkinson's disease and isolated rapid eye movement sleep behavior disorder. Sleep 2021;44. zsab167.
[46] Markov K, Elgendi M, Birrer V, Menon C. Interpretable feature-based machine learning for automatic sleep detection using photoplethysmography. npj Biosensing 2025;2:24.
[47] Feuerstein S, Stefani A, Angerbauer R, Egger K, Ibrahim A, Holzknecht E, et al. Sleep structure discriminates patients with isolated REM sleep behavior disorder: a deep learning approach. 2024 46th annual international conference of the IEEE engineering in medicine and biology Society (EMBC). IEEE; 2024. p. 1–4.
[48] Cesari M, Portscher A, Stefani A, Angerbauer R, Ibrahim A, Brandauer E, et al. Machine learning predicts phenoconversion from polysomnography in isolated REM sleep behavior disorder. Brain Sci 2024;14:871.
[49] Cesari M, Christensen JA, Sorensen HB, Jennum P, Mollenhauer B, Muntean ML, et al. External validation of a data-driven algorithm for muscular activity identification during sleep. J Sleep Res 2019;28:e12868.
[50] Kempfner J, Sorensen G, Nikolic M, Frandsen R, Sorensen H, Jennum P. Rapid eye movement sleep behavior disorder as an outlier detection problem. J Clin Neurophysiol 2014;31:86–93.
[51] Jung Y, Kim S, Kim S, Kim W, Ryu D, Choi Y, et al. A deep learning approach for classifying REM sleep behavior disorder using EEGNet, vol. 134. Parkinsonism & Related Disorders; 2025.
[52] Sharma M, Lodhi H, Yadav R, Acharya UR. Sleep disorder identification using wavelet scattering on ECG signals. Int J Imag Syst Technol 2024;34:e22980.
[53] Urtnasan E, Joo EY, Lee KH. Ai-enabled algorithm for automatic classification of sleep disorders based on single-lead electrocardiogram. Diagnostics 2021;11:2054.
[54] Widasari ER, Tanno K, Tamura H. Automatic sleep disorders classification using ensemble of bagged tree based on sleep quality features. Electronics 2020;9:512.
[55] Salsone M, Vescio B, Quattrone A, Marelli S, Castelnuovo A, Casoni F, et al. Periodic leg movements during sleep associated with REM sleep behavior disorder: a machine learning Study. Diagnostics 2024;14:363.
[56] Buettner R, Grimmeisen A, Gotschlich A. High-performance diagnosis of sleep disorders: a novel, accurate and fast machine learning approach using electroencephalographic data. 2020.
[57] Kazemi A, McKeown MJ, Mirian MS. Sleep staging using semi-unsupervised clustering of EEG: application to REM sleep behavior disorder. Biomed Signal Process Control 2022;75:103539.
[58] Srivastava S. Classifying REM sleep behavior disorder through CNNs with image-based representations of EEGs. medRxiv 2022. 2022.04. 03.22273365.
[59] Rechichi I, Amato F, Cicolin A, Olmo G. Single-Channel EEG detection of REM sleep behaviour disorder: the influence of REM and slow wave sleep. International work-conference on bioinformatics and biomedical engineering. Springer; 2022. p. 381–94.
[60] Dimitriadis SI, Salis CI, Liparas D. An automatic sleep disorder detection based on EEG cross-frequency coupling and random forest model. J Neural Eng 2021;18:046064.
[61] Wadichar A, Murarka S, Shah D, Bhurane A, Sharma M, Mir HS, et al. A hierarchical approach for the diagnosis of sleep disorders using convolutional recurrent neural network. IEEE Access 2023;11:125244–55.
[62] Murarka S, Wadichar A, Bhurane A, Sharma M, Acharya UR. Automated classification of cyclic alternating pattern sleep phases in healthy and sleep-disordered subjects using convolutional neural network. Comput Biol Med 2022;146:105594.
[63] Christensen JA, Kempfner J, Zoetmulder M, Leonthin HL, Arvastson L, Christensen SR, et al. Decreased sleep spindle density in patients with idiopathic REM sleep behavior disorder and patients with Parkinson's disease. Clin Neurophysiol 2014;125:512–9.
[64] Sharma M, Patel V, Tiwari J, Acharya UR. Automated characterization of cyclic alternating pattern using wavelet-based features and ensemble learning techniques with eeg signals. Diagnostics 2021;11:1380.
[65] Lo C, Arora S, Ben-Shlomo Y, Barber TR, Lawton M, Klein JC, et al. Olfactory testing in Parkinson disease and REM behavior disorder: a machine learning approach. Neurology 2021;96:e2016–27.
[66] Fereshtehnejad S-M, Yao C, Pelletier A, Montplaisir JY, Gagnon J-F, Postuma RB. Evolution of prodromal Parkinson's disease and dementia with Lewy bodies: a prospective study. Brain 2019;142:2051–67.
[67] Jeancolas L, Mangone G, Petrovska-Delacrétaz D, Benali H, Benkelfat B-E, Arnulf I, et al. Voice characteristics from isolated rapid eye movement sleep behavior disorder to early Parkinson's disease. Park Relat Disord 2022;95:86–91.
[68] Zhang X, Liang D, Ma L, Huang Y, Wan Y, Zhou K, et al. Cognitive and motor profiles as prodromal markers in predicting phenoconversion and phenotype in isolated REM sleep behavior disorder. Sleep Med 2023;112:262–72.
[69] Arora S, Baig F, Lo C, Barber TR, Lawton MA, Zhan A, et al. Smartphone motor testing to distinguish idiopathic REM sleep behavior disorder, controls, and PD. Neurology 2018;91:e1528–38.
[70] Brink-Kjaer A, Gupta N, Marin E, Zitser J, Sum-Ping O, Hekmat A, et al. Ambulatory detection of isolated rapid-eye-movement sleep behavior disorder combining actigraphy and questionnaire. Mov Disord 2023;38:82–91.
[71] Cochen De Cock V, Dotov D, Lacombe S, Picot MC, Galtier F, Driss V, et al. Classifying idiopathic rapid eye movement sleep behavior disorder, controls, and mild Parkinson's disease using gait parameters. Mov Disord 2022;37:842–6.
[72] Cesari M, Ruzicka L, Högl B, Ibrahim A, Holzknecht E, Heidbreder A, et al. Improved automatic identification of isolated rapid eye movement sleep behavior disorder with a 3D time-of-flight camera. Eur J Neurol 2023;30:2206–14.
[73] Rusz J, Hlavnička J, Tykalová T, Bušková J, Ulmanová O, Růžička E, et al. Quantitative assessment of motor speech abnormalities in idiopathic rapid eye movement sleep behaviour disorder. Sleep Med 2016;19:141–7.
[74] Rusz J, Hlavnička J, Novotný M, Tykalová T, Pelletier A, Montplaisir J, et al. Speech biomarkers in rapid eye movement sleep behavior disorder and Parkinson disease. Ann Neurol 2021;90:62–75.
[75] Rusz J, Hlavnička J, Tykalová T, Novotný M, Dušek P, Šonka K, et al. Smartphone allows capture of speech abnormalities associated with high risk of developing Parkinson's disease. IEEE Trans Neural Syst Rehabil Eng 2018;26:1495–507.
[76] Benba A, Jilbab A, Sandabad S, Hammouch A. Voice signal processing for detecting possible early signs of Parkinson's disease in patients with rapid eye movement sleep behavior disorder. Int J Speech Technol 2019;22:121–9.
[77] Barber TR, Lawton M, Rolinski M, Evetts S, Baig F, Ruffmann C, et al. Prodromal parkinsonism and neurodegenerative risk stratification in REM sleep behavior disorder. Sleep 2017;40. zsx071.
[78] Simonet C, Pérez-Carbonell L, Galmés-Ordinas MA, Huxford BF, Chohan H, Gill A, et al. The motor dysfunction seen in isolated REM sleep behavior disorder. Mov Disord 2024;39(6):1054–9.
[79] Abdelfattah M, Zhou L, Sum-Ping O, Hekmat A, Galati J, Gupta N, et al. Automated detection of isolated REM sleep behavior disorder using computer vision. Ann Neurol 2025;97:860–72.
[80] Guarín DL, Acevedo G, Calonge C, Wong JK, McFarland NR, Ramirez-Zamora A, et al. Video analysis reveals early signs of Bradykinesia in REM sleep behavior disorder and Parkinson's disease. npj Parkinson's Dis 2025;11:222.
[81] Shin Y-W, Byun J-I, Sunwoo J-S, Rhee C-S, Shin J-H, Kim H-J, et al. Predicting phenoconversion in isolated RBD: machine learning and explainable AI approach. Clocks & Sleep 2025;7:19.
[82] Fereshtehnejad S, Bahador N, Delva A, Pelletier A, Montplaisir J, Gagnon J, et al. Predicting time-to-phenoconversion to parkinson's Disease/Dementia with lewy bodies in idiopathic REM sleep behavior disorder (iRBD) using machine learning, vol. 134. Parkinsonism & Related Disorders; 2025.
[83] Stiasny-Kolster K, Doerr Y, Möller JC, Höffken H, Behr TM, Oertel WH, et al. Combination of 'idiopathic' REM sleep behaviour disorder and olfactory dysfunction as possible indicator for alpha-synucleinopathy demonstrated by dopamine transporter FP-CIT-SPECT. Brain 2005;128:126–37.
[84] Postuma RB, Iranzo A, Hu M, Högl B, Boeve BF, Manni R, et al. Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study. Brain 2019;142:744–59.
[85] Arnaldi D, Chincarini A, Hu MT, Sonka K, Boeve B, Miyamoto T, et al. Dopaminergic imaging and clinical predictors for phenoconversion of REM sleep behaviour disorder. Brain 2021;144:278–87.
[86] Feng S, Ge J, Zhao S, Xu Q, Lin H, Li X, et al. Dopaminergic damage pattern predicts phenoconversion time in isolated rapid eye movement sleep behavior disorder. Eur J Nucl Med Mol Imag 2023;51:159–67.
[87] Howell M, Avidan AY, Foldvary-Schaefer N, Malkani RG, During EH, Roland JP, et al. Management of REM sleep behavior disorder: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med : JCSM : official publication of the American Academy of Sleep Medicine 2023;19:759–68.
[88] Jeong E, Woo Shin Y, Byun J-I, Sunwoo J-S, Roascio M, Mattioli P, et al. EEG-based machine learning models for the prediction of phenoconversion time and subtype in isolated rapid eye movement sleep behavior disorder. Sleep 2024;47. zsae031.
[89] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ (Clinical research ed) 2024;385:e078378.
[90] van Kolfschooten H, van Oirschot J. The EU Artificial Intelligence Act (2024): implications for healthcare. Health Policy 2024;149:105152.