Machine and deep learning in REM sleep behavior disorder: a scoping review and analysis of reporting quality

Sleep Medicine Reviews 88 (2026) 102299

Contents lists available at ScienceDirect

Sleep Medicine Reviews

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

Technical Review

# Machine and deep learning in REM sleep behavior disorder: a scoping review and analysis of reporting quality

Andreas Brink-Kjaer [a,1], Irene Rechichi [b,1], Dario Arnaldi [c,d], Olivia Cesarone [e], Emmanuel During [f], Simon Feuerstein [g], Katarina M. Gunter [h], Birgit Högl [g], Abubaker Ibrahim [g], Poul Jennum [i], Emmanuel Mignot [j], Gabriella Olmo [b], Monica Roascio [k], Ambra Stefani [g], Qi Tang [g], Matteo Cesari [g,*]

[a] Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark
[b] Department of Control and Computer Engineering, Politecnico di Torino, Turin, Italy
[c] DINGOMI, University of Genoa, Genoa, Italy
[d] Clinical Neurophysiology, IRCCS Ospedale Policlinico San Martino, Genoa, Italy
[e] Department of Neurology, Mayo Clinic Rochester, Minnesota, USA
[f] Department of Neurology, Medicine, Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, USA
[g] Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria
[h] Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, Oxford, UK
[i] Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Rigshospitalet, Copenhagen, Denmark
[j] Stanford Sleep Sciences and Medicine, Department of Psychiatry, Stanford University, Palo Alto, USA
[k] Department of Informatics, Bioengineering, Robotics and System Engineering (DIBRIS), University of Genoa, Genoa, Italy

ARTICLE INFO

*Keywords:*
Artificial intelligence
AI
DL
ML
RBD
Reporting quality

ABSTRACT

Rapid eye movement (REM) sleep behavior disorder (RBD) is a parasomnia, and its isolated form is of particular interest, as it is an early phase alpha-synucleinopathy. Machine learning (ML) and deep learning (DL) models offer potential for automated detection, prediction of phenoconversion, and phenotyping. This scoping review identified 75 studies applying ML/DL in RBD and evaluated their methodological and reporting quality using the APPRAISE-AI tool. Most studies (73.3%) focused on RBD detection and mainly used polysomnographic data for this, while 16% addressed prediction of phenoconversion, with imaging data being the most employed modality. Sample sizes were generally small (most studies including only 20–100 individuals). According to APPRAISE-AI scores, 80% of studies had moderate overall methodological and reporting quality. Common deficiencies included lack of transparency in data and code sharing (23.3%), and poor reporting of hyperparameter tuning (17.1%), bias assessment (26.9%), and error analysis (0.66%). Data leakage was observed in 32% of studies. These issues hinder clinical translation and prevent incremental progress between research groups. Without transparent reporting and shared resources, replication and model comparisons become nearly impossible. Future work should adopt open science principles and rigorous validation to advance AI-based tools in sleep medicine.

## 1. Introduction

The term artificial intelligence (AI) refers to machine-based systems operating with varying degrees of autonomy and that, based on the inputs received, can generate outputs such as classification, predictions, recommendations, or other types of decisions [1]. Those systems typically use machine learning (ML) and deep learning (DL) algorithms to generate the desired outputs [2]. Compared to rule-based systems, ML and DL algorithms learn patterns and make decisions directly from data, without hard-coded instructions [2].

The role of ML/DL is expanding tremendously in every field of healthcare [3], including sleep medicine and research [4,5]. Among the various subfields of sleep research, investigations of rapid eye movement (REM) sleep behavior disorder (RBD) are particularly well positioned to benefit from the use of ML and DL. First, diagnosis of RBD is a

* Corresponding author. Department of Neurology, Medical University of Innsbruck Anichstrasse 35, Innsbruck, 6020, Austria.
*E-mail address:* matteo.cesari@i-med.ac.at (M. Cesari).
[1] Shared first authorship.

https://doi.org/10.1016/j.smrv.2026.102299
Received 22 August 2025; Received in revised form 10 March 2026; Accepted 20 April 2026
Available online 29 April 2026

**Glossary of terms**

| | |
|---|---|
| **AI** | artificial intelligence |
| **DL:** | deep learning |
| **DLB** | dementia with Lewy bodies |
| **iRBD** | isolated RBD |
| **ML:** | machine learning |
| **MSA** | multiple system atrophy |
| **RBD** | REM sleep behavior disorder |
| **REM** | rapid eye movement |
| **RWA** | REM sleep without atonia |
| **PD** | Parkinson's disease |
| **PRISMA-ScR** | Preferred Reporting Items for Systematic Reviews and Meta-Analyses extended for Scoping Reviews |
| **PSG** | polysomnography |

time-consuming task, requiring documentation of REM sleep without atonia (RWA) from night polysomnography (PSG) and careful analysis of nocturnal video [6,7]. ML/DL algorithms might therefore offer a great opportunity to quickly and objectively analyze sleep recordings and support clinicians in the diagnostic process. Second, patients with isolated RBD (iRBD) are recognized to be in the early stages of an alpha-synuclein-related neurodegeneration (i.e., Parkinson's disease - PD, dementia with Lewy bodies – DLB, and multiple system atrophy – MSA) [8,9], but a single biomarker that reliably indicates the proximity to phenoconversion to overt alpha-synucleinopathy is still missing [9]. As ML/DL algorithms can potentially find complex patterns in large data, such solutions might contribute to identifying novel biomarkers or biomarker combinations predicting phenoconversion. In this regard, prediction of the phenotype (PD, DLB or MSA) could also be accomplished by these solutions. Finally, RBD remains largely underdiagnosed and ML/DL algorithms applied to data collected with wearable and unobtrusive technologies might be useful, after appropriate extensive validation, for screening in the targeted populations [10].

Although there is great excitement related to the application of ML and DL in the healthcare sector, recent reviews have highlighted concerns related to quality of reporting of ML/DL studies in several fields of medicine [11], including oncology [12], neuroimaging [13], and urology [14]. For ML/DL applications, in fact, it is highly important that methods and results are reported in a comprehensive, reproducible and transparent manner. Only high-quality reporting can ensure that proposed AI solutions are not affected by flaws such as data leakage, overfitting and poor generalizability. To the best of our knowledge, no study has ever evaluated the quality of reporting of ML/DL studies in the field of sleep medicine.

With the aim of improving the quality of reporting of ML/DL research in medicine, several guidelines applicable to various research phases (preclinical, clinical or translational) have been recently suggested [15]. Among them, the APPRAISE-AI stands out as the only tool with demonstrated methodological robustness for assessing the reporting quality of AI studies designed for clinical decision support [16]. It uniquely provides empirical evidence of reliability - showing low inter-rater reliability- and it is the only framework to show strong correlations with independent indicators of study quality, such as 3-year citation rate [16]. Given that ML/DL applications in the field of RBD are primarily aimed at clinical decision support—such as screening, diagnosis, and prediction of phenoconversion timing and phenotype—the APPRAISE-AI tool offers the most rigorous and appropriately validated framework for evaluating the reporting quality of these studies.

The aim of this review is to evaluate literature studies with applications of ML/DL in RBD cohorts and to critically appraise their methodological and reporting quality using the APPRAISE-AI framework. Ultimately, through this critical evaluation, we aim to highlight the strengths and limitations of current research in this field and to identify key areas for future improvement.

## 2. Methods

### 2.1. Registration and reporting guidelines

This scoping review was registered in osf.io (https://doi.org/10.17605/OSF.IO/MWJD4) and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, extended for Scoping Reviews (PRISMA-ScR) guideline [17].

### 2.2. Search strategy

We searched PubMed and Web of Science for studies where ML/DL was applied in the field of RBD research published until 30 April 2025. The detailed search strategies are provided in Tables S1 and S2 in the Supplemental Material.

### 2.3. Eligibility criteria

We included studies that: (i) focused on populations with RBD (diagnosed with video-PSG); (ii) used either RBD as input variable or outcome of a model; (iii) employed ML/DL for the purposes of items (i) or (ii). Studies were excluded if: (i) the study type was a systematic/narrative/scoping review or a guidance document, commentary, editorial; (ii) no patients with RBD were included in the study; (iii) diagnosis of RBD was not performed with video-PSG according to international criteria [6]; (iv) no abstract was available; (v) they were not in English language; (vi) their text was not available to the authors; and (vii) no ML/DL model was used in the study. Furthermore, we decided to include in the review only studies where the primary focus of the ML/DL models was on RBD, either as predictor or outcome of the ML/DL models. In this way, we excluded studies using for instance ML/DL for sleep staging applied to RBD. For the scope of this review, we defined a ML/DL as follows either as (i) a model with >1 variable (i.e., predictor): fitted to some set of data and evaluated in another (usually a train-test, or train-validation, or cross-validation split); or (ii) a clustering approach that assigns groups to input data and evaluates these groups with regards to some metrics or known variables.

### 2.4. Study selection

A 2-level study selection process was conducted. Initial selection was based on title and abstract screening, and carried out in the online Rayyan platform. Prior to the main screening phase, a calibration procedure was conducted to verify the clarity and consistency of the inclusion and exclusion criteria. The three leading authors (ABK, IR, and MC) first established ground truth inclusion or exclusion decisions for a subset of 50 abstracts and ten authors independently screened them. Agreement among authors was moderate (Fleiss' κ: 0.38, averaged pairwise Cohen's κ: 0.46), but was considered acceptable for our purpose. Following this step, each title and abstract were evaluated independently by two authors; conflicts were resolved by one of the leading authors of the review (ABK, IR, and MC). For the articles that passed the initial screening, full-text screening was carried out by two of the three leading authors. Conflicts at this stage were resolved through consensus among all three leading authors.

### 2.5. Data extraction and assessment of quality of reporting

For each of the studies ultimately included in the review, general descriptive study characteristics were extracted. These included: i) year of publication, ii) country/countries of the study population, iii) sample size of the study populations, iv) aim of the proposed ML/DL models, v)

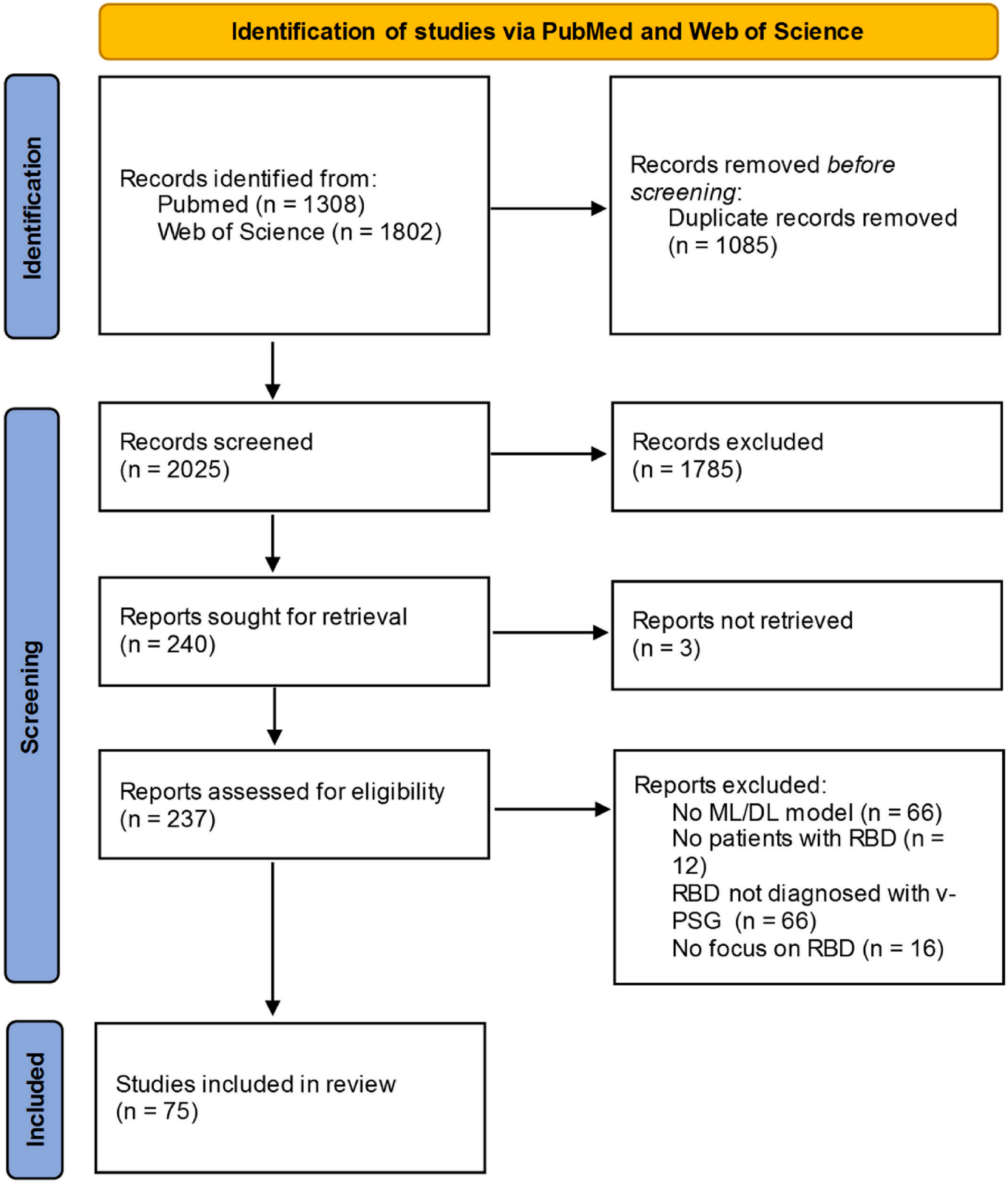


**Fig. 1.** PRISMA-ScR diagram of this scoping review.

type of algorithm/architecture employed, vi) input data of the model, and vii) main findings. Data extraction was performed for each study by one of the three leading authors.

The APPRAISE-AI tool [16] was used to assess the quality of AI methodology and reporting of the included studies. This instrument consists of 24 different items with a maximum overall score of 100 points. These items can be grouped in six main domains: 1) clinical relevance (maximum of 4 points), 2) data quality (maximum of 24 points), 3) methodological conduct (maximum of 20 points), 4) robustness of results (maximum of 20 points), 5) reporting quality

**Table 1**
**Overview of the studies included in the review.** Input data, aim of the ML/DL model, main findings and overall score according to the APPRAISE-AI tool [16] are shown.

| Study | Input data | Aim of ML/DL model | Main findings | Overall score |
|---|---|---|---|---|
| Abdelfattah M et al., 2025 [18] | Imaging (2D nocturnal video) | Detection of RBD | Using 5 features describing short movements in REM sleep, accuracy of 92% for RBD detection. | 61 |
| Alushaj E et al., 2025 [19] | Imaging (MRI) | Detection of RBD | Quantitative susceptibility mapping showed iron changes in SNc, GPe, GPi, and VTA that differentiated iRBD, PD, and HC with 81–86% accuracy. | 54 |
| Arnaldi D et al., 2021 [20] | Imaging (DaT-SPECT) | Prediction of phenoconversion | Highest risk linked to age over 70, constipation, and reduced putamen dopaminergic activity, yielding a hazard ratio of 5.71. | 57 |
| Arnaldi D et al., 2024 [21] | Imaging (DaT scan), motor, cognitive, olfaction, constipation. | Prediction of phenoconversion | Imaging data predicted conversion with up to 85% specificity and 77–85% sensitivity. Imaging superior over clinical data alone. | 52 |
| Arora S et al., 2018 [22] | Motor (smartphone motor test) | Detection of RBD | iRBD distinguished from controls and PD with 85-92% sensitivity and specificity. | 57 |
| Arora S et al., 2021 [23] | Voice | Detection of RBD | iRBD distinguished from controls and PD with up to 75% sensitivity and 73% specificity. | 63 |
| Benba A et al., 2019 [24] | Voice | Detection of RBD | PD distinguished from controls with 85% accuracy, 66% of iRBD classified as PD with this model. | 49 |
| Bisgaard S et al., 2015 [25] | PSG (EEG) | Detection of RBD | iRBD distinguished from controls with 78% accuracy. | 43 |
| Brink-Kjaer A et al., 2022 [26] | PSG (EEG, EOG, EMG, ECG) | Detection of RBD | Model classified RBD with 91.4% sensitivity and 86.3% specificity. | 60 |
| Brink-Kjaer A et al., 2023 [27] | Actigraphy, olfaction, constipation, autonomic, other (RBD questionnaire) | Detection of RBD | iRBD detected from actigraphy with 95% sensitivity and 91% precision, reaching 100% specificity when combining actigraphy with questionnaire (assessing RBD and non-motor symptoms). | 64 |
| Brink-Kjaer A et al., 2023 [28] | Actigraphy | Detection of RBD | Combining nighttime and 24-h actigraphy markers, detection of iRBD with 79% sensitivity and 96% specificity. | 62 |
| Byeon H, 2020 [29] | Motor, cognitive, other (clinical history). | Detection of RBD | Detection of RBD in PD with 79% sensitivity (age, motor scores and cognitive measures best predictors). | 43 |
| Cesari M et al., 2018 [30] | PSG (EMG) | Detection of RBD | REM atonia index the most sensitive method for RBD detection, though no single method optimal across all conditions. | 51 |
| Cesari M et al., 2021 [31] | PSG (EEG, EOG) | Detection of RBD | RBD identified with accuracy of 82% in PD patients (micro-sleep instability key predictor). | 58 |
| Cesari M et al., 2019 [32] | PSG (EMG) | Detection of RBD | EMG algorithm distinguishes RBD from controls and patients with PLMD with accuracy up to 70.8% and outperforms other EMG methods. | 60 |
| Cesari M et al., 2019 [33] | PSG (EMG) | Detection of RBD | Previously developed EMG algorithm detects RBD with 84.2% accuracy in an external center. | 63 |
| Cesari M et al., 2024 [34] | PSG (EEG, EMG) | Prediction of phenoconversion | Using REM EEG features, prediction of phenoconversion with a Harrel's C-index of 0.723 (EEG slowing key predictor). | 55 |
| Cesari M et al., 2023 [35] | Imaging (3D time-of-flight camera) | Detection of RBD | Short movements across head, hands, upper and lower body during REM sleep distinguish iRBD patients from controls with 86.6% accuracy. | 54 |
| Chahine et al., 2025 [36] | Other (electronic health record data) | Detection of RBD | RBD ICD code had a positive predictive value of 73.25% for actual iRBD. | 50 |
| Christensen JAE et al., 2013 [37] | PSG (EOG) | Detection of RBD | EOG-based model distinguished iRBD and PD patients from controls with 90% accuracy, 95% sensitivity, and 80% specificity. | 45 |
| Christensen JAE et al., 2014 [38] | PSG (EEG, EOG) | Detection of RBD | Unsupervised topic modelling classified PD and iRBD from controls with 91.4% sensitivity and 68.8% specificity (unstable N3 and REM sleep key predictors). | 52 |
| Cochen De Cock V et al., 2022 [39] | Motor (gait) | Detection of RBD | Gait-based model that distinguished iRBD patients from healthy controls and PD patients with 95% accuracy, 100% sensitivity, and 90% specificity. | 47 |
| Cooray N et al., 2019 [40] | PSG (EEG, EOG, EMG) | Detection of RBD | Detection of RBD with 92% accuracy, even when incorporating automatic REM sleep detection. | 54 |
| Cooray N et al., 2021 [41] | PSG (ECG, EOG, EMG) | Detection of RBD | One EOG and one EMG channels achieved 90% accuracy in detection of RBD. | 60 |
| Cooray N et al., 2018 [42] | PSG (EMG) | Detection of RBD | EMG combined with sleep architecture achieved 88% accuracy, 91% sensitivity, and 86% specificity for RBD detection. | 49 |
| Feng S et al., 2023 [43] | Imaging (PET) | Prediction of phenoconversion | Dopaminergic damage patterns predicted conversion timing with high accuracy (AUC up to 0.94). | 53 |
| Fereshtehnejad S-M et al., 2015 [44] | Motor, autonomic, olfaction, ophtalmic, cognitive, other (psychiatric, sleep). | Phenotyping | Three PD subtypes identified. PD patients with the diffuse/malignant subtype (mild cognitive impairment, orthostatic hypotension and RBD) experienced faster progression. | 53 |
| Feuerstein S et al., 2024 [45] | PSG (EEG) | Detection of RBD | DL applied to hypnodensity graphs identify iRBD with F1-score of 0.784. | 53 |
| Gunter KM et al., 2023 [46] | PSG (EEG, EMG, EOG) | Detection of RBD | Spectral vision transformer applied to PSG scalograms detected RBD with F1 score of 0.87 (up to 0.93 using EEG and EOG). | 67 |
| Hällqvist J et al., 2024 [47] | Biofluids (Plasma, serum) | Detection of RBD | A blood-based multiplex mass identified PD patients and 79% of pre-motor iRBD individuals up to 7 years before motor symptoms onset. | 53 |
| Hansen IH et al., 2013 [48] | PSG (EEG) | Detection of RBD | Detection of RBD from PSG EEG with up to 90% sensitivity and specificity. | 46 |
| Huang B et al., 2023 [49] | Biofluids (stool) | Detection of RBD | Gut microbiota detects RBD with AUC of 0.75. | 57 |

(*continued on next page*)

**Table 1** (*continued*)

| Study | Input data | Aim of ML/DL model | Main findings | Overall score |
|---|---|---|---|---|
| Illner V et al., 2024 [50] | Voice | Detection of RBD | Analysis of just 18 min of smartphone-recorded speech can screen for iRBD with AUC of 0.85. | 56 |
| Jeancolas L et al., 2022 [51] | Voice | Detection of RBD | Automated voice analysis detected early Parkinson's with 89% accuracy in males and 70% in females, and identified iRBD with up to 70% accuracy. | 47 |
| Jeong E et al., 2024 [52] | Electrophysiological signals (Resting state EEG) | Prediction of phenoconversion | Baseline EEG features predicted phenoconversion in iRBD with 0.775 concordance for time and 0.901 AUC for subtype (EEG slowing key predictor). | 64 |
| Jiang et al., 2020 [53] | Biofluids (serum) | Detection of RBD | RBD detected with AUC of 0.81 when considering exosomal α-synuclein combined with clusterin. | 55 |
| Kempfner J et al., 2013 [54] | PSG (EMG) | Detection of RBD | High REM muscle activity in iRBD was accurately detected using outlier detection, leading to 100% accuracy for RBD detection. | 47 |
| Kempfner J et al., 2010 [55] | PSG (EMG) | Detection of RBD | Detection of RBD in patients with PD with 100% sensitivity and specificity. | 46 |
| Kempfner J et al., 2014 [56] | PSG (EMG, EOG) | Detection of RBD | Using EMG and EOG data, iRBD patients detected with an AUC of 0.993. EOG signals were more effective than EMG. | 59 |
| Kempfner J et al., 2014 [57] | PSG (EEG, EOG, EMG) | Detection of RBD | Using EEG and EOG to identify REM sleep and EMG to measure muscle activity as outliers, patients with iRBD identified with 100% accuracy. | 56 |
| Kim H et al., 2023 [58] | Electrophysiological signals (Event-related potential EEG) | Detection of RBD | 3D spatiotemporal brain activity during attention task distinguished iRBD from controls with 99.8% accuracy. | 44 |
| Kim H et al., 2023 [59] | Electrophysiological signals (Event-related potential EEG) | Detection of RBD | Brain activity patterns describing impaired visuospatial attention detected iRBD with accuracy of 90.9%. | 46 |
| Koch H et al., 2013 [60] | PSG (EEG) | Detection of RBD | Data-driven sleep EEG model distinguished iRBD and Parkinson's patients from controls, achieving 95% sensitivity and 80% specificity. | 43 |
| Laguna A et al., 2021 [61] | Biofluids (serum) | Prediction of phenoconversion | Nuclear magnetic resonance spectroscopy (for analysis of serum lipoprotein and glycosylation profiles) identified phenoconverters with AUC of 0.759. | 56 |
| Lee DA et al., 2022 [62] | Imaging (Diffusion tensor MRI) | Detection of RBD | DTI data identified iRBD patients with accuracy 0.875. Conventional DTI measures outperformed structural connectomic profiles. | 52 |
| Levendowski DJ et al., 2024 [63], | PSG (EEG, EMG) | Other (detection of neurodegeneration) | Age and sleep biomarkers identified RBD and other neurodegenerative diseases with 75% accuracy, showing strong test-retest variability. | 41 |
| Li Y et al., 2024 [64] | Biofluids (plasma) | Detection of RBD | Next-generation sequencing of extracellular vesicle-derived microRNAs distinguished iRBD from controls (AUC of 0.969) and from PD (AUC of 0.929). | 47 |
| Lo C et al., 2021 [65] | Olfaction | Detection of RBD | 3-item Sniffin' Stick odor test combined with an RBD questionnaire identified iRBD patients from controls with 65% sensitivity and 100% specificity. | 65 |
| Mahlknecht P et al., 2016 [66] | Olfaction | Phenotyping | 8-item Sniffin' Stick odor test detects PD with AUC of 0.95. Incident PD cases among patients with iRBD predicted with 100% sensitivity. | 55 |
| Matsushima T et al., 2024 [67] | Imaging (MRI) | Detection of RBD | Resting state functional connectivity distinguished iRBD with accuracy of 64.9%. Important connections involved motor, somatosensory, parietal, temporal cortices, thalamus, and cerebellum. | 58 |
| Mattioli P et al., 2023 [68] | Imaging (PET) | Prediction of phenoconversion | Identification of a brain glucose metabolism pattern in iRBD patients that predicted phenoconversion with 87% sensitivity, 72% specificity. | 64 |
| Mombelli S et al., 2023 [69] | Cognition | Phenotyping | Identification of three distinct neuropsychological phenotypes in iRBD patients. | 49 |
| Nishiwaki H et al., 2020 [70] | Biofluids (stool) | Phenotyping | Unsupervised clustering of gut microbiota revealed four enterotypes in controls, iRBD, and PD. | 58 |
| Ophey A et al., 2025 [71] | Actigraphy | Phenotyping | Clustering analysis revealed that individuals with iRBD, higher physical activity and more stable circadian rhythms were associated with fewer non-motor symptoms and better cognitive performance. | 55 |
| Orso B et al., 2024 [72] | Imaging (PET) | Prediction of phenoconversion | Baseline brain PET distinguished iRBD who later phenoconverted with hazard ratio of 3.52. | 58 |
| Park K et al., 2024 [73], | Electrophysiological signals (Resting state EEG) | Prediction of phenoconversion | EEG spectro-spatial patterns in iRBD patients predicted phenoconversion with AUC up to 0.93. | 56 |
| Possti D et al., 2024 [74] | PSG (EMG) | Detection of RBD | In-home detection of RBD with portable system and semi-automatic EMG analysis (83% sensitivity and 79% specificity). | 42 |
| Pyatigorskaya N et al., 2017 [75] | Imaging (MRI) | Detection of RBD | iRBD exhibited significant substantia nigra damage, yielding to accuracy of 92% to distinguish them from controls. | 48 |
| Raschellà F et al., 2023 [76] | Actigraphy | Detection of RBD | Wrist actigraphy detected RBD in PD patients, achieving up to 100% accuracy in 2-week home recordings. | 49 |
| Rechichi I et al., 2021 [77] | PSG (EMG) | Detection of RBD | Automated method to detect REM sleep without atonia, achieving 87% accuracy to detect RBD. | 68 |
| Rusz J et al., 2016 [78] | Voice | Detection of RBD | Combination of four speech measures achieved 96% sensitivity and 79% specificity in distinguishing RBD from controls. | 40 |
| Rusz J et al., 2018 [79] | Voice | Detection of RBD | Smartphone-recorded speech distinguished iRBD from controls with AUC of 0.69. | 47 |
| Ryoo HG et al., 2022 [80] | Imaging (PET) | Phenotyping | Deep learning-based cognitive dysfunction score derived from PET scans effectively distinguished iRBD patients with MCI from those without (AUC = 0.70). | 68 |

(*continued on next page*)

**Table 1** (*continued*)

| Study | Input data | Aim of ML/DL model | Main findings | Overall score |
|---|---|---|---|---|
| Saeda S et al., 2024 [81] | Electrophysiological signals (ECG) | Detection of RBD | Heart rate variability during supine rest that accurately distinguished iRBD patients with orthostatic hypotension from those without (81% accuracy and 100% sensitivity). | 42 |
| Salsone M et al., 2022 [82] | PSG (ECG) | Detection of RBD | 24-h heart rate variability identified patients with iRBD with up to 94% accuracy, 95% sensitivity, and 92% specificity. | 54 |
| Salsone M et al., 2024 [83] | PSG (ECG) | Phenotyping | Distinction of patients with iRBD with and without PLMS using heart rate variability data (86% accuracy, 96% sensitivity, and 74% specificity) | 54 |
| Seger A et al., 2023 [84] | Other (sleep questionnaires) | Detection of RBD | Questionnaire-based algorithm identifying iRBD with over 80% accuracy, reducing unnecessary PSGs by nearly 70% while detecting 81% of iRBD cases. | 51 |
| Shin Y-W et al., 2025 [85] | Motor, cognition, autonomic, other (demographic, medication, clinical history) | Prediction of phenoconversion | Model to predict time and type of phenoconversion in iRBD with clinical data. Time predicted with concordance index of 0.823 and subtype prediction with Matthews correlation coefficient of 0.697. | 56 |
| Soto M et al., 2022 [86] | Imaging (DaT-SPECT) | Detection of RBD | Specific serum microRNA biosignatures deregulated across the progression of iRBD disorder and detect iRBD, PD and DLB with AUC of 0.980. | 56 |
| Tripathi S et al., 2023 [87] | Imaging (PET) | Prediction of phenoconversion | Self-supervised learning approach using a novel hemisphere dissimilarity loss predicts phenoconversion with AUC of 0.720. | 40 |
| Urtnasan E et al., 2021 [88] | PSG (ECG) | Detection of RBD | Deep convolutional neural network applied to ECG detected RBD with 95% F1-score. | 45 |
| Vaish A, 2024 [89] | Voice | Detection of RBD | Detection of RBD from voice recordings with accuracy of 70% (vs controls) and 55% (vs PD patients) | 42 |
| Valappil RA et al., 2010 [90] | PSG (ECG) | Detection of RBD | Heart rate variability in wakefulness discriminated RBD from controls with 95.5% accuracy. | 41 |
| van Veen et al., 2022 [91] | Imaging (PET) | Prediction of phenoconversion | GMLVQ applied to PET predicted phenoconversion, with progression velocity in GMLVQ space correlating significantly with motor symptom changes (Spearman's rho = 0.62). | 49 |
| Yang Y et al., 2021 [92] | Imaging (diffusion weighted MRI) | Detection of RBD | Connectome-wide association analysis of diffusion MRI data detected RBD from controls with accuracy of 83.3%. | 56 |

AUC: area under the curve; DaT: dopamine transporter; DL: deep learning; DLB: dementia with Lewy bodies; DTI: diffusion tensor imaging; ECG: electrocardiography; EEG: electroencephalography; EMG: electromyography; EOG: electrooculography; GMLVQ: Generalized Matrix Learning Vector Quantization; GPe: globus pallidus externa; GPi: globus pallidus interna; HC: healthy controls; ICD: International Classification of Disorders; iRBD: isolated RBD; MCI: mild cognitive impairment; ML: machine learning; MRI: magnetic resonance imaging; PD: Parkinson's disease; PET: positron emission tomography; PLMD: periodic leg movement disorder; PLMS: periodic leg movements in sleep; PSG: polysomnography; RBD: REM sleep behavior disorder; REM: rapid eye movement; RNA: ribonucleic acid; SNc: substantia nigra pars compacta; SPECT: Single-Photon Emission Computed Tomography; VTA: ventral tegmental area.

(maximum of 12 points), and 6) reproducibility (maximum of 20 points). Depending on the overall score, studies were categorized into the following pre-defined categories [16]: very low quality (overall score: 0-19), low quality (overall score: 20-39), moderate quality (overall score: 40-59), high quality (overall score: 60-79), and very high quality (overall score: 80-100). For sake of completeness, the APPRAISE-AI tool is reported in Table S3 in the Supplemental Material.

In order to reduce inter-rater variability, three studies were randomly selected for the three leading authors (ABK, IR, MC) to perform both the extraction of general study characteristics and evaluation according to the APPRAISE-AI tool. This process allowed the authors to detect discrepancies in the interpretation of both general characteristics data extraction and the APPRAISE-AI tool. Initial inter-rater agreement appeared moderate (Fleiss' kappa: 0.43, average pairwise Cohen's kappa: 0.49); the discrepancies were discussed and clarified in a structured calibration round to harmonize interpretation criteria. Following this consultation, agreement was reached and improved substantially (Fleiss' kappa: 0.89, average pairwise Cohen's kappa: 0.94). After the calibration round each remaining study was appraised by only one of the three leading authors. None of the leading authors assessed the reporting quality of their own studies, thus ensuring fairness in the evaluation.

### 2.6. Data synthesis

#### 2.6.1. General study characteristics

To determine whether there was an increase in the number of publications over the years, we calculated the distribution studies over time. With regard to countries, we calculated the frequency and relative percentage of each of them. To facilitate the description of the study populations size, we defined the following categories: patients with iRBD, patients with RBD (not otherwise specified whether primary or secondary) patients with PD and RBD, patients with PD (either without RBD or unspecified), patients with other synucleinopathies or neurodegenerative disease (either with RBD, without RBD or unspecified), and controls (including both healthy individuals and those with other disorders, except from neurodegenerative diseases). For each study, the sample size of each group was categorized into the following bins: 1-20, 21-50, 51-100, >100. The frequency and respective percentage of the group sizes across all studies were then obtained.

Regarding ML/DL models, the following information was synthesized: aims, types of algorithms/architectures and input data used by the models. To facilitate the description of the aims, the following aim categories were defined: detection of RBD, prediction of phenoconversion, phenotyping, or other. The frequency and relative percentage of each category was calculated. Additionally, we reported the frequency and percentage of different ML/DL algorithms and architectures employed in the studies. Finally, the input data were grouped according to the following classes: data from PSG (with further specification on the biosignals analysed), actigraphy, imaging data (with further specification on the imaging modality), other electrophysiological signals other than PSG, other biomarkers (with further subcategorization into motor, cognition, olfaction, autonomic, ophthalmic, and biofluids), and others. The frequency and relative percentages of each class across the studies were derived. In this review, we did not aim to provide a data synthesis of the main findings of each study.

#### 2.6.2. Quality of AI reporting

Distribution of the overall APPRAISE-AI score was used to evaluate

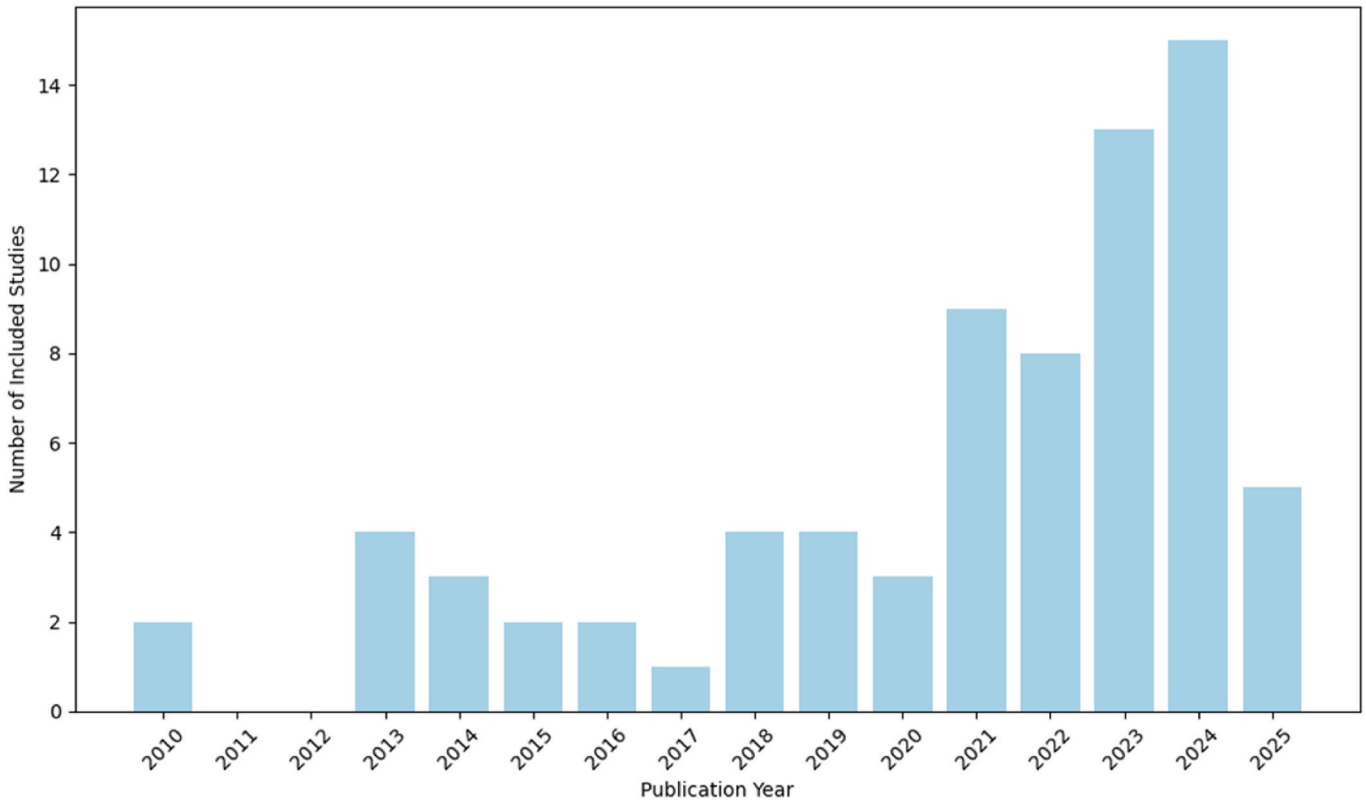


**Fig. 2.** Distribution of number of publications per year.

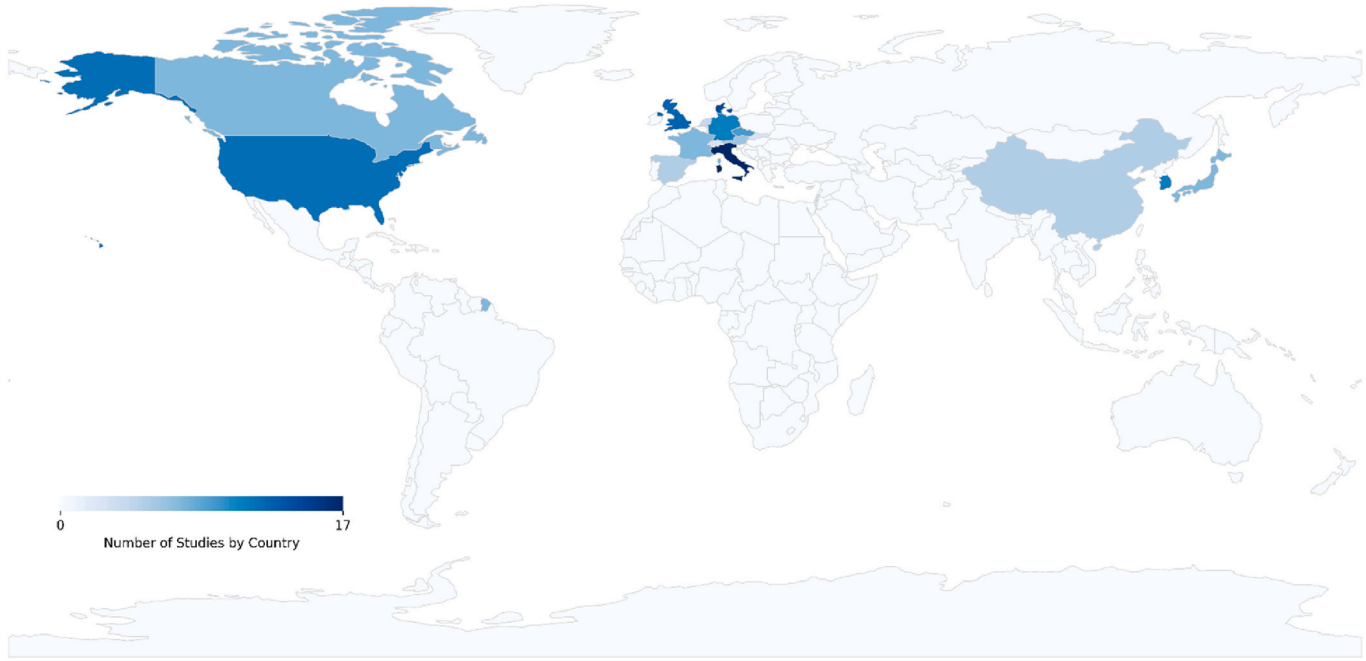


**Fig. 3.** Distribution of population by country in the included studies.

the overall quality of AI reporting for the studies included. To provide a more detailed analysis and to allow comparison of scores across items and domains, we normalized the scores, by calculating the percentage of the maximum possible score for each item and domain. To visualize the distribution across item scores, we used bar plots highlighting the mean with error bars indicating the standard error of the mean.

## 3. Results

The search yielded in 2025 unique records. Of these, 237 were assessed for eligibility, and 75 were ultimately included in this scoping review. Fig. 1 reports the detailed PRISMA-ScR diagram. Table 1 summarizes the studies included, reporting input data, aim of the ML/DL model, main study findings and the overall score according to the

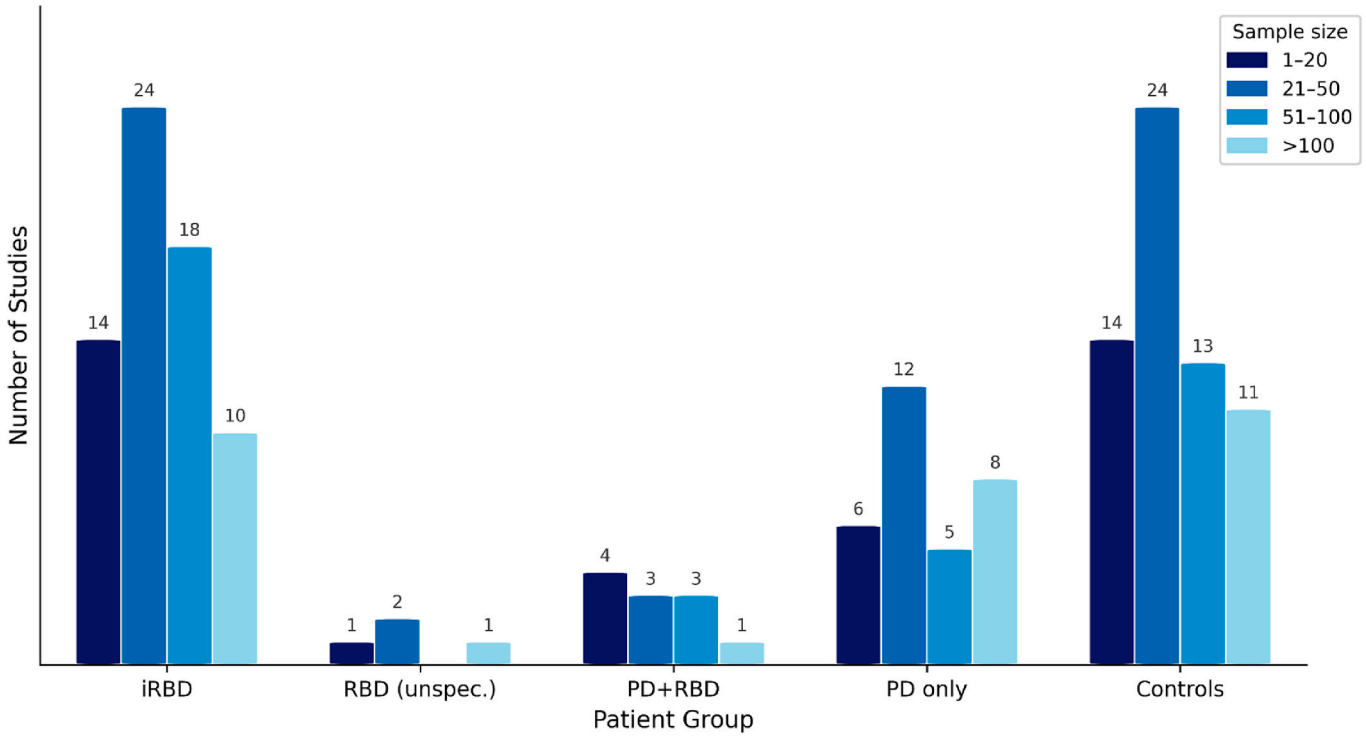


**Fig. 4. Sample size of the populations included in the studies.** The cohorts are grouped into sample size ranges and categorized as iRBD, unspecified RBD, PD with RBD (PD + RBD), and control groups. Other synucleinopathies and neurodegenerative disorders (NDD) were less frequently represented and are therefore not reported in this chart. These included Dementia with Lewy bodies (DLB; n = 1 in the 21–50 range and n = 1 in the 51–100 range), multiple system atrophy (MSA; n = 1 in the 1–20 range), and generalized NDD cohorts (n = 1 in the 21–50 range, n = 3 in the 51–100 range, and n = 1 in the >100 range).

APPRAISE-AI tool. Table S4 in the Supplemental Material reports the PRISMA-ScR checklist. The next sections describe in detail our findings.

### 3.1. General study characteristics

Table S5 in the Supplemental Material reports detailed information on the general study characteristics of each of the included studies. Fig. 2 shows the distribution of the number of studies over the years, which reflects a clear trend towards an increase over time.

The studies included in this review were conducted across a broad international landscape and study participants were recruited from medical centers across multiple countries, with Europe and North America contributing the majority of studies. The countries that were most represented are Italy (22.7%), Denmark (17.3%), the United Kingdom (14.7%), United States of America (13.3%), and Germany/South Korea (12% each). Additional contributions came from Canada, France, Japan, the Czech Republic, and Spain. Fig. 3 displays the distribution of population under study by country.

The sample sizes and composition of patient and control cohorts across the included studies are shown in Fig. 4. As appreciable, sample sizes varied widely, ranging from small-scale observational studies to larger studies, with sample sizes overcoming 100 participants, though these were the minority. As regards cohorts of patients with iRBD, the majority of studies (32%) included a sample size in the range 21–50, followed by 24% of studies reporting cohorts of 51–100 subjects.

Regarding the primary aims of the ML/DL models, we found that the majority of studies (73.3%) focus on RBD detection, highlighting the efforts directed toward the development of automated diagnostic tools. A smaller but growing proportion of studies (16%) addressed the prediction of phenoconversion, often relying on longitudinal and retrospective datasets to identify individuals at higher risk of developing overt alpha-synucleinopathies. Only a few studies (9.3%) explored phenotyping, with a variety of aims, including identification of clinically relevant subgroups. Table S6 in the Supplemental Material summarizes this information.

To address these aims, various ML and DL algorithms and architectures were adopted across the studies, reflecting choices based on dataset size and study objectives. Traditional (or "classical") ML approaches (logistic regression, random forest, and support vector machines) were adopted in more than half of the studies, often for their relative interpretability and the adoption of hand-engineered features, which was considered a key factor for clinical applications. DL approaches – such as convolutional or recurrent neural networks, and long short-term memory networks – were adopted in 12% of studies. A subset of studies (8%) employed survival models, particularly in the context of predicting phenoconversion. Ensemble or hybrid models were also explored, often with the aim of improving classification robustness. Out of the algorithms employed in the included studies, only a minority (11.9%) were unsupervised, while the remaining were supervised. A summary of the algorithms and architectures employed in the included studies is displayed in Supplementary Table S7.

We also synthesized data related to the input data types across the included studies. Given the multimodal nature of sleep and neurological assessments, input data varied widely across studies, as displayed in Fig. 5. The most frequently used data sources were derived from PSG (36%), with electromyography being the most common (21.3%), followed by EEG (16%), electrooculography (12%), and electrocardiography (ECG) (8.0%). Imaging data were also often employed (21.3%), with positron emission tomography being the most widely used modality (8%), followed by magnetic resonance imaging (6.7%). Additional electrophysiological measures, such as resting-state EEG, event-related potentials and heart rate-variability derived from ECG, were also adopted in 16% of the studies. A portion of studies employed actigraphy (5.3%), and voice recordings (9.3%, in some cases recorded directly with smartphones), reflecting the applicability of wearable and portable technologies. In addition, 21.3% of the studies focused on other biomarkers of neurodegeneration. Of these, motor markers were the most investigated ones (9.3%), followed by biofluids (8%), cognition

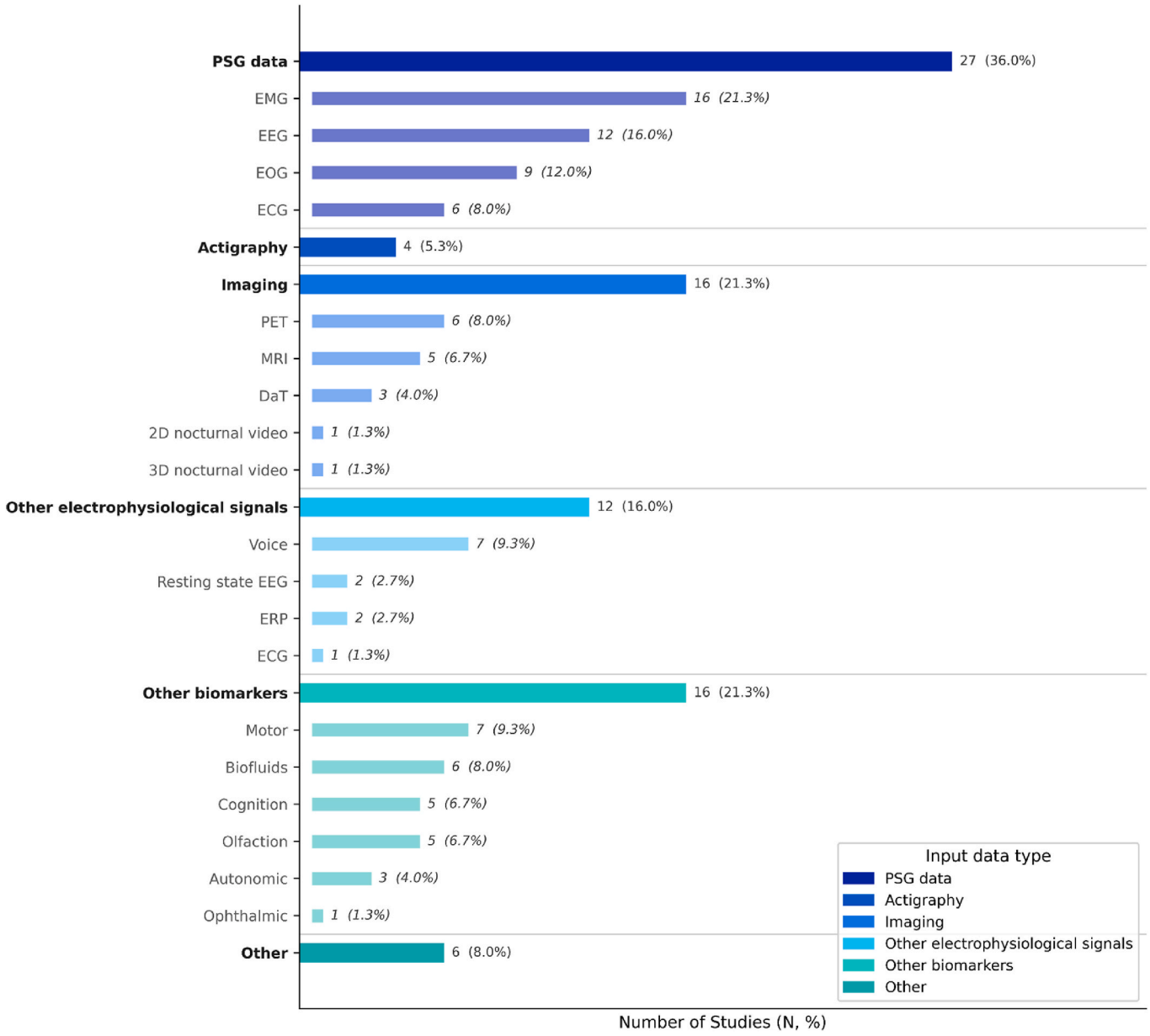


**Fig. 5. Distribution of input data types in the included studies.** Please note that multiple data types are possible for each included study. For reference: DaT: Dopamine transporter; ECG: electrocardiography; EEG: electroencephalography; EHR: electronic health record; EMG: electromyography; EOG: electrooculography; ERP: event-related potential; HRV: heart rate variability; MRI: magnetic resonance imaging; PD: Parkinson's disease; PET: positron emission tomography; PSG: polysomnography.

(6.7%) and olfaction (6.7%).

Out of the studies employing PSG data as input, most of them (88.9%) investigated the problem of automatic RBD detection. The performance of the proposed ML/DL algorithms are promising, with accuracy often reaching over 80%. Sample sizes, however, were generally relatively small, with only 22.2% of these studies including iRBD cohorts with more than 50 patients and 7.4% of studies including more than 50 PD patients with RBD. Imaging studies, on the other hand, had a wider variety of aims: in half of them, ML/DL algorithms were employed for detecting RBD, while prediction of phenoconversion was the aim in 37.5% of the studies. Their performance for RBD detection was similar to the models using PSG data (accuracy >80%). Performance for prediction of phenoconversion could not be easily compared between the studies, due to the variety of performance measures employed. Interestingly, ML/DL models using imaging data as input had generally larger cohorts, with 62.5% of the studies having iRBD cohorts with over 50 patients. Most of the studies analyzing actigraphy and voice (90.9%) aimed to detect RBD, with good performance metrics (accuracy >70%), but the cohorts were generally small (only 18.2% of the studies included over 50 patients with iRBD). Studies that used motor, cognition, olfaction, autonomic, ophthalmic, and biofluids biomarkers as input data of ML/DL models mainly aimed to detect RBD (57.1%). Nevertheless, these biomarkers were also employed in 28.6% of these studies to investigate new patient phenotypes. Sample sizes were generally large, with 85.7% of the studies having iRBD cohorts with over 50 patients.

### 3.2. *Quality of AI reporting*

Table S8 in the Supplemental Material reports the scores of the APPRAISE-AI tool for each item. Of the studies included, 80 % have a moderate overall score and 20 % have a high overall score (Table 1).

A more detailed overview of the quality of AI reporting as assessed

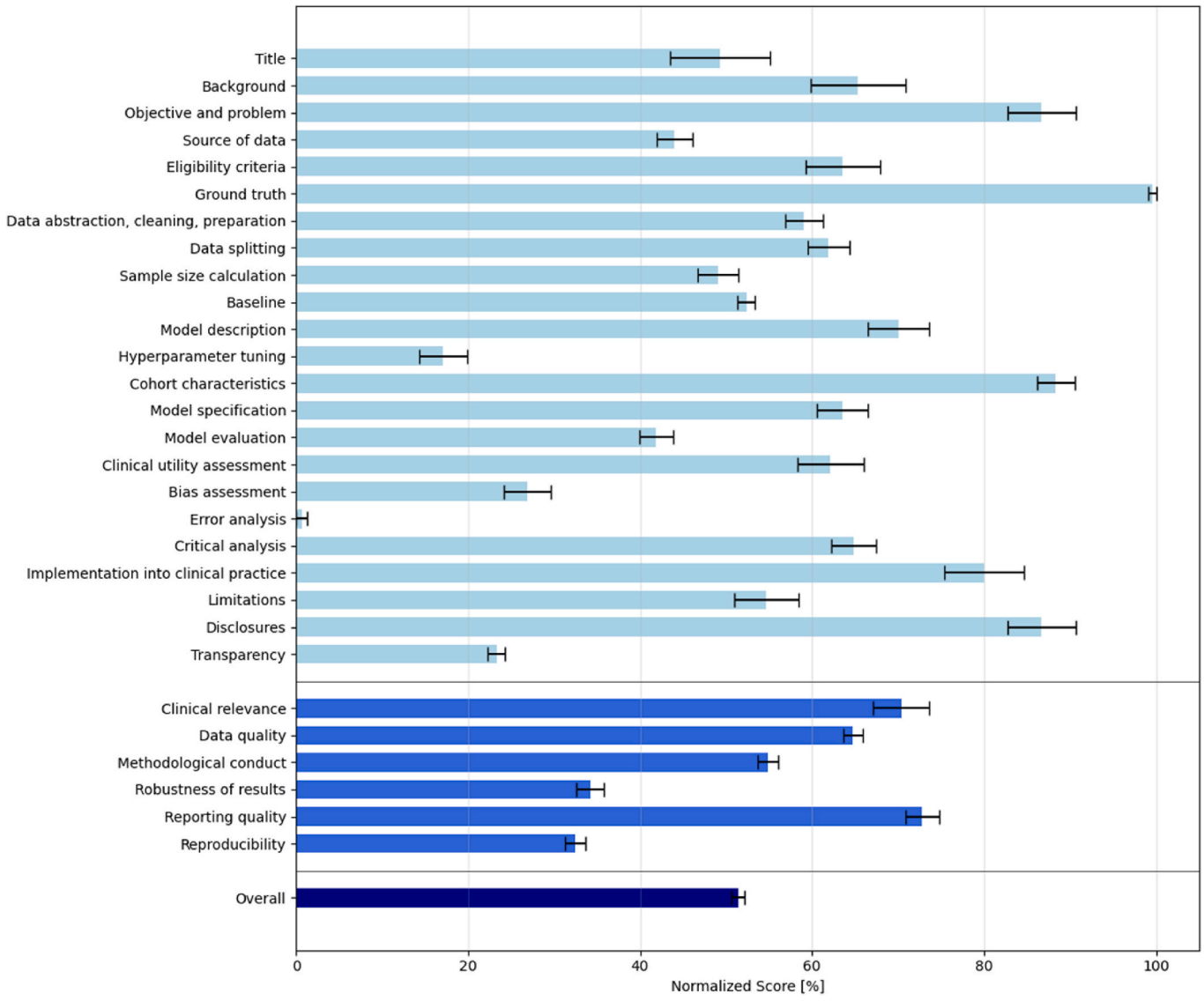


**Fig. 6. Normalized score of the APPRAISE-AI items, domains, and overall score**. Each bar represent the means across the included studies and error bars indicate the standard error of the mean.

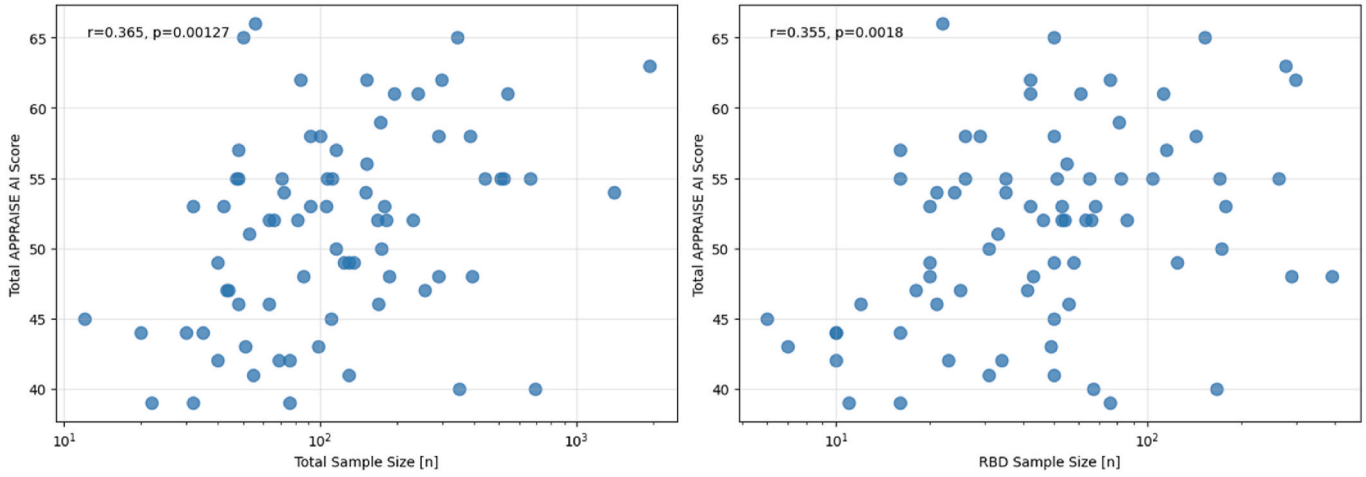


**Fig. 7. Association between reported sample size and total APPRAISE-AI total score.** Correlation was computed as the Spearman's correlation coefficient and visualized using a log-scale for the sample size.

using the APPRAISE-AI is visualized in Fig. 6, which displays normalized scores of each item, domain, and overall score. The total APPRAISE-AI score was correlated with the reported sample size ($r = 0.37$, $p = 0.0013$) as visualized in Fig. 7. The domains with the highest normalized scores are reporting quality (72.8 ± 2.0%) and clinical relevance (70.3 ± 3.2%). The single items with the highest scores were the definition of the ground truth (99.6 ± 0.4%), the description of the cohort characteristics (88.3 ± 2.2%), and considerations about implementation in clinical practice (80.0 ± 4.6%). The domains with the lowest average normalized values were related to the robustness of the results (34.2 ± 1.6%) and the reproducibility (32.5 ± 1.2%). The single items with the lowest average normalized scores were error analysis (0.66 ± 0.66%), hyperparameter tuning (17.1 ± 2.8%), transparency (23.3 ± 1.0%), bias assessment (26.9 ± 2.8%), and model evaluation (41.9 ± 2.0%). One noteworthy finding of the APPRAISE-AI assessment was the frequency of data leakage (a subquestion of the data splitting item), which we found to be present in 32 % of the included studies. This finding was present across ML modeling types ranging from logistic regression, through Random Forest, to deep learning methods (see Fig. S1).

## 4. Discussion

In this scoping review, we provided an overview of studies where ML/DL models were applied to the field of RBD research, whether for detection, phenotyping or prognostication. Seventy-five studies were included, showing a tendency of increase in the past years. Studies originated mainly from Europe and North America and included mostly cohorts of medium size (i.e., up to 100 subjects). Traditional ML algorithms were primarily used, with PSG serving as the primary input modality for most models, and detection of RBD being the primary aim. Concerning methodological and reporting quality evaluated with the APPRAISE-AI tool, most of the studies showed overall moderate quality. The overall quality of reporting was correlated to the study's sample size, thereby, pointing to a higher quality for larger studies. Generally, the studies showed high scores in domains of reporting quality and clinical relevance, but lower scores on robustness of results and reproducibility.

The increase of ML/DL publications in the field of RBD research mirrors what is seen in other fields of medicine [93], highlighting the increasing use of AI in clinical research across specialties. Noteworthy, most of the research carried out on RBD originates from Europe and North America, which likely introduce a regional bias and does not guarantee generalization to other world regions. Most of the included works employed classical ML algorithms and only a minority investigated DL approaches. This is probably due to the small-middle sample sizes of the cohorts that were included in the studies, as DL architectures generally need larger numbers for efficient training.

Interestingly, most of the included studies focused on RBD detection and employed PSG data as input. This reflects the need for faster and more objective tools to support physicians in the diagnostic process in the sleep lab. The classification performance of these algorithms are consistently over 80% in accuracy, despite a variety of models, biosignals and features being employed. Notably none of these algorithms was tested on a general sleep lab population, reflecting true RBD prevalence [94]. Although assessed only in a limited number of studies, ML/DL algorithms show potential when applied to actigraphy and voice signals for RBD screening, though no study on the general population has been carried out so far. Interestingly, no detection study investigated multimodal integration of different objective modalities. A minority of ML/DL models aimed at predicting phenoconversion in patients with iRBD, with most of them employing imaging data as input. Whether imaging data represents the optimal prognostic biomarker remains uncertain, due to the scarcity of comparative studies across different modalities. It is noteworthy that imaging studies were generally large and multicentric, indicating the promise of this modality combined with ML/DL models for prognosis. Only few studies focused on identification of phenotypes of RBD and this is reflected by the small number of studies using unsupervised learning. Finally, it is worth noticing that none of the AI algorithms proposed in the studies included in this review is currently available for clinical applications.

Regarding AI methodological and reporting quality, high scores were found in the domains of clinical relevance and reporting quality (which covers the items of description of cohort characteristics, critical analysis, limitations and disclosures). This indicates high attention of the studies' authors towards good description of the included cohort and the possible implementation of the proposed models in clinical practice. The item "ground truth" had very high scores over all the studies due to the inclusion criteria (only patients with PSG confirmed RBD were included). In this context, we would like to highlight that, although we applied strict inclusion criteria, this does not ensure the homogeneity of populations included in the studies reviewed. This is particularly evident regarding the demonstration of RWA, for which various qualitative and quantitative approaches were used across studies. Similarly, although we did not systematically investigate this aspect, the definition of control populations is likely highly heterogeneous across the studies, limiting the generalizability of the proposed ML/DL algorithms.

One pervasive limitation across studies was the lack of transparency in both data and model availability. Few studies made their code or pretrained models publicly accessible, and details on hyperparameter tuning strategies were often missing or insufficiently described. This lack of transparency hinders replication and makes it difficult for other research groups to build upon existing work or benchmark new approaches. Without open access to source code and datasets, the development of cumulative knowledge in the field is severely constrained.

Another critical shortcoming was the near absence of rigorous error analysis and bias assessments. Only a single study (1.33 %) performed any error analyses but only received partial score resulting in an overall mean of 0.66 % in this category. Moreover, just 26.9% of studies reported any form of bias assessment. This omission limits insight into when and why models fail, which is essential for improving model generalizability and clinical safety. Similarly, only a minority of studies reported model calibration, which is especially relevant in clinical settings where the probability of an event (e.g., phenoconversion) should be interpretable and actionable. Without these evaluations, progress is stifled, leaving future work without a solid basis for refining methodologies and advancing the field.

The most concerning finding was the high prevalence of data leakage, identified in 32% of studies. This effect was present both when studies relied on simple ML models such as logistic regression and for more advanced ML and DL models. Data leakage undermines the credibility of performance metrics and invalidates conclusions about model bias or robustness. The most common form of leakage involved feature selection performed across the entire dataset, including the test set. We also observed several cases where hyperparameter tuning was based on test set performance, and others where data from the same subjects were included in both training and testing sets. These practices artificially inflate reported accuracy and can create a false sense of confidence in model reliability. In the context of detecting iRBD and predicting phenoconversion, such distortions are particularly problematic, as they obscure which algorithms or input modalities truly offer clinical utility.

Based on the findings presented in our review, we propose several recommendations to improve the quality, transparency, and impact of future AI research in the field of RBD:

## Practice points

- Over the last years, there was an increasing number of applications of ML/DL in the field of RBD research, though cohort sizes were generally relatively small.
- Most available studies focus on detecting RBD using classical ML approaches with PSG data, while prognostic applications of ML/DL models rely primarily on imaging data. The use of ML/DL algorithms to analyze actigraphy and voice recordings shows promise as a non-invasive screening approach.
- According to the APPRAISE-AI tool, 80% of the study showed overall moderate methodological and reporting quality.
- Studies had higher scores in the domains of quality of reporting and clinical relevance, but lower scores on reproducibility and robustness of results.

- Open science practices should be adopted more widely, including detailed documentation of model development processes, sharing of source code, model weights, and, when possible, datasets. Large consortia or societies such as the International RBD Study Group [95] should consider collecting multi-centric benchmark datasets where ML/DL algorithms can be tested. In the process of acquiring these datasets, efforts should be made to minimize regional bias by including cohorts from regions beyond Europe and North America.
- Future detection studies should consider investigating populations that reflect the true prevalence of RBD, allowing a correct estimation of their sensitivity and specificity. This is particularly important for studies that will evaluate novel wearable, nearable and airable technologies.
- The potential of ML/DL in the field of RBD could be further enhanced through the integration of multimodal data, such as PSG and neuroimaging. This approach has the potential not only to improve detection accuracy but also to enable more precise phenotyping and longitudinal assessment of neurodegenerative progression.
- The RBD research community should prioritize harmonizing and standardizing study cohorts. For research purposes, we encourage the use of existing open-source, automated RWA detection algorithms [96,97] to ensure consistent, objective quantification of RWA across studies. For clinical diagnostics, we strongly recommend that the RBD community and international sleep societies reach consensus on standardized measurement protocols for RWA, replacing the current variety of approaches so that future AI algorithms can be developed in conformity with them. In addition, the definition and selection of control cohorts should be improved and standardized to enhance comparability and reproducibility across studies.
- Studies must employ rigorous validation procedures that explicitly avoid data leakage, with clearly defined and separated training, validation, and test sets. For this purpose, we refer to published recommendations [98,99].
- Systematic error analysis, bias assessment, and model calibration should become standard components of ML/DL research in RBD. Authors should clearly report a comprehensive set of validation/test performance metrics, with explicit presentation of the trade-offs between sensitivity and specificity of the proposed solution. These practices are essential to identify model limitations, guide methodological innovation, and build trustworthy systems that can eventually support clinical trials and decision-making. We believe that editorial boards of clinical sleep medicine journals should consider implementing rigorous procedures to ensure improved quality of AI methodology and reporting, also enforcing availability of datasets. For example, authors should be encouraged to follow the APPRAISE-AI [16] or similar tools [15]. Including AI specialists on editorial boards may help ensure higher standards of methodological and reporting quality in AI research.

We believe that adherence to these recommendations will enable future RBD research to develop AI algorithms with improved robustness and quality, ultimately facilitating their integration into routine clinical practice.

This scoping review has some limitations. First, we explored only two databases for retrieval of studies, but we believe that the broad search terms we included prevented us from missing significant studies. Second, we employed the APPRAISE-AI tool to evaluate the quality of reporting of the studies, as most of them were designed for clinical decision support. However, the APPRAISE-AI tool is mainly designed to evaluate quality of reporting for classical supervised ML algorithms. We encountered some difficulties to answer some of the items for studies using DL or proposing unsupervised architectures. Each issue was solved through a consensus procedure between the leading authors. Furthermore, the evaluation of some items of the APPRAISE-AI tool is prone to subjective interpretation, leading to possible biases.

## 5. Conclusions

This scoping review highlights both the growing interest and the significant methodological gaps in the application of ML and DL for RBD. Most studies applied ML/DL models to PSG data for RBD detection. Prognostic uses of MD/DL models primarily involved imaging data, while their application to actigraphy and voice recordings suggests promising potential for screening. Overall methodological and reporting quality - as assessed by the APPRAISE-AI tool - was only moderate for

## Research agenda

- Investigate ML/DL models in several types of input data derived from multi-centric cohorts.
- Validate the models on larger and homogenized populations that reflect the true prevalence of RBD.
- Increase the quality of reporting of AI research in RBD, with a particular focus on enhancing transparency and reproducibility.
- Require authors to utilize tools such as APPRAISE-AI as a condition for submitting manuscripts with ML/DL analyses to clinical sleep medicine journals.

most studies, with key deficiencies in transparency, reproducibility, bias assessment, and error analysis. Notably, data leakage was present in nearly a third of studies, undermining model credibility and translational potential.

These limitations restrict the clinical applicability of current models and hinder collaborative progress in the field. Without transparency in data and code availability, it is nearly impossible for research groups to replicate findings or incrementally improve existing models. To move toward robust, generalizable, and clinically useful AI tools, future studies must adopt open science practices, implement rigorous validation protocols, and follow structured reporting guidelines such as APPRAISE-AI. Furthermore, future studies should validate detection models in larger populations that reflect the true prevalence of RBD, integrate data collected with various modalities, use harmonized protocols for cohort definitions and investigate further the role of AI in prognostication.

Raising the standards of AI research in RBD will require coordinated efforts from the research community, editorial boards, and clinical societies. Such efforts are essential to ensure that promising innovations can mature into reliable tools for diagnosis, risk prediction, and personalized clinical care.

## Authors' contributions

Conceptualization: A. Brink-Kjaer, I. Rechichi, M. Cesari; Methodology: A. Brink-Kjaer, I. Rechichi, M. Cesari; Formal analysis: A. Brink-Kjaer, I. Rechichi, M. Cesari; Investigation: A. Brink-Kjaer, I. Rechichi, D. Arnaldi, O. Cesarone, E. During, S. Feuerstein, K.M. Gunter, B. Högl, A. Ibrahim, P. Jennum, E. Mignot, G. Olmo, M. Roascio, A. Stefani, Q. Tang, M. Cesari; Data curation: A. Brink-Kjaer, I. Rechichi, M. Cesari; Writing – original draft: A. Brink-Kjaer, I. Rechichi, M. Cesari; Writing – review & editing: D. Arnaldi, O. Cesarone, E. During, S. Feuerstein, K.M. Gunter, B. Högl, A. Ibrahim, P. Jennum, E. Mignot, G. Olmo, M. Roascio, A. Stefani, Q. Tang; Visualization: A. Brink-Kjaer, I. Rechichi, M. Cesari; Supervision: M. Cesari; Project administration: M. Cesari.

## Conflict of interest

E. Mignot has received consulting, speaker, board engagements and travel funds from Avadel, Harmony, Idorsia, Jazz Pharmaceuticals, Paladin Labs and Takeda; and has stock options in Centessa. All remaining authors do not report any conflict of interest related to this work.

## Acknowledgements

E. Mignot's work on REM sleep behavior disorder is funded by U19 AG071754.

M. Roascio is currently supported by the project HubLife Science – Digital Health (LSH-DH) PNC-E3-2022-23683267 - Progetto DHEALCOM – CUP:D33C22001980001, founded by Ministero della Salute within "Piano Nazionale Complementare al PNRR Ecosistema Innovativo della Salute - Codice univoco investimento: PNC-E.3".

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.smrv.2026.102299.

## References

[1] EU AI act; 2024. [URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689#art_3] Accessed 30.04.2026.

[2] Zan X, Liu F, Xian X, Pardalos PM. Empowering sleep health: unleashing the potential of artificial intelligence and data science in sleep disorders. In: Berry RB, Pardalos PM, Xian X, editors. Handbook of AI and data sciences for sleep disorders. Cham: Springer Nature Switzerland; 2024. p. 1–44. https://doi.org/10.1007/978-3-031-68263-6_1.

[3] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022;28:31–8. https://doi.org/10.1038/s41591-021-01614-0.

[4] Goldstein CA, Berry RB, Kent DT, Kristo DA, Seixas AA, Redline S, et al. Artificial intelligence in sleep medicine: an American academy of sleep medicine position statement. J Clin Sleep Med 2020;16:605–7. https://doi.org/10.5664/jcsm.8288.

[5] Bandyopadhyay A, Oks M, Sun H, Prasad B, Rusk S, Jefferson F, et al. Strengths, weaknesses, opportunities, and threats of using AI-enabled technology in sleep medicine: a commentary. J Clin Sleep Med 2024;20:1183–91. https://doi.org/10.5664/jcsm.11132.

[6] American Academy of Sleep Medicine. The international classification of sleep disorders (ICSD-3). third ed. Darien, IL: American Academy of Sleep Medicine; 2014.

[7] Cesari M, Heidbreder A, St Louis EK, Sixel-Döring F, Bliwise DL, Baldelli L, 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 2022;45. https://doi.org/10.1093/sleep/zsab257. zsab257.

[8] Högl B, Stefani A, Videnovic A. Idiopathic REM sleep behaviour disorder and neurodegeneration - an update. Nat Rev Neurol 2018;14:40–55. https://doi.org/10.1038/nrneurol.2017.157.

[9] Miglis MG, Adler CH, Antelmi E, Arnaldi D, Baldelli L, Boeve BF, et al. Biomarkers of conversion to α-synucleinopathy in isolated rapid-eye-movement sleep behaviour disorder. Lancet Neurol 2021;20:671–84. https://doi.org/10.1016/S1474-4422(21)00176-9.

[10] Cesari M, Brink-Kjaer A, During E, Ganglberger W, Gnarra O, Huang B, et al. Isolated REM sleep behaviour Disorder-Is screening possible? J Sleep Res 2025: e70109. https://doi.org/10.1111/jsr.70109.

[11] Kapoor S, Cantrell EM, Peng K, Pham TH, Bail CA, Gundersen OE, et al. REFORMS: consensus-Based recommendations for machine-learning-based science. Sci Adv 2024;10. https://doi.org/10.1126/sciadv.adk3452. eadk3452.

[12] Krepper D, Cesari M, Hubel NJ, Zelger P, Sztankay MJ. Machine learning models including patient-reported outcome data in oncology: a systematic literature review and analysis of their reporting quality. J Patient-Rep Outcomes 2024;8:126. https://doi.org/10.1186/s41687-024-00808-7.

[13] Dzialas V, Doering E, Eich H, Strafella AP, Vaillancourt DE, Simonyan K, et al. Houston, we have AI problem! quality issues with neuroimaging-based artificial intelligence in parkinson's disease: a systematic review. Mov Disord 2024;39: 2130–43. https://doi.org/10.1002/mds.30002.

[14] Khondker A, Kwong JCC, Rickard M, Erdman L, Kim JK, Ahmad I, et al. Application of STREAM-URO and APPRAISE-AI reporting standards for artificial intelligence studies in pediatric urology: a case example with pediatric hydronephrosis. J Pediatr Urol 2024;20:455–67. https://doi.org/10.1016/j.jpurol.2024.01.020.

[15] Kolbinger FR, Veldhuizen GP, Zhu J, Truhn D, Kather JN. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. Commun Med 2024;4:1–10. https://doi.org/10.1038/s43856-024-00492-0.

[16] Kwong JCC, Khondker A, Lajkosz K, McDermott MBA, Frigola XB, McCradden MD, et al. APPRAISE-AI tool for quantitative evaluation of AI studies for clinical decision support. JAMA Netw Open 2023;6:e2335377. https://doi.org/10.1001/jamanetworkopen.2023.35377.

[17] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169:467–73. https://doi.org/10.7326/M18-0850.

[18] 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. https://doi.org/10.1002/ana.27170.

[19] Alushaj E, Kuurstra A, Menon RS, Ganjavi H, Morava A, Sharma M, et al. Midbrain and pallidal iron changes identify patients with REM sleep behaviour disorder and Parkinson's disease. npj Parkinson's Dis 2025;11:84. https://doi.org/10.1038/s41531-025-00916-1.

[20] 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. https://doi.org/10.1093/brain/awaa365.

[21] Arnaldi D, Mattioli P, Raffa S, Pardini M, Massa F, Iranzo A, et al. Presynaptic dopaminergic imaging characterizes patients with REM sleep behavior disorder due to synucleinopathy. Ann Neurol 2024;95:1178–92. https://doi.org/10.1002/ana.26902.

[22] 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. https://doi.org/10.1212/WNL.0000000000006366.

[23] Arora S, Lo C, Hu M, Tsanas A. Smartphone speech testing for symptom assessment in rapid eye movement sleep behavior disorder and parkinson's disease. IEEE Access 2021;9:44813–24. https://doi.org/10.1109/ACCESS.2021.3057715.

[24] 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. https://doi.org/10.1007/s10772-018-09588-0.

[25] Bisgaard S, Duun-Christensen B, Kempfner L, Sorensen HBD, Jennum P. EEG recordings as a source for the detection of IRBD. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:606–9. https://doi.org/10.1109/EMBC.2015.7318435.

[26] Brink-Kjaer A, Gunter KM, Mignot E, During E, Jennum P, Sorensen HBD. End-to-end deep learning of polysomnograms for classification of REM sleep behavior

disorder. Annu Int Conf IEEE Eng Med Biol Soc 2022;2022:2941–4. https://doi.org/10.1109/EMBC48229.2022.9871576.
[27] 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. https://doi.org/10.1002/mds.29249.
[28] Brink-Kjaer A, Winer J, Zeitzer JM, Sorensen HBD, Jennum P, Mignot E, et al. Fully automated detection of isolated rapid-eye-movement sleep behavior disorder using actigraphy. Annu Int Conf IEEE Eng Med Biol Soc 2023;2023:1–5. https://doi.org/10.1109/EMBC40787.2023.10341133.
[29] Byeon H. Exploring the predictors of rapid eye movement sleep behavior disorder for parkinson's disease patients using classifier ensemble. Healthcare (Basel) 2020;8:121. https://doi.org/10.3390/healthcare8020121.
[30] Cesari M, Christensen JAE, Kempfner L, Olesen AN, Mayer G, Kesper K, et al. Comparison of computerized methods for rapid eye movement sleep without atonia detection. Sleep 2018;41. https://doi.org/10.1093/sleep/zsy133.
[31] Cesari M, Christensen JAE, Muntean M-L, Mollenhauer B, Sixel-Döring F, Sorensen HBD, et al. A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease. Sleep Med 2021;77:238–48. https://doi.org/10.1016/j.sleep.2020.04.010.
[32] Cesari M, Christensen JAE, Sixel-Döring F, Trenkwalder C, Mayer G, Oertel WH, et al. Validation of a new data-driven automated algorithm for muscular activity detection in REM sleep behavior disorder. J Neurosci Methods 2019;312:53–64. https://doi.org/10.1016/j.jneumeth.2018.11.016.
[33] Cesari M, Christensen JAE, Sorensen HBD, Jennum P, Mollenhauer B, Muntean M-L, et al. External validation of a data-driven algorithm for muscular activity identification during sleep. J Sleep Res 2019;28:e12868. https://doi.org/10.1111/jsr.12868.
[34] 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. https://doi.org/10.3390/brainsci14090871.
[35] 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. https://doi.org/10.1111/ene.15822.
[36] Chahine LM, Ratner D, Palmquist A, Dholakia G, Newman AB, Boyce RD, et al. REM sleep behavior disorder diagnostic code accuracy and implications in the real-world setting. Neurol, Clin Pract 2025;15:e200387. https://doi.org/10.1212/CPJ.0000000000200387.
[37] Christensen JAE, Koch H, Frandsen R, Kempfner J, Arvastson L, Christensen SR, et al. Classification of iRBD and Parkinson's disease patients based on eye movements during sleep. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:441–4. https://doi.org/10.1109/EMBC.2013.6609531.
[38] Christensen JAE, 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. https://doi.org/10.1016/j.jneumeth.2014.07.014.
[39] 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. https://doi.org/10.1002/mds.28894.
[40] Cooray N, Andreotti F, Lo C, Symmonds M, Hu MTM, De Vos M. Detection of REM sleep behaviour disorder by automated polysomnography analysis. Clin Neurophysiol 2019;130:505–14. https://doi.org/10.1016/j.clinph.2019.01.011.
[41] Cooray N, Andreotti F, Lo C, Symmonds M, Hu MTM, De Vos M. Proof of concept: screening for REM sleep behaviour disorder with a minimal set of sensors. Clin Neurophysiol 2021;132:904–13. https://doi.org/10.1016/j.clinph.2021.01.009.
[42] Cooray N, Andreotti F, Lo C, Symmonds M, Hu MTM, De Vos M. Automating the detection of REM sleep behaviour disorder. 2018 40th annual international conference of the IEEE engineering in medicine and biology society. EMBC); 2018. p. 1460–3. https://doi.org/10.1109/EMBC.2018.8512539.
[43] 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. https://doi.org/10.1007/s00259-023-06402-1.
[44] Fereshtehnejad S-M, Romenets SR, Anang JBM, Latreille V, Gagnon J-F, Postuma RB. New clinical subtypes of parkinson disease and their longitudinal progression: a prospective cohort comparison with other phenotypes. JAMA Neurol 2015;72:863–73. https://doi.org/10.1001/jamaneurol.2015.0703.
[45] 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); 2024. p. 1–4. https://doi.org/10.1109/EMBC53108.2024.10782600.
[46] Gunter KM, Brink-Kjaer A, Mignot E, Sorensen HBD, During E, Jennum P. SViT: a spectral vision transformer for the detection of REM sleep behavior disorder. IEEE J Biomed Health Inform 2023;27:4285–92. https://doi.org/10.1109/JBHI.2023.3292231.
[47] Hällqvist J, Bartl M, Dakna M, Schade S, Garagnani P, Bacalini M-G, et al. Plasma proteomics identify biomarkers predicting Parkinson's disease up to 7 years before symptom onset. Nat Commun 2024;15:4759. https://doi.org/10.1038/s41467-024-48961-3.
[48] Hansen IH, Marcussen M, Christensen JAE, Jennum P, Sorensen HBD. Detection of a sleep disorder predicting Parkinson's disease. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:5793–6. https://doi.org/10.1109/EMBC.2013.6610868.
[49] Huang B, Chau SWH, Liu Y, Chan JWY, Wang J, Ma SL, et al. Gut microbiome dysbiosis across early Parkinson's disease, REM sleep behavior disorder and their first-degree relatives. Nat Commun 2023;14:2501. https://doi.org/10.1038/s41467-023-38248-4.
[50] Illner V, Novotný M, Kouba T, Tykalová T, Šimek M, Sovka P, et al. Smartphone voice calls provide early biomarkers of parkinsonism in rapid eye movement sleep behavior disorder. Mov Disord 2024;39:1752–62. https://doi.org/10.1002/mds.29921.
[51] 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. Parkinsonism Relat Disorders 2022;95:86–91. https://doi.org/10.1016/j.parkreldis.2022.01.003.
[52] 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. https://doi.org/10.1093/sleep/zsae031. zsae031.
[53] Jiang C, Hopfner F, Katsikoudi A, Hein R, Catli C, Evetts S, et al. Serum neuronal exosomes predict and differentiate Parkinson's disease from atypical parkinsonism. J Neurol Neurosurg Psychiatry 2020;91:720–9. https://doi.org/10.1136/jnnp-2019-322588.
[54] Kempfner J, Jennum P, Nikolic M, Christensen JAE, Sorensen HBD. Sleep phenomena as an early biomarker for parkinsonism. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:5773–6. https://doi.org/10.1109/EMBC.2013.6610863.
[55] Kempfner J, Sorensen G, Zoetmulder M, Jennum P, Sorensen HBD. REM behaviour disorder detection associated with neurodegenerative diseases. Annu Int Conf IEEE Eng Med Biol Soc 2010;2010:5093–6. https://doi.org/10.1109/IEMBS.2010.5626212.
[56] Kempfner J, Sorensen GL, Nikolic M, Frandsen R, Sorensen HBD, Jennum P. Rapid eye movement sleep behavior disorder as an outlier detection problem. J Clin Neurophysiol 2014;31:86–93. https://doi.org/10.1097/WNP.0000000000000021.
[57] Kempfner J, Sorensen HBD, Nikolic M, Jennum P. Early automatic detection of Parkinson's disease based on sleep recordings. J Clin Neurophysiol 2014;31:409–15. https://doi.org/10.1097/WNP.0000000000000065.
[58] Kim H, Seo P, Byun J-I, Jung K-Y, Kim KH. Spatiotemporal characteristics of cortical activities of REM sleep behavior disorder revealed by explainable machine learning using 3D convolutional neural network. Sci Rep 2023;13:8221. https://doi.org/10.1038/s41598-023-35209-1.
[59] Kim H, Seo P, Kim MJ, Huh JI, Sunwoo J-S, Cha KS, et al. Characterization of attentional event-related potential from REM sleep behavior disorder patients based on explainable machine learning. Comput Methods Progr Biomed 2023;234:107496. https://doi.org/10.1016/j.cmpb.2023.107496.
[60] Koch H, Christensen JAE, Frandsen R, Arvastson L, Christensen SR, Sorensen HBD, et al. Classification of iRBD and Parkinson's patients using a general data-driven sleep staging model built on EEG. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:4275–8. https://doi.org/10.1109/EMBC.2013.6610490.
[61] Laguna A, Xicoy H, Tolosa E, Serradell M, Vilas D, Gaig C, et al. Serum metabolic biomarkers for synucleinopathy conversion in isolated REM sleep behavior disorder. npj Parkinson's Dis 2021;7:40. https://doi.org/10.1038/s41531-021-00184-9.
[62] Lee DA, Lee H-J, Kim HC, Park KM. Application of machine learning analysis based on diffusion tensor imaging to identify REM sleep behavior disorder. Sleep Breath 2022;26:633–40. https://doi.org/10.1007/s11325-021-02434-9.
[63] Levendowski DJ, Tsuang D, Chahine LM, Walsh CM, Berka C, Lee-Iannotti JK, et al. Concordance and test-retest consistency of sleep biomarker-based neurodegenerative disorder profiling. Sci Rep 2024;14:31234. https://doi.org/10.1038/s41598-024-82528-y.
[64] Li Y, Cao Y, Liu W, Chen F, Zhang H, Zhou H, et al. Candidate biomarkers of EV-microRNA in detecting REM sleep behavior disorder and Parkinson's disease. npj Parkinson's Dis 2024;10:18. https://doi.org/10.1038/s41531-023-00628-4.
[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. https://doi.org/10.1212/WNL.0000000000011743.
[66] Mahlknecht P, Pechlaner R, Boesveldt S, Volc D, Pinter B, Reiter E, et al. Optimizing odor identification testing as quick and accurate diagnostic tool for Parkinson's disease. Mov Disord 2016;31:1408–13. https://doi.org/10.1002/mds.26637.
[67] Matsushima T, Yoshinaga K, Wakasugi N, Togo H, Hanakawa T, Japan Parkinson's Progression Markers Initiative (J-PPMI) study group. Functional connectivity-based classification of rapid eye movement sleep behavior disorder. Sleep Med 2024;115:5–13. https://doi.org/10.1016/j.sleep.2024.01.019.
[68] Mattioli P, Orso B, Liguori C, Famà F, Giorgetti L, Donniaquio A, et al. Derivation and validation of a phenoconversion-related pattern in idiopathic rapid eye movement behavior disorder. Mov Disord 2023;38:57–67. https://doi.org/10.1002/mds.29236.
[69] Mombelli S, Leitner C, D'Este G, Sforza M, Marelli S, Castelnuovo A, et al. A data-driven approach to neuropsychological features in isolated REM behaviour disorder: a latent class analysis. J Neuropsychol 2023;17:161–79. https://doi.org/10.1111/jnp.12292.
[70] Nishiwaki H, Hamaguchi T, Ito M, Ishida T, Maeda T, Kashihara K, et al. Short-chain fatty acid-producing gut microbiota is decreased in parkinson's disease but not in rapid-eye-movement sleep behavior disorder. mSystems 2020;5:e00797. https://doi.org/10.1128/mSystems.00797-20. 20.
[71] Ophey A, Vinod V, Röttgen S, Scharfenberg D, Fink GR, Sommerauer M, et al. Accelerometry-derived features of physical activity, sleep and circadian rhythm

relate to non-motor symptoms in individuals with isolated REM sleep behavior disorder. J Neurol 2025;272:201. https://doi.org/10.1007/s00415-025-12931-6.
[72] Orso B, Mattioli P, Yoon E-J, Kim YK, Kim H, Shin JH, et al. Progression trajectories from prodromal to overt synucleinopathies: a longitudinal, multicentric brain [18F]FDG-PET study. npj Parkinson's Dis 2024;10:200. https://doi.org/10.1038/s41531-024-00813-z.
[73] Park K, Shin JH, Byun J-I, Jeong E, Kim H-J, Jung K-Y. Electroencephalographic spectro-spatial covariance patterns related to phenoconversion in isolated rapid eye movement sleep behavior disorder and their longitudinal trajectories in α-synucleinopathies. Sleep 2024;47. https://doi.org/10.1093/sleep/zsae052. zsae052.
[74] Possti D, Oz S, Gerston A, Wasserman D, Duncan I, Cesari M, et al. Semi automatic quantification of REM sleep without atonia in natural sleep environment. npj Digit Med 2024;7:1–11. https://doi.org/10.1038/s41746-024-01354-8.
[75] Pyatigorskaya N, Gaurav R, Arnaldi D, Leu-Semenescu S, Yahia-Cherif L, Valabregue R, et al. Magnetic resonance imaging biomarkers to assess substantia nigra damage in idiopathic rapid eye movement sleep behavior disorder. Sleep 2017;40. https://doi.org/10.1093/sleep/zsx149.
[76] Raschellà F, Scafa S, Puiatti A, Martin Moraud E, Ratti P-L. Actigraphy enables home screening of rapid eye movement behavior disorder in parkinson's disease. Ann Neurol 2023;93:317–29. https://doi.org/10.1002/ana.26517.
[77] 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. https://doi.org/10.3390/ijerph19010248.
[78] 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. https://doi.org/10.1016/j.sleep.2015.07.030.
[79] Rusz J, Hlavnicka J, Tykalova T, Novotny M, Dusek P, Sonka 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. https://doi.org/10.1109/TNSRE.2018.2851787.
[80] Ryoo HG, Byun J-I, Choi H, Jung K-Y. Deep learning signature of brain [18F]FDG PET associated with cognitive outcome of rapid eye movement sleep behavior disorder. Sci Rep 2022;12:19259. https://doi.org/10.1038/s41598-022-23347-x.
[81] Saeda S, Fujiwara K, Sumi Y, Kadotani H. Heart rate variability-based model for estimating the severity of orthostatic hypotension in patients with REM sleep behavior disorder. Annu Int Conf IEEE Eng Med Biol Soc 2024;2024:1–4. https://doi.org/10.1109/EMBC53108.2024.10781845.
[82] 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. https://doi.org/10.3390/diagnostics12112689.
[83] 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. https://doi.org/10.3390/diagnostics14040363.
[84] Seger A, Ophey A, Heitzmann W, Doppler CEJ, Lindner M-S, Brune C, et al. Evaluation of a structured screening assessment to detect isolated rapid eye movement sleep behavior disorder. Mov Disord 2023;38:990–9. https://doi.org/10.1002/mds.29389.
[85] 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. https://doi.org/10.3390/clockssleep7020019.
[86] Soto M, Iranzo A, Lahoz S, Fernández M, Serradell M, Gaig C, et al. Serum MicroRNAs predict isolated rapid eye movement sleep behavior disorder and lewy body diseases. Mov Disord 2022;37:2086–98. https://doi.org/10.1002/mds.29171.
[87] Tripathi S, Mattioli P, Liguori C, Chiaravalloti A, Arnaldi D, Giancardo L. Brain hemisphere dissimilarity, a self-supervised learning approach for alpha-synucleinopathies prediction with FDG PET. Proc IEEE Int Symp Biomed Imaging 2023;2023. https://doi.org/10.1109/isbi53787.2023.10230560.
[88] Urtnasan E, Joo EY, Lee KH. AI-Enabled algorithm for automatic classification of sleep disorders based on single-lead electrocardiogram. Diagnostics 2021;11:2054. https://doi.org/10.3390/diagnostics11112054.
[89] Vaish A. A machine learning approach for early identification of prodromal parkinson's disease. Cureus 2024;16:e63240. https://doi.org/10.7759/cureus.63240.
[90] Valappil RA, Black JE, Broderick MJ, Carrillo O, Frenette E, Sullivan SS, et al. Exploring the electrocardiogram as a potential tool to screen for premotor Parkinson's disease. Mov Disord 2010;25:2296–303. https://doi.org/10.1002/mds.23348.
[91] van Veen R, Meles SK, Renken RJ, Reesink FE, Oertel WH, Janzen A, et al. FDG-PET combined with learning vector quantization allows classification of neurodegenerative diseases and reveals the trajectory of idiopathic REM sleep behavior disorder. Comput Methods Progr Biomed 2022;225:107042. https://doi.org/10.1016/j.cmpb.2022.107042.
[92] Yang Y, Ye C, Sun J, Liang L, Lv H, Gao L, et al. Alteration of brain structural connectivity in progression of Parkinson's disease: a connectome-wide network analysis. Neuroimage, Clin 2021;31:102715. https://doi.org/10.1016/j.nicl.2021.102715.
[93] Senthil R, Anand T, Somala CS, Saravanan KM. Bibliometric analysis of artificial intelligence in healthcare research: trends and future directions. Future Healthc J 2024;11:100182. https://doi.org/10.1016/j.fhj.2024.100182.
[94] Frauscher B, Gschliesser V, Brandauer E, Marti I, Furtner MT, Ulmer H, et al. REM sleep behavior disorder in 703 sleep-disorder patients: the importance of eliciting a comprehensive sleep history. Sleep Med 2010;11:167–71. https://doi.org/10.1016/j.sleep.2009.03.011.
[95] IRBDSG n.d. https://irbdsg.com/. [Accessed 2 March 2026].
[96] Cesari M, Ferri R, Högl B, Stefani A, Silvani A. SOMAS – an open-source software for the analysis of muscle activity during sleep. Sleep Med 2026:108791. https://doi.org/10.1016/j.sleep.2026.108791.
[97] Röthenbacher A, Cesari M, Doppler CEJ, Okkels N, Willemsen N, Sembowski N, et al. RBDtector: an open-source software to detect REM sleep without atonia according to visual scoring criteria. Sci Rep 2022;12:20886. https://doi.org/10.1038/s41598-022-25163-9.
[98] Bradshaw TJ, Huemann Z, Hu J, Rahmim A. A guide to cross-validation for artificial intelligence in medical imaging. Radiol Artif Intell 2023;5:e220232. https://doi.org/10.1148/ryai.220232.
[99] Apicella A, Isgrò F, Prevete R. Don't push the button! exploring data leakage risks in machine learning and transfer learning. Artif Intell Rev 2025;58:339. https://doi.org/10.1007/s10462-025-11326-3.