Journal of Integrative Neuroscience

J. Integr. Neurosci. **2026**; 25(11): 50816
https://doi.org/10.31083/JIN50816

Review

Machine Learning Methods for Automatic Sleep Stage Classification: Current Status and Future Challenges

Yingying Shao[1], Yi Liu[1], Weidong Gao[1,*], Fen Miao[2,*], Xiaomao Fan[3], Ruihan Yu[1], Siying Wu[1]

[1]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, 100876 Beijing, China
[2]Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, 518000 Shenzhen, Guangdong, China
[3]School of Artificial Intelligence, Shenzhen Technology University, 518118 Shenzhen, Guangdong, China
*Correspondence: gaoweidong@bupt.edu.cn (Weidong Gao); fenmiao@uestc.edu.cn (Fen Miao)
Academic Editors: Davide Borra and Bettina Platt
Submitted: 7 February 2026 Revised: 23 April 2026 Accepted: 18 May 2026 Published: 19 November 2026

Abstract

The intricacies of sleep architecture hold significant implications for understanding human health, making automated sleep staging a critical frontier in sleep research. While existing literature predominantly delves into deep learning methods for sleep staging, the current landscape of lightweight, home-use sleep monitoring devices necessitates a closer examination of traditional machine learning algorithms for sleep pattern recognition and analysis. In response to this need, in this paper we provide a comprehensive review of machine learning methods for automatic sleep staging. By synthesizing state-of-the-art approaches, we traverse the entire process, from data collection devices to feature extraction and selection, culminating in diverse sleep stage classification methods. These methods encompass single-modality, dual-modality, and multi-modality approaches, reflecting the evolving complexity of sleep stage classification methodologies. Moreover, we delve into the challenges, identify emerging trends, and outline future directions in the field of sleep staging. By offering these insights, we aim to equip researchers, clinicians, and technologists with a roadmap for advancing automated sleep staging methodologies within the realm of machine learning.

Keywords: sleep stages; signal processing, computer-assisted; machine learning; polysomnography; algorithms; wearable electronic devices

1. Introduction

Sleep plays a vital role in human health and well-being, supporting physiological restoration, cognitive function, emotional regulation, and physical performance. Adequate sleep is necessary for tissue repair, immune function, hormonal balance, and memory consolidation [1], while sleep deprivation can lead to cognitive issues, mood disorders, and metabolic diseases [2]. Sleep stage classification is vital for diagnosing and monitoring sleep disorders [3], as different stages indicate various physiological processes. Accurate identification of these stages helps clinicians diagnose conditions like insomnia and sleep apnea, and monitor treatment effectiveness of interventions, ultimately promoting better sleep health and overall quality of life.

There are two widely used standards in sleep scoring, one developed by Rechtschaffen and Kales (R&K) in 1968 [4], and the other is a revised guideline issued by the American Academy of Sleep Medicine (AASM) in 2007 [5]. According to the R&K criteria, sleep was initially categorized into seven stages: rapid eye movement (REM), sleep stage 1 (S1), sleep stage 2 (S2), sleep stage 3 (S3), sleep stage 4 (S4), wakefulness (W), and movement stage. In contrast, the AASM rules simplify this categorization by eliminating movement stages and combining stages 3 and 4 into one stage, N3. A total of five stages are identified in the AASM standard: W, stage N1, stage N2, stage N3, and REM sleep. The AASM criteria are now recognized as the standard for sleep scoring.

Sleep scoring is the foundation for diagnosing and treating sleep disorders and is increasingly gaining attention. Sleep stage classification has seen significant development, transitioning from manual scoring methods to automated techniques driven by machine learning and deep learning algorithms. Initially, the classification of sleep stages relied on manual scoring conducted by proficient professionals using polysomnography (PSG), which is still the gold standard for sleep analysis [6]. PSG entails recording many physiological signals simultaneously, such as electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), electrocardiography (ECG), respiratory signals, and sometimes additional modalities like pulse oximetry. Sleep experts visually analyze these recorded signals and assign sleep stages according to standardized guidelines, a process that is subjective and time-consuming.

With the advancement of computational methods and the development of machine learning and deep learning techniques, more efficient and objective alternatives for automated sleep stage classification have emerged.

Copyright: © 2026 The Author(s). Published by IMR Press.
This is an open access article under the CC BY 4.0 license.

Publisher's Note: IMR Press stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

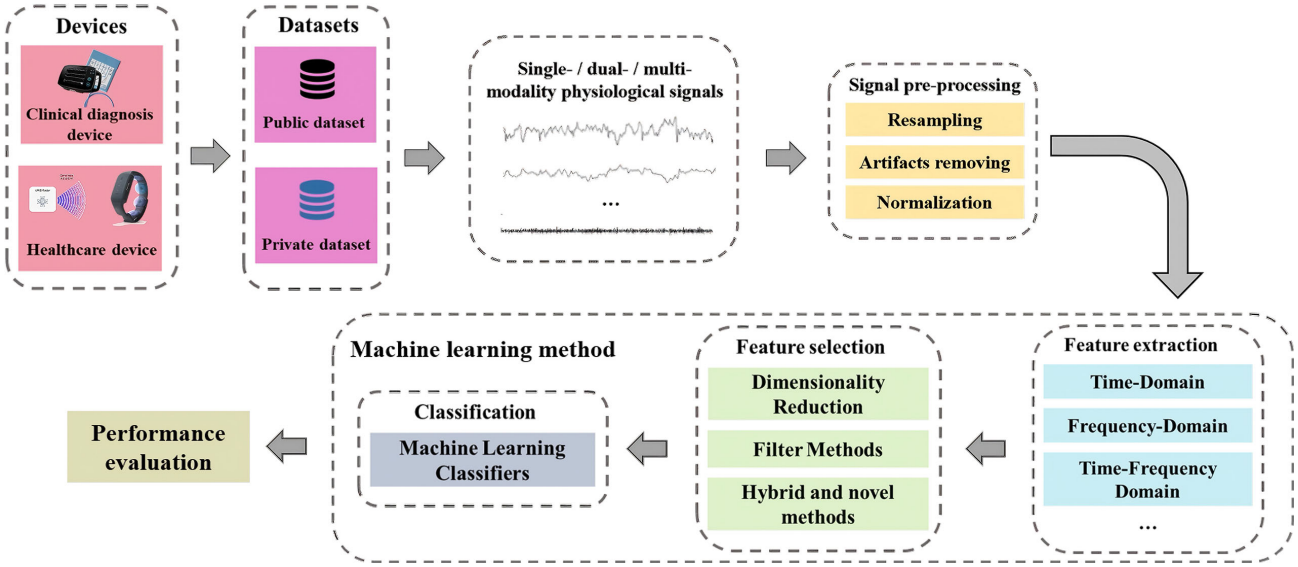


Fig. 1. The pipe line of sleep stage classification methods based on machine learning. It includes data collection, signal pre-processing, feature extraction, feature selection, classification and performance evaluation.

Initially, researchers focused on extracting relevant features from physiological signals to distinguish different sleep stages. These features encompassed time-domain, frequency-domain, and statistical measures, capturing temporal, spectral, and statistical characteristics. Feature extraction methods mainly included fast Fourier transform (FFT), wavelet transform (WT), and statistical analysis. Machine learning algorithms such as support vector machine (SVM) [7,8,9,10,11,12,13,14,15,16,17,18,19], decision tree (DT) [11,20,21,22,23,24], and k-nearest neighbor (kNN) [25,26] were commonly used to construct classification models for sleep staging. These models utilized the extracted features as input and employed supervised learning techniques to train models on labeled datasets. Subsequently, the models were capable of automatically classifying sleep stages based on learned patterns. The pipeline of automatic sleep stage classification methods based on machine learning is illustrated in Fig. 1.

As sleep datasets expanded, deep learning emerged as a powerful method for sleep stage classification. Deep learning algorithms, particularly convolutional Neural networks (CNNs) [27,28,29] and recurrent neural networks (RNNs) [30,31,32], exhibited superior performance in capturing complex temporal and spatial dependencies in sleep signals. Deep learning models can learn directly from raw or preprocessed signals without the need to manually extract features. This capability allowed for discovering intricate patterns and subtle variations that were challenging to capture using traditional feature-based approaches. Furthermore, the accessibility of sizable annotated databases fueled the advancement of sleep stage classification research, providing researchers with ample opportunities to train and validate models, resulting in improved accuracy and generalization.

Sleep stage classification is a systematic research encompassed from physiological signal processing, physiological signal representation to learning methods. Worldwide researchers have devoted their efforts to this field. Chriskos et al. [33] have provided an overview of automated sleep staging methods based on bio-signal recordings from 2014 to the first quarter of 2019. Fiorillo et al. [34] have comprehensively described the advanced deep learning methods for enhancing sleep scoring, while also discussed the obstacles to adopting automated sleep scoring in clinical practice. Djanian's work [35] provided a systematic review of the current AI technologies utilized for sleep stage classification through consumer sleep technologies (CSTs), comparing AI models and their performance to synthesize the current state of the art. Sarkar et al. [36] have analyzed and compared the performance of the automatated sleep scoring methods emerged between 2016 and 2021.

While several reviews have summarized automated sleep staging methods, their emphases differ from the focus of this paper. Existing surveys tend to emphasize deep neural networks, with comparatively less attention to classical and lightweight machine-learning methods for sleep staging. With the development of wearable electronics, sleep monitoring is transitioning from primarily in-laboratory assessment to home-care and long-term daily monitoring. This transition introduces practical constraints that strongly influence algorithm selection, including limited computation and memory on edge devices and strict power budgets in real-world applications. Although deep-learning approaches have achieved strong performance in many sleep-staging benchmarks, large deep models can be difficult to deploy in resource-constrained wearable and home-care settings due to their computational and storage demands. In this context, lightweight machine-learning approaches—

emphasizing efficient feature engineering, compact model structures, and low inference overhead—remain attractive for deployable sleep-staging systems on wearable and edge devices. Therefore, this review focuses on lightweight machine-learning pipelines for sleep stage classification that are readily implementable in wearable and home-care scenarios. In this review, we use the term "lightweight" in an operational sense, referring to sleep-staging pipelines that are compatible with real-world edge and wearable deployment constraints. We define "lightweight" using three criteria:

(1) Computational cost: Lightweight methods should execute efficiently on low-power processors (e.g., Microcontroller Unit (MCU)-class cores or mobile Central Processing Units (CPUs)) and avoid inference procedures with high or rapidly growing compute demand. In practice, deployable pipelines often combine low-cost feature computation (e.g., simple time/frequency features or compact physiological indices) with computationally efficient classifiers (e.g., decision trees, linear models, Naïve Bayes, or SVM variants). Methods that require expensive end-to-end representation learning on high-rate, multi-channel waveforms—unless accompanied by clear efficiency measures—are generally less aligned with the lightweight objective in this review.

(2) Memory footprint: Lightweight methods should operate under strict memory constraints, including both non-volatile storage (e.g., Flash/Read-Only Memory (ROM) for code and model parameters) and runtime memory (e.g., Random Access Memory (RAM) for intermediate buffers). Lightweight designs typically favor compact model structures and a parsimonious yet informative feature set, and avoid long-duration buffering of high-sampling-rate, multi-channel signals when possible. With these design choices, both model size and runtime memory can generally be kept in the kilobyte (KB) to low megabyte (MB) range, which is compatible with many wearable and embedded platforms.

(3) On-device deployability: Beyond the model itself, the complete pipeline—including preprocessing, feature extraction, and inference—should be feasible to run locally on edge/wearable hardware under energy constraints, without continuous cloud offloading. In this review, a method is considered deployable when its end-to-end workflow can be implemented on-device with predictable latency and memory usage, and when required processing steps can be executed in a resource-aware manner consistent with embedded/TinyML deployment practices.

Guided by these criteria, we summarize representative studies across data acquisition devices, signal modalities, preprocessing, feature extraction and selection, and classification. Deep-learning methods are discussed where relevant, particularly when they address deployability requirements (e.g., efficient architectures or compression strategies), but broader deep-learning surveys are beyond the central scope of this paper. A comparison between this review and existing surveys is provided in Table 1 (Ref. [33,34,35,36]).

The structure of this paper is outlined as follows. Section 2 outlines the methodology employed for article selection. Section 3 introduces available public databases for sleep staging research. Section 4 delineates physiological signals involved in sleep stage classification, including types and devices. Physiological signal processing methods are presented in section 5. Section 6 comprises sleep scoring methods based on machine learning and evaluation metrics. The challenges and future prospects are discussed in section 7. Finally, the paper is concluded in section 8.

2. Methodology

This narrative review summarizes machine-learning-based methods for automatic sleep stage classification and synthesizes key information across the research pipeline, including data sources and datasets, physiological signals (types and acquisition devices), signal preprocessing and representation, and classification algorithms. Based on the included studies, we also discuss current limitations and potential directions for future research.

This narrative review was informed by a structured literature search. The following subsections describe the literature search, selection, and information extraction procedures used to support the thematic and descriptive synthesis.

2.1 Article Search Strategy

Electronic searches were conducted in the Web of Science and PubMed databases to identify relevant studies published between 2018 and 2026. This period was chosen to focus on contemporary research on lightweight, computationally efficient machine-learning approaches for automatic sleep stage classification, particularly in wearable or portable monitoring scenarios.

The search strategy combined sleep staging-related terms ("sleep stage classification" OR "sleep stage identification" OR "sleep staging" OR "automatic sleep stage scoring") with classical machine learning-related terms ("machine learning" OR "support vector machine" OR "SVM" OR "random forest" OR "decision tree" OR "k-nearest neighbor" OR "k-NN" OR "naive Bayes" OR "ensemble learning" OR "feature selection") using the Boolean operator AND. To broaden coverage, we included both general machine-learning keywords and algorithm-specific terms commonly used in classical machine-learning-based sleep staging. Reference lists of relevant papers were also screened to identify additional eligible studies when applicable.

2.2 Literature Screening and Inclusion Criteria

All retrieved records were exported to a reference-management tool and duplicate records were removed before screening. Two reviewers independently screened ti-

Table 1. Comparison of this paper with existing surveys.

Reference	Year	Content	Databases	Devices	Physiological signals	Signal pre-processing	Feature extraction	Feature selection	Lightweight machine learning method
Fiorillo et al. [34]	2019	Provided a review of the most recent approaches utilizing deep learning for automated sleep staging.	≈	✗	✗	✗	≈	≈	≈
Chriskos et al. [33]	2021	Provided an overview of current developments in automatic sleep scoring with bio-signal recordings and future challenges.	✗	✗	✓	✓	✓	✗	≈
Djanian et al. [35]	2022	Analyzed 27 papers that integrate CSTs with Artificial Intelligence (AI) models for the classification of sleep stages.	≈	✓	✗	✗	✗	✗	≈
Sarkar et al. [36]	2022	Analyzed and compared the efficacy of different automated sleep scoring methods.	≈	✗	✓	✓	✓	✓	≈
Our paper	2026	Presents a comprehensive survey of lightweight machine learning methods for automatic sleep stage classification.	✓	✓	✓	✓	✓	✓	✓

Footnote: The term physiological signals refers to commonly used signals for automatic sleep staging, including standard clinical PSG signals, wearable-recorded physiological signals (e.g., single-channel EEG, wrist-worn actigraphy, and ECG), and non-contact sensing signals (e.g., radar- and pressure sensor–based signals) for home sleep monitoring. (✓: included, ✗: not included, ≈: partially included). PSG, polysomnography; EEG, electroencephalography; ECG, electrocardiography.

tles and abstracts according to the predefined inclusion and exclusion criteria. Reports considered potentially eligible by either reviewer were sought for retrieval and full-text assessment. The same two reviewers then independently assessed the available full texts for eligibility. Disagreements at any stage were resolved through discussion; if consensus could not be reached, a third senior author adjudicated.

The narrative synthesis discusses 77 studies.

The detailed inclusion criteria were defined as follows:

• The articles are original studies published in peer-reviewed journals or conference proceedings.

• The publication years range from 2018 to 2026.

• The articles are written in English and the full text is accessible.

• The studies propose, evaluate, or compare machine-learning-based approaches for automatic sleep stage classification, and report sufficient methodological details (e.g., signal modalities, preprocessing, feature extraction/selection, model settings, and evaluation).

• Studies evaluating lightweight machine-learning methods or selected rule-based comparators for automatic sleep–wake detection or sleep-stage classification, with sufficient methodological information for extraction.

The exclusion criteria were as follows:

• Review papers, editorials, tutorials, and non-research items.

• Studies using non-machine-learning methods, except selected rule-based approaches used for comparison.

• Deep-learning studies outside the lightweight scope defined in the Introduction.

• Studies without accessible full text or with insufficient methodological information for extraction.

2.3 Information Extraction and Narrative Synthesis

For each included study, we extracted the following information: publication year; physiological signal acquisition device(s); signal modalities; dataset and sample size (where reported); preprocessing methods; feature extraction and selection techniques; classification models; eval-

uation protocols and metrics; and the main challenges and limitations discussed by the authors. The extracted information was organized and synthesized to support a qualitative comparison across studies.

The extracted information was organized thematically across the sleep-staging pipeline and synthesized narratively. No formal methodological quality or risk-of-bias assessment was conducted. Given the absence of study-level appraisal and differences in datasets and validation protocols, cross-study performance comparisons are descriptive and do not establish the general superiority of any algorithm or modality.

3. Available Public Databases

For automatic sleep stage classification studies, several well-known and comprehensively annotated sleep databases are widely used. The classical datasets include the "Sleep-EDF database" (SLEDF) [37], "Sleep-EDF expanded database" (XSLEDF) [37], "Sleep Heart Healthy Study database" (SHHS) [38], "University College Dublin Sleep Apnea database" (UCD) [39], "DREAMS subject database" (DRMS) [40], "Massachusetts Institute of Technology Beth Israel Hospital database" (MIT-BIH) [41], "ISRUC-Sleep database" (ISRUC-Sleep) [42], and "Montreal Archive of Sleep Studies database" (MASS) [43].

Beyond these foundational datasets, the growing demand to evaluate the generalizability and scalability of machine learning models has driven the adoption of large-scale cohort studies and specialized clinical databases. For instance, the Multi-Ethnic Study of Atherosclerosis (MESA) [44] and the Osteoporotic Fractures in Men Study (MrOS) [45], both hosted on the National Sleep Research Resource (NSRR), offer thousands of PSG recordings. To test algorithmic robustness across specific sleep disorders, the CAP Sleep Database (CAPSLPDB) [46] provides a rich collection of pathological recordings with detailed microstructure annotations, while the Haaglanden Medisch Centrum (HMC) database [47] represents a modern clinical routine PSG collection. Furthermore, as the field shifts toward lightweight and home-based monitoring, datasets specifically targeting wearable technologies have emerged. The Dreem Open Datasets (DOD-H and DOD-O) [48] and the recently introduced Bitbrain Open Access Sleep (BOAS) database [49] provide essential resources for training algorithms on dry-EEG and wearable sensor data.

Notably, most of the earlier datasets (e.g., SLEDF, XSLEDF, UCD, MIT-BIH, SHHS) adhere to the classic R&K staging criteria. In contrast, modern collections (ISRUC-Sleep, MASS, MESA, MrOS, BOAS, DOD, HMC) typically adopt the updated AASM standards, while CAPSLPDB contains recordings scored under both systems depending on the collection year. The DRMS database uniquely provides dual annotations compatible with both frameworks. Regarding data accessibility, the official access route for each database is indicated after its name in Table 2. SHHS, MESA Sleep, and MrOS Sleep are accessed through NSRR following registration and approval of a data-access request. Key specifications of these fourteen databases, including subject demographics, physiological signals, epoch length, and number of recordings, are comprehensively summarized in Table 2.

4. Physiological Signals & Devices

4.1 Physiological Signals

Physiological signals underpin sleep stage classification by capturing unique biomarkers of sleep architecture. For example, a PSG study reported differences in EEG activity and HRV during N3 sleep between patients with Parkinson's disease and healthy controls [50]. Fig. 2 illustrates the characteristics of physiological signals during various sleep stages. However, lightweight machine learning algorithms rarely process these raw waveforms directly. Instead, they extract specific features to balance classification accuracy with low computational cost. The following sections outline the computational utility, common noise sources, and stage-specific marginal value of each signal in the context of lightweight modeling.

4.1.1 Electroencephalography (EEG)

The EEG signal monitors electrical activity in the brain and is the primary modality for sleep staging. Rather than tracking classic waveforms (e.g., sleep spindles or K-complexes), lightweight algorithms typically rely on efficient metrics such as power spectral density, differential entropy, or sample entropy. While using a single-channel EEG reduces the computational load for wearable applications, it remains sensitive to movement artifacts and baseline wander. Despite these noise issues, EEG is essential because it provides the highest marginal value for detecting subtle stage transitions, particularly the difficult Wake-to-N1 and N2-to-N3 shifts.

4.1.2 Electrooculography (EOG)

EOG captures eye movements and is primarily used as an auxiliary channel. Its main computational value is improving the detection of N1 (slow rolling eye movements) and REM (rapid eye movements)—two stages that are notoriously difficult to isolate using single-channel EEG alone. For lightweight models, extracting basic time-domain energy or simple wavelet features from the EOG is usually sufficient and requires minimal processing overhead.

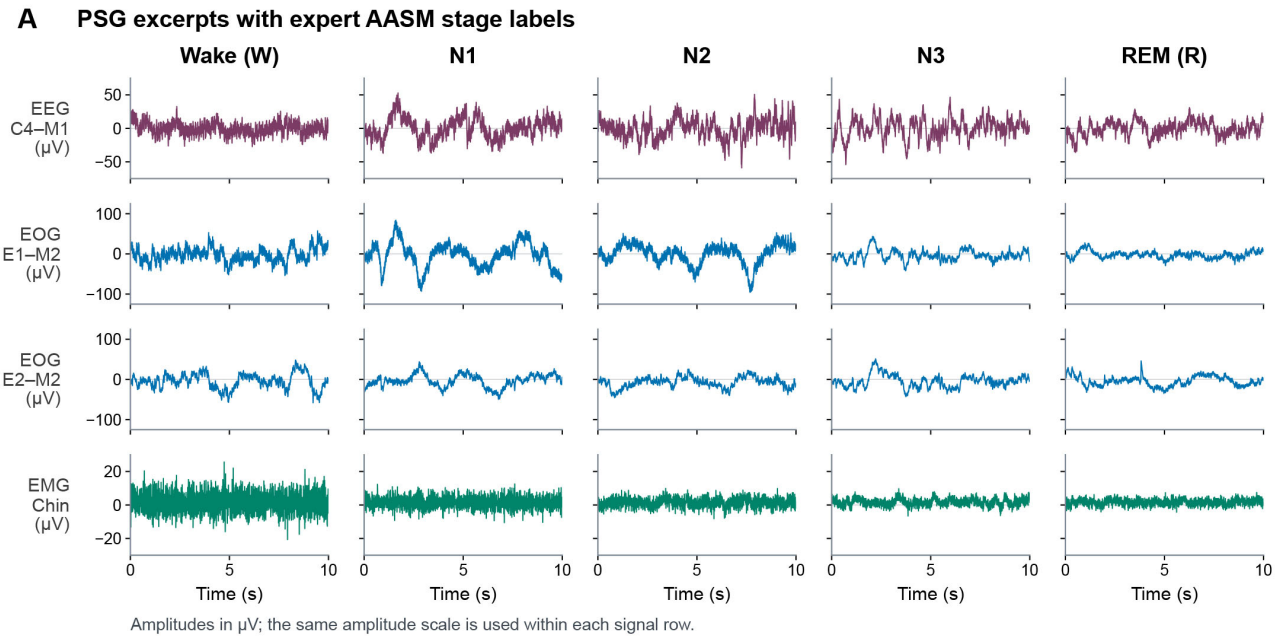
4.1.3 Electromyography (EMG)

EMG tracks muscle tone. In algorithm design, its main contribution is establishing a clear boundary between Wakefulness (high muscle activity) and REM sleep (muscle atonia). To keep computational costs low, models usually extract basic time-domain statistics, such as root mean

Table 2. Summary of well-known sleep databases.

Database	Subsets	Subjects	Recordings	Age	Gender	Physiological signals/ Channels	Epoch length
SLEDF (PhysioNet)	-	8 (4 healthy, 4 mild difficulty in falling asleep)	8	21–35	4M, 4F	EEG (Fpz-Cz, Pz-Oz), EOG, EMG	30s
XSLEDF (PhysioNet)	SC	78 (healthy)	153	25–101	37M, 41F	EEG (Fpz-Cz, Pz-Oz), EOG, EMG	30s
	ST	22 (mild difficulty in falling asleep)	44	18–79	7M, 15F		30s
SHHS (NSRR)	SHHS-1	6441	-	≥40	-	EEG (C3-A2, C4-A1), EOG, EMG, ECG	30s
	SHHS-2	3295	-	≥40	-		30s
UCD (PhysioNet)	-	25 (sleep-disordered breathing)	-	18	21M, 4F	EEG (C3-A2), EEG (C4-A1), EOG, EMG, ECG	30s
DRMS (Zenodo)	DRMS-SUB	20 (healthy)	20	20–65	4M, 16F	EEG, EOG, EMG, ECG	20s/30s
	DRMS-PAT	20 (having sleep disorders)	27	18–75	4M, 16F		20s/30s
MIT-BIH (PhysioNet)	-	16 (14 with sleep apnea, 2 healthy)	18	32–56	-	EEG, EOG, EMG, ECG	30s
ISRUC-Sleep (Official website)	Subgroup-I	100 (having sleep disorders)	100	20–85	55M, 45F	EEG, EOG, EMG, ECG	30s
	Subgroup-II	8 (having sleep disorders)	16	26–79	6M, 2F		30s
	Subgroup-III	10 (healthy)	10	30–58	9M, 1F		30s
MASS (Borealis)	SS1	53	53	55–76	34M, 19F	EEG, EOG, EMG, ECG	30s
	SS2	19	19	18–33	8M, 11F		20s
	SS3	62	62	20–69	29M, 33F		30s
	SS4	40	40	18–35	14M, 26F		20s
	SS5	26	26	20–59	13M, 13F		20s
MESA (NSRR)	-	2237	2056	45–84	-	EEG, EOG, EMG, ECG	30s
MrOS (NSRR)	Sleep Visit 1	3135	2907	≥65	3135M	EEG, EOG, EMG, ECG	30s
CAPSLPDB (PhysioNet)	-	108	108	≥18	-	EEG, EOG, EMG, ECG, respiration signals	30s
HMC (PhysioNet)	-	151	151	39–68	85M, 66F	EEG, EOG, EMG, ECG	30s
DOD (Zenodo)	DOD-H	25 (healthy)	25	≥18	-	EEG	30s
	DOD-O	55 (sleep apnea)	55	≥18	-	EEG	30s
BOAS (OpenNeuro)	-	128	128	≥18	-	EEG	30s

Abbreviations: M, male; F, female; SLEDF, Sleep-EDF database; XSLEDF, Sleep-EDF expanded database; SHHS, Sleep Heart Healthy Study database; UCD, University College Dublin Sleep Apnea database; DRMS, DREAMS subject database; MIT-BIH, Massachusetts Institute of Technology Beth Israel Hospital database; ISRUC-Sleep, ISRUC-Sleep database; MASS, Montreal Archive of Sleep Studies database; MESA, Multi-Ethnic Study of Atherosclerosis; MrOS, Osteoporotic Fractures in Men Study; CAPSLPDB, CAP Sleep Database; HMC, Haaglanden Medisch Centrum; DOD, Dreem Open Dataset; BOAS, Bitbrain Open Access Sleep; EEG, electroencephalography; EOG, electrooculography; EMG, electromyography; ECG, electrocardiography.



B **Typical stage-related physiological features**

Stage	EEG	EOG	Chin EMG
Wake (W)	Mixed-frequency activity; alpha with eyes closed	Blinks and voluntary eye movements	Relatively high and variable chin tone
N1	Low-amplitude mixed frequencies; theta activity	Slow rolling eye movements may occur	Chin tone declines from wakefulness
N2	Sleep spindles and K-complexes may occur	Rapid eye movements typically absent	Reduced chin muscle tone
N3	Prominent slow-wave (delta) activity	Eye movements generally absent or infrequent	Reduced chin muscle tone
REM (R)	Low-amplitude mixed-frequency activity	Bursts of rapid eye movements may occur	Very low tonic chin EMG; brief twitches may occur

HMC Sleep Staging Database v1.1 · SN001 · CC BY 4.0

10-s excerpts from 30-s scored epochs

Fig. 2. Characteristics of physiological signals (EEG, EOG, EMG) during sleep.

square (RMS) or zero-crossing rate, avoiding complex frequency transforms. In wearable setups, the primary noise challenges for EMG are motion artifacts and electrode detachment.

4.1.4 Electrocardiography (ECG)

Rather than processing raw ECG waveforms, lightweight models typically extract derivative metrics like heart rate variability (HRV) or cardiopulmonary coupling (CPC). Because HRV data is low-dimensional and easily acquired via smartwatches or chest patches, it is well-suited for ultra-low-power, long-term home monitoring. While ECG alone struggles to pinpoint exact N1/REM boundaries compared to EEG, it provides a highly efficient alternative for coarse sleep/wake or NREM/REM estimations.

4.1.5 Other Signals

Beyond conventional polysomnography, emerging sensors offer alternative data sources. For instance, radar-based technologies extract respiration and movement [26,51], while wrist-worn actigraphy captures acceleration [52,53]. From a machine learning perspective, these kinematic signals generate very low-dimensional data streams. They avoid typical electrophysiological noise (like skin impedance changes) and require minimal preprocessing. This makes them ideal inputs for edge-computing and TinyML applications where processing power is strictly limited.

Table 3. Devices for classifying sleep stages and their advantages and disadvantages.

Type		Device	Signals	Advantages	Disadvantages
Clinical diagnosis	PSG	Standard system with physiological signal electrode	EEG EOG EMG ECG	• Comprehensive data collection • Gold standard in sleep studies	• High cost • Time-consuming • Limited availability
	Implantable	Intracranial depth electrode Subdermal scalp EEG electrode Subthalamic local field potentials	iEEG	• Precise brain activity measurement • Long-term monitoring	• Invasiveness • High skill requirement • Patient discomfort • Ethical considerations
Healthcare	Contact	Wrist-worn sensors (Actigraphy) Flexible sensors attached to the skin	Acceleration Respiration Skin temperature	• Convenience • Long-term monitoring • More accurate signals • Reduced movement artifacts • Multi-modal sensing	• Limited accuracy • Sensitivity to movement • Uncomfortable • Limited coverage • Strict device placement
		Wearable in-ear EEG	EEG	• Portability • Long-term monitoring	• Subject to electromagnetic interference from the surrounding environment
		Sensors attached to the chest	ECG	• More accurate signals	• Sense of bondage • More movement artifacts
	Non-contact	Continuous wave doppler radar	Radar	• Continuous undisturbed monitoring • Movement detection	• Interfered by multiple persons in the same space
		Ultra-Wide band (UWB) radar	Radar	• High precision in detecting small movements • Multi-modal sensing	• Interfered by multiple persons in the same space
		Microphone	Respiratory sound	• Respiratory sounds detection	• Environmental noise • Dependency on audio quality
		Pressure sensor	Micromovement signals	• Natural sleep environment	• Limited accuracy

4.2 Devices

Sleep stage classification involves the utilization of data collection devices, which can be classified into two primary types according to the approach to data acquisition: clinical diagnosis and healthcare. Clinical diagnosis devices encompass traditional PSG and implantable devices. PSG involves attaching electrodes directly to the subject's skin to detect signals, while implantable devices are characterized by electrodes that invade the subject's body. On the other hand, healthcare devices include contact and non-contact sensors. Contact devices are placed on the individual's body, usually on the skin's surface, to collect data such as EEG, EMG, and ECG. Non-contact devices capture data from the surroundings without direct contact with the subject. Some common devices used for sleep monitoring, along with their advantages and disadvantages and the signals collected by the devices are summarized in Table 3.

4.2.1 Clinical Diagnosis Devices

4.2.1.1 Polysomnography (PSG). PSG is a comprehensive measurement of various physiological signals for sleep stage classification, which is of high reliability. Researchers usually select one or several physiological signals collected by PSG for sleep staging. In most studies, the data collection process takes place within a hospital sleep laboratory, and EEG is the primary signal selected.

4.2.1.2 Implantable Sensor. In some specific scenarios, conventional PSG cannot meet the requirement of long-term monitoring. For instance, in cases where an EEG is conducted as a component of the presurgical assessment of patients with drug-resistant focal epilepsy, it is necessary to continuously obtain data for 1 week or longer. Collodion glued standard scalp EEG electrodes used by PSG not only increase the patient's discomfort but also may lead to inconsistent signal quality, since contacts may drop off or

may be unrecoverable during the collection. In addition, the traditional scalp EEG detection principle is not applicable to patients with epilepsy. To deal with such problems, some studies use implantable sensors in sleep stage classification tasks, such as intracranial depth electrode [54], subthalamic local field potentials [13], and subdermal scalp EEG electrode [55], which can intrusively obtain intracranial electroencephalograph (iEEG) signal from subjects.

4.2.2 Healthcare Devices

Clinical diagnosis devices pose several limitations and constraints on both the signal acquisition environment and subjects. Typically conducted in a laboratory or sleep center, these procedures involve specialists attaching multiple electrodes to the subject's body according to medical guidelines, or in the case of implantable electrodes, requiring surgical implantation before the experiment. The emergence of the internet of things (IoT) has prompted significant research in the field of healthcare devices aimed at classifying sleep stages. Healthcare devices provide a more comfortable method for collecting signals from subjects. They are well-suited for home environments, and have the potential to offer sleep stage categorization services to large populations. These devices can be broadly categorized as contact or non-contact, each with unique advantages in terms of convenience and scalability. Due to their portability as well as accessibility, the use of healthcare devices for sleep stage classification has become a trend.

4.2.2.1 Contact Device. Contact devices, because of their compact, portable nature, can integrate a variety of sensors to capture physiological signals associated with sleep stages. These signals include ECG signals, respiratory signals, acceleration signals (notably utilized in actigraphy), and skin temperature. Actigraphy, typically worn on the wrist, relies on accelerometers to record motor activity. Because it provides an exceptionally lightweight data stream, actigraphy is widely employed as an ultra-low-power solution for sleep/wake classification and is increasingly being integrated with machine learning for multi-stage estimation. Additionally, EEG signals can also be recorded by these devices. These devices can be easily worn on the skin [56], wrist [56,57], ear [14], or chest [58]. Due to their versatility and user-friendliness, contact devices have become an essential tool for non-invasive monitoring of sleep-related physiological parameters, providing an important support for the development of personalized sleep monitoring technology.

4.2.2.2 Contact-Free Device. With the development of sensor technology and signal processing, researchers are gradually turning their attention to exploring the use of non-contact devices in sleep staging, such as radar [26,51], microphones [59] and so on. These non-contact methods provide completely undisturbed detection, offering greater comfort to subjects compared to contact devices and enabling the observation of subjects in their natural sleep state over the long term. In many studies, high-frequency radar signals are employed to sample micromovement signals generated by subjects during sleep. Researchers extract at least two types of physiological data from these micromovement signals, typically involving parameters related to breathing, heartbeat, and body movement. This approach not only introduces a more comfortable and unobtrusive method for sleep monitoring but also broadens the range of available tools for comprehensive sleep stage classification.

The physical characteristics of these devices directly determine the design constraints and input dimensions of the subsequent machine learning algorithms. When processing high-dimensional clinical PSG data, algorithms typically rely on strict feature selection and channel reduction to remain lightweight. In contrast, healthcare devices (like actigraphy or single-channel in-ear EEG) generate narrow data streams. Models built for these sensors are naturally compact, making them well-suited for edge computing and TinyML frameworks. Non-contact devices (such as radar and pressure sensors) maximize user comfort but introduce significant environmental noise. As a result, models using these inputs must prioritize noise-resilient feature extraction alongside lightweight classification. Overall, the hardware choice directly shapes the specific lightweight strategy employed.

5. Physiological Signal Processing Methods

5.1 Signal Pre-processing Methods

Signal pre-processing plays a crucial role in the initial preparation of raw data for subsequent analysis and applications, including feature extraction and selection, and sleep stage classification. Common techniques employed in signal pre-processing for sleep staging contain resampling, artifacts removing, and normalization. These methods are described in detail as follows.

5.1.1 Resampling

Resampling involves altering the sampling rate of a signal, either to increase or decrease it (i.e., downsampling), based on specific task requirements. Downsampling is commonly employed to streamline computational processes and optimize storage utilization. In the context of sleep scoring, data collected from various devices is often downsampled to a designated sampling frequency. For instance, EEG signals, downsampling to frequencies such as 100 Hz [60], 120 Hz [61], 125 Hz [61], or 128 Hz [14,62] has been observed. In a study [57], data obtained from BioStampRC sensors underwent interpolation and resampling to 62 Hz for accelerometer data and 1000 Hz for ECG data. This process aimed to rectify occasional instances of duplicated or missing data points.

5.1.2 Artifacts Removing

Artifacts originate from a variety of sources and can compromise signal quality. The effective identification and mitigation of artifacts are essential for reliable signal analysis. Artifactual interference in physiological signals is mainly classified into the following five categories: (1) electrode artifacts, which include baseline noise and signal distortion caused by poor electrode contact, displacement shedding, or impedance fluctuations; (2) motion artifacts, characterized by transient signal interruptions and high-frequency noise resulting from body or head movements and muscle activity; (3) power line interference, which manifests itself in the form of 50/60 Hz industrial frequency noise; (4) environmental noise, including signal attenuation caused by electromagnetic interference from electronic devices and background noise; (5) physiological artifacts, such as blinking EOG signals, EMG activity noise, and ECG rhythm interference mixed in the EEG acquisition. As shown in Table 4 (Ref. [8,10,12,14,21,23,53,57,58,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79]), artifacts can be addressed using pre-processing methods such as filtering, wavelets, and independent component analysis (ICA).

5.1.3 Normalization

To mitigate the impact of variations in devices and physiological differences among subjects, normalization is performed on the extracted features or epochs derived from raw or filtered data. The most frequently utilized normalization technique in the literature is Z-score standardization. This process involves subtracting the mean and dividing by the standard deviation.

Based on the methods reviewed above and detailed in Table 4, the choice of pre-processing strategy depends heavily on the deployment environment. For offline clinical analysis, high-order Butterworth filters (e.g., 10th–20th order) combined with ICA are standard because they maximize artifact removal. However, these methods are computationally heavy and unsuitable for real-time wearable devices. For lightweight edge deployment, lower-order IIR filters (e.g., 2nd to 4th order Butterworth) provide a better balance between signal integrity and memory constraints. Median filtering may be considered for baseline-wander correction, with the window length selected to preserve the physiological features of interest. Its suitability and computational cost depend on the signal characteristics, window length, and implementation.

5.2 Feature Extraction Methods

Feature extraction techniques can be categorized into four main groups: signal transformation methods, decomposition methods, morphological methods, and statistical methods. These methods are described in the following sections.

5.2.1 Signal Transformation Methods

Signal transformation methods are of vital importance in the feature extraction for sleep staging. These methods allow for the conversion of raw signals into different representations that highlight specific characteristics or properties of the data. The selection of a signal transformation method typically relies on the characteristics of the signal under analysis. Some methods analyzed the frequency characteristic of sleep signals, such as FFT [10]. By converting time domain signals into frequency domain signals, it is possible to capture crucial information pertaining to various sleep stages and physiological phenomena, such as sleep spindles and slow-wave activity. Time-frequency analysis methods enable the dynamic resolution of signal frequency over time. These methods include discrete wavelet transform (DWT) [11], optimize flexible analytic wavelet transform (OFAWT) [80], tunable Q factor wavelet transform (TQWT) [81], Hilbert-huang transform (HHT) [21], and short-time Fourier transform (STFT) [13].

5.2.2 Decomposition Methods

Decomposition methods analyze signals from different time and frequency scales by decomposing complex physiological signals into multiple components. For example, wavelet decomposition is utilized to decompose EEG signals into multiple frequency bands, to extract frequency features specific to different sleep stages. In addition, the method effectively realizes signal noise reduction by disassembling the original signal into different components, removing environmental noise, individual motion artifacts and other interferences, and improving the signal quality. Common decomposition methods include singular value decomposition (SVD) [21], variational mode decomposition (VMD) [21], multimodal decomposition which involves frequency-domain decomposition using a set of FIR band-pass filters and time-domain decomposition employing EMD [22], ensemble empirical mode decomposition (EEMD) [81], and wavelet decomposition [82].

5.2.3 Morphological Methods

Morphological methods focus on the morphology and structure of sleep signals and are designed to accurately capture the unique patterns and waveform characteristics that correspond to different stages of sleep. In time domain waveform analysis, the method quantifies parameters such as amplitude, slope, peaks, troughs and duration of the signal. Additionally, the morphological method is also able to detect characteristic events, such as the Pan-Tompkins algorithm for identifying QRS clusters [23]. Furthermore, this method is advantageous for extracting features related to signal amplitude and energy. For example, the amplitude-time mapping model [9] clearly characterizes the transition from high frequency, low amplitude during wakefulness to low frequency, high amplitude during deep sleep. Although morphological methods have significant advantages

Table 4. Artifacts removal approaches.

Signals	Butterworth filter	Chebyshev filters	High-pass filter (HP)	Low-pass filter (LP)	Notch filter	Band-pass filter (BP)	Others
EEG	20th order, 0.5 Hz (HP) [12] 0.3–35 Hz (BP) [63] 10th order, 0.5–49.5 Hz (BP) [64] 10th order, 0.5–32 Hz (BP) [10] 0.5–40 Hz (BP) [65]; 0.5–45 Hz (BP) [66] 4th order, 0.5–30 Hz (BP) [14,67] 0.5–30 Hz (BP) [68] 8th order, 0.5–35 Hz (BP) [69] 14th order, 0.5–18 Hz (BP) [53] 0.5–70 Hz (BP) [21] 2nd order, 1–50 Hz (BP) [70]	50th order, 30 Hz [12]	0.3 Hz [23]	30 Hz [23]	50 Hz [23,71] 48–52 Hz [21] 60 Hz [70]	0.3–35 Hz [72] 0.5–31.5 Hz [73] 0–40 Hz [74] 0.5–50 Hz [75,76] 3rd order, 0.1–40 Hz [71]	Artifact-blocking algorithm [76] ICA-Wavelet [77] Wavelet [8,68] ICA [78]
EOG	-	-	0.3 Hz [23]	30 Hz [23]	50 Hz [23,71]	3rd order, 0.1–40 Hz [71]	ICA [78]
EMG	-	-	10 Hz [23]	75 Hz [23]	50 Hz [23,71]	3rd order, 10–100 Hz [71]	ICA [78]
ECG	5th order, 1 Hz (HP) [57]	-	0.3 Hz [23]	30 Hz [23]	50 Hz [23]	-	Rule threshold [57]
Accelerometer signals	1st order, 1 Hz (HP) [57] 4th order, 0.16–0.3 Hz (BP) [53]	-	-	-	-	-	-
Pulse oximeter signals	4th order, 3 Hz (LP) [53]	-	-	-	-	-	-
Respiration	5th order, 2 Hz (LP) [79]	-	-	-	-	-	Median filter [58]

in waveform characterization, such as sleep spindle wave, K-complex wave detection and QRS cluster identification, their application is limited by the requirement for a high signal-to-noise ratio.

5.2.4 Statistical Methods

The statistical analysis method quantifies signal characteristics from the perspectives of data distribution and dynamics. Specifically, the method first analyzes the overall trend and local change characteristics of sleep signals by calculating basic statistics. For example, it assesses the signal steady-state characteristics using basic statistics such as mean [11,21], variance [23,56], standard deviation [8,16,23], skewness [16,17], and kurtosis [77,81]. Additionally, it performs entropy analysis using sample entropy [7], approximate entropy [23,51], etc., to analyze the complex evolution of sleep stages. Furthermore, it uses fractal analysis to assess the signal's irregularity and long-range correlation using metrics like Hurst index [23] and fractal dimension [7,8,17]. This method is characterized by two advantages: strong interpretability and outstanding computing efficiency. Therefore, it is utilized to create lightweight sleep staging models, such as in situations requiring high real-time requirements or small data volume.

5.3 Features

Features for sleep stage classification can be categorized into six main groups: time-domain features, frequency-domain features, time-frequency features, entropy features, fractal features, and

nonlinearity features. These features are described in the following sections. Table 5 (Ref. [7,8,9,10,11,12,13,14,16,17,20,21,22,23,25,26,44,51,54,55,56,58,60,62,77,80,81,82,83,84]) summarizes these features, including feature types and representative feature names.

5.3.1 Time-Domain Features

Time-domain features quantify temporal patterns in physiological signals through statistical and structural descriptors. Basic statistical descriptors, including central tendency (e.g., mean, median), dispersion (e.g., variance, standard deviation), and distribution properties (e.g., skewness, kurtosis), capture signal magnitude and variability. Signal dynamics are characterized by Hjorth parameters (activity, mobility, complexity) and zero-crossing rates, while autoregressive coefficients model temporal dependencies in neural oscillations. Transient autonomic shifts, such as HRV or muscle tone changes, are analyzed via NN interval counts (e.g., NN20, NN50) and R-peak detection. Time-domain measures include root mean square and Teager energy. Chen et al. [13] also examined the mean absolute difference between successive samples, termed MAE in that study, as a candidate feature reflecting within-segment signal variation. Physiological markers like respiratory depth and body movement indices (e.g., body movement index, amplitude difference accumulation) correlate with stage-specific events (e.g., suppressed breathing in N3, REM-associated muscle atonia). Inter-epoch consistency is further assessed through similarity (e.g., Itakura distance) and correlation (e.g., Pearson's coefficients) measures. Collectively, these features decode sleep physiology through cortical, autonomic, and motor features that distinguish different sleep stages.

5.3.2 Frequency-Domain Features

Frequency-domain features describe the distribution of physiological signal energy in different frequency bands, providing static spectral features necessary for sleep stage analysis. Core metrics include power spectral density (PSD), which quantifies the energy intensity of a signal. Absolute spectral power refers to the total energy of a signal within a specific frequency band. For example, the absolute spectral power of δ waves is usually dominant in stage N3, reflecting the slow-wave activity characteristic of deep sleep. In contrast, relative spectral power denotes the proportion of total power in a particular frequency band. During wakefulness, the relative spectral power of the β band is generally higher, indicating increased high-frequency brain activity. Conversely, during the NREM stage, particularly in deep sleep (e.g., N3), the relative spectral power of the θ wave and the δ wave dominate, marking the characteristics of slow-wave sleep. Additionally, other significant metrics, such as spectral edge frequency (SEF) and Teagle energy, can further enhance the analysis and understanding of sleep signals.

5.3.3 Time-Frequency Features

Time-frequency features can effectively resolve transient physiological patterns in sleep signals by combining time and frequency information, which is crucial for processing non-stationary sleep signals. The phase mean (PM) metric aids in identifying the time-frequency feature demarcation between REM and NREM sleep stages by quantifying the average phase difference of the signals. Conversely, the phase standard deviation (PhSD) reveals the nonlinear dynamic characteristics of sleep stages by analyzing the degree of phase fluctuation. Fourier synchronous compression techniques can pinpoint asynchronous transient events associated with N2 or REM stages, such as sleep spindle waves. Additionally, multiple fractal logarithmic accumulators further analyze scale-dependent energy fluctuations, revealing microstructural changes in the signal during stage transitions. These methods capture dynamic interactions in sleep signals, accurately characterizing the sleep microstructure.

5.3.4 Entropy Features

Entropy-based metrics reveal sleep stage transitions by analyzing the irregularity and complexity of physiological signals. Sample entropy (SampEn) and approximate entropy (ApEn) measure the unpredictability of signals over time, helping to distinguish between the low-complexity EEG patterns of slow-wave sleep and the irregular EEG activity of wakefulness. Alignment entropy (Pen) tracks changes in sequential patterns of neural oscillations, such as the transition from sleep spindle waves to slow waves. Multiscale entropy, such as multiscale sample entropy (MSE) and complex dispersion entropy (RCMDE), quantifies the complexity on different time scales and captures the phenomenon of abrupt changes during REM versus NREM cycling. Through the adaptive parameter adjustment of Tsallis entropy and Rényi entropy, the identification of occasional characteristic waves (such as K-composite waves) and steady-state rhythms (like spindle wave clusters) can be optimized, enhancing robustness against inter-individual variability.

5.3.5 Fractal Features

Fractal features quantify the self-similarity and complexity of physiological signals using scale-invariant geometric analysis. The Higuchi fractal dimension (HFD) and Katz fractal dimension (KFD) evaluate the degree of irregularity in signals, effectively capturing non-stationary patterns of change, such as a significant decrease in the complexity of the EEG during N3. The Hurst index and its expanded form, the generalized Hurst exponent (GHE), provide important tools for signal long-range correlation analysis. The index can effectively distinguish two types of physiological dynamic characteristics: physiological signals such as heart rate show continuous stability during NREM, while REM periods show fluctuating characteris-

Table 5. Feature taxonomy and representative features for sleep stage classification.

Feature types	Categories	Feature names
Time-domain features	Basic statistical descriptors	Mean [9,11,17,21,22,23,56,77], median, variance [9,11,16,17,23,51,56,77,81], average of normal-to-normal (NN) intervals (AVNN) [83], mode, percentiles (first/second/third quartile) [16,17], coefficient of variation [17], maximum, minimum [9,16,17,22,23,56,77,81], mean absolute deviation (MAD), mean detrend, mean nondetrend [23], standard deviation (SD) [8,16,17,21,22,23,62], inter-quartile range (IQR) [10,23,56,80,81], range [16,56], skewness [8,11,16,17,21,22,23,77,84], kurtosis [8,11,16,17,21,22,23,56,77,81,84]
	Signal dynamics & Complexity	Hjorth parameters (activity, mobility, complexity) [8,9,20,22,23,77,82], zero cross rate [17,77], zero cross mean [77], singular value, slope, shape factor [21], autoregressive model parameters [11,13,23], statistical behavior of local extrema (SBLE) [12], fast movement ratio [51]
	Event & Interval analysis	NN20, PNN20 [56], NN50, PNN50 [56,83], number of R peaks, number of repetitions [21], root mean square of successive differences (RMSSD) [51,83], standard deviation of NN Intervals (SDNN), standard deviation of the successive differences (SDSD) [83]
	Energy & Time-series metrics	Energy [8,21,22,23,77], log-energy [7], relative energy ratio [22], teager energy, area [44], mean teager energy (MTE) [23], root mean square (RMS) [13,23,56], mean absolute difference between successive samples (termed MAE in [13]; candidate time-domain feature)
	Physiological-specific features	Mean respiratory depth of exhalation, median respiratory volume during Inhalation, body movement (BM) [58], body movement index (BMI), amplitude difference accumulation (ADA) [26]
	Distance & Correlation	Itakura distance [77], total distance [17], Pearson's coefficient, Pearson's *p*-value [56], maximum-minimum distance (MMD) [20], difference between max and min [23]
Frequency-domain features	-	Spectral edge frequency (SEF) [14,17,23], power [9,13,21,54,55], relative spectral powers [17,23], normalized sub-band power [10], power spectral density (PSD), absolute spectral powers, power ratios [23], very low frequency (VLF), low frequency (LF), high frequency (HF), LF/HF ratio, total power (TP), low frequency normalized power (Lfnorm), high frequency normalized power (Hfnorm) [56], teager energy in the frequency domain [51], proportion of power, variability of the power [55]
Time-frequency features	-	Coherent oscillations [9], fourier synchrosqueezed feature [10], PSD absolute averages [21], average power of the coefficients [84], the first three log-cumulants of the multifractal spectrum [55], phase mean (PM), phase standard deviation [20]
Entropy features	-	Sample entropy [7], Approximate entropy (ApEn) [23,51], information entropy [51], multiscale sample entropy (MSE) [14,51], Fuzzy entropy, Tsallis entropy [8], Kraskov entropy (KE) [20,22], refined composite multiscale dispersion entropy (RCMDE) [11], Shannon entropy, log-energy entropy, norm entropy, threshold entropy, "SURE" entropy [11], spectral entropy (SE) [17,23], permutation entropy (Pen) [17,23], normalized spectral entropy (NSE), Rényi entropy (RE) [20]
Fractal features	-	Fractal dimensions (FD) [7,8,17], Katz fractal dimension (KFD) [22,51], Petrosian fractal dimension (PFD) [20,22,23,77], Higuchi fractal dimension (HFD) [22,77], Hurst exponent (HE) [23], generalized HE [20]
Nonlinearity features	-	Detrended fluctuation (DF) [22], normalized line length (NLL), log root sum of sequential variations (LRSSV) [20], HRV triangular index [56], graph's characteristics (betweenness centrality, degree distribution) [60], curve length, mean curve length (MCL) [23], visibility graph features [25], average amplitude change (AAC), clearance factor (CLF), simple square integral (SSI) [80]

tics of the autonomic nervous system. Petrosian fractal dimension (PFD) is suitable for real-time monitoring scenarios, as it can rapidly estimate signal complexity through binary sequence transformation. The Generalized fractal dimension resolves the hierarchical structure of physiological signals through multi-scale density functions. It can quantify the spatial and temporal distribution pattern of sleep spindle wave in cortical EEG analysis, and characterize the coupling strength of respiration-heart rate in cardiorespiratory testing.

5.3.6 Nonlinearity Features

Nonlinear characterization captures chaotic dynamics and transient interactions in sleep-related signals. DFA quantifies remote correlations in nonstationary sequences, revealing changes in EEG amplitude during transitions between REM and NREM sleep. Converting time series into complex network structures by using graphic techniques (including visible graphic features and intermediate centrality) is helpful to reveal the nonlinear relationship between heart rate variations and muscle activities. Amplitude variability indicators (AAC, CLF) and curve length indices (NLL, MCL) can be used to track rapid changes in respiratory or movement data. By combining geometric and cumulative nonlinear effects, HRV delta indices and logarithmic roots of sequence variation (LRSSV) increase sensitivity to sleep fragmentation patterns.

5.4 Feature Selection Techniques

In certain instances, the extracted features may be redundant and result in a dataset with a too high dimensionality. Including irrelevant or redundant features in the classification model may introduce unnecessary complexity, which may also negatively impact the performance. Feature selection techniques are utilized to identify and choose a subset of the most informative features from the initial feature set, eliminating redundant features and enabling the model to concentrate on pertinent information. These techniques have the potential to enhance the predictive accuracy, interpretability, and efficiency of sleep scoring models. The three primary categories of feature selection techniques are elaborated upon below.

5.4.1 Dimensionality Reduction Methods

Dimensionality reduction methods map high-dimensional features into a low-dimensional space, preserving the main information while reducing noise and redundancy. Principal component analysis (PCA) [63] and SVD [85] are common linear methods for signal denoising and compression, such as fusing multi-channel EEG signals into a few principal components. These methods are computationally efficient and enhance model generalization, but they may obscure the physical interpretability of the features and complicate direct connections to physiological mechanisms.

5.4.2 Filter Methods

Filtering methods perform feature selection by independently evaluating the statistical correlation between features and target variables, without relying on model training. Typical methods include sample nearest-neighbor discriminative weighting (ReliefF) [66,71], minimum redundancy maximum relevance (mRMR) [65,71,76] and mutual information (MI) [86]. These methods are suitable for the initial screening of high-dimensional data, such as quickly removing redundant band energy features from EEG signals. Their advantages include fast computation and high interpretability; however, they may overlook interaction effects among features. For example, these methods are unable to capture the impact of synergistic changes across multiple frequency bands on sleep staging.

5.4.3 Hybrid and Novel Methods

Hybrid approaches integrate classical algorithms with innovative strategies to address complex feature selection challenges. For instance, combining generalized canonical correlation analysis (CCA) with ReliefF [59] enhances discriminative power for multimodal data, such as EEG and ECG fusion. Additionally, online streaming feature selection (OSFS) [87] adapts to dynamic data streams in real-time monitoring. Advanced techniques like multi-instance InfoNCE [74] optimize contrastive learning for noisy labels. While these methods provide flexibility, they require domain-specific customization to balance performance and practicality.

5.5 Summary and Recommendations for Lightweight Feature Engineering

Designing a lightweight sleep staging model requires trading off computational efficiency against noise robustness and clinical interpretability. Based on the preceding review, several specific strategies emerge for home monitoring environments:

Feature robustness in noisy environments: Time-domain metrics (such as variance or zero-crossing rate) demand minimal processing power but degrade rapidly under motion artifacts. For wearable deployment, frequency-domain features—specifically frequency band power calculated via fast Fourier transform—provide a much better balance. They resist baseline wander and environmental noise far better than raw time-domain statistics while keeping computational overhead acceptable for microcontrollers.

Computational cost of non-linear features: Entropy measures (like sample entropy) excel at detecting subtle stage shifts but typically require $O(N^2)$ time complexity. This makes them impractical for continuous real-time monitoring on edge devices. If tracking non-linear dynamics is strictly necessary, computationally simplified metrics like the Petrosian Fractal Dimension (PFD) should replace multiscale entropy to save CPU cycles.

Table 6. Sleep stage classification methods (single-modality).

Signals	Ref	Classifiers	Subjects/Recordings	Datasets	Performances		
					Acc	Kp	Others
EEG	Zhao et al. [8]	SVM	20 subjects (20H)	XSLEDF	4-85.93%	-	-
	An et al. [9]	C-SVM, OC-SVM	20 subjects (20H)	XSLEDF	6-93%	-	Sen / Pre / F1 95.45%/99.42%/97.40%(W) 65.99%/46.52%/54.57%(S1) 89.29%/91.10%/90.19%(S2) 77.12%/79.11%/78.10%(S3) 83.73%/92.05%/87.70%(S4) 92.40%/65.27%/76.50%(REM)
	Zaidi et al. [10]	SVM	8 subjects (4H, 4P) 20 subjects (20H) 20subjects (20H)	SLEDF XSLEDF DRMS	SLEDF/XSLEDF/DRMS: 2-98.2%/97.7%/97.1% 3-94.9%/95.0%/91.4% 4-92.4%/93.2%/87.4% 5-91.2%/92.4%/85.6% 6-89.8%/91.4%/82.5%	-	-
	Seifpour et al. [12]	Multi-class SVM	8 subjects (4H, 4P) 20 subjects (20H)	SLEDF DRMS	SLEDF/DRMS: 2-97.9/- 3-94.5/- 4-92.8/- 5-91.8/83.3 6-90.6/83.4	SLEDF/DRMS: 2-0.9600/- 3-0.9100/- 4-0.8800/- 5-0.8700/0.7700 6-0.8500/0.7700	SLEDF: Sen / Spe / Acc 98.6%/96.4%/97.7%(W) 41.1%/99.0%/96.6%(S1) 91.2%/96.4%/95.1%(S2) 58.6%/98.1%/96.3%(S3) 79.9%/99.2%/98.3%(S4) 84.7%/97.4%/96.0%(REM)
	Diykh et al. [16]	LS-SVM	18 subjects (10H, 8P) 13 subjects	ISRUC-Sleep XSLEDF	XSLEDF: 6-96.0% ISRUC-Sleep: 5-96.74%	XSLEDF/ISRUC-Sleep: 0.8700/0.9000	XSLEDF: Acc/Sen/Kp 99.0%/0.95/0.97(W) 97.0%/0.83/0.90(S1) 97.0%/0.94/0.89(S2) 94.1%/0.86/0.82(S3) 91.0%/0.85/0.80(S4) 98.0%/0.87/0.89(REM) ISRUC-Sleep: Acc/Sen/Kp 99.5%/0.99/0.98(W) 97.0%/0.86/0.87(S1) 95.5%/0.95/0.90(S2) 95.0%/0.93/0.89(SWS) 96.7%/0.92/0.87(REM)
	Huang et al. [17]	SVM	30 subjects (30H)	XSLEDF	2-98.28% 3-95.50% 4-94.28% 5-93.08% 6-92.34%	6-0.8407	Sen 99.05%(W) 37.26%(S1) 88.28%(S2) 57.90%(S3) 76.74%(S4) 75.63%(REM)
	Chung et al. [19]	SVM	20 subjects (20H) 25 subjects (25P)	DRMS UCD	DRMS/UCD: 2-74.7%/70.6% 3-66.3%/57.1%	DRMS/UCD: 2-0.3080/0.2380 3-0.3720/0.2440	DRMS: Sen/Spe/F1 2-66.9%/77.6%/0.445 UCD: Sen/Spe/F1 2-57.6%/75.3%/0.407

Table 6. Continued.

Signals	Ref	Classifiers	Subjects/Recordings	Datasets	Performances			
					Acc	Kp	Others	
	Memar et al. [20]	RF	8 subjects (4H, 4P) 25 subjects (25P)	SLEDF UCD	SLEDF/UCD: 2-98.98%/95.75% 3-98.37%/95.37% 4-98.24%/95.51% 5-98.44%/95.31% 6-98.50%/95.85%	-	SLEDF: Sen/Spe/Acc 99.68%/98.10%/98.94%(W) 68.11%/99.66%/98.41%(S1) 95.96%/98.15%/97.63%(S2) 78.43%/99.31%/98.40%(S3) 91.78%/99.62%/99.29%(S4) 92.18%/99.04%/98.32%(REM) 87.69%/98.98%/98.50%(Overall)	UCD: Sen/Spe/Acc 93.10%/96.17%/95.51%(W) 73.35%/97.22%/93.29%(S1) 93.13%/92.24%/92.54%(S2) 57.83%/99.89%/98.50%(S3) 89.91%/99.31%/98.40%(S4) 87.51%/98.50%/96.85%(REM) 82.47%/97.22%/95.85%(Overall)
	Jiang et al. [22]	RF+HMM	30 subjects (20H, 10P) 61 subjects (61H)	XSLEDF MASS	XSLEDF(20H/10H,10P)/ MASS 5-0.930/0.894/0.793	XSLEDF(20H/10H,10P)/ MASS 5-0.8620/0.8350/0.6900	XSLEDF(20H/10H,10P)/MASSF1 5-0.930/0.895/0.783	
	Supriya et al. [25]	KNN	8 subjects (4H, 4P)	SLEDF	2-98.78% 3-98.21% 4-98.13% 5-97.73% 6-97.91%	-	Acc: W/S1/S2/S3/S4/REM 93.4%/87.9%/99.5%/99.61%/99.16%/97.34%	
	Alickovic et al. [84]	RotSVM	30 subjects (20H, 10P)	XSLEDF	5-91.10%	5-0.88	Sen: W/N1/N2/SWS/REM/Overall 98.59%/48.78%/95.54%/97.94%/94.59%/84.46%	
	Huang et al. [81]	Bagging	8 subjects (4H, 4P) 16 subjects (8H, 8P)	SLEDF XSLEDF	SLEDF/XSLEDF 2-98.16%/97.37% 3-94.72%/92.89% 4-94.71%/90.66% 5-91.29%/89.37% 6-90.11%/88.14%	-	-	
	Sharma et al. [82]	Ensemble bagged trees	2431 recordings	CAP, UCD, SLEDF, ISRUC, MIT-BIH	5-83.2%	5-0.7345	Spe: 95.8% Sen: W/N1/N2/SWS/REM/Overall 91.5%/20.8%/92.7%/65.0%/73.6%/82.9%	
	Abdulla et al. [60]	Ensemble model	8 subjects (4H, 4P) 25 subjects (25P)	SLEDF UCD	SLEDF/UCD: 6-93.1%/94%	SLEDF/UCD: 6-0.9150/0.9300	SLEDF: Acc/Kp/Sen 98%/0.95/0.85(W) 97%/0.93/0.84(S1) 90%/0.91/0.82(S2) 89%/0.83/0.79(S3) 83%/0.81/0.77(S4) 97%/0.96/0.88(REM)	UCD: Acc/Kp/Sen 96%/97%/0.87(W) 98%/92%/0.89(S1) 96%/94%/0.86(S2) 84%/90%/0.80(S3) 88%/87%/0.76(S4) 98%/97%/0.86(REM)
	Taran et al. [80]	Ensemble bagged trees	8 subjects	-	2-99.36% 3-97.56% 4-96.48% 5-96.39% 6-96.03%	2-0.9527 3-0.9386 4-0.9438 5-0.9477 6-0.9443	-	

Table 6. Continued.

Signals	Ref	Classifiers	Subjects/Recordings	Datasets	Performances		
					Acc	Kp	Others
	Palotti et al. [88]	Extra trees	1817 subjects	MESA	2-86.7	-	Sen/Spe/Pre/F1 95.4%/88.3%/67.9%/77.6%
	Abdulla et al. [89]	Ensemble model	180 recordings	XSLEDF (termed 'Sleep-EDF' in [89]) UCD	XSLEDF/UCD LOSO:93.21%/92.93%	-	-
	Ghasemzadeh et al. [72]	SVM	25 subjects (25P) 13 subjects 10 subjects	XSLEDF ISRUC-Sleep	XSLEDF/ISRUC-Sleep 2-99.30%/- 3-97.38%/- 4-95.75%/- 5-94.83%/82.25% 6-93.92%/-	SLEDF/ISRUC-Sleep 2-0.9794/- 3-0.9336/- 4-0.9002/- 5-0.8876/0.7702 6-0.8680/-	SLEDF/ISRUC-Sleep 99.37%/88.73%(W) 44.39%/50.33%(N1) 93.59%/84.96%(N2) 84.24%/97.35%(N3) 91.86%/88.38%(REM)
	Satapathy et al. [64]	Ensemble model	8 subjects (4H,4P) 61 recordings (20H)	SLEDF XSLEDF	SLEDF/XSLEDF 6-91.10%/81.32%	SLEDF/XSLEDF 6-0.8700/0.7700	SLEDF: Sen/Pre/F1 95.65%/98.01%/96.81%(W) 86.22%/83.97%/85.08%(S1) 92.51%/90.38%/91.43%(S2) 90.32%/80.72%/85.25%(S3) 79.65%/84.96%/82.21%(S4) 71.93%/71.61%/71.76%(REM)
	Karimzadeh et al. [69]	Distributed DT	20 subjects (20H) 140 subjects (140P)	XSLEDF SHHS	XSLEDF/SHHS 6-88.97%/5-83.17%	-	XSLEDF: Acc/Spe/Sen 97.4%/97.39%/97.43%(W) 91.37%/93.56%/81.84%(S1) 88.05%/90.81%/74.70%(S2) 79.75%/93.18%/47.04%(S3) 82.88%/88.21%/56.85%(S4) 94.37%/94.75%/92.76%(REM) 88.97%/92.98%/75.10%(Total) SHHS: Acc/Spe/Sen 88.17%/86.99%/69.35%
	An et al. [90]	Hierarchical weighted SVM	20 subjects (10H,10P)	XSLEDF	6-88.52%	-	Sen/Pre/F1 97.75%/96.86%/97.31%(W) 29.32%/42.11%/34.57%(S1) 89.41%/85.64%/87.48%(S2) 51.35%/66.67%/58.02%(S3) 79.07%/72.86%/75.84%(S4) 69.26%/68.31%/68.78%(REM)
	Liu et al. [61]	XGBoost	8 subjects (4H,4P) 20 subjects (20H) 111 subjects	SLEDF DRMS SHHS	SLEDF/DRMS/SHHS 4-93.1%/86.4%/87.5% 5-91.9%/83.4%/85.8%		

Table 6. Continued.

Signals	Ref	Classifiers	Subjects/Recordings	Datasets	Performances		
					Acc	Kp	Others
	Satapathy et al. [66]	DT, KNN, RF	8 subjects (4H, 4P)	ISRUC-Sleep	DT/KNN/RF 2-93.42%/94.40%/96.70%(P) 2-92.40%/93.73%/95.33%(H)	-	-
	Liang et al. [91]	Ensemble model	20 subjects (20H) 20 subjects (20H)	XSLEDF(SC) ear-Feature	XSLEDF(SC)/ear-Feature 5-84.3%/84.1%	XSLEDF(SC)/ear-Feature 5-0.7820/0.7790	-
	Liu et al. [92]	RF	20 subjects (20H) 154 recordings 10 subjects (10H)	SleepEDF-20 XSLEDF ISRUC-Sleep-III	SleepEDF-20/XSLEDF/ ISRUC-Sleep-III 5-96.33%/94.6%/88.5%	SleepEDF-20/XSLEDF/ISRUC- Sleep-III 5-0.9400/0.9200/0.8300	SleepEDF-20/XSLEDF/ISRUC-Sleep-III 94.2%/89.3%/86.5%(MF1) F1: 99.1%/99.5%/95.2%(W) 87.4%/75.1%/78.9%(N1) 96.0%/93.4%/91.6%(N2) 94.5%/84.0%/89.2%(N3) 93.9%/94.5%/77.3%(REM)
	Yilmaz et al. [93]	RF	128 subjects (128H)	BOAS	5-0.87	5-0.7700	F1: W/N1/N2/N3/REM 0.86/0.19/0.93/0.70/0.82
	Salfi et al. [94]	Ensemble model	10 subjects (10 H) 35 recordings	Private	5-88.83%	5-0.8410	F1: W/N1/N2/N3/REM 86.37%/53.20%/87.93%/89.53%/93.99%
EOG	Rahman et al. [11]	RUSBoost	8 subjects (4H, 4P) 20 subjects (20H) 10 subjects	SLEDF XSLEDF ISRUC-Sleep	SLEDF/XSLEDF/ISRUC-Sleep 2-96.31%/-/- 3-93.44%/-/- 4-91.11%/-/- 5-89.10%/-/- 6-86.61%/90.0%/84.7%	-	Acc: SLEDF/XSLEDF/ISRUC-Sleep 63.5%/64.6%/64.4%(S1)
ECG	Yücelbas et al. [21]	RF	28 subjects (21H, 7P)	Private	3-87.11%(H)/78.08%(P)	3-0.8690(H)/0.7820(P)	F1: 3-73.69(H)/57.15%(P)
	Surantha et al. [83]	ELM	18 recordings	MIT-BIH	2-82.10% 3-76.77% 4-71.52% 6-62.66%	-	-
	Chung et al. [19]	SVM	83 subjects 20 subjects (20H) 25 subjects (25P)	CGMH DRMS UCD	CGMH/DRMS/UCD 2-75.8%/74.7%/70.6% 3-67.6%/66.3%/57.1%	CGMH/DRMS/UCD 2-0.3220/0.3080/0.2380 3-0.3900/0.3720/0.2440	-
Respiratory sound	Xue et al. [59]	Ensemble model	13 subjects (13H)	Private	2-84.4% 3-75.2%	2-0.4900 3-0.4200	-

Abbreviations: H, health; P, patient; KNN, k-nearest neighbor. The numerical prefixes (2–6) preceding the reported performance values indicate the number of classes in the corresponding classification task. For example, "5-" denotes five-class classification.

Balancing interpretability and efficiency: Algorithms deployed in healthcare must remain clinically interpretable. Projection-based dimensionality reduction methods like PCA effectively compress data but destroy its physical meaning. To strike a better balance, offline filter methods (such as mRMR or ReliefF) are preferable. They allow developers to identify a minimal, physiologically meaningful feature subset (e.g., delta-band power or HRV) that the edge device can compute directly, avoiding complex matrix inversions during inference.

In short, effective lightweight deployment relies on extracting a narrow set of noise-resistant frequency bands combined with basic time-domain statistics, rather than using complex non-linear calculations or dense data projection on the device.

6. Sleep Stage Classification Methods

In sleep research, sleep stage classification can be categorized into three principal groups according to the variety and quantity of physiological signals used: single-modality, dual-modality, and multi-modality sleep stage classification. Single-modality methods prioritize computational efficiency by analyzing individual signal types, while multi-modality methods enhance classification robustness by capturing cross-system physiological interactions. A comprehensive comparison of these methods is presented in Table 6 (Ref. [8,9,10,11,12,16,17,19,20,21,22,25,59,60,61,64,66,69,72,80,81,82,83,84,88,89,90,91,92,93,94]), Table 7 (Ref. [18,52,58,65,67,77,95,96,97,98]), and Table 8 (Ref. [7,23,26,31,71,78,99,100,101]), which includes signal modalities, datasets, and performance results. The details of these methods and evaluation metrics are elaborated upon below.

6.1 Single-Modality Sleep Stage Classification

Single-modality approaches rely on information from a single type of signal to classify sleep stages. Because they eliminate the need for cross-channel synchronization and enforce the smallest possible memory footprint, single-modality models are inherently well-suited to the demands of strict edge deployability. These include EEG signals [7,12,60,61,63,64,66,69,72,74,76,86,87,89,90,102,103,104,105], HRV data [19,83,106,107,108], respiratory sound [59] and airflow signal [79].

Since EEG can directly reflect the electrophysiological activities of the cerebral cortex in different sleep stages, most existing studies focus on the analysis of single-channel EEG signals, resulting in the following representative methods. Seifpour's team [12] proposed a novel automated sleep staging system based on the statistical behavior of local extremes (SBLE). Extracting SBLE relies entirely on basic time-domain operations, satisfying the low computational complexity requirement of lightweight frameworks. Abdulla's research group [60] established a machine-learning framework based on correlation graphs. By converting 30-second EEG segments into subgraph structures and extracting modular features as inputs to an ensemble classifier, the method effectively captures the temporal correlation properties of EEG signals. Supriya et al. [25] proposed an EEG analysis method based on weighted visibility graphs, by converting the signal into a complex network, extracting topological features such as modularity coefficients and average weighted degree, and combining them with a KNN classifier to achieve sleep staging. The method utilizes the graph structure to characterize the signal dynamics and effectively preserves the nonlinear dynamics of EEG. Huang et al. [81] proposed a hybrid method combining EEMD and TQWT. The method effectively improves the classification robustness of sleep staging by combining multiscale feature extraction with Bagging integration learning. However, from a lightweight perspective, iterative decomposition methods like EEMD approach the computational limits of standard microcontrollers, demanding careful hardware optimization prior to edge deployment. For wearable applications, Wang et al. [109] evaluated TinySleepNet using single-channel EEG, reporting a mean five-stage accuracy of 74.81% and Cohen's κ of 0.6328 across eight separately evaluated EEG channels. The headband also recorded inertial and temperature signals, but these were not used as model inputs.

Although single-channel EEG analysis dominates the field of sleep staging, research on multichannel fusion methods has gradually emerged. An's team [9] designed a cascade SVM framework to address the difficulty of classifying S1 and REM sleep. Firstly, C-SVM was used to identify the merged stages of S1 and REM, and then fine differentiation was achieved by OC-SVM. Satapathy et al. [64] constructed a hybrid model of random forest (RF) and eXtreme gradient boosting (XGBoost), which improved the overall accuracy to 91.1% by feature complementation of dual-channel EEG signals.

Sleep staging studies based on EEG signals have yielded competitive performance. However, EEG signal acquisition can be cumbersome and inconvenient. Therefore, there has been a push towards using EOG signals for sleep staging, which are easier to obtain. Rahman et al. [11] introduced an automatic sleep scoring method using EOG signals. The classification of sleep stages was performed using SVM, RF, and random under-sampling boosting (RUSBoost). EOG signals provide valuable information about eye movements during sleep, offering a practical alternative for sleep staging applications.

As wearable devices grow by leaps and bounds, it is comfortable and convenient to automatically analyze the sleep stages by measuring continuous physiological parameters, such as a single-lead ECG and respiratory sound. From a deployment perspective, these particular modalities produce constrained, low-dimensional data streams that align well with the resource limits of TinyML platforms.

Table 7. Sleep stage classification methods (dual-modality).

Signals	Ref	Classifiers	Subjects/ Recordings	Datasets	Performances		
					Acc	Kp	Others
Acceleration & pressure signals (body movement, heart rate)	He et al. [52]	-Rule-based comparator (non-ML)	-	Private	2-95.1% 4-79.7%	-	Sen: 2-90.2%(W)/100%(Sleep) 4-90.2%(W)/78.1%(LS) 73.5%(SWS)/77.1%(REM)
Radar sensing signals (body movement, respiration)	Gaiduk et al. [58]	Multinomial logistic regression	35 recordings	Private	3-73%	3-0.44	Sen/Pre 46.1%/48.9%(W) 79.9%/87.4%(NREM) 65.5%/43.8%(REM)
EEG, EOG	Sekkal et al. [77]	SVM	19 subjects	XSLEDF	5-89.1%	-	Pre/Sen/F1 0.95/0.95/0.95(W) 0.65/0.49/0.56(N1) 0.89/0.92/0.91(N2) 0.90/0.90/0.90(N3) 0.86/0.87/0.86(REM)
	Zhou et al. [67]	Bagging	8 subjects 20 subjects (20H) 95 recordings	SLEDF DRMS XSLEDF	SLEDF/DRMS/XSLEDF 5-93.57%/86.36%/88.93% 6-92.56%/84.18%/88.42%	-	-
	Wei et al. [18]	LightGBM	78 subjects (78H) 20 subjects (20H) 25 subjects (25P)	XSLEDF-78 XSLEDF-20 UCD	XSLEDF-20/XSLEDF-78/UCD 5-85.3%/80.8%/79.1%	XSLEDF-20/XSLEDF-78/UCD 5-0.80/0.74/0.72	XSLEDF-20/XSLEDF-78/UCD MF1: 78.7%/73.6%/76.0% XSLEDF-20: F1 W/N1/N2/N3/REM 92.7%/44.3%/87.2%/84.9%/84.5%
	Xu et al. [96]	XGBoost	75 subjects (75H)	XSLEDF	5-87.0%	5-0.81	F1: 5-86.6% W/N1/N2/N3/REM 93.64%/48.54%/88.85%/85.98%/ 80.54%

Table 7. Continued.

Signals	Ref	Classifiers	Subjects/ Recordings	Datasets	Performances		
					Acc	Kp	Others
EEG, ECG	Lyu et al. [65]	SVM	10 subjects (10H)	ISRUC-Sleep	ISRUC-Sleep/MASS 5-84.3%/82.2%	ISRUC-Sleep/MASS 5-0.794/0.734	F1: ISRUC-Sleep/MASS: 0.819/ 0.754
			62 subjects (62H)	MASS			F1: W/N1/N2/N3/REM 0.904/0.511/0.830/0.933/0.876
	Rashidi et al. [95]	Ensemble model	18 recordings	MIT-BIH	3-97.14% 4-95.18%	-	-
			77 subjects (77H)	XSLEDF	5-92.7% 6-85.64%		
PPG&Acceleration	Silva et al. [98]	RNN	1552 recordings	Private	4-71.6%	4-0.56	Sen W/LS/SWS/REM 71.3%/67.5%/75.9%/72.2%
ECG&Acceleration	Davidson et al. [97]	Wav2sleep	13 subjects (13P) 105 recordings	Private	4-77.1%	4-0.679	-

Abbreviations: H, health; P, patient; RNN, recurrent neural network. The numerical prefixes (2–6) preceding the reported performance values indicate the number of classes in the corresponding classification task. For example, "5-" denotes five-class classification.

Table 8. Sleep stage classification methods (multi-modality).

Signals	Ref	Classifiers	Subjects/ Recordings	Datasets	Performances		
					Acc	Kp	Others
EEG, EOG, EMG, ECG	Yan et al. [23]	RF	6 subjects	CAP	6-86.24%	-	Sen/Pre/Spe 80.2%/87.8%/99.0%(W) 66.3%/45.0%/97.4%(S1) 84.2%/89.3%/93.8%(S2) 77.6%/60.2%/95.9%(S3) 93.3%/95.0%/98.8%(S4) 90.9%/92.4%/97.7%(REM)
EEG, EOG, EMG	Cao et al. [71]	RF	28 subjects	Private	5-93.9%	-	Pre/Sen/F1 82.5%/83.1%/82.8%
	Jiang et al. [99]	Bagging	61 recordings	MASS	5-81.2%	5-0.722	Sen/Pre/F1 0.753/0.837/0.793(W) 0.386/0.472/0.425(N1) 0.872/0.861/0.866(N2) 0.773/0.728/0.750(N3) 0.905/0.850/0.877(REM)
	Hamouda et al. [78]	Ensemble model	20 subjects (20H)	XSLEDF	6-89.2%	6-0.87	-
	Jain et al. [31]	RUSBoost	-	SLEDF, XSLEDF, ISRUC, DRMS	SLEDF/XSLEDF/ISRUC-S1/-S2/-S3/DRMS-SUB/ DRMS-PAT 5-83.1%/90.0%/84.4%/82.1%/81.5%/79.9%/73.7%	-	-
	Wang et al. [101]	XGBoost	200 recordings	SHHS1	5-83.24%	-	-
Respiration, heartbeat, and body movement	Hong et al. [26]	Subspace KNN	11 recordings	Private	2-89.6% 3-85.8% 4-81%	-	-
	Pérez-Ávila et al. [7]	KNN	12 subjects (12H) 12 recordings	Private	3-69.0%	-	Sen/Pre 52.5%/48.3%(W) 39.0%/52.5%(REM) 88.2%/73.1%(NREM)
19 channels of PSG device	Arslan et al. [100]	Extra trees	50 patients	Private	5-95.2%	-	Pre/Sen/F1 0.953/0.952/0.945

Abbreviations: H, health; P, patient. The numerical prefixes (2–6) preceding the reported performance values indicate the number of classes in the corresponding classification task. For example, "5-" denotes five-class classification.

Surantha et al. [83] presented a study on sleep stage classification using HRV features extracted from ECG signals. The authors developed a joint model that combined an extreme learning machine (ELM) with particle swarm optimisation (PSO) for feature selection and determining the optimal number of hidden nodes. Xue et al. [59] studied a new method for non-contact sleep stage detection using respiratory sound. The detection was performed using classifiers such as DT, SVM, KNN, and ensemble classifiers.

6.2 Dual-Modality Sleep Stage Classification

Dual-modality approaches for sleep stage classification often involve integrating information from two distinct types of signals or sensors to enhance the accuracy and depth of the analysis. Under the lightweight framework, dual-modality systems face a trade-off between pursuing greater clinical accuracy and adhering to stringent memory limitations. One example is the fusion of EEG and ECG signals for sleep scoring, as seen in certain research studies [77]. This combined approach recognizes that sleep manifests not only in changes in EEG patterns but also in variations in heart rate patterns. The integration of physiological data from both EEG and ECG signals provides a more comprehensive view of sleep dynamics.

Additionally, recognizing the role of body movement as a valuable indicator of sleep patterns, some studies have explored combinations such as body movement with heart rate [52,97], body movement with respiratory signals [58] or accelerometer with photoplethysmography (PPG) signal [53,98]. The incorporation of accelerometer data offers distinct advantages for edge computing; the derivation of gross motor activity metrics imposes minimal computational burden while substantially enhancing the detection accuracy of wakefulness. By incorporating information on body movement, researchers aim to enhance the robustness of sleep staging algorithms, acknowledging that movement patterns can provide valuable supplementary data to ECG-based or respiratory-based sleep stage classification. These dual-modality approaches leverage the complementary nature of different physiological signals to improve the overall accuracy and reliability of sleep stage classification methodologies.

6.3 Multi-Modality Sleep Stage Classification

Multi-modality approaches in sleep stage classification involve the integration of data from three or more different types of signals or sensors, aiming to enhance the accuracy and granularity of sleep staging. The combination of multiple signals allows for a more detailed characterization of sleep stages, distinguishing between wakefulness, REM sleep, and non-REM sleep stages with greater precision. The integration of three or more continuous data streams, however, inherently pushes against the memory constraints typical of microcontroller-based systems. In various studied papers, two common combinations of different modality signals were identified. One is the combination of EEG, EOG, and EMG signals [99,101,110,111,112]. Integrating EEG, EOG, and EMG signals offers a comprehensive perspective on sleep architecture. EEG captures brainwave activity, EOG monitors eye movements, and EMG measures muscle activity. This combination enables a more nuanced and accurate assessment of sleep stages, considering both neural and muscular aspects. Hamouda et al. [78] presented an ensemble learning approach for automated sleep staging using multi-channel PSG recordings, such as EEG, EOG, and EMG. Implementing full clinical PSG analysis on an edge device necessitates aggressive dimensionality reduction. Approaches capable of local deployment typically extract only a minimal set of statistical features per signal channel to circumvent memory limitations [26,109,113]. Another approach combines cardiac, respiratory, and movement-related information for sleep staging [26,113]. A related approach extracts several physiological feature groups from a single sensing modality. Du et al. [85] combined raw triaxial accelerometry with HRV, respiratory rate variability (RRV), and activity features derived from the same recordings for sleep-stage classification. The discrimination of sleep staging is enhanced by the multidimensional feature fusion of respiratory rhythm, heartbeat variability and movement patterns. This approach compensates for the information limitations of single signals, and especially improves the efficacy of differentiating light sleep (stage N1) from wakefulness significantly. Since these particular vital signs can be effectively summarized into informative scalar metrics (e.g., heart rate in beats per minute, respiration rate), this multi-modality strategy maintains a low parameter footprint. This characteristic renders it highly viable for continuous and unobtrusive monitoring in home settings.

6.4 Clinical Applicability: Healthy vs. Patient Cohorts

An analysis of Tables 6,7,8 indicates a recurring performance gap between models evaluated on healthy cohorts (H) and those tested on patient data (P). While many lightweight pipelines report accuracies above 90% on datasets dominated by healthy individuals (e.g., the healthy subset of Sleep-EDF), performance often degrades on clinical databases such as UCD (sleep apnea) or ISRUC (mixed sleep disorders).

This degradation is commonly associated with greater heterogeneity and fragmentation in patient sleep. For example, frequent micro-arousals, event-driven transitions (e.g., respiratory events in OSA), and comorbidities or medication effects can increase signal non-stationarity and artifact burden, weakening the separability of standard sleep-stage patterns and thereby reducing the reliability of automated sleep staging. Under these conditions, EEG-only pipelines may generalize less reliably, particularly when trained primarily on healthy data.

We therefore highlight that multi-modal designs (e.g., EEG combined with ECG and/or respiratory signals) can be advantageous in some patient settings, as they provide complementary autonomic and respiratory information that remains informative even when cortical EEG patterns are transiently disrupted. Nevertheless, modality robustness is condition-dependent, and models intended for home-care or clinical use should be validated on patient cohorts matching the target disorder(s) and acquisition conditions. Accordingly, performance metrics obtained from healthy populations should not be directly extrapolated to clinical settings; empirical evaluation on patient datasets is essential.

6.5 Common Performance Evaluation Metrics

Standard metrics like overall accuracy, sensitivity, specificity, and precision provide a basic evaluation of sleep staging models. However, relying on overall accuracy often overstates a model's true capability because sleep datasets are inherently imbalanced. The N1 stage, for example, makes up a very small fraction of total sleep time but remains consistently the most difficult to classify. A model can easily achieve a high overall accuracy by defaulting to majority classes, such as N2 or N3, while performing poorly or failing entirely on N1.

To prevent this class imbalance from masking poor performance, evaluations must rely on metrics that treat all sleep stages equally. Balanced accuracy, calculated as the arithmetic mean of per-class recall, ensures that minority classes contribute equally to the final score. Similarly, the Macro-F1 score takes the unweighted mean of the F1 scores across all stages, heavily penalizing algorithms that ignore minority classes like N1. Beyond single-number summaries, examining per-class recall alongside a confusion matrix is necessary to diagnose specific error patterns, such as how often N1 is confused with Wake or N2. Cohen's Kappa is also widely used alongside these approaches, as it evaluates the agreement between predicted and actual sleep stages while directly correcting for chance.

7. Challenges and Future Prospects

7.1 Challenges

7.1.1 Lack of Large-Scale Benchmark Labeled Datasets

To assess the robustness of automatic sleep scoring model, a large-scale benchmark labeled dataset is required. However, obtaining large labeled datasets can be difficult due to the need for expert annotation and time-consuming manual scoring. Besides, the manual judgment based expert annotation can be subjective and prone to inter-rater variability. The majority of studies on sleep stage classification rely on publicly available datasets, while only a few studies create and share their own datasets. The limited dataset size and the absence of standardized ground truth may hinder the training and generalization capabilities of automatic sleep stage classification methods and makes it challenging to evaluate different methods.

7.1.2 Data Imbalance

In a normal sleep cycle, sleep stages are often imbalanced, with certain stages occurring more frequently than others. For example, N2 makes up the largest proportion of sleep time, ranging from 45% to 55% of a sleep cycle, while the proportion of W and N1 is typically around 5%. This imbalance has the potential to result in biased models that exhibit a preference for the majority class, making it challenging to effectively classify the minority classes. Consequently, this situation may lead to high overall accuracy but lower accuracy specifically for the minority class. In addition, stages W and N1 are inherently difficult to distinguish between each other, because except for the difference in brainwave activity, the characteristics of these two sleep stages are closely similar. These issues make data imbalance a problem that researchers need to solve. However, the temporal dependencies of the sleep cycle is complex, which often includes both short-term and long-term correlations, making it inappropriate to address data imbalance simply by random sampling. The aforementioned challenges render addressing the data imbalance issue in the automated sleep staging task challenging.

7.1.3 Complex Temporal Dependencies

The transition of sleep stages exhibits a potentially temporal dependency, characterized by a complex pattern that cannot be fully captured solely through segmented physiological signal characteristics. At present, quite a number of studies have used models with temporal feature extraction ability to address this issue, such as HMM, RNN, temporal CNN, etc., to realize the classification from sequence to sequence. Nonetheless, modeling the intricate dynamics and transitions among sleep stages remains a challenge, especially when considering the long-term dependencies. Consequently, the current outcomes do not fully satisfy meeting criteria requirements for clinical diagnosis.

7.1.4 High Individual Variability in Sleep Patterns

A generalized strategy for classifying sleep stages is difficult to design because individual sleep patterns can differ greatly. Variability in sleep disorders, age, gender, and other factors further complicates the classification process. Many studies only selected a certain ethnic group of subjects in a certain age range to form a dataset to develop automatic sleep stage classification methods, and obtained high accuracy. However, when the same method is applied to another dataset, its classification performance tend to decline, reflecting poor generalization ability. It is challenging to study cross-domain automatic sleep stage classification methods with high generalization ability.

7.1.5 Clinical and Practical Constraints

Traditional PSG is recognized for its high cost, requirement for specialized equipment, and usual implemen-

tation in clinical settings, which may disrupt normal sleep patterns. Present healthcare alternatives provide a more convenient option, but they may compromise on accuracy. Furthermore, the challenge lies in ensuring that sleep stage classification methods adhere to regulatory benchmarks and are verified against recognized clinical standards.

7.2 Future Prospects

7.2.1 Developing Large-Scale Benchmark Datasets

Given the constraints of a small dataset and standardized ground truth, building a more comprehensive dataset is one of the future prospects. This dataset would provide a more thorough training set for sleep staging algorithms. The creation of this dataset will adhere to rigorous standards. For instance, to ensure high reliability in annotation, the sleep physiological signals should be analyzed by at least two experts under double-blind conditions. Additionally, another expert is required to further evaluate the sleep segments with different sleep stage classification results to achieve accurate labeling. Moreover, the dataset should comprise a large number of individuals from diverse ethnicities, groups, ranges, and genders. Furthermore, the equipment used for signal acquisition should be diverse, and the signal processing should be carried out in a wide range of scenarios to enhance the robustness of the analysis. The optimized dataset can serve as a robust foundation for higher classification performance and domain adaptability of next-generation sleep stage classification methods.

7.2.2 Integration of Multimodal Data

Among existing sleep stage classification methods, those integrating multi-sensor data have demonstrated significantly better performance. For instance, the fusion of EEG and ECG offers a more comprehensive representation of sleep stages, facilitating downstream models to extract informative features from different perspectives. With the advancement of sensor technology, signals from radar, Wi-Fi, and pressure sensors are increasingly being used in sleep stage classification tasks. These emerging signals contain complementary information not captured by traditional physiological signals. Recent studies have shown that sleep apnea is closely associated with factors such as sleep posture, which can be captured through these novel signals. Incorporating such additional features has been proven to enhance classification accuracy. Looking ahead, leveraging multimodal signals and constructing multitask learning frameworks to further improve the sleep stage classification accuracy is expected to be a key direction in future research.

7.2.3 Domain Adaptation

Domain adaptation technology can enhance the generalization ability and clinical application value of sleep staging models by addressing the challenges posed by cross-domain data distribution differences. The essence of this technology lies in employing mechanisms such as adversarial training and feature decoupling to effectively separate common sleep features in the source domain data from disturbances specific to the target domain (e.g., individual physiological differences). This approach facilitates adaptive migration across various acquisition devices, populations, and environments. Exploring domain adaptation holds great promise for advancing sleep monitoring and personalized healthcare.

7.2.4 Bridging Deep Learning and Lightweight Machine Learning

While deep learning architectures—such as CNNs, RNNs, and Transformers—achieve state-of-the-art performance in sleep staging benchmarks, their high computational and memory demands preclude direct deployment on edge devices. Instead of discarding lightweight machine learning approaches, a promising direction is to bridge the two paradigms. One practical strategy is knowledge distillation, in which a large, accurate model serves as a teacher to train a compact conventional classifier. Such distillation allows the lightweight model to capture sophisticated representations without incurring substantial inference overhead. Alternatively, hybrid systems can integrate aggressively pruned neural networks for feature extraction with efficient statistical classifiers—such as SVMs or decision trees—for the final staging decision. This combination enables wearable monitoring systems to leverage the powerful feature representation capability of deep learning, while simultaneously satisfying the stringent hardware constraints and interpretability requirements of real-time sleep analysis in home-based settings.

7.2.5 Sleep Stage Classification for Enhanced Patient Care and Treatment

Sleep stage classification plays a crucial role in the automatic control of ventilators in patients with sleep apnea. Accurate classification helps real-time adjustment of the ventilator, which improves the therapeutic effect and patient comfort. Combining machine learning-based sleep stage classification with ventilator control systems is a promising research direction. In addition, combining alarm systems for specific sleep stages can enhance patient safety. For example, the system can notify medical personnel in a timely manner when abnormalities are detected during critical stages. Integrating these systems with telemedicine platforms not only enables remote monitoring and immediate intervention, but also provides a proactive solution for sleep health management.

8. Conclusions

This review has critically evaluated lightweight machine learning methodologies for automatic sleep stage classification, focusing on their applicability to edge computing and wearable devices. Several actionable take-

home messages emerge from this analysis. First, regarding algorithmic design, ensemble tree-based classifiers (e.g., Random Forest, XGBoost) combined with FFT-based frequency features and filter-based feature selection currently offer the most robust trade-off. Unlike resource-intensive deep learning or exact entropy calculations, this combination strictly adheres to memory constraints while remaining highly resilient to motion artifacts. Second, for unobtrusive, long-term home monitoring, modalities generating low-dimensional data streams—specifically actigraphy (acceleration) and HRV—should be prioritized. If direct cortical staging is required, single-channel EEG paired with an accelerometer provides the optimal balance between clinical accuracy and hardware deployability. Finally, comparative performance metrics in the literature must be interpreted with extreme caution. High accuracies achieved exclusively on healthy cohorts do not reliably translate to clinical populations, where micro-arousals severely blur physiological boundaries. Furthermore, studies reporting only overall accuracy without class-aware metrics (e.g., macro-F1) frequently mask profound classification failures on the minority N1 stage. Ultimately, advancing lightweight sleep staging relies not on accumulating heavier models, but on selecting low-dimensional signal combinations validated rigorously on imbalanced clinical datasets.

Author Contributions

YYS, YL, FM, and XMF contributed to the conception and design of this review. FM and XMF also contributed to the analysis and interpretation of the reviewed findings. YYS, RHY, and SYW conducted the literature search, data extraction, and synthesis. WG made substantial contributions to developing the framework for evaluating and categorizing lightweight machine-learning methods and to the analysis and interpretation of findings across the reviewed studies. WG also provided funding support. YYS and YL drafted the initial manuscript. FM, and XMF critically revised the manuscript for important intellectual content. All authors contributed to editorial changes. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research was funded by the Key R&D Program of Xizang Autonomous Region (Grant No. CGZH2026000510), the National Natural Science Foundation of China (Grant No. 62473267), and the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2025A1515011614).

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration of AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used GPT-5.2, DeepSeek V3.2 and Gemini 3.1 Pro in order to check spelling and grammar. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Chokroverty S. Overview of sleep & sleep disorders. The Indian Journal of Medical Research. 2010; 131: 126–140.

[2] Newman AB, Nieto FJ, Guidry U, Lind BK, Redline S, Pickering TG, et al. Relation of sleep-disordered breathing to cardiovascular disease risk factors: the Sleep Heart Health Study. American Journal of Epidemiology. 2001; 154: 50–59. https://doi.org/10.1093/aje/154.1.50

[3] Boostani R, Karimzadeh F, Nami M. A comparative review on sleep stage classification methods in patients and healthy individuals. Computer Methods and Programs in Biomedicine. 2017; 140: 77–91. https://doi.org/10.1016/j.cmpb.2016.12.004

[4] Rechtschaffen A, Kales A (eds.). A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. U.S. National Institute of Neurological Diseases and Blindness, Neurological Information Network: Bethesda, MD. 1968.

[5] Iber C, Ancoli-Israel S, Chesson AL Jr, Quan SF. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. 1st ed. Westchester, IL: American Academy of Sleep Medicine. 2007.

[6] Jafari B, Mohsenin V. Polysomnography. Clinics in Chest Medicine. 2010; 31: 287–297. https://doi.org/10.1016/j.ccm.2010.02.005

[7] Pérez-Ávila AJ, Ruiz-Herrera N, Martínez-Olmos A, Carvajal MA, Capitán-Vallvey LF, López-Ruiz N, et al. Smartphone-based and non-invasive sleep stage identification system with piezo-capacitive sensors. Sensors and Actuators A: Physical. 2024; 376: 115659. https://doi.org/10.1016/j.sna.2024.115659

[8] Zhao D, Wang Y, Wang Q, Wang X. Comparative analysis of different characteristics of automatic sleep stages. Computer Methods and Programs in Biomedicine. 2019; 175: 53–72. https://doi.org/10.1016/j.cmpb.2019.04.004

[9] An P, Yuan Z, Zhao J, Jiang X, Du B. An effective multi-model fusion method for EEG-based sleep stage classification. Knowledge-Based Systems. 2021; 219: 106890. https://doi.org/10.1016/j.knosys.2021.106890

[10] Zaidi TF, Farooq O. EEG sub-bands based sleep stages classification using Fourier Synchrosqueezed transform features. Expert Systems with Applications. 2023; 212: 118752. https://doi.org/10.1016/j.eswa.2022.118752

[11] Rahman MM, Bhuiyan MIH, Hassan AR. Sleep stage classification using single-channel EOG. Computers in Biology and Medicine. 2018; 102: 211–220. https://doi.org/10.1016/j.compbiomed.2018.08.022

[12] Seifpour S, Niknazar H, Mikaeili M, Nasrabadi AM. A new automatic sleep staging system based on statistical behavior of local extrema using single channel EEG signal. Expert Systems

with Applications. 2018; 104: 277–293. https://doi.org/10.1016/j.eswa.2018.03.020

[13] Chen Y, Gong C, Hao H, Guo Y, Xu S, Zhang Y, et al. Automatic Sleep Stage Classification Based on Subthalamic Local Field Potentials. IEEE Transactions on Neural Systems and Rehabilitation Engineering: a Publication of the IEEE Engineering in Medicine and Biology Society. 2019; 27: 118–128. https://doi.org/10.1109/TNSRE.2018.2890272

[14] Nakamura T, Alqurashi YD, Morrell MJ, Mandic DP. Hearables: Automatic Overnight Sleep Monitoring With Standardized In-Ear EEG Sensor. IEEE Transactions on Bio-medical Engineering. 2020; 67: 203–212. https://doi.org/10.1109/TBME.2019.2911423

[15] Dhok S, Pimpalkhute V, Chandurkar A, Bhurane AA, Sharma M, Acharya UR. Automated phase classification in cyclic alternating patterns in sleep stages using Wigner-Ville Distribution based features. Computers in Biology and Medicine. 2020; 119: 103691. https://doi.org/10.1016/j.compbiomed.2020.103691

[16] Diykh M, Li Y, Abdulla S. EEG sleep stages identification based on weighted undirected complex networks. Computer Methods and Programs in Biomedicine. 2020; 184: 105116. https://doi.org/10.1016/j.cmpb.2019.105116

[17] Huang W, Guo B, Shen Y, Tang X, Zhang T, Li D, et al. Sleep staging algorithm based on multichannel data adding and multifeature screening. Computer Methods and Programs in Biomedicine. 2020; 187: 105253. https://doi.org/10.1016/j.cmpb.2019.105253

[18] Wei J, Chu R, Lu W, Chen Y. Effective contextual feature fusion and individualized information for automated sleep staging. Computerized Medical Imaging and Graphics: the Official Journal of the Computerized Medical Imaging Society. 2026; 127: 102678. https://doi.org/10.1016/j.compmedimag.2025.102678

[19] Chung YM, Hu CS, Lo YL, Wu HT. A Persistent Homology Approach to Heart Rate Variability Analysis With an Application to Sleep-Wake Classification. Frontiers in Physiology. 2021; 12: 637684. https://doi.org/10.3389/fphys.2021.637684

[20] Memar P, Faradji F. A Novel Multi-Class EEG-Based Sleep Stage Classification System. IEEE Transactions on Neural Systems and Rehabilitation Engineering: a Publication of the IEEE Engineering in Medicine and Biology Society. 2018; 26: 84–95. https://doi.org/10.1109/TNSRE.2017.2776149

[21] Yücelbaş Ş, Yücelbaş C, Tezel G, Özşen S, Yosunkaya Ş. Automatic sleep staging based on SVD, VMD, HHT and morphological features of single-lead ECG signal. Expert Systems with Applications. 2018; 102: 193–206. https://doi.org/10.1016/j.eswa.2018.02.034

[22] Jiang D, Lu Yn, Yu M, Yuanyuan W. Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement. Expert Systems with Applications. 2019; 121: 188–203. https://doi.org/10.1016/j.eswa.2018.12.023

[23] Yan R, Zhang C, Spruyt K, Wei L, Wang Z, Tian L, et al. Multi-modality of polysomnography signals' fusion for automatic sleep scoring. Biomedical Signal Processing and Control. 2019; 49: 14–23. https://doi.org/10.1016/j.bspc.2018.10.001

[24] Zaman A, Kumar S, Shatabda S, Dehzangi I, Sharma A. SleepBoost: a multi-level tree-based ensemble model for automatic sleep stage classification. Medical & Biological Engineering & Computing. 2024; 62: 2769–2783. https://doi.org/10.1007/s11517-024-03096-x

[25] Supriya S, Siuly S, Wang H, Zhang Y. EEG sleep stages analysis and classification based on weighed complex network features. IEEE Transactions on Emerging Topics in Computational Intelligence. 2018; 5: 236–246. https://doi.org/10.1109/TETCI.2018.2876529

[26] Hong H, Zhang L, Gu C, Li Y, Zhou G, Zhu X. Noncontact sleep stage estimation using a CW Doppler radar. IEEE Journal on Emerging and Selected Topics in Circuits and Systems. 2018; 8: 260–270. https://doi.org/10.1109/JETCAS.2017.2789278

[27] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. Pattern Recognition. 2018; 77: 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

[28] Fei K, Wang J, Pan L, Wang X, Chen B. A sleep staging model on wavelet-based adaptive spectrogram reconstruction and light weight CNN. Computers in Biology and Medicine. 2024; 173: 108300. https://doi.org/10.1016/j.compbiomed.2024.108300

[29] Wei Y, Zhu Y, Zhou Y, Yu X, Luo Y. Automatic Sleep Staging Based on Contextual Scalograms and Attention Convolution Neural Network Using Single-Channel EEG. IEEE Journal of Biomedical and Health Informatics. 2024; 28: 801–811. https://doi.org/10.1109/JBHI.2023.3332503

[30] Salehinejad H, Sankar S, Barfett J, Colak E, Valaee S. Recent advances in recurrent neural networks. arXiv. 2017. https://doi.org/10.48550/arXiv.1801.01078 (preprint)

[31] Jain R, G RA. Modality-Specific Feature Selection, Data Augmentation and Temporal Context for Improved Performance in Sleep Staging. IEEE Journal of Biomedical and Health Informatics. 2024; 28: 1031–1042. https://doi.org/10.1109/JBHI.2023.3339713

[32] Pei W, Li Y, Wen P, Yang F, Ji X. An automatic method using MFCC features for sleep stage classification. Brain Informatics. 2024; 11: 6. https://doi.org/10.1186/s40708-024-00219-w

[33] Chriskos P, Frantzidis CA, Nday CM, Gkivogkli PT, Bamidis PD, Kourtidou-Papadeli C. A review on current trends in automatic sleep staging through bio-signal recordings and future challenges. Sleep Medicine Reviews. 2021; 55: 101377. https://doi.org/10.1016/j.smrv.2020.101377

[34] Fiorillo L, Puiatti A, Papandrea M, Ratti PL, Favaro P, Roth C, et al. Automated sleep scoring: A review of the latest approaches. Sleep Medicine Reviews. 2019; 48: 101204. https://doi.org/10.1016/j.smrv.2019.07.007

[35] Djanian S, Bruun A, Nielsen TD. Sleep classification using Consumer Sleep Technologies and AI: A review of the current landscape. Sleep Medicine. 2022; 100: 390–403. https://doi.org/10.1016/j.sleep.2022.09.004

[36] Sarkar D, Guha D, Tarafdar P, Sarkar S, Ghosh A, Dey D. A comprehensive evaluation of contemporary methods used for automatic sleep staging. Biomedical Signal Processing and Control. 2022; 77: 103819. https://doi.org/10.1016/j.bspc.2022.103819

[37] Kemp B, Zwinderman AH, Tuk B, Kamphuisen HA, Oberyé JJ. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE Transactions on Biomedical Engineering. 2000; 47: 1185–1194. https://doi.org/10.1109/10.867928

[38] Quan SF, Howard BV, Iber C, Kiley JP, Nieto FJ, O'Connor GT, et al. The Sleep Heart Health Study: design, rationale, and methods. Sleep. 1997; 20: 1077–1085.

[39] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000; 101: E215–E220. https://doi.org/10.1161/01.cir.101.23.e215

[40] Devuyst S, Dutoit T, Kerkhofs M. The DREAMS databases and assessment algorithm. Zenodo: Geneva, Switzerland. 2005.

[41] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine: the Quarterly Magazine of the Engineering in Medicine & Biology Society. 2001; 20: 45–50. https://doi.org/10.1109/51.932724

[42] Khalighi S, Sousa T, Santos JM, Nunes U. ISRUC-Sleep: A

comprehensive public dataset for sleep researchers. Computer Methods and Programs in Biomedicine. 2016; 124: 180–192. https://doi.org/10.1016/j.cmpb.2015.10.013

[43] O'Reilly C, Gosselin N, Carrier J, Nielsen T. Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research. Journal of Sleep Research. 2014; 23: 628–635. https://doi.org/10.1111/jsr.12169

[44] Chen X, Wang R, Zee P, Lutsey PL, Javaheri S, Alcántara C, et al. Racial/Ethnic Differences in Sleep Disturbances: The Multi-Ethnic Study of Atherosclerosis (MESA). Sleep. 2015; 38: 877–888. https://doi.org/10.5665/sleep.4732

[45] Blackwell T, Redline S, Ancoli-Israel S, Schneider JL, Surovec S, Johnson NL, et al. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. Sleep. 2008; 31: 283–291. https://doi.org/10.1093/sleep/31.2.283

[46] Terzano MG, Parrino L, Sherieri A, Chervin R, Chokroverty S, Guilleminault C, et al. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. Sleep Medicine. 2001; 2: 537–553. https://doi.org/10.1016/s1389-9457(01)00149-6

[47] Alvarez-Estevez D, Rijsman R. Haaglanden Medisch Centrum sleep staging database (version 1.1). PhysioNet. 2022. RRID:SCR_007345. https://doi.org/10.13026/t79q-fr32

[48] Guillot A, Sauvet F, During EH, Thorey V. Dreem Open Datasets: Multi-Scored Sleep Datasets to Compare Human and Automated Sleep Staging. IEEE Transactions on Neural Systems and Rehabilitation Engineering: a Publication of the IEEE Engineering in Medicine and Biology Society. 2020; 28: 1955–1965. https://doi.org/10.1109/TNSRE.2020.3011181

[49] López-Larraz E, Sierra-Torralba M, Clemente S, Fierro G, Oriol D, Minguez J, et al. The Bitbrain Open Access Sleep (BOAS) dataset (version 1.1.1). OpenNeuro. 2025. https://doi.org/10.18112/openneuro.ds005555.v1.1.1

[50] Cheng Z, Li Q, Zou X, Zhong Z, Ouyang Q, Gan C, et al. Cyclic Alternating Pattern of EEG Activities and Heart Rate Variability in Parkinson's Disease Patients during Deep Sleep. Journal of Integrative Neuroscience. 2025; 24: 26397. https://doi.org/10.31083/JIN26397

[51] de Goederen R, Pu S, Silos Viu M, Doan D, Overeem S, Serdijn WA, et al. Radar-based sleep stage classification in children undergoing polysomnography: a pilot-study. Sleep Medicine. 2021; 82: 1–8. https://doi.org/10.1016/j.sleep.2021.03.022

[52] He C, Tan J, Jian X, Zhong G, Cheng L, Lin J. A smart flexible vital signs and sleep monitoring belt based on MEMS triaxial accelerometer and pressure sensor. IEEE Internet of Things Journal. 2022; 9: 14126–14136. https://doi.org/10.1109/JIOT.2022.3146926

[53] Dogaheh SB, Moradi MH. Automatic sleep stage classification based on Dreem headband's signals. In 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME) (pp. 259–263). Tehran, Iran. IEEE. 2020. https://doi.org/10.1109/ICBME51989.2020.9319415

[54] Kremen V, Brinkmann BH, Van Gompel JJ, Stead M, St Louis EK, Worrell GA. Automated unsupervised behavioral state classification using intracranial electrophysiology. Journal of Neural Engineering. 2019; 16: 026004. https://doi.org/10.1088/1741-2552/aae5ab

[55] von Ellenrieder N, Peter-Derex L, Gotman J, Frauscher B. SleepSEEG: automatic sleep scoring using intracranial EEG recordings only. Journal of Neural Engineering. 2022; 19: 026057. https://doi.org/10.1088/1741-2552/ac6829

[56] Boe AJ, McGee Koch LL, O'Brien MK, Shawen N, Rogers JA, Lieber RL, et al. Automating sleep stage classification using wireless, wearable sensors. NPJ Digital Medicine. 2019; 2: 131. https://doi.org/10.1038/s41746-019-0210-1

[57] Melo MC, da Silva Vallim JR, Garbuio S, Soster LA, Sousa KMM, Bonaldi RR, et al. Validation of a sleep staging classification model for healthy adults based on two combinations of a single-channel EEG headband and wrist actigraphy. Journal of Clinical Sleep Medicine: JCSM. 2024; 20: 983–990. https://doi.org/10.5664/jcsm.11082

[58] Gaiduk M, Perea JJ, Seepold R, Martinez Madrid N, Penzel T, Glos M, et al. Estimation of Sleep Stages Analyzing Respiratory and Movement Signals. IEEE Journal of Biomedical and Health Informatics. 2022; 26: 505–514. https://doi.org/10.1109/JBHI.2021.3099295

[59] Xue B, Deng B, Hong H, Wang Z, Zhu X, Feng DD. Non-Contact Sleep Stage Detection Using Canonical Correlation Analysis of Respiratory Sound. IEEE Journal of Biomedical and Health Informatics. 2020; 24: 614–625. https://doi.org/10.1109/JBHI.2019.2910566

[60] Abdulla S, Diykh M, Laft RL, Saleh K, Deo RC. Sleep EEG signal analysis based on correlation graph similarity coupled with an ensemble extreme machine learning algorithm. Expert Systems with Applications. 2019; 138: 112790. https://doi.org/10.1016/j.eswa.2019.07.007

[61] Liu C, Tan B, Fu M, Li J, Wang J, Hou F, et al. Automatic sleep staging with a single-channel EEG based on ensemble empirical mode decomposition. Physica A: Statistical Mechanics and its Applications. 2021; 567: 125685. https://doi.org/10.1016/j.physa.2020.125685

[62] Allocca G, Ma S, Martelli D, Cerri M, Del Vecchio F, Bastiani S, et al. Validation of 'Somnivore', a Machine Learning Algorithm for Automated Scoring and Analysis of Polysomnography Data. Frontiers in Neuroscience. 2019; 13: 207. https://doi.org/10.3389/fnins.2019.00207

[63] Saghayan MH, Seifpour S, Khadem A. Automated Sleep Stage Scoring Using Brain Effective Connectivity and EEG Signals. In 2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS) (pp. 1–5). Shiraz, Iran. IEEE. 2021. https://doi.org/10.1109/ICSPIS54653.2021.9729377

[64] Satapathy SK, Bhoi AK, Loganathan D, Khandelwal B, Barsocchi P. Machine learning with ensemble stacking model for automated sleep staging using dual-channel EEG signal. Biomedical Signal Processing and Control. 2021; 69: 102898. https://doi.org/10.1016/j.bspc.2021.102898

[65] Lyu J, Shi W, Zhang C, Yeh CH. A Novel Sleep Staging Method Based on EEG and ECG Multimodal Features Combination. IEEE Transactions on Neural Systems and Rehabilitation Engineering: a Publication of the IEEE Engineering in Medicine and Biology Society. 2023; 31: 4073–4084. https://doi.org/10.1109/TNSRE.2023.3323892

[66] Satapathy S, Loganathan D, Kondaveeti HK, Rath R. Performance analysis of machine learning algorithms on automated sleep staging feature sets. CAAI Transactions on Intelligence Technology. 2021; 6: 155–174. https://doi.org/10.1049/cit2.12042

[67] Zhou X, Ling BWK, Ahmed W, Zhou Y, Lin Y, Zhang H. Multivariate phase space reconstruction and Riemannian manifold for sleep stage classification. Biomedical Signal Processing and Control. 2024; 88: 105572. https://doi.org/10.1016/j.bspc.2023.105572

[68] Zhou T, Wang G, Choi KS, Wang S. Recognition of Sleep-Wake Stages by Deep Takagi-Sugeno-Kang Fuzzy Classifier with Random Rule Heritage. IEEE Transactions on Emerging Topics in Computational Intelligence. 2023; 7: 1458–1469. https://doi.org/10.1109/TETCI.2022.3233045

[69] Karimzadeh F, Boostani R, Seraj E, Sameni R. A Distributed Classification Procedure for Automatic Sleep Stage Scoring Based on Instantaneous Electroencephalogram Phase and Envelope Features. IEEE Transactions on Neural Systems and Re-

habilitation Engineering: a Publication of the IEEE Engineering in Medicine and Biology Society. 2018; 26: 362–370. https://doi.org/10.1109/TNSRE.2017.2775058

[70] Munoz JP, Rivera LA. Towards Improving Sleep Quality Using Automatic Sleep Stage Classification and Binaural Beats. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2020; 2020: 4982–4985. https://doi.org/10.1109/EMBC44109.2020.9176385

[71] Cao T, Lian Z, Du H, Shen J, Fan Y, Lyu J. A sleep staging model for the sleep environment control based on machine learning. Building Simulation. 2023; 16: 1409–1423. https://doi.org/10.1007/s12273-023-1049-6

[72] Ghasemzadeh P, Kalbkhani H, Sartipi S, Shayesteh MG. Classification of sleep stages based on LSTAR model. Applied Soft Computing. 2019; 75: 523–536. https://doi.org/10.1016/j.asoc.2018.11.007

[73] Huang S, Zhu J, Chen Y, Wang T, Ma T. Analysis and Classification of Sleep Stages Based on Common Frequency Pattern From a Single-Channel EEG Signal. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2020; 2020: 3711–3714. https://doi.org/10.1109/EMBC44109.2020.9176024

[74] Ye J, Xiao Q, Wang J, Zhang H, Deng J, Lin Y. Cosleep: A multi-view representation learning framework for self-supervised learning of sleep stage classification. IEEE Signal Processing Letters. 2021; 29: 189–193. https://doi.org/10.1109/LSP.2021.3130826

[75] Kang C, An S, Kim HJ, Devi M, Cho A, Hwang S, et al. Age-integrated artificial intelligence framework for sleep stage classification and obstructive sleep apnea screening. Frontiers in Neuroscience. 2023; 17: 1059186. https://doi.org/10.3389/fnins.2023.1059186

[76] Ravan M, Begnaud J. Investigating the Effect of Short Term Responsive VNS Therapy on Sleep Quality Using Automatic Sleep Staging. IEEE Transactions on Bio-medical Engineering. 2019; 66: 3301–3309. https://doi.org/10.1109/TBME.2019.2903987

[77] Sekkal RN, Bereksi-Reguig F, Ruiz-Fernandez D, Dib N, Sekkal S. Automatic sleep stage classification: From classical machine learning methods to deep learning. Biomedical Signal Processing and Control. 2022; 77: 103751. https://doi.org/10.1016/j.bspc.2022.103751

[78] Hamouda GB, Rejeb L, Said LB. Ensemble learning for multi-channel sleep stage classification. Biomedical Signal Processing and Control. 2024; 93: 106184. https://doi.org/10.1016/j.bspc.2024.106184

[79] Chung YM, Huang WK, Wu HT. Topological data analysis assisted automated sleep stage scoring using airflow signals. Biomedical Signal Processing and Control. 2024; 89: 105760. https://doi.org/10.1016/j.bspc.2023.105760

[80] Taran S, Sharma PC, Bajaj V. Automatic sleep stages classification using optimize flexible analytic wavelet transform. Knowledge-Based Systems. 2020; 192: 105367. https://doi.org/10.1016/j.knosys.2019.105367

[81] Huang Z, Ling BWK. Joint ensemble empirical mode decomposition and tunable Q factor wavelet transform based sleep stage classifications. Biomedical Signal Processing and Control. 2022; 77: 103760. https://doi.org/10.1016/j.bspc.2022.103760

[82] Sharma M, Makwana P, Chad RS, Acharya UR. A novel automated robust dual-channel EEG-based sleep scoring system using optimal half-band pair linear-phase biorthogonal wavelet filter bank. Applied Intelligence (Dordrecht, Netherlands). 2023; 1–19. https://doi.org/10.1007/s10489-022-04432-0

[83] Surantha N, Lesmana TF, Isa SM. Sleep stage classification using extreme learning machine and particle swarm optimization for healthcare big data. Journal of Big Data. 2021; 8: 14. https://doi.org/10.1186/s40537-020-00406-6

[84] Alickovic E, Subasi A. Ensemble SVM method for automatic sleep stage classification. IEEE Transactions on Instrumentation and Measurement. 2018; 67: 1258–1265. https://doi.org/10.1109/TIM.2018.2799059

[85] Du Y, Gu J, Duan S, Trueb J, Tzavelis A, Shin HS, et al. A skin-interfaced wireless wearable device and data analytics approach for sleep-stage and disorder detection. Proceedings of the National Academy of Sciences of the United States of America. 2025; 122: e2501220122. https://doi.org/10.1073/pnas.2501220122

[86] Tabar YR, Mikkelsen KB, Rank ML, Hemmsen MC, Kidmose P. Investigation of low dimensional feature spaces for automatic sleep staging. Computer Methods and Programs in Biomedicine. 2021; 205: 106091. https://doi.org/10.1016/j.cmpb.2021.106091

[87] Satapathy SK, Kondaveeti HK. Automated sleep stage analysis and classification based on different age specified subjects from a single-channel of EEG signal. In 2021 IEEE Madras section conference (MASCON) (pp. 1–7). Chennai, India. IEEE. 2021. https://doi.org/10.1109/MASCON51689.2021.9563485

[88] Palotti J, Mall R, Aupetit M, Rueschman M, Singh M, Sathyanarayana A, et al. Benchmark on a large cohort for sleep-wake classification with machine learning techniques. NPJ Digital Medicine. 2019; 2: 50. https://doi.org/10.1038/s41746-019-0126-9

[89] Abdulla S, Diykh M, Siuly S, Ali M. An intelligent model involving multi-channels spectrum patterns based features for automatic sleep stage classification. International Journal of Medical Informatics. 2023; 171: 105001. https://doi.org/10.1016/j.ijmedinf.2023.105001

[90] An P, Yuan Z, Zhao J. Unsupervised multi-subepoch feature learning and hierarchical classification for EEG-based sleep staging. Expert Systems with Applications. 2021; 186: 115759. https://doi.org/10.1016/j.eswa.2021.115759

[91] Liang H, Wang Y, Yang L, Wu M, Wang D, Wang X, et al. Automatic Sleep Staging of Single-Channel Ear-EEG Signals With a Probabilistic Ensemble Learning Approach. IEEE Journal of Biomedical and Health Informatics. 2026; 30: 2603–2616. https://doi.org/10.1109/JBHI.2025.3599874

[92] Liu Q, Wu D, Gao Y, Wu G, Vidal PP, Cao J, et al. Sleep Staging Algorithm Incorporating Multi-threhold Neighborhood Polar Pattern Statistics. IEEE Journal of Biomedical and Health Informatics. 2025; PP: 10.1109/JBHI.2025.3631565. https://doi.org/10.1109/JBHI.2025.3631565

[93] Yilmaz G, Martinez CS, Braun F, Gnarra O, Seiler A, Schindler K, et al. Sleep Staging for Wearable Electroencephalography Leveraging Machine Learning and Conventional Polysomnography Datasets. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2025; 2025: 1–7. https://doi.org/10.1109/EMBC58623.2025.11254304

[94] Salfi F, Corigliano D, Amicucci G, Mombelli S, D'Atri A, Axelsson J, et al. The Potential of Ensemble-Based Automated Sleep Staging on Single-Channel EEG Signal From a Wearable Device. Journal of Sleep Research. 2026; e70282. https://doi.org/10.1111/jsr.70282

[95] Rashidi S, Asl BM. Strength of ensemble learning in automatic sleep stages classification using single-channel EEG and ECG signals. Medical & Biological Engineering & Computing. 2024; 62: 997–1015. https://doi.org/10.1007/s11517-023-02980-2

[96] Xu X, Zhang B, Xu T, Tang J. An Effective and Interpretable Sleep Stage Classification Approach Using Multi-Domain Elec-

troencephalogram and Electrooculogram Features. Bioengineering (Basel, Switzerland). 2025; 12: 286. https://doi.org/10.3390/bioengineering12030286

[97] Davidson S, Carter JF, Stanyer EC, Sharman R, Roman C, Kyle SD, et al. Longitudinal cardiorespiratory wearable sleep staging in the home. Frontiers in Neuroscience. 2026; 20: 1693860. https://doi.org/10.3389/fnins.2026.1693860

[98] Silva FB, Uribe LFS, Cepeda FX, Alquati VFS, Guimarães JPS, Silva YGA, et al. Sleep staging algorithm based on smartwatch sensors for healthy and sleep apnea populations. Sleep Medicine. 2024; 119: 535–548. https://doi.org/10.1016/j.sleep.2024.05.033

[99] Jiang D, Ma Y, Wang Y. Sleep stage classification using covariance features of multi-channel physiological signals on Riemannian manifolds. Computer Methods and Programs in Biomedicine. 2019; 178: 19–30. https://doi.org/10.1016/j.cmpb.2019.06.008

[100] Arslan RS, Ulutaş H, Köksal AS, Bakır M, Çiftçi B. Automated sleep scoring system using multi-channel data and machine learning. Computers in Biology and Medicine. 2022; 146: 105653. https://doi.org/10.1016/j.compbiomed.2022.105653

[101] Wang Y, Ye S, Xu Z, Chu Y, Zhang J, Yu W. Research on Sleep Staging Based on Support Vector Machine and Extreme Gradient Boosting Algorithm. Nature and Science of Sleep. 2024; 16: 1827–1847. https://doi.org/10.2147/NSS.S467111

[102] Shao Z, Guo J, Zhu H, Ding Z, Bai M. Prediction and Simulation of sleep stages based on EEG signals. In 2022 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS) (pp. 587–591). Xiamen, China. IEEE. 2022. https://doi.org/10.1109/ICITBS55627.2022.00129

[103] Eldele E, Ragab M, Chen Z, Wu M, Kwoh CK, Li X, et al. ADAST: Attentive cross-domain EEG-based sleep staging framework with iterative self-training. IEEE Transactions on Emerging Topics in Computational Intelligence. 2022; 7: 210–221. https://doi.org/10.1109/TETCI.2022.3189695

[104] Melek M. Automatic Sleep Scoring System Based On Autoregressive Model. In 2021 Medical Technologies Congress (TIPTEKNO) (pp. 1–4). Istanbul, Turkey. IEEE. 2021. https://doi.org/10.1109/TIPTEKNO53239.2021.9633004

[105] Yeh PL, Ozgoren M, Chen HL, Chiang YH, Lee JL, Chiang YC, et al. Automatic Wake and Deep-Sleep Stage Classification Based on Wigner-Ville Distribution Using a Single Electroencephalogram Signal. Diagnostics (Basel, Switzerland). 2024; 14: 580. https://doi.org/10.3390/diagnostics14060580

[106] Biczuk B, Żurek S, Jurga S, Turska E, Guzik P, Piskorski J. Sleep Stage Classification Through HRV, Complexity Measures, and Heart Rate Asymmetry Using Generalized Estimating Equations Models. Entropy (Basel, Switzerland). 2024; 26: 1100. https://doi.org/10.3390/e26121100

[107] Hermans L, van Meulen F, Anderer P, Ross M, Cerny A, van Gilst M, et al. Performance of cardiorespiratory-based sleep staging in patients using beta blockers. Journal of Clinical Sleep Medicine: JCSM. 2024; 20: 575–581. https://doi.org/10.5664/jcsm.10938

[108] Svensson T, Madhawa K, Nt H, Chung UI, Svensson AK. Validity and reliability of the Oura Ring Generation 3 (Gen3) with Oura sleep staging algorithm 2.0 (OSSA 2.0) when compared to multi-night ambulatory polysomnography: A validation study of 96 participants and 421,045 epochs. Sleep Medicine. 2024; 115: 251–263. https://doi.org/10.1016/j.sleep.2024.01.020

[109] Wang Z, Ding Y, Chen H, Chen C, Chen W. A Novel Multi-Modal Flexible Headband System for Sleep Monitoring. Bioengineering (Basel, Switzerland). 2025; 12: 1103. https://doi.org/10.3390/bioengineering12101103

[110] Klok AB, Edin J, Cesari M, Olesen AN, Jennum P, Sorensen HBD. A New Fully Automated Random-Forest Algorithm for Sleep Staging. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2018; 2018: 4920–4923. https://doi.org/10.1109/EMBC.2018.8513413

[111] Gunnarsdottir KM, Gamaldo CE, Salas RME, Ewen JB, Allen RP, Sarma SV. A Novel Sleep Stage Scoring System: Combining Expert-Based Rules with a Decision Tree Classifier. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference. 2018; 2018: 3240–3243. https://doi.org/10.1109/EMBC.2018.8513039

[112] Satapathy SK, Brahma B, Panda B, Barsocchi P, Bhoi AK. Machine learning-empowered sleep staging classification using multi-modality signals. BMC Medical Informatics and Decision Making. 2024; 24: 119. https://doi.org/10.1186/s12911-024-02522-2

[113] Krauss D, Richer R, Küderle A, Jukic J, German A, Leutheuser H, et al. Incorporating respiratory signals for machine learning-based multimodal sleep stage classification: a large-scale benchmark study with actigraphy and heart rate variability. Sleep. 2025; 48: zsaf091. https://doi.org/10.1093/sleep/zsaf091