Received: 30 April 2024 | Revised: 13 September 2024 | Accepted: 16 September 2024

DOI: 10.1111/jsr.14368

RESEARCH ARTICLE

A custom-built single-channel in-ear electroencephalography sensor for sleep phase detection: an interdependent solution for at-home sleep studies

Daniel Filipe Borges[1,2,3] | Joana Isabel Soares[4,5] | Heloísa Silva[6] | João Felgueiras[3,6] | Carla Batista[3,6] | Simão Ferreira[1] | Nuno Barbosa Rocha[1] | Alberto Leal[7,8]

[1]Center for Translational Health and Medical Biotechnology Research (TBIO) | Health Research and Innovation (RISE-Health), E2S, Polytechnic University of Porto, Porto, Portugal

[2]Department of Neurophysiology, E2S, Polytechnic University of Porto, Porto, Portugal

[3]Faculty of Medicine, University of Porto, Porto, Portugal

[4]Polytechnic University of Coimbra, Coimbra, Portugal

[5]H&TRC – Health and Technology Research Center, Coimbra Health School, Polytechnic University of Coimbra, Coimbra, Portugal

[6]Department of Neurology, Unidade Local de Saúde de Matosinhos, Hospital Pedro Hispano, Matosinhos, Portugal

[7]Department of Neurophysiology, Unidade Local de Saúde de S. José, Centro Hospitalar Psiquiátrico de Lisboa, Lisbon, Portugal

[8]Evolutionary Systems and Biomedical Engineering Lab (LaSEEB), Institute for Systems and Robotics (ISR) – Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

Correspondence
Daniel Filipe Borges, Center for Translational Health and Medical Biotechnology Research (TBIO) | Health Research and Innovation (RISE-Health), E2S, Polytechnic University of Porto, Rua Dr. António Bernardino de Almeida, 400, 4200-072 Porto, Portugal.
Email: dfb@ess.ipp.pt

Summary

Sleep is vital for health. It has regenerative and protective functions. Its disruption reduces the quality of life and increases susceptibility to disease. During sleep, there is a cyclicity of distinct phases that are studied for clinical purposes using polysomnography (PSG), a costly and technically demanding method that compromises the quality of natural sleep. The search for simpler devices for recording biological signals at home addresses some of these issues. We have reworked a single-channel in-ear electroencephalography (EEG) sensor grounded to a commercially available memory foam earplug with conductive tape. A total of 14 healthy volunteers underwent a full night of simultaneous PSG, in-ear EEG and actigraphy recordings. We analysed the performance of the methods in terms of sleep metrics and staging. In another group of 14 patients evaluated for sleep-related pathologies, PSG and in-ear EEG were recorded simultaneously, the latter in two different configurations (with and without a contralateral reference on the scalp). In both groups, the in-ear EEG sensor showed a strong correlation, agreement and reliability with the 'gold standard' of PSG and thus supported accurate sleep classification, which is not feasible with actigraphy. Single-channel in-ear EEG offers compelling prospects for simplifying sleep parameterisation in both healthy individuals and clinical patients and paves the way for reliable assessments in a broader range of clinical situations, namely by integrating Level 3 polysomnography devices. In addition, addressing the recognised overestimation of the apnea-hypopnea index, due to the lack of an EEG signal, and the sparse information on sleep metrics could prove fundamental for optimised clinical decision making.

KEYWORDS
in-ear electroencephalography, home recording, polysomnography, sleep staging, wearable devices

1 | INTRODUCTION

Sleep is a complex and essential process for health and quality of life, and the importance of its stability has recently been addressed and emphasised (Sletten et al., 2023). Sleep is critical for optimal brain development (Frank, 2011), cognitive function (Chee & Chuah, 2008; Killgore, 2010; Lim & Dinges, 2010; Maggio et al., 2013; Mishra et al., 2020; Wild et al., 2018) and the clearance of neurotoxic waste

products in the brain that accumulate during wakefulness (Xie et al., 2013). People who obtain sufficient high-quality sleep at proper times were found to have better general health (Jean-Louis et al., 2000) and overall quality of life (Groeger et al., 2004). Sleep deprivation has detrimental consequences for several human systems (Abbott et al., 2020; Thompson et al., 2022), including a significantly higher risk of cardiovascular disease and overall cardiometabolic health (Chan et al., 2022; Chellappa et al., 2019; Kerkhofs & Boudjeltia, 2012; Kohansieh & Makaryus, 2015; Pan et al., 2023; St-Onge et al., 2016), impaired mental health (Murray et al., 2017; Song et al., 2015; Uccella et al., 2023) and also cancer (Lingas, 2023). In view of the increasing importance of sleep for human health and the consequences of sleep deprivation and susceptibility to disease, various sensor modalities for recognising and classifying sleep have been researched that go beyond the 'gold standard' of polysomnography (PSG) and which we will discuss in more detail below. The developments consisted mainly in wearable and sensing technology devices (Imtiaz, 2021) to automatic sleep stage classification (Kwon et al., 2022; Olesen et al., 2021; You et al., 2022) and the use of artificial intelligence (Bandyopadhyay & Goldstein, 2023; Goldstein et al., 2020; Verma et al., 2023). To date, electroencephalography (EEG) is by far the most popular pick, as it largely complies with international standards and recent advances in wearable technology and offline post-processing methods. In general, EEG-based systems have the highest accuracy and kappa values, while others based on accelerometers and non-contact sensors have the lowest values for all sleep stages and also for wakefulness (Danzig et al., 2020; Rapoport, 2019), which underpins our study design and chosen methodology. PSG is based on various biological signals such as brain electrical activity (EEG), eye movements (electro-oculography [EOG]), electrical muscle activity (electromyography [EMG]) of the chin and lower limb anterior tibial muscle, and additional information such as airflow, thoracic and abdominal breathing, oxygen saturation, body position and cardiac activity (electrocardiography [ECG]) (Berry et al., 2017). The significant number of sensors and the wiring means that monitoring is uncomfortable and therefore has a negative impact on sleep. In addition, sleep studies often take place in a hospital or clinic rather than in the patient's natural home environment. Although this method provides high-quality recordings, it requires expert assistance and hours of monitoring/supervision (Fischer et al., 2012), as well as expensive equipment. These facts combined with the need for clinical referral and long waiting lists (Flemons et al., 2004) limit its use outside of specialised clinics. The solution to these major bottlenecks has been the widespread use of at-home sleep testing (Arnardottir et al., 2016). This improved accessibility is apparently not enough to solve all the problems, as for example for obstructive sleep apnea, one of the most prevalent sleep disorders. The waiting time for an initial examination at a specialised centre can easily be 1 year and for continuous positive airway pressure titration up to 2 years (Arnardottir et al., 2022). In addition, sleep classification requires up to 4 h of manual inspection and labelling of an entire night's sleep by experts, a known wearying task (Arnardottir et al., 2022; Ronzhina et al., 2012). This process is achieved by classifying sleep according to a set of guidelines from the American Academy of Sleep Medicine (AASM), currently in version 3, which categorise sleep into four distinct stages: N1, N2, N3, corresponding to non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep, where physiological characteristics, particularly EEG waveforms and patterns, are most important. The normal percentage of each stage is given by the number of total REM sleep cycles recorded overnight. In adults, stage N1 accounts for ~5% of total sleep time (TST), stage N2 for 50%, and stage N3 for 20%. The remaining 25% are REM sleep stages (Troester et al., 2023). Also, other sleep metrics need to be determined for a rigorous procedure of sleep classification and reporting, such as total recording time (TRT), TST including wake after sleep onset (WASO), sleep efficiency (SE), and sleep latency (SL), which are illustrated and defined in Figure 1.

The massive growth of sleep medicine, with demand for studies far outstripping supply, has led to the search for alternative methods that can be carried out at home and are based on a smaller number of electrophysiological signals. So, if we want to understand the effects of different sleep states on a wider population, we need methods and devices that are readily available, inexpensive and unobtrusive. Developing accurate, low-threshold sleep monitoring solutions that can be self-administered and used at home could help avoid these issues of convenience, accessibility, and reproducibility. Currently, one of the most validated methods for studying sleep characteristics is actigraphy (ACT), in which movements are measured to assess sleep or wakefulness. This method makes it possible to assess the binary presence of sleep or wakefulness based on the presence/absence of movement and light by attaching an accelerometer to a device worn on the wrist. The major advantage is that it enables and facilitates prolonged data recording in the home environment at relatively low cost. However, it does not provide accurate sleep analysis, as it does not allow classification of the different sleep stages, and over- or underestimates sleep and wake times. These aspects, as well as accuracy, sensitivity and specificity of ACT, are well described in the literature (Ancoli-Israel et al., 2003; Kim et al., 2022; Marino et al., 2013; Peterson et al., 2012; Tabar et al., 2021). Methods such as sleep questionnaires and sleep diaries are also important for sleep assessment, as they primarily provide information on sleep patterns, sleep problems and sleep rhythms and are often used in combination with PSG or ACT (Tabar et al., 2021). In recent years, attention has increasingly turned to wearable platforms for sleep monitoring (Imtiaz, 2021). Portable and small sleep devices that enable long-term monitoring can facilitate the objective of recognising and establishing a connection between sleep disorders and daytime activity. In addition, sleep tests conducted in a clinical setting cannot accurately quantify the safety and occupational risks of sleepiness (caused by sleep deprivation) (Gavriloff et al., 2018). There are many gadgets, devices and applications that promise information about sleep duration, sleep quality, or even sleep apnea. However, most of these solutions are not considered medical devices and have not been validated by direct comparison and simultaneous recording with 'gold standard' methods. In view of these aspects, there has recently been an increased search for high-quality ambulatory sleep monitoring devices that can be used over a longer period of time, are clinically validated, and can directly monitor the long-term EEG in a natural environment, preferably enabling automatic recognition of sleep patterns

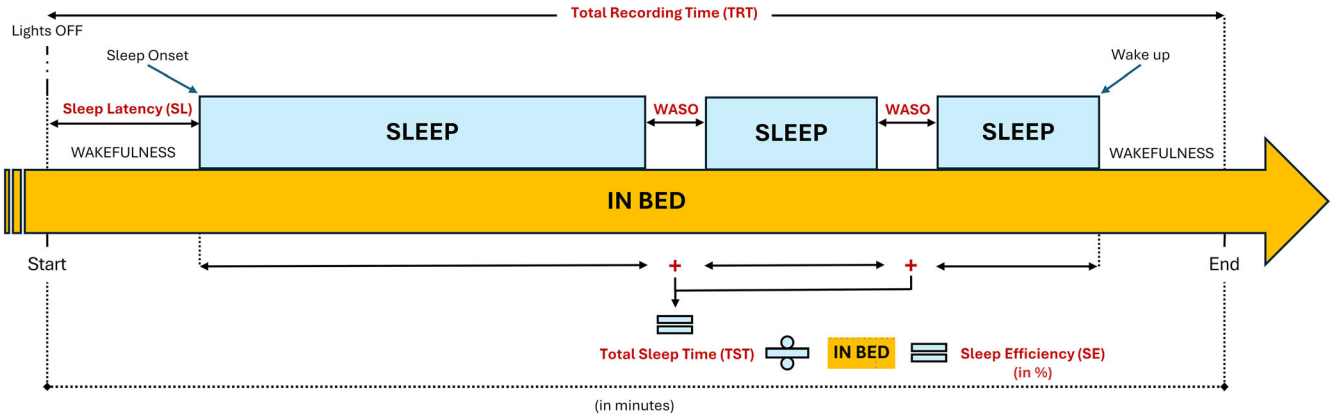

FIGURE 1 Conceptual scheme for the definition of TRT, TST and WASO, SL, SE (in min). SE, sleep efficiency (percentage of total time in bed actually spent asleep; SL, sleep latency (duration of time between turning off the light [lights out] while the patient attempts to sleep, to the time when the patient actually falls asleep); TRT, total recording time (total time during which the patient lies in bed with the recording device activated); TST, total sleep time (total time scored during the total recording time); WASO, wake after sleep onset (total time of wake phases occurring after the defined sleep onset). It is calculated as the sum of the N1, N2, N3 and rapid eye movement sleep stages, divided by the total time in bed and multiplied by 100).

(Looney et al., 2016; Mikkelsen et al., 2017; Nakamura et al., 2017; Stochholm et al., 2016). One of these methods is the so-called in-ear EEG, which has emerged (Looney et al., 2012) and shown great potential due to factors such as portability, very little to no impact on sleep quality as reported by the users (Mikkelsen, Tabar, et al., 2019), sleep staging capabilities (Looney et al., 2016), and the prospect of implementation by the user, which facilitates daily use (Mikkelsen et al., 2017). Another method is devices that are worn around the ear. They offer overall better signal quality and improved topographical coverage also by enabling multi-channel recordings (Meiser et al., 2024), when compared with in-ear sensors and are therefore more suitable for comprehensive monitoring of brain activity in clinical and research settings or in controlled environments, although they can be less comfortable and more obtrusive, which can be a huge contraindication in the case of sleep studies (Kaongoen et al., 2023). As such, regarding sleep monitoring, the placement of the electrodes should naturally be chosen according to the criteria of inconspicuousness, reliability, and comfort. Of course, the in-ear EEG concept assumes that the electrodes must be placed in the external auditory canal or concha (at least the active electrode), although several studies have been conducted (Kaongoen et al., 2023; Mohamed et al., 2023) with a variety of electrode placements, they show very similar performance in detecting sleep stages (Mikkelsen, Phan, et al., 2022), with a trend towards greater electrode spacing leading to better classifiers. It is assumed that differences in the signal-to-noise ratio of the EEG channels speak in favour of this aspect, as it is better at greater distances. Sleep, and sleep stage-specific oscillations, are likely to be global phenomena (Brancaccio et al., 2020) and as long as the derivations used have a sufficiently wide sensitivity range, it is possible to distinguish between sleep stages. The hypothetical need for a contralateral cephalic reference electrode (as occurs in the 'gold standard' PSG with M1/M2) should be addressed and evaluated simultaneously, because if it is not absolutely necessary, the bilateral setup with only in-ear EEG sensors or ideally with single-sided sensors is particularly attractive for self-administered sleep recordings (Mikkelsen, Tabar, et al., 2022) without compromising performance in sleep metrics when participants carried out the experiment themselves (versus expert-applied techniques) (Mikkelsen, Ebajemito, et al., 2019). Of course, to ensure the generalisability of the performance of EEG-based wearables beyond the controlled sleep laboratory environment, further research in real-world settings, such as at home, is needed (Kaongoen et al., 2023); hence, the study design here proposed. Naturally, there are legitimate concerns about the performance of this method compared to scalp EEG. Initially, Looney et al. (2011) demonstrated a high coherence between the temporal electrode T7 and mastoid electrode M1 of the international 10–20 electrode system with an in-ear electrode, indicating a common activity of scalp and in-ear electrodes. More recently, these relationships have been thoroughly analysed (Meiser et al., 2020; Nakamura et al., 2017; Pazuelo et al., 2024; Zibrandtsen et al., 2016). Furthermore, it is still unclear to what extent pathological populations can benefit from such studies conducted in a healthy physiological baseline state, which is a common appointed recommendation for future research.

The aim of the present study was to clinically validate the suitability and reliability of a custom-built single-channel in-ear EEG sensor (with two different electrode configurations) for sleep monitoring and classification by comparing various sleep metrics and staging decisions with simultaneously recorded PSG, the current 'gold standard' in sleep medicine, in both healthy individuals and clinical patients.

2 | METHODS

2.1 | Methods

We conducted a prospective cross-sectional study with a sample of 28 participants divided into two groups: 14 healthy individuals and

14 clinical patients. The latter were consecutively recruited from a sleep clinic after having been previously prescribed a Level 1 PSG. All participants signed an informed consent form. Anonymity was maintained and all results obtained were used solely for academic purposes and without financial interest. We began by validating the methodology in the healthy individuals who had no known neurological or sleep history, were not taking any medications that affect the central nervous system, had not experienced partial or complete sleep deprivation the night before, and had not consumed caffeine, tobacco or energy drinks up to 24 h prior to the study. Recordings were conducted at participants' homes, although in the first phase monitoring and set-up took place in the laboratory. The cortical electrodes were placed in the F3, F4, C3, C4, O1, O2 and M1 and M2 positions of the international 10–20 system, as was the EOG according to the latest AASM recommendations (Troester et al., 2023). The chin EMG did not prove to be stable enough during ambulatory signal acquisition, so we decided not to include it in this preliminary validation. The electrodes were fixed to the scalp with collodion to ensure proper adhesion. A conductive paste was then applied to achieve a maximum impedance of 5 kΩ. The in-ear EEG system, which can be put on and taken off by the user, was then connected. For this purpose, each participant received the necessary instructions for the correct placement of the in-ear sensors, which were inserted shortly before falling asleep and removed again after waking up. The actigraph was also attached to the left wrist and is activated by pressing a button when the subject falls asleep and in the morning when they wake up. Both the PSG and in-ear EEG electrodes were connected to the Morpheus™ ambulatory device, where the different channels were sampled at 256 Hz (bandwidth 0.5–70 Hz) using 16-bit digitisation.

Sleep was assessed independently for the three electrode configurations, all of which were co-registered and connected to the same EEG amplifier:

1. Standard AASM recommendation for EEG derivations (F3-M2, C3-M2, O1-M2, F4-M1, C4-M1, O2-M1) – from now on referred as PSG;
2. One single channel (in-ear electrode referenced to the contralateral in-ear electrode) – from now on referred as ELR;
3. Two bipolar channels (left in-ear sensor referenced to T4 electrode and right in-ear sensor referenced to T3 electrode) – from now on referred as ET. Both the T4 and T3 electrodes are independent of the sensor, i.e., they are commercially available EEG electrodes that have been placed in these 10/20 International System topographies.

Recordings were randomised and blindly evaluated by a European Sleep Research Society (ESRS)-certified somnologist/technologist (H.S.) with >15 years of experience and in accordance with current AASM guidelines (Troester et al., 2023). The masking procedure was specifically achieved by anonymously and randomly distributing the three electrode configurations to the ESRS-certified technologist in such a way that would be possible to identify which of the three electrode configurations was analysed (the channels must be identified according to the AASM sleep classification requirements), but not to which patient it belonged. Each sleep dataset was analysed taking into account the EEG characteristics of the recordings, i.e., the identification of typical phasic elements and in accordance with the EEG frequencies as follows: (a) slow wave activity with a frequency of 0.5–2.0 Hz and a peak-to-peak amplitude of >75 μV measured over the frontal regions; (b) delta waves are 0–3.99 Hz; (c) theta waves are 4–7.99 Hz; (d) alpha waves are 8–13 Hz; (e) beta waves are >13 Hz. In view of this, the classification was carried out according to the following parameters: (1) assessment and scoring in consecutive 30-s epochs, starting at the beginning of the study; (2) assignment of a stage to each epoch; (3) if two or more stages coexist during a single epoch, assignment of the stage that accounts for most of the epoch; (4) if three or more segments of an epoch fulfil criteria for different stages (stage wakefulness [W], N1, N2, N3, REM [R]): (4a) score the epoch as sleep if most of the epoch meets criteria for stage N1, N2, N3, or R; (4b) assign the sleep stage that occurs for most of the sleep within the epoch. In this way, hypnograms were generated for each patient for both the EEG and in-ear EEG methods for the entire night. The parameters related to each sleep stage (N1, N2, N3 and REM), such as the time in each sleep stage (min) and the percentage of total sleep time in each stage – by dividing the time in each stage by the TST × 100 (%) – are only compared between the PSG, ET, ELR recordings, as they cannot be obtained with ACT. For each of the analysed methods, the same 30-s epochs were selected to allow a visual correlation of the characteristics of the EEG waves and patterns of the different sleep phases, but also of the waking state (see Section 3.2.1 – visual correlations). In this way, it was possible to visually compare the morphology and amplitude features and attempt to standardise the classification. After validating the results and performance of the in-ear sensor in the healthy individuals, we analysed the same variables with the equivalent methods in a group of clinical patients ($n = 14$), adding an automatic sleep scoring algorithm to this dataset in the three-electrode configuration, aiming to achieve epoch-by-epoch agreement between the methods and with the ESRS-certified technologist. This group underwent a Level 1 PSG sleep study in a clinic prescribed by a physician with a formal indication. Therefore, all sensors and compliant devices recommended by the AASM were used. The software used for sleep recording of the patient group and sleep staging was Sleepware G3, version 3.9.6.0 and the computerised automatic scoring algorithm, which is validated and AASM compliant, was Somnolyzer 24x7 (Anderer et al., 2010), both from Philips Respironics®. All biosignals and other data were collected and organised in a database, in which all information from the reports of the different sleep analysis procedures was collected and properly anonymised, and which, at this preliminary stage, was only accessible to the first author (D.F.B.).

2.2 | Materials

In-ear sensor (healthy individuals and patient group): a custom-built single-channel in-ear sensor developed by the authors based on an

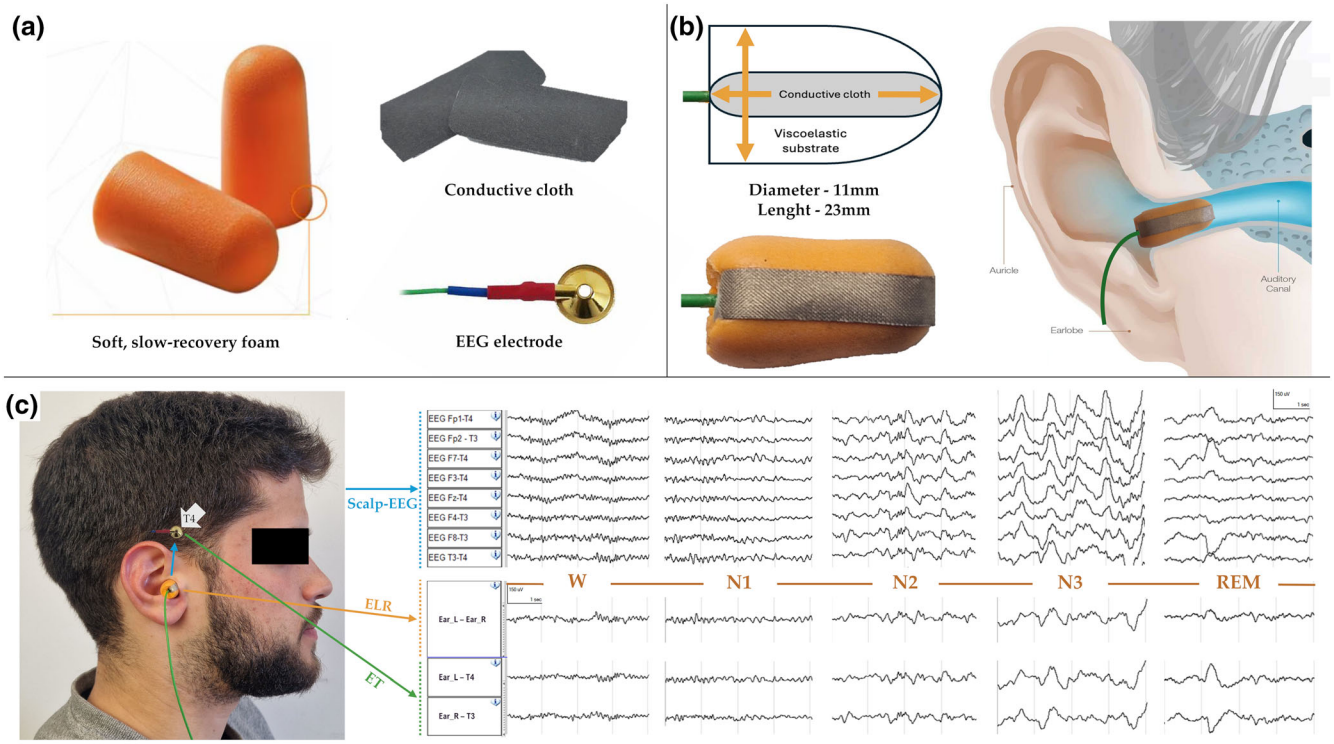

FIGURE 2 In-ear electroencephalography (EEG) sensor configuration, placement and EEG data. (a) The three main components of the sensor are a 3 M™ Foam Earplug 1100, a reusable Grass® cup electrode and a conductive cloth. (b) Schematic diagram of the in-ear EEG and relating dimensions. The corresponding close-up of the sensor shows the viscoelastic substrate (orange) with conductive tape (grey) embedded in a thin film of conductive paste (gold shimmer) bonded with collodion to a touch-safe commercial EEG cable. Please note the anatomical position of the sensor. (c) Photograph of a subject wearing the sensor and T4 electrode for illustration. On the right are segments of the 5-stage sleep study classification acquired from the scalp EEG as the 'gold standard' (only eight leads are shown for this purpose), from the stand-alone in-ear EEG, and with the addition of an independent temporal cephalic electrode that served as a contralateral reference for each (left and right) in-ear EEG sensor. ELR, one (single) bipolar channel (left in-ear sensor referenced to right in-ear sensor); ET, two bipolar channels (left in-ear sensor referenced to T4 electrode and right in-ear sensor referenced to T3 electrode); REM, rapid eye movement; W, wakefulness.

original concept and configurations by Goverdovsky et al. (2016). It consists of a tinned copper and polyvinyl chloride lead wire (Grass® reusable cup electrodes) encased in a slow-resetting polyurethane foam earplug (3 M™ Foam Earplugs 1100), reducing potential mechanical artefacts caused by muscle movement and heart pulsations. We bonded them with collodion to a conductive fabric of silver-coated nylon interwoven with elastic fibres to ensure low impedances and ensure electrical connection with the surface through the use of specific tape with electrically conductive properties (Figure 2). To reduce the impedances, a thin film of conductive paste was applied (Elefix™ by Nihon Kodhen®). This ensured a good fit in the ear canal and created a stable interface between the skin and the electrodes, which could also withstand all the deformations necessary to increase wearing comfort without losing its electrical properties.

Ambulatory EEG/PSG device (healthy individuals): Morpheus™ model of the Micromed® brand running with System Plus Evolution™ software version 1.04.0212. It includes 26 analog channels, two channels for the integrated flow pressure sensor, three oximeter channels, three accelerometer channels and two channels for XActTrace belts. Size: 3.5 cm (height) × 7.5 cm (length) × 11.4 cm (width); Weight: 200 g in a robust ABS plastic housing. The device complies with all AASM standards.

The ACT device (healthy individuals): Camtech® brand, MotionWatch 8™ model with the following specifications: Size mm (length) × 28.2 mm (width) × 9.4 mm (depth); Weight: 9.1 g (including battery, excluding wristband) Triaxial sensor, MEMs technology. Amplitude from 0.01 to 8 g, 3–11 Hz; CR2032 lithium coin cell; Battery life: 3 months with typical use.

The Level 1 PSG device (patient group): Philips Respironics® brand, model Alice-6 LDxS™, with Sleepware G3™ software version 3.9.6.0 with Somnolyzer® 24x7. It includes 55 channels, 19 EEG inputs, five dedicated EMG channels, three chin EMG, two EOG channels and seven ECG channels (three physical and four derived). Signal resolution of 16-bit, maximum sampling rate of 2000 Hz, maximum storage rate of 500 Hz with continuous impedance recording for quality control. The device fulfils all AASM requirements.

2.3 | Statistical analysis

For the group comparisons, we used both the Mann–Whitney *U* test and *t*-tests for independent samples, depending on the nature and distribution of the data. For categorical demographic variables, we used

the chi-square test to ascertain the distributional differences between groups. Correlation analyses were performed to assess the similarity between the collected minutes in the different EEG approaches. Following the suggestion of Gignac and Szodorai (2016), we interpreted the magnitude of these correlations as small (≥0.10), moderate (≥0.20) or large (≥0.30). We employed Bland–Altman plots to provide a clear visualisation of the correspondence between the variables. In these diagrams, the differences between paired measurements were plotted against their means. The limits of agreement were calculated as the mean difference ± 1.96 times the standard deviation (SD) of the differences so that we could identify any systematic bias and the range in which most of the differences between the measurements lay (Doğan, 2018). To assess the performance of an automated sleep staging algorithm in processing the different datasets of the subgroup of patients seeking epoch-by-epoch comparison with the 'gold standard' (manual staging of the human PSG expert), we used a method that evaluates models with categorical or binary outcomes, referred to as confusion matrix (Marom et al., 2010), with correlated kappa scores to assess epoch-wise agreement between sleep staging methods. Kappa scores were interpreted as follows: <0 as less than chance agreement, 0.01–0.20 as slight agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1.00 as near perfect agreement (McHugh, 2012). All statistical analyses were performed using the IBM Statistical Package for the Social Sciences (SPSS) for Windows, version 26.0 (IBM Corp., Armonk, NY, USA) and Python, version 3.7. For these analyses, we mainly used libraries such as 'pandas', 'scipy.stats', 'seaborn' and 'matplotlib.pyplot'. We set the significance level for multiple comparisons at 0.01, with p values between 0.01 and 0.05 considered marginally significant.

TABLE 1 Demographic data of the healthy individuals and patient group and the latter clinical indication for polysomnography.

Variable	Healthy individuals ($n = 14$)	Patient group ($n = 14$)
Age, years, mean (SD, range)	44.29 (2.46, 42–51)	53.21 (17.43, 25–78)
Weight, kg, mean (SD, range)	67.57 (9.60, 52–84)	78.75 (13.11, 61–115)
Height, cm, mean (SD, range)	166.50 (9.67, 151–183)	163.50 (6.38, 152–173)
BMI, kg/m^2, mean (SD, range)	23 (3.90, 19.10–33.20)	29.50 (4.86, 24.34–42.20)
ESS score, mean (SD, range)	3.43 (2.07, 0–7)	5.29 (3.17, 0–9)
Sex, *n* (%)		
Female	9 (64.3)	8 (57.1)
Male	5 (35.7)	6 (42.9)
Clinical indication for PSG		
Obstructive sleep apnea	–	7 (50)
Intermediate insomnia; roncopathy	–	1 (7.14)
REM behaviour disorder; Parkinson's disease	–	1 (7.14)
Loss of consciousness	–	1 (7.14)
Excessive daytime sleepiness	–	1 (7.14)
NREM parasomnia	–	1 (7.14)
Narcolepsy	–	1 (7.14)
White matter disease	–	1 (7.14)

Abbreviations: BMI, body mass index; ESS, Epworth Sleepiness Scale; NREM, non-rapid eye movement; PSG, polysomnography; REM, rapid eye movement.

3 | RESULTS

Our results comprise a total of 258 h (30,960 epochs) of mainly nocturnal EEG recordings presented in three different datasets according to the three methods used, allowing us to measure a total of 774 h of sleep stage. For the 14 participants in the healthy individual group, we will also present ACT data allowing comparison between referenced EEG, in-ear EEG and ACT for TRT, TST, SL, and SE. For the entire sample, we consider all other relevant variables such as N1, N2, N3 and REM phases in minutes and the corresponding latencies. TRT was determined as: Light OFF – Light ON (min); TST: total time during sleep – time in N1 + N2 + N3 + REM (min); SL: time between light OFF and the first sleep epoch (min); SE: division between total sleep time and TRT × 100 (%) to compare between the different evaluation methods PSG, ET, ELR, and ACT.

3.1 | Clinical demographic data

Our participant pool was divided into two different groups: healthy individual controls and patients, so that we could examine the EEG measurements in different health conditions. The healthy control group, consisting of 14 individuals, was between 42 and 51 years old, with a mean (SD) age of 44.29 (2.46) years. Their weight ranged from 52 to 84 kg, with a mean (SD) weight of 67.57 (9.60) kg, and their height ranged from 151 to 183 cm, with an mean (SD) of 166.50 (9.67) cm. Participants had a mean (SD) body mass index (BMI) of 23 (3.90) kg/m^2, which was broadly within the normal weight range, a metric known to correlate with sleep duration (Garfield, 2019). On the Epworth Sleepiness Scale (ESS), the group scored a low mean (SD) of 3.43 (2.07), indicating limited daytime sleepiness. In terms of sex, the group was predominantly female: 64.30% of participants were female and 35.70% male. The patient group, which also included 14 participants, had a wider age range of 25 to 78 years, resulting in a mean (SD) age of 53.21 (17.43) years. Weight ranged from 61 to 115 kg, with a mean (SD) of 78.75 (13.11) kg, and height ranged from 152 to 173 cm, with a mean (SD) of 163.50 (6.38) cm. The mean (SD) BMI of this group was 29.50 (4.86) kg/m^2. The mean (SD) ESS score was 5.29 (3.17), indicating slightly higher daytime sleepiness. The sex distribution was almost equal with 57.10% female and 42.90% male participants. The group of patients included people with various clinical indications for sleep

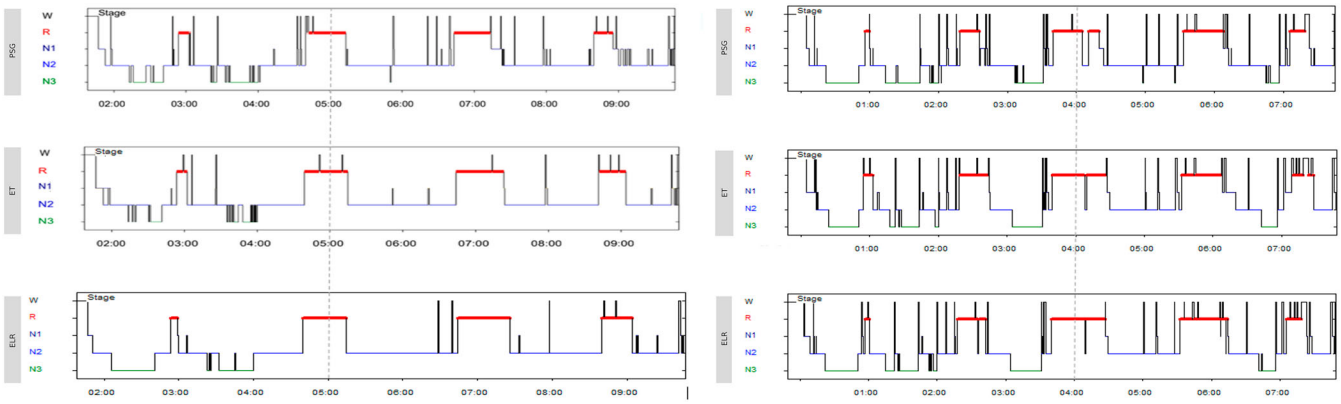

FIGURE 3 Examples of two healthy individuals' all-night hypnograms using the three methods. ELR, in-ear electroencephalography (EEG) in the left ear as active electrode and in-ear EEG in the right ear as reference electrode; ET, two bipolar channels, in which each in-ear electrode serves as active electrode with the corresponding contralateral temporal cephalic electrode as reference; N1, non-rapid eye movement (NREM) N1 sleep; N2, NREM N2 sleep; N3, NREM N3 sleep; PSG, polysomnography (complete arrangement of 10/20 IS electrodes; R, rapid eye movement sleep; W, wakefulness.

staging, mainly with obstructive sleep apnea. The discriminated demographic variables for both groups can be found in Table 1. Regarding the assumption of equal variances in our first analysis, subsequent tests showed that this assumption was not valid for all variables. The Shapiro–Wilk test showed that the distributions for age and weight were not normal, so non-parametric Mann–Whitney *U* tests were performed. These tests revealed no significant difference in age between the groups (*U* statistic = 61.0, $p = 0.09$), but a significant difference in weight (*U* statistic = 46.5, $p = 0.02$), suggesting that the patient group had different weight characteristics compared to the healthy individuals. Height did not differ significantly between groups (*t*-statistic = 1.01, $p = 0.32$), which is confirmed by Levene's test for equal variances ($p = 0.05$), suggesting that this variable was evenly distributed among participants. A significant difference was observed for BMI (*t*-statistic = −3.04, $p = 0.01$), indicating a divergent BMI distribution between the healthy individuals and patient group. The ESS scores did not differ significantly between groups (*t*-statistic = −1.836, $p = 0.08$), indicating comparable levels of daytime sleepiness. Finally, the chi-square test confirmed that there was no significant difference in the sex distribution between the groups (chi-square statistic = 0.00, $p = 1.00$), ensuring that gender did not play a role in the comparison of sleep staging results between the healthy individuals and patient group. This gender balance increases the robustness of the study and provides a comprehensive understanding of the effectiveness and accuracy of single-channel in-ear EEG for sleep classification.

3.2 | Correlations

3.2.1 | Visual correlations (healthy individuals)

In a first phase, to validate the in-ear EEG sensor (single channel or two bipolar channels with independent contralateral temporal cephalic reference), the hypnograms of the entire night resulting from the randomised, blind sleep staging were visually superimposed and compared with the 'gold standard' of PSG, as shown in two examples in Figure 3. Due to the similarity of the hypnograms, the rhythmic and morphological characteristics of the EEG signals obtained with the three methods were analysed in detail by superimposing synchronised 30-s epochs, regarding wakefulness (Figure 4) and N1 + N2 + N3 + REM sleep (Figure 5).

3.2.2 | Statistical correlations (healthy individuals)

We investigated the correlations between minutes of TRT, TST, SE and SL using three different data collection methods. As already mentioned, this included a single-channel in-ear EEG and two bipolar channels with cephalic reference (ET) as well as ACT, each of which was compared with the 'gold standard' of PSG. In TRT, all methods showed globally strong positive correlations with PSG. ACT, on the other hand, showed a relatively weaker correlation ($r[12] = 0.14$, $p = 0.637$), suggesting that the TRT is less accurately represented than with the above-mentioned methods. As for the TST, each method again showed significant positive correlations with the PSG. ET stood out with a strong correlation ($r[12] = 0.98$, $p < 0.001$), underlining its effectiveness in capturing sleep time. ELR also showed a high correlation ($r[12] = 0.92$, $p < 0.001$), while the correlation of ACT was comparatively weaker ($r[12] = 0.81$, $p < 0.001$). In the evaluation of SE, ET again showed a robust correlation with PSG ($r[12] = 0.96$, $p < 0.001$), which is in close agreement with the efficiency measurements of PSG. ELR was also highly correlated ($r[12] = 0.82$, $p < 0.001$), but ACT fell behind with a lower correlation ($r[12] = 0.57$, $p < 0.001$). For SL, the ET method was most congruent with the PSG ($r[12] = 0.97$, $p < 0.001$), which effectively measures time to sleep. ELR also showed a high correlation ($r[12] = 0.96$, $p < 0.001$), but the accuracy of ACT was relatively lower ($r[12] = 0.83$, $p < 0.001$).

We further investigated the correlations for different sleep stages—N1, N2, N3 and REM—but now only with ELR and ET, all

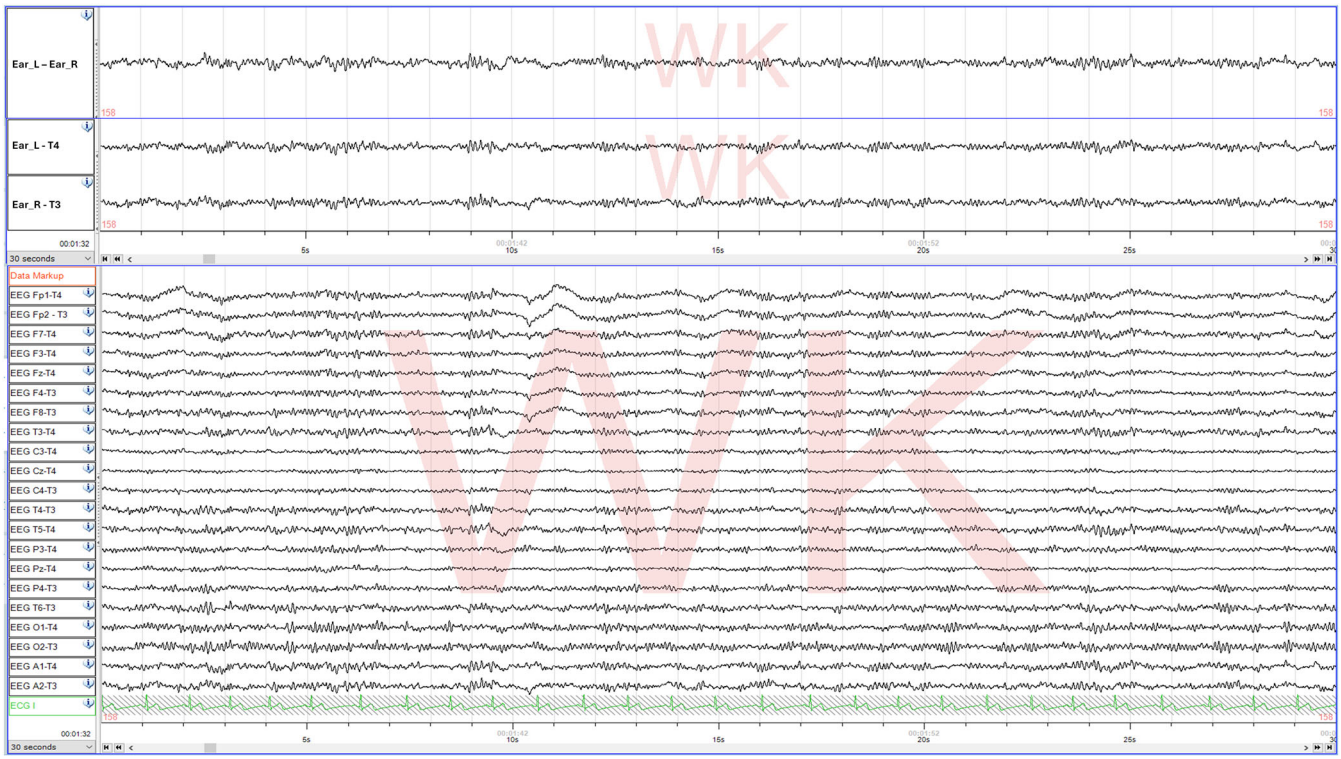

FIGURE 4 Wakefulness (WK) with the three methods in a healthy individual. The 30-s epoch of the 'gold standard' polysomnography with a complete array of 10/20 IS, synchronised with ELR (Ear_L – Ear_D) and ET (Ear_L – T4 / Ear_R – T3). ECG, electrocardiography; ELR, in-ear electroencephalography (EEG) in the left ear as active electrode and in-ear EEG in the right ear as reference electrode; ET, two bipolar channels, in which each in-ear electrode serves as active electrode with the corresponding contralateral temporal cephalic electrode as reference.

compared to the 'gold standard' of PSG. For the N1 sleep stage, the ET method showed a significant correlation with PSG ($r[12] = 0.80$, $p = 0.001$), and the ELR method showed an even stronger correlation ($r[12] = 0.87$, $p < 0.001$). During the N2 sleep stage, ET exhibited a very strong correlation with PSG ($r[12] = 0.94$, $p < 0.001$), while the ELR method also showed a solid correlation ($r[12] = 0.83$, $p < 0.001$), indicating its reliability in measuring deeper sleep. In the N3 sleep stage, often referred to as deep or slow-wave sleep, the ET method again showed an impressive correlation with PSG ($r[12] = 0.95$, $p < 0.001$), while the ELR method also showed a strong correlation ($r[12] = 0.81$, $p < 0.001$), confirming its effectiveness in capturing deep sleep stages. Finally, in REM sleep, the ET method maintained a high degree of agreement with the PSG ($r[12] = 0.94$, $p < 0.001$), and the ELR method was also highly correlated ($r[12] = 0.91$, $p < 0.001$), demonstrating its usefulness in capturing REM sleep. Overall, these results across all sleep stages emphasise that both the ET and ELR methods are highly effective in detecting sleep stages and consistently show strong correlations with PSG. These methods therefore have significant potential for accurate monitoring and analysis of sleep stages.

3.2.3 | Statistical correlations (patient group)

Following the promising results, we wanted to infer whether the demonstrated performance of the in-ear EEG sensor could also be observed in a clinical context by analysing a group of patients with a formal indication for PSG on medical prescription. Therefore, we replicated the methodology and recorded the brain electrical activity signal throughout the night with ELR and ET simultaneously with the PSG. Focusing now on the patient group from our study, we examined the TRT and found that both the ET and the ELR showed a high level of agreement. The results for the TST were similarly encouraging for both methods. In particular, ELR showed a robust correlation ($r[12] = 0.99$, $p < 0.001$), underlining its efficacy in measuring sleep duration. The ET method also exhibited a high correlation ($r[12] = 0.97$, $p < 0.001$), which highlights its reliability in the patient group. In the assessment of SE, the ELR method again agreed closely with the PSG and showed a strong correlation ($r[12] = 0.99$, $p < 0.001$). This is an indication that the method accurately reflects the efficiency values determined with the PSG. The ET method also showed a high correlation ($r[12] = 0.96$, $p < 0.001$), which further confirms the effectiveness of these in-ear EEG methods. For SL, the ELR method continued to show high agreement with PSG ($r[12] = 0.97$, $p < 0.001$), effectively capturing the time of falling asleep. The ET method was not far behind and also showed a significant correlation ($r[12] = 0.96$, $p < 0.001$).

For the N1 sleep stage within the patient group, the ET method revealed a significant correlation with the PSG ($r[12] = 0.84$, $p = 0.001$). However, the ELR method showed a lower

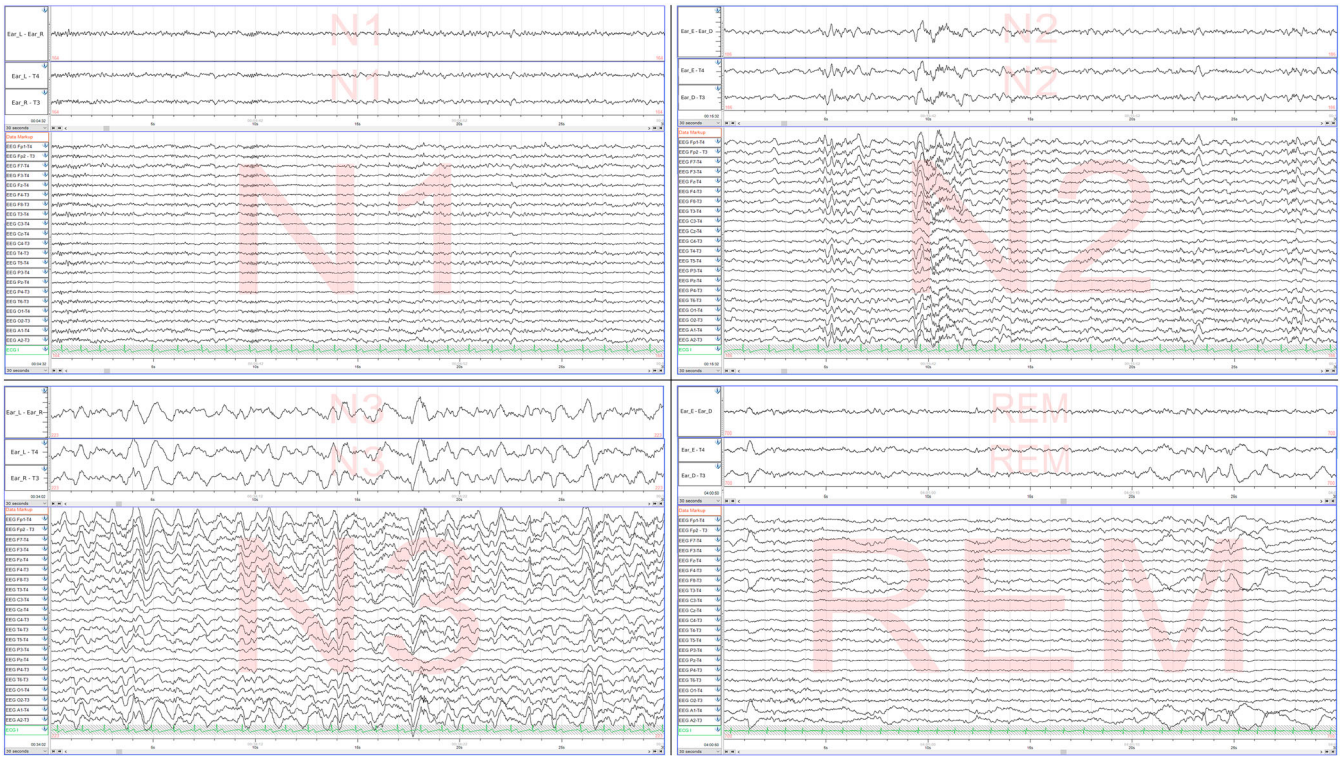

FIGURE 5 Sleep stages N1, N2, N3 and rapid eye movement (REM) sleep with the three methods in a healthy individual. The 30-s epoch of the 'gold standard' polysomnography with a complete array of the international 10/20 system, synchronised with ELR (Ear_L – Ear_D) and ET (Ear_L – T4 / Ear_R – T3). ECG, electrocardiography; ELR, in-ear electroencephalography (EEG) in the left ear as active electrode and in-ear EEG in the right ear as reference electrode; ET, two bipolar channels, in which each in-ear electrode serves as active electrode with the corresponding contralateral temporal cephalic electrode as reference; N1, non-REM (NREM) N1 sleep; N2, NREM N2 sleep; N3, NREM N3 sleep; REM: REM sleep.

correlation ($r[12] = 0.60$, $p = 0.023$), indicating a difference in the effectiveness of these methods in identifying the N1 sleep stage in patients. In the N2 sleep stage, ET showed a very high correlation with PSG ($r[12] = 0.94$, $p < 0.001$), and the ELR method also showed a similarly strong correlation ($r[12] = 0.93$, $p < 0.001$). This indicates that both methods are very reliable in detecting deeper sleep stages in the patient group. During the N3 sleep stage, which is known for deep or slow-wave sleep, the ET method maintained an impressive correlation with PSG ($r[12] = 0.93$, $p < 0.001$). The ELR method also exhibited a robust correlation ($r[12] = 0.88$, $p < 0.001$), highlighting its effectiveness in detecting deep sleep stages in patients. Finally, for the REM sleep stage, the ET method showed extremely high agreement with the PSG ($r[12] = 0.97$, $p < 0.001$). The ELR method was not far behind and showed a strong correlation ($r[12] = 0.96$, $p < 0.001$), underlining the usefulness of these methods in detecting REM sleep in patients. Overall, these results suggest that both the ET and ELR methods are highly effective in detecting sleep stages in this sample, highlighting their potential utility for accurately monitoring and analysing sleep stages in clinical populations. Table 2 shows the direct comparison of correlations between the healthy individuals and the group of patients in relation to the eight variables measured by the different methods.

3.3 | Sleep stage detection and duration by the human expert (healthy individuals and patient group)

As previously mentioned, Bland–Altman plots were used to assess the agreement between PSG and three different measurement methods: ET, ELR and ACT across different sleep metrics. Each plot included the mean difference and limits of agreement (1.96 SD) to highlight the agreement between methods. For this understanding, it is important to clarify that the data colour points used represent the different methods of recording the EEG signal. In particular, they illustrate individual measurements of each sleep metric within the same patient recording and therefore do not correspond to each participant. For TRT, the mean difference between PSG and ET was small, indicating good agreement. The limits of agreement were within acceptable ranges, indicating slender bias. The comparisons between PSG and ELR and between PSG and ACT also showed mean differences close to zero with narrow limits of agreement, indicating strong agreement between the methods. In the case of TST, the mean differences between PSG and the other methods (ET, ELR, ACT) were also extremely acceptable. The Bland–Altman plots showed that the limits of agreement were relatively close, confirming the consistency of these methods in measuring TST. For SL, the mean differences between PSG and ET, ELR and ACT were small, with limits of

TABLE 2 Correlation values between the methods used in each group compared to the 'gold standard' (polysomnography) across four sleep metrics and four sleep stages. Please note that higher correlation values per variable, per method, per group are in bold.

Variable	Method	Healthy individuals ($n = 14$)	Patient group ($n = 14$)
TRT	ET	**1.00**	**1.00**
	ELR	**1.00**	**1.00**
	ACT	0.14	n/a
TST	ET	**0.98**	**0.97**
	ELR	0.92	0.99
	ACT	0.81	n/a
SL	ET	**0.97**	**0.95**
	ELR	0.96	0.97
	ACT	0.83	n/a
SE	ET	**0.96**	**0.96**
	ELR	0.82	0.99
	ACT	0.57	n/a
N1	ET	0.80	**0.84**
	ELR	**0.87**	0.60
N2	ET	**0.94**	**0.94**
	ELR	0.83	0.93
N3	ET	**0.95**	**0.93**
	ELR	0.81	0.88
REM	ET	**0.94**	**0.97**
	ELR	0.91	0.96

Abbreviations: ACT, actigraphy; ELR, single-channel in-ear electroencephalography; ET, two bipolar channels with contralateral temporal cephalic reference; n/a, not available; REM, rapid eye movement; SL, sleep latency; SE, sleep efficiency; TRT, total recording time; TST, total sleep time.

agreement indicating reasonable consistency between methods. The graphs showed a slight deviation in agreement for ACT, but this remained within a satisfactory range. Comparisons of SE showed very close mean differences between PSG and ET, ELR and ACT, with limits of agreement indicating high consistency across methods, as illustrated in Figure 6.

For the N1 sleep stage, the mean differences between PSG and ET and PSG and ELR were minimum. The limits of agreement were slightly wider than for the other metrics, but still within acceptable limits, indicating a reasonable level of agreement. In the N2 sleep phase, the Bland–Altman plots revealed small mean differences between PSG and ET and PSG and ELR even narrow limits of agreement. For sleep stage N3, the comparisons between PSG and ET as well as PSG and ELR showed minimal mean differences and narrow limits of agreement, suggesting high consistency between these measurement methods. REM sleep comparisons showed small mean differences between PSG and ET as well as PSG and ELR with narrow limits of agreement, i.e., high reliability between these methods. The comparison between PSG and ACT also showed good agreement, albeit with slightly wider limits of agreement. The Bland–Altman plots (Figure 7) showed that all methods provided comparable measurements of REM latency, with acceptable limits of agreement and bias. The time spent awake and in the different sleep stages is of course not part of the ACT data, as it does not encompass EEG signal, and was therefore not included. Overall, these sensor data show no obvious outliers or large deviations in our sample, which, even if rare, could be a major drawback in the validation of a new method intended for single-case diagnosis (Figure 8).

3.4 | Epoch-by-epoch analysis with automatic sleep staging algorithm (patient group)

The analysis comparing automatic sleep staging methods (ELR_aut, ET_aut, and PSG_aut) to the 'gold standard' manual sleep staging (PSG_man) reveals several key insights. All automatic methods exhibit a slight systematic bias when compared to PSG_man, with each method measuring slightly higher or lower. Despite this bias, most data points fall within the 95% limits of agreement, indicating acceptable overall agreement with PSG_man. PSG_aut shows strong overall agreement with PSG_man, with a narrow variation band indicating high precision. ET_aut and ELR_aut also show small systematic biases and good precision, suggesting accurate and dependable measurements.

The analysis of the mean absolute error (MAE) and root mean squared error (RMSE), which are widely recognised as robust metrics for evaluating model performance (Hodson, 2022), confirms these results. PSG_aut exhibits the lowest errors (MAE: 0.21, RMSE: 0.57), indicating superior accuracy and reliability compared to the other methods. In contrast, both ELR_aut and ET_aut have slightly higher

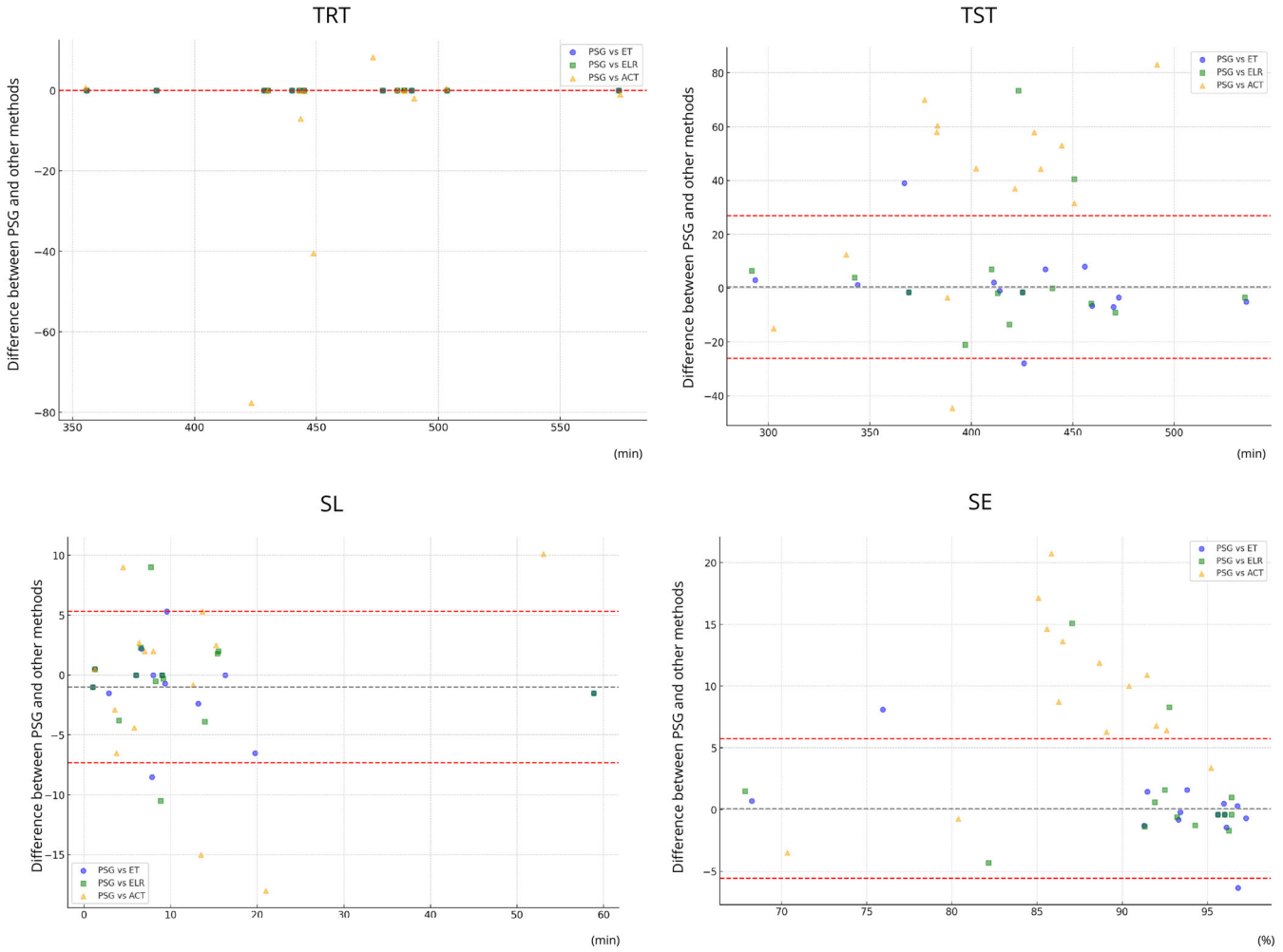

FIGURE 6 Bland–Altman plots for the sleep metrics classification. Each point represents the difference between the measurement methods plotted against their average value. The grey dashed line indicates the mean difference (bias) between the methods and thus shows whether one method tends to over- or underestimate compared to the other. The red dashed lines are the limits of agreement (mean difference ± 1.96 standard deviation), indicating the range within which 95% of the differences between the methods are likely to lay. ACT, actigraphy; ELR, one bipolar channel (left in-ear sensor referenced to right in-ear sensor); ET, two bipolar channels (left in-ear sensor referenced to T4 electrode and right in-ear sensor referenced to T3 electrode); PSG, polysomnography; SE, sleep efficiency (percentage of total time in bed actually spent asleep; SL, sleep latency (duration of time between turning off the light [lights out] while the patient attempts to sleep, to the time when the patient actually falls asleep); TRT, total recording time (total time during which the patient lies in bed with the recording device activated); TST, total sleep time (total time scored during the total recording time).

but still minimal error values (MAE: 0.23, RMSE: 0.59 and 0.59 respectively), emphasising their strong overall performance. However, these small deviations indicate that PSG_aut provides slightly better agreement with the 'gold standard' and is therefore the most accurate of the three methods.

The confusion matrices reveal that certain sleep stages, such as N1, are more prone to misclassification and are often misclassified as W or N2, leading to certain sleep stages being over- or underestimated. ELR_aut, ET_aut, and PSG_aut all show similar patterns of misclassification, though PSG_aut performs better overall. For example, W stages are often confused with N1 and N2, while N1 stages are frequently misclassified as W and N2. Despite these misclassifications, the sensitivity and specificity values indicate that all methods perform well, with slight variations in their accuracy.

Kappa scores quantify the agreement between each method and PSG_man, with values ranging from 0.74 to 0.76, indicating very strong agreement. These scores support the findings from the Bland–Altman plots and confusion matrices, reinforcing the high levels of agreement with the manual expert. The automatic sleep staging algorithm shows consistent performance across ELR, ET, and PSG signals, with similar biases and misclassification patterns observed across methods. While PSG_aut performs slightly better with regard to the kappa values, the differences in MAE and RMSE are not statistically significant (MAE $p = 0.40$; RMSE $p = 0.35$). The comparable consistency of the deviations for all three methods indicates that the automatic analyses process the signals in a very similar way. This close agreement with the 'gold standard' of PSG for all methods reflects the high reliability of the

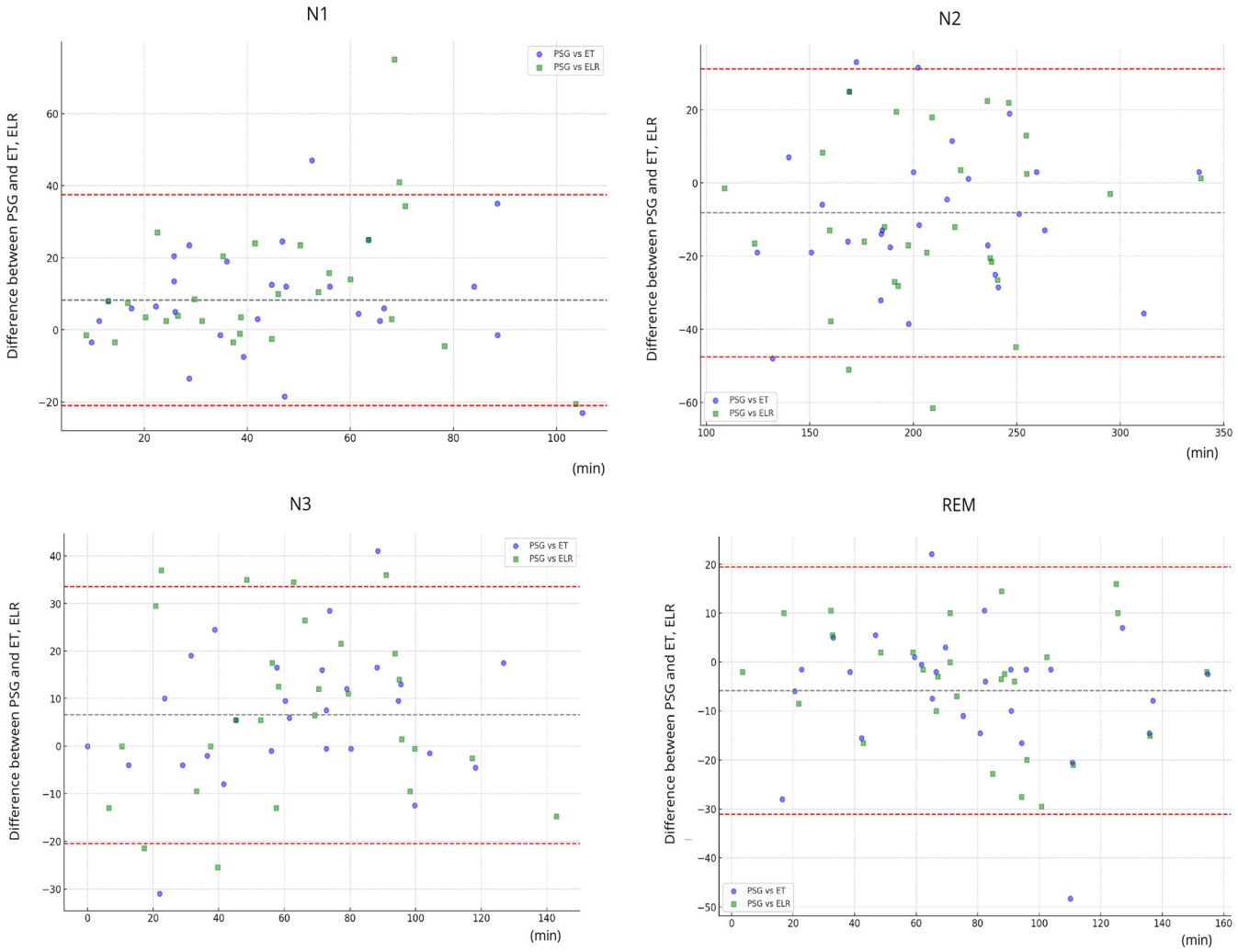

FIGURE 7 Bland–Altman plots for the sleep stage classification. Each point represents the difference between the measurement methods plotted against their average value. The grey dashed line indicates the mean difference (bias) between the methods and thus shows whether one method tends to over- or underestimate compared to the other. The red dashed lines are the limits of agreement (mean difference ± 1.96 standard deviations), indicating the range within which 95% of the differences between the methods are likely to lay. ACT, actigraphy; ELR, one bipolar channel (left in-ear sensor referenced to right in-ear sensor); ET, two bipolar channels (left in-ear sensor referenced to T4 electrode and right in-ear sensor referenced to T3 electrode); N1, non-rapid eye movement (NREM) N1 sleep; N2, NREM N2 sleep; N3, NREM N3 sleep; PSG, polysomnography; REM, rapid eye movement sleep.

automatic scoring algorithm, with PSG_aut having a slight edge in overall accuracy.

While PSG_aut shows slightly better performance, all three methods (ELR_aut, ET_aut, PSG_aut) are reliable and exhibit high levels of agreement with the manual sleep staging. The consistent performance across methods highlights areas for further improvement to increase the accuracy of sleep staging algorithms.

4 | DISCUSSION

Our comparative analysis of sleep parameterisation and staging of two configurations of in-ear EEG electrodes outperformed ACT in all assessed sleep metrics and showed robust correlation with PSG, so we investigated their applicability in clinical patients with formal indications for sleep assessment. Informal feedback from the participants revealed that this custom-made sensor had little to no negative impact on their sleep, as reported also in another study with a similar sample size (Mikkelsen, Tabar, et al., 2019). Some subjects even reported that the diminished auditory sensitivity/acoustics (albeit subjective) aided sleep function; hence, this acceptability is very attractive for future applications. Of course, if the aim is clinical validation, which is mainly concerned with co-registration of different methods, the degree of agreement expected between the chosen 'gold standard' and the new method to be tested must of course be determined. It must also be discussed what values, metrics or statistical results are indeed sufficient for the validation to be considered successful. In our opinion, regarding correlations (Hung et al., 2017),

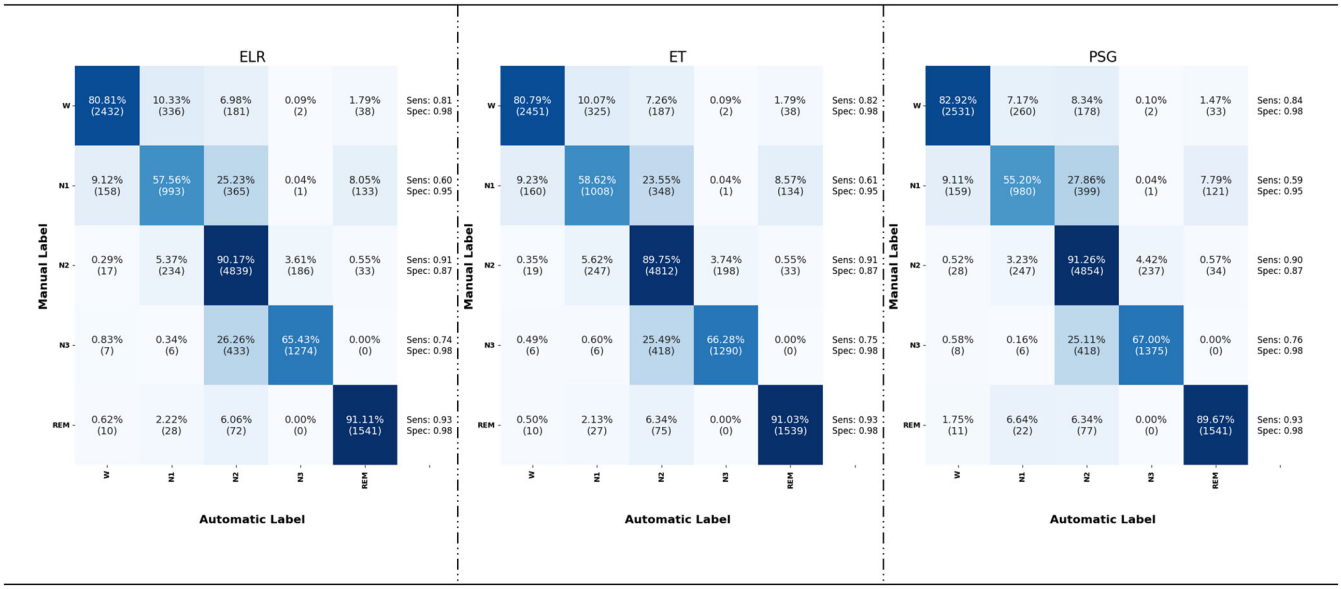

FIGURE 8 Confusion matrices of the epoch-wise concordance comparisons of the different methods using a manual human expert and an automatic algorithm to classify sleep stages. The classes are: W (wakefulness), REM (rapid eye movement sleep), N1 (non-REM [NREM] N1sleep), N2 (NREM N2 sleep), N3 (NREM N3 sleep). The numbers indicate how many 30-s epochs were classified with this combination of scores. The '%' indicates the percentage of epochs relative to the total number of epochs. Numbers in the different shades of blue diagonal grades agreement. Sensitivity, or the true positive rate (recall), is calculated as true positives/(true positives + false negatives). Specificity, or the true negative rate, is calculated as true negatives/(true negatives + false positives). These metrics are indicated for each class per method. The global kappa scores, averaged across all patients and for each confusion matrix, are as follows: ELR: 0.74, ET: 0.74, PSG: 0.76. ELR, one bipolar channel (left in-ear sensor referenced to right in-ear sensor); ET, two bipolar channels (left in-ear sensor referenced to T4 electrode and right in-ear sensor referenced to T3 electrode); PSG, polysomnography.

values >0.80 are generally considered strong, representing a high level of agreement between methods. For the Bland–Altman plot analysis (Doğan, 2018), the difference between the methods should be within ±1.96 SD of the mean difference for most data points to present consistent results. The level of agreement between the in-ear EEG sensor and the 'gold standard' of PSG should be >0.70, ideally >0.81, for the Cohen's kappa coefficient (Cohen, 1960a, 1960b; Vanacore & Pellegrino, 2022) to ensure successful validation. Mikkelsen, Tabar, et al. (2019, 2022) agree with this view as they point out that for most home recording scenarios, a kappa value of 0.73–0.76 is sufficient – 0.74 to 0.76 in our study. This means that for a simple electrode setup like ours with single-channel recordings (ELR) or two bilateral bipolar channels with contracephalic reference (M2-left in-ear and/or M1-right in-ear as in their study), in our study ET, this new method is reliable for clinical and research applications. When we analyse the correlations, as shown in Table 2, all variables of the different methods (except ACT) show a correlation value of >0.80, which indicates a high level of agreement. In all the enrolled subjects (healthy individuals and patient group), the Bland–Altman plots show that the ET, ELR and ACT methods have high agreement with the 'gold standard' PSG measures on several sleep metrics, with minimal bias, and acceptable limits of agreement. These results demonstrate the reliability of these alternative methods for assessing sleep parameters compared to the 'gold standard' of PSG. The overall mean agreement across all comparisons revealed a mean difference (bias) of ~1.87, indicating that most comparisons had a high level of agreement with minimal bias, but this particular comparison had the greatest variability despite having a zero-mean difference, likely due to highly consistent measurements. The analytical measures further support the reliability of these alternative methods for assessing sleep parameters compared to the 'gold standard' of PSG. Additional processing was performed specifically for the subgroup of patients, namely the use of an automatic sleep staging algorithm in all datasets and electrode configurations with the aim of a 30-s epoch-wise concordance – using Cohen's kappa – with the 'gold standard' of PSG. The application of the algorithm in this subset of consecutive patients referred to a Level 1 medical PSG was used to thoroughly analyse the performance of the in-ear sensor under pathological conditions and in theoretically less cooperative patients due to behaviour or comorbidities affecting the quality of signal acquisition. In our opinion, the results show the notional worst-case scenario and supplement the literature on this topic with non-normative data (i.e., from healthy populations), as has already been published several times. The confusion matrices and the corresponding kappa values of the three datasets (PSG, ELR and ET) processed by the automatic sleep stage algorithm yielded values between 0.74 and 0.76, which were plotted against our chosen ground truth – PSG_man, indicating a very strong agreement. Nevertheless, it is very important and of course has potential clinical implications that the NREM sleep stage N1 is most often misinterpreted towards W or N2, with sensitivities reaching around 60% for all three

methods, while the higher sensitivities were achieved for the identification of N2 and REM sleep. The good performance of in-ear EEG electrodes has been extended to the clinical group beyond the healthy populations usually addressed, opening promising perspectives for use in the field of neurophysiological assessment of sleep disorders.

4.1 | In-ear EEG validation in healthy individuals

Both of the in-ear EEG electrode configurations showed strong predictive abilities for different sleep parameters. In the healthy individuals, the ET method consistently showed determination coefficients around or above 0.90. In contrast, we found that ACT performed poorly in healthy, collaborative patients. Consequently, it was not meaningful to transfer the ACT method to clinical patients, as both in-ear EEG methods correlated better with the 'gold standard' of PSG in all sleep metrics, as also addressed recently (Mikkelsen et al., 2017). Looney et al. (2016) in a preliminary study achieved an average accuracy of 0.83 and 0.84 (in four healthy individuals during 360 epochs of 30 s of recording) for discriminating between NREM sleep and wakefulness (Scenario 1) and between N2/N3 and wakefulness/N1 (Scenario 2), respectively. These results stress out the good performance of ear-EEG devices such as our own, as we achieved an average correlation of 0.85 for the evaluation of four phases. Quite similar results were obtained in other studies by Stochholm et al. (2016), Nakamura et al. (2017), Jorgensen et al. (2020) and Kjaer et al. (2022), although the first study used a single subject as a sample and had the limitation that the ear-EEG recordings were unilateral and referenced to one electrode in the same ear, while the scalp electrodes referenced to the opposite side. Therefore, it was postulated that a better comparison of sleep stages using scalp EEG and in-ear EEG would be of great benefit it would be advantageous if further tests (with larger samples) were carried out in which the in-ear EEG recordings were not only referenced to the scalp, but also to the opposite ear. These and other considerations were taken into account when planning and conducting our study. Zibrandtsen et al. (2016) compared the signals from in-ear EEG and scalp EEG during different sleep stages. Two experienced sleep scorers were recruited to analyse both signals. The reported results showed that the two methods were in 90.9% agreement and had similar characteristics. Nakamura et al. (2017) compared an automatic classification algorithm between scalp and in-ear EEG recordings in four subjects and obtained an accuracy of 76.8% and a kappa of 0.65 in the classification of W, N1, N2 and N3, which according to the interpretation of Cohen's kappa represents a substantial agreement. Mikkelsen, Tabar, et al. (2019) also performed a direct comparison of the performance of an automatic sleep classification algorithm on scalp EEG and in-ear EEG datasets using 80 PSG and obtained similar results, in particular an average Cohen's kappa of 0.81 for classification into five stages. Tabar et al. (2021), obtained a kappa and an accuracy of 0.72 and 80% using 80 PSG recordings with TST, SL and SE values using in-ear EEG of 381.6 min (PSG = 386.9 min), 9.41 min (PSG = 10.79 min) and 81.1% (PSG = 82.1%), respectively. One study (Kjaer et al., 2022) found good performance with high accuracy and precision between scorers and the automated sleep assessment for the sleep metrics TST, SL and SE and poor performance with high accuracy and low precision between scorers and the automated sleep assessment for REM latency and the REM fraction of sleep. These aspects emphasise the continued importance of manually scored PSG for sleep assessment. Against this background, Tabar et al. (2023) compared automatic sleep evaluation based on in-ear EEG with PSG-based sleep evaluation by a professionally trained sleep scorer. Cohen's kappa was used to assess the agreement between the manual and automatic sleep evaluation, resulting in an average kappa value of 0.71. Our study provides promising visual and statistical correlations for each variable, in line with the literature described above. This suggests that the designed sensor and configuration used are reliable and fully capable of capturing the essential EEG data for correct sleep classification. Nevertheless, there was little difference between the studies in which the sleep classification results were categorised as either automatic or manual. This suggests that there is probably little scope for increasing accuracy within the same framework. Furthermore, we are convinced that the improvement, if any, is not necessarily clinically meaningful.

4.2 | In-ear EEG in clinical patients

It is known that the PSGs of healthy individuals are easier to read and analyse than those of patients with sleep disorders. When methods are used to diagnose and classify diseases, it is important that they are validated on relevant populations. Our extension of the methodology to patients was therefore primarily motivated by two main premises: (i) the promising results and thus the efficiency of the in-ear EEG technique in sleep classification in the group of healthy individuals; (ii) the smaller number of scientific publications with a sample that did not consist of healthy individuals (Jorgensen et al., 2020). We had hoped that the results seen in our healthy volunteers would be equally robust and comparable to the 'gold standard' of PSG, which would open the door for the use of this technique in a clinical context. With this in mind, and considering the heterogeneity of the clinical indication of the patients who underwent PSG (OSA for seven and seven different reasons; Table 1), the ELR method was characterised by perfect performance for TRT (alongside ET), near perfect results for TST and SE, and exceptional predictive accuracy for SL. When we analyse the ability to detect sleep stages, the scenario reverses, i.e., ET performs better than single-channel in-ear EEG (ELR) in all four stages (N1–0.70 versus 0.36; N2–0.88 versus 0.87; N3–0.86 versus 0.78; REM – 0.94 versus 0.92). Thus, the cited studies and other investigations (Alqurashi et al., 2018; Mikkelsen, Kappel, et al., 2019; Mikkelsen, Phan, et al., 2022; Nakamura et al., 2020) have demonstrated the usefulness of in-ear EEG for sleep detection in healthy subjects, with not too different results. In relation to epoch-by-epoch agreement and the use of the automatic sleep staging software illustrated by the confusion matrices, it became clear that the identified biases associated with the algorithm could affect the accuracy of sleep stage distribution reports and influence clinical decision-making when

used in the clinical setting. Accurate classification of sleep stages is crucial for the diagnosis of sleep disorders, and misclassification may lead to misdiagnosis or incorrect assessment of sleep quality. In sleep research, the accuracy of data is of paramount importance, and bias can distort study results and affect the reliability of research findings and subsequent recommendations. Understanding bias can help improve the algorithms used in these devices. Addressing the tendency to misclassify N1 stages (sensitivity ~0.60 for each of the three methods, although specificity is 0.95) may improve the overall accuracy of sleep staging. To mitigate the effects of bias, the algorithms can be refined to better differentiate between similar stages such as N1 and N2. This distinction requires the identification of sleep spindles and/or K-complexes, which in-ear EEG has already been shown to be able to do (Mikkelsen, Kappel, et al., 2019; Pazuelo et al., 2024; Zibrandtsen et al., 2017). Nevertheless, a recognised lower spatial resolution, as well as an unfavourable signal-to-noise ratio, could impair the ability of this sensor to detect sleep spindles, even if it is a spatially circumscribed phenomenon (Andrillon et al., 2011). Increasing the amount and variety of training data may help the methods to learn more accurate classifications. Likewise, using rigorous cross-validation techniques can ensure that the models generalise well across different datasets and populations. While PSG_aut, ET_aut and ELR_aut perform well overall, the presence of bias, particularly in the classification of N1, indicates areas in need of improvement, which is corroborated by other studies (Jorgensen et al., 2020; Looney et al., 2016; Mikkelsen et al., 2017). Eliminating these biases is critical to improving the accuracy and reliability of sleep stage classification, thereby improving both clinical outcomes and research validity. As the method of sleep classification has traditionally been based on systematic sleep classification and absolute amplitudes and frequencies, one must argue and probably agree that it is an inaccurate process. This may explain why sleep experts do not achieve agreement >0.85 on Cohen's kappa (Rosenberg & Van Hout, 2013). Nonetheless, kappa values in patients with OSA were 0.82 for the ET and ELR methods, while they reached 0.84 for the PSG, which certainly encourages its use in this particular pathological population. Overall, both in-ear EEG methods (ET and ELR) show remarkable potential as reliable tools for sleep monitoring and can consequently produce all-night sleep hypnograms similar to those of PSG, with the statistically recognised additional capabilities of the two bipolar channels underlining the need for a cephalic contralateral reference for optimal prediction accuracy, regardless of its unknown clinical implications.

4.3 | Potential benefits of in-ear EEG for sleep disorders

The techniques used should also be applicable to non-textbook sleep phenotypes, where the structure, timing and external characteristics of individual sleep may either be masked by disease-related artefacts or where the sleep EEG itself and the physiological characteristics of the different phases may be radically altered by the patient's condition. The results to date show that in-ear EEG with cephalic reference is a method that can be used to reliably classify sleep in sleep disorders. The overall results are comparable to those obtained with the 'gold standard' method PSG, but with much simpler tools and better tolerability. This opens up the possibility of wider application of quantitative sleep staging in the clinical realm where the use of standard PSG is not an option, whether due to significant cost, poor availability or significant sleep disturbance (e.g., insomnia). Given its current performance and general design, we believe that in-ear EEG can make a valuable contribution to the treatment and diagnosis of sleep disorders where the amount, timing or type of sleep is of interest, such as narcolepsy, sleep-related breathing disorders, insomnia or idiopathic hypersomnia. Despite the potential for the use of simplified recording methods, the extraction of robust metrics, as we obtained in the present study, requires a high level of expertise for the visual staging process. Whether the verification process can also be simplified has not yet been investigated and remains to be explored.

5 | CONCLUSIONS

Overall, the in-ear EEG methods (ET and ELR) showed consistently strong positive correlations with PSG for all sleep parameters in both the healthy individuals and the patient group. These methods proved to be particularly effective in accurately capturing TRT, TST, SE and SL and often achieved near perfect correlation values. This high level of agreement with PSG suggests that they are reliable tools for assessing sleep in different contexts. In contrast, ACT, although useful, showed comparatively weaker correlations, suggesting that although widely used, it may not be as accurate as one might assume, especially when compared to in-ear EEG methods. We have therefore confirmed both visually and through extensive statistical testing, that this configuration of an active electrode in the external auditory canal in combination with a mid-temporal contralateral cephalic electrode captures sufficient EEG information to allow accurate classification of sleep stages. However, it is important to point out that a more detailed analysis of the EEG signals was not performed, which prevents electrophysiological validation of the in-ear EEG for the detection of various features relevant to sleep EEG, such as delta power modulation and/or sleep spindles. Nonetheless, in a real-world setting, this sensor has shown the potential to be integrated into the extensively used Level 3 PSG devices for home use, so that in addition to the conventional limited recording of biosignals in sleep medicine, the most important sleep metrics can also be documented in these devices with a single-channel EEG. The recordings from the self-applied in-ear EEG sensor (ELR) were slightly inferior to the contralateral EEG-referenced (ET) recordings on the scalp for most of the variables tested. This could pose a problem for the autonomous use of the in-ear EEG by the patient. However, whether the lower reliability has clinical implications needs to be investigated in more comprehensive studies, preferably with a multicentre design involving several observers and a larger sample size.

AUTHOR CONTRIBUTIONS

Daniel Filipe Borges: Conceptualization; investigation; methodology; visualization; writing – original draft; writing – review and editing; data curation; software. **Joana Isabel Soares:** Methodology; formal analysis. **Heloísa Silva:** Validation; software. **João Felgueiras:** Investigation. **Carla Batista:** Investigation. **Simão Ferreira:** Formal analysis; visualization; writing – original draft; data curation. **Nuno Barbosa Rocha:** Project administration; writing – review and editing. **Alberto Leal:** Conceptualization; methodology; project administration; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

The authors would like to thank Neuroevolution® for the technical and medical equipment support and Phillips Respironics® for the software licence and support.

FUNDING INFORMATION

This research received no external funding.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

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

INSTITUTIONAL REVIEW BOARD STATEMENT

The study was conducted in accordance with the Declaration of Helsinki.

INFORMED CONSENT STATEMENT

Written informed consent has been obtained from all subjects involved in the study.

ORCID

Daniel Filipe Borges https://orcid.org/0000-0003-0189-7908

Nuno Barbosa Rocha https://orcid.org/0000-0002-3139-2786

REFERENCES

Abbott, S. M., Malkani, R. G., & Zee, P. C. (2020). Circadian disruption and human health: A bidirectional relationship. *The European Journal of Neuroscience*, *51*(1), 567–583. https://doi.org/10.1111/ejn.14298

Alqurashi, Y. D., Nakamura, T., Goverdovsky, V., Moss, J., Polkey, M. I., Mandic, D. P., & Morrell, M. J. (2018). A novel in-ear sensor to determine sleep latency during the multiple sleep latency test in healthy adults with and without sleep restriction. *Nature and Science of Sleep*, *10*, 385–396. https://doi.org/10.2147/NSS.S175998

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, *26*(3), 342–392. https://doi.org/10.1093/sleep/26.3.342

Anderer, P., Moreau, A., Woertz, M., Ross, M., Gruber, G., Parapatics, S., Loretz, E., Heller, E., Schmidt, A., Boeck, M., Moser, D., Kloesch, G., Saletu, B., Saletu-Zyhlarz, G. M., Danker-Hopfe, H., Zeitlhofer, J., & Dorffner, G. (2010). Computer-assisted sleep classification according to the standard of the American Academy of sleep medicine: Validation study of the AASM version of the Somnolyzer 24 × 7. *Neuropsychobiology*, *62*(4), 250–264. https://doi.org/10.1159/000320864

Andrillon, T., Nir, Y., Staba, R. J., Ferrarelli, F., Cirelli, C., Tononi, G., & Fried, I. (2011). Sleep spindles in humans: Insights from intracranial EEG and unit recordings. *The Journal of Neuroscience*, *31*(49), 17821–17834. https://doi.org/10.1523/jneurosci.2604-11.2011

Arnardottir, E. S., Islind, A. S., Óskarsdóttir, M., Ólafsdóttir, K. A., August, E., Jónasdóttir, L., Hrubos-Strøm, H., Saavedra, J. M., Grote, L., Hedner, J., Höskuldsson, S., Ágústsson, J. S., Jóhannsdóttir, K. R., McNicholas, W. T., Pevernagie, D., Sund, R., Töyräs, J., & Leppänen, T. (2022). The sleep revolution project: The concept and objectives. *Journal of Sleep Research*, *31*(4), e13630. https://doi.org/10.1111/jsr.13630

Arnardottir, E. S., Verbraecken, J., Gonçalves, M., Gjerstad, M. D., Grote, L., Puertas, F. J., Mihaicuta, S., McNicholas, W. T., & Parrino, L. (2016). Variability in recording and scoring of respiratory events during sleep in Europe: A need for uniform standards. *Journal of Sleep Research*, *25*(2), 144–157. https://doi.org/10.1111/jsr.12353

Bandyopadhyay, A., & Goldstein, C. (2023). Clinical applications of artificial intelligence in sleep medicine: A sleep clinician's perspective. *Sleep & Breathing*, *27*(1), 39–55. https://doi.org/10.1007/s11325-022-02592-4

Berry, R. B., Brooks, R., Gamaldo, C., Harding, S. M., Lloyd, R. M., Quan, S. F., Troester, M. T., & Vaughn, B. V. (2017). AASM scoring manual updates for 2017 (version 2.4). *Journal of Clinical Sleep Medicine*, *13*(5), 665–666. https://doi.org/10.5664/jcsm.6576

Brancaccio, A., Tabarelli, D., Bigica, M., & Baldauf, D. (2020). Cortical source localization of sleep-stage specific oscillatory activity. *Scientific Reports*, *10*(1), 6976. https://doi.org/10.1038/s41598-020-63933-5

Chan, K., Wong, F. S., & Pearson, J. A. (2022). Circadian rhythms and pancreas physiology: A review. *Frontiers in Endocrinology (Lausanne)*, *13*, 920261. https://doi.org/10.3389/fendo.2022.920261

Chee, M. W., & Chuah, L. Y. (2008). Functional neuroimaging insights into how sleep and sleep deprivation affect memory and cognition. *Current Opinion in Neurology*, *21*(4), 417–423. https://doi.org/10.1097/WCO.0b013e3283052cf7

Chellappa, S. L., Vujovic, N., Williams, J. S., & Scheer, F. (2019). Impact of circadian disruption on cardiovascular function and disease. *Trends in Endocrinology and Metabolism*, *30*(10), 767–779. https://doi.org/10.1016/j.tem.2019.07.008

Cohen, J. (1960a). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46. https://doi.org/10.1177/001316446002000104

Danzig, R., Wang, M., Shah, A., & Trotti, L. M. (2020). The wrist is not the brain: Estimation of sleep by clinical and consumer wearable actigraphy devices is impacted by multiple patient- and device-specific factors. *Journal of Sleep Research*, *29*(1), e12926. https://doi.org/10.1111/jsr.12926

Doğan, N. (2018). Bland-Altman analysis: A paradigm to understand correlation and agreement. *Turkish Journal of Emergency Medicine*, *18*(4), 139–141. https://doi.org/10.1016/j.tjem.2018.09.001

Fischer, J., Dogas, Z., Bassetti, C. L., Berg, S., Grote, L., Jennum, P., Levy, P., Mihaicuta, S., Nobili, L., Riemann, D., Puertas Cuesta, F. J., Raschke, F., Skene, D. J., Stanley, N., & Pevernagie, D. (2012). Standard procedures for adults in accredited sleep medicine centres in Europe. *Journal of Sleep Research*, *21*(4), 357–368. https://doi.org/10.1111/j.1365-2869.2011.00987.x

Flemons, W. W., Douglas, N. J., Kuna, S. T., Rodenstein, D. O., & Wheatley, J. (2004). Access to diagnosis and treatment of patients with suspected sleep apnea. *American Journal of Respiratory and Critical Care Medicine*, *169*(6), 668–672. https://doi.org/10.1164/rccm.200308-1124PP

Frank, M. G. (2011). Sleep and developmental plasticity not just for kids. *Progress in Brain Research*, *193*, 221–232. https://doi.org/10.1016/B978-0-444-53839-0.00014-4

Garfield, V. (2019). The association between body mass index (BMI) and sleep duration: Where are we after nearly two decades of epidemiological research? *International Journal of Environmental Research and Public Health*, *16*(22), 4327. https://doi.org/10.3390/ijerph16224327

Gavriloff, D., Sheaves, B., Juss, A., Espie, C. A., Miller, C. B., & Kyle, S. D. (2018). Sham sleep feedback delivered via actigraphy biases daytime symptom reports in people with insomnia: Implications for insomnia disorder and wearable devices. *Journal of Sleep Research*, *27*(6), e12726. https://doi.org/10.1111/jsr.12726

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Goldstein, C. A., Berry, R. B., Kent, D. T., Kristo, D. A., Seixas, A. A., Redline, S., & Westover, M. B. (2020). Artificial intelligence in sleep medicine: Background and implications for clinicians. *Journal of Clinical Sleep Medicine*, *16*(4), 609–618. https://doi.org/10.5664/jcsm.8388

Goverdovsky, V., Looney, D., Kidmose, P., & Mandic, D. P. (2016). In-ear EEG from viscoelastic generic earpieces: Robust and unobtrusive 24/7 monitoring. *IEEE Sensors Journal*, *16*(1), 271–277. https://doi.org/10.1109/Jsen.2015.2471183

Groeger, J. A., Zijlstra, F. R., & Dijk, D. J. (2004). Sleep quantity, sleep difficulties and their perceived consequences in a representative sample of some 2000 British adults. *Journal of Sleep Research*, *13*(4), 359–371. https://doi.org/10.1111/j.1365-2869.2004.00418.x

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Hung, M., Bounsanga, J., & Voss, M. W. (2017). Interpretation of correlations in clinical research. *Postgraduate Medicine*, *129*(8), 902–906. https://doi.org/10.1080/00325481.2017.1383820

Imtiaz, S. A. (2021). A systematic review of sensing technologies for wearable sleep staging. *Sensors (Basel)*, *21*(5), 1562. https://doi.org/10.3390/s21051562

Jean-Louis, G., Kripke, D. F., & Ancoli-Israel, S. (2000). Sleep and quality of well-being. *Sleep*, *23*(8), 1115–1121.

Jorgensen, S. D., Zibrandtsen, I. C., & Kjaer, T. W. (2020). Ear-EEG-based sleep scoring in epilepsy: A comparison with scalp-EEG. *Journal of Sleep Research*, *29*(6), e12921. https://doi.org/10.1111/jsr.12921

Kaongoen, N., Choi, J., Woo Choi, J., Kwon, H., Hwang, C., Hwang, G., Kim, B. H., & Jo, S. (2023). The future of wearable EEG: A review of ear-EEG technology and its applications. *Journal of Neural Engineering*, *20*(5), 051002. https://doi.org/10.1088/1741-2552/acfcda

Kerkhofs, M., & Boudjeltia, K. Z. (2012). From total sleep deprivation to cardiovascular disease: A key role for the immune system? *Sleep*, *35*(7), 895–896. https://doi.org/10.5665/sleep.1938

Killgore, W. D. (2010). Effects of sleep deprivation on cognition. *Progress in Brain Research*, *185*, 105–129. https://doi.org/10.1016/B978-0-444-53702-7.00007-5

Kim, H., Kim, D., & Oh, J. (2022). Automation of classification of sleep stages and estimation of sleep efficiency using actigraphy. *Frontiers in Public Health*, *10*, 1092222. https://doi.org/10.3389/fpubh.2022.1092222

Kjaer, T. W., Rank, M. L., Hemmsen, M. C., Kidmose, P., & Mikkelsen, K. (2022). Repeated automatic sleep scoring based on ear-EEG is a valuable alternative to manually scored polysomnography. *PLOS Digit Health*, *1*(10), e0000134. https://doi.org/10.1371/journal.pdig.0000134

Kohansieh, M., & Makaryus, A. N. (2015). Sleep deficiency and deprivation leading to cardiovascular disease. *International Journal of Hypertension*, *2015*, 615681. https://doi.org/10.1155/2015/615681

Kwon, K., Kwon, S., & Yeo, W. H. (2022). Automatic and accurate sleep stage classification via a convolutional deep neural network and nanomembrane electrodes. *Biosensors (Basel)*, *12*(3), 155. https://doi.org/10.3390/bios12030155

Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin*, *136*(3), 375–389. https://doi.org/10.1037/a0018883

Lingas, E. C. (2023). A narrative review of the carcinogenic effect of night shift and the potential protective role of melatonin. *Cureus*, *15*(8), e43326. https://doi.org/10.7759/cureus.43326

Looney, D., Goverdovsky, V., Rosenzweig, I., Morrell, M. J., & Mandic, D. P. (2016). Wearable in-ear encephalography sensor for monitoring sleep. Preliminary observations from nap studies. *Annals of the American Thoracic Society*, *13*(12), 2229–2233. https://doi.org/10.1513/AnnalsATS.201605-342BC

Looney, D., Kidmose, P., Park, C., Ungstrup, M., Rank, M., Rosenkranz, K., & Mandic, D. (2012). The in-the-ear recording concept: User-centered and wearable brain monitoring. *IEEE Pulse*, *3*(6), 32–42. https://doi.org/10.1109/MPUL.2012.2216717

Looney, D., Park, C., Kidmose, P., Rank, M. L., Ungstrup, M., Rosenkranz, K., & Mandic, D. P. (2011). An in-the-ear platform for recording electroencephalogram. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *2011*, 6882–6885. https://doi.org/10.1109/iembs.2011.6091733

Maggio, M., Colizzi, E., Fisichella, A., Valenti, G., Ceresini, G., Dall'Aglio, E., Ruffini, L., Lauretani, F., Parrino, L., & Ceda, G. P. (2013). Stress hormones, sleep deprivation and cognition in older adults. *Maturitas*, *76*(1), 22–44. https://doi.org/10.1016/j.maturitas.2013.06.006

Marino, M., Li, Y., Rueschman, M. N., Winkelman, J. W., Ellenbogen, J. M., Solet, J. M., Dulin, H., Berkman, L. F., & Buxton, O. M. (2013). Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*, *36*(11), 1747–1755. https://doi.org/10.5665/sleep.3142

Marom, N. D., Rokach, L., & Shmilovici, A. (2010). Using the confusion matrix for improving ensemble classifiers. 2010 IEEE 26-th Convention of Electrical and Electronics Engineers in Israel.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica (Zagreb)*, *22*(3), 276–282. https://www.biochemia-medica.com/assets/images/upload/xml_tif/McHugh_ML_Interrater_reliability.pdf

Meiser, A., Lena Knoll, A., & Bleichner, M. G. (2024). High-density ear-EEG for understanding ear-centered EEG. *Journal of Neural Engineering*, *21*(1), 016001. https://doi.org/10.1088/1741-2552/ad1783

Meiser, A., Tadel, F., Debener, S., & Bleichner, M. G. (2020). The sensitivity of ear-EEG: Evaluating the source-sensor relationship using forward modeling. *Brain Topography*, *33*(6), 665–676. https://doi.org/10.1007/s10548-020-00793-2

Mikkelsen, K. B., Ebajemito, J. K., Bonmati-Carrion, M. A., Santhi, N., Revell, V. L., Atzori, G., Della Monica, C., Debener, S., Dijk, D. J., Sterr, A., & de Vos, M. (2019). Machine-learning-derived sleep-wake staging from around-the-ear electroencephalogram outperforms manual scoring and actigraphy. *Journal of Sleep Research*, *28*(2), e12786. https://doi.org/10.1111/jsr.12786

Mikkelsen, K. B., Kappel, S. L., Hemmsen, M. C., Rank, M. L., & Kidmose, P. (2019). Discrimination of sleep spindles in ear-EEG. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *2019*, 6697–6700. https://doi.org/10.1109/EMBC.2019.8857114

Mikkelsen, K. B., Phan, H., Rank, M. L., Hemmsen, M. C., de Vos, M., & Kidmose, P. (2022). Sleep monitoring using ear-centered setups: Investigating the influence from electrode configurations. *IEEE Transactions on Biomedical Engineering*, *69*(5), 1564–1572. https://doi.org/10.1109/TBME.2021.3116274

Mikkelsen, K. B., Tabar, Y. R., Kappel, S. L., Christensen, C. B., Toft, H. O., Hemmsen, M. C., Rank, M. L., Otto, M., & Kidmose, P. (2019). Accurate whole-night sleep monitoring with dry-contact ear-EEG. *Scientific Reports*, *9*(1), 16824. https://doi.org/10.1038/s41598-019-53115-3

Mikkelsen, K. B., Tabar, Y. R., Toft, H. O., Hemmsen, M. C., Rank, M. L., & Kidmose, P. (2022). Self-applied ear-EEG for sleep monitoring at home. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *2022*, 3135–3138. https://doi.org/10.1109/EMBC48229.2022.9871076

Mikkelsen, K. B., Villadsen, D. B., Otto, M., & Kidmose, P. (2017). Automatic sleep staging using ear-EEG. *Biomedical Engineering Online*, *16*(1), 111. https://doi.org/10.1186/s12938-017-0400-5

Mishra, P., Panigrahi, M., & Ankit, D. (2020). Cognition and alertness in medical students: Effects of a single night of partial sleep deprivation. *Annals of Neurosciences*, *27*(2), 57–62. https://doi.org/10.1177/0972753120965083

Mohamed, M., Mohamed, N., & Kim, J. G. (2023). Advancements in wearable EEG Technology for Improved Home-Based Sleep Monitoring and Assessment: A review. *Biosensors (Basel)*, *13*(12), 1019. https://doi.org/10.3390/bios13121019

Murray, J. M., Sletten, T. L., Magee, M., Gordon, C., Lovato, N., Bartlett, D. J., Kennaway, D. J., Lack, L. C., Grunstein, R. R., Lockley, S. W., & Rajaratnam, S. M. (2017). Prevalence of circadian misalignment and its association with depressive symptoms in delayed sleep phase disorder. *Sleep*, *40*(1), zsw002. https://doi.org/10.1093/sleep/zsw002

Nakamura, T., Alqurashi, Y. D., Morrell, M. J., & Mandic, D. P. (2020). Hearables: Automatic overnight sleep monitoring with standardized in-ear EEG sensor. *IEEE Transactions on Biomedical Engineering*, *67*(1), 203–212. https://doi.org/10.1109/TBME.2019.2911423

Nakamura, T., Goverdovsky, V., Morrell, M. J., & Mandic, D. P. (2017). Automatic sleep monitoring using ear-EEG. *IEEE Journal of Translational Engineering in Health and Medicine*, *5*, 1-8. https://doi.org/10.1109/JTEHM.2017.2702558

Olesen, A. N., Jørgen Jennum, P., Mignot, E., & Sorensen, H. B. D. (2021). Automatic sleep stage classification with deep residual networks in a mixed-cohort setting. *Sleep*, *44*(1), zsaa161. https://doi.org/10.1093/sleep/zsaa161

Pan, Y., Zhou, Y., Shi, X., He, S., & Lai, W. (2023). The association between sleep deprivation and the risk of cardiovascular diseases: A systematic meta-analysis. *Biomedical Reports*, *19*(5), 78. https://doi.org/10.3892/br.2023.1660

Pazuelo, J., Juez, J. Y., Moumane, H., Pyrzowski, J., Mayor, L., Segura-Quijano, F. E., Valderrama, M., & Le Van Quyen, M. (2024). Evaluating the electroencephalographic signal quality of an in-ear wearable device. *Sensors (Basel)*, *24*(12), 3973. https://doi.org/10.3390/s24123973

Peterson, B. T., Chiao, P., Pickering, E., Freeman, J., Zammit, G. K., Ding, Y., & Badura, L. L. (2012). Comparison of actigraphy and polysomnography to assess effects of zolpidem in a clinical research unit. *Sleep Medicine*, *13*(4), 419–424. https://doi.org/10.1016/j.sleep.2011.12.003

Rapoport, D. M. (2019). Non-contact sleep monitoring: Are we there yet? *Journal of Clinical Sleep Medicine*, *15*(7), 935–936. https://doi.org/10.5664/jcsm.7864

Ronzhina, M., Janoušek, O., Kolářová, J., Nováková, M., Honzík, P., & Provazník, I. (2012). Sleep scoring using artificial neural networks. *Sleep Medicine Reviews*, *16*(3), 251–263. https://doi.org/10.1016/j.smrv.2011.06.003

Rosenberg, R. S., & Van Hout, S. (2013). The American Academy of sleep medicine inter-scorer reliability program: Sleep stage scoring. *Journal of Clinical Sleep Medicine*, *9*(1), 81–87. https://doi.org/10.5664/jcsm.2350

Sletten, T. L., Weaver, M. D., Foster, R. G., Gozal, D., Klerman, E. B., Rajaratnam, S. M. W., Roenneberg, T., Takahashi, J. S., Turek, F. W., Vitiello, M. V., Young, M. W., & Czeisler, C. A. (2023). The importance of sleep regularity: A consensus statement of the National Sleep Foundation sleep timing and variability panel. *Sleep Health*, *9*, 801–820. https://doi.org/10.1016/j.sleh.2023.07.016

Song, H. T., Sun, X. Y., Yang, T. S., Zhang, L. Y., Yang, J. L., & Bai, J. (2015). Effects of sleep deprivation on serum cortisol level and mental health in servicemen. *International Journal of Psychophysiology*, *96*(3), 169–175. https://doi.org/10.1016/j.ijpsycho.2015.04.008

Stochholm, A., Mikkelsen, K., & Kidmose, P. (2016). Automatic sleep stage classification using ear-EEG. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *2016*, 4751–4754. https://doi.org/10.1109/EMBC.2016.7591789

St-Onge, M. P., Grandner, M. A., Brown, D., Conroy, M. B., Jean-Louis, G., Coons, M., Bhatt, D. L., American Heart Association Obesity, B. C. D, Nutrition Committees of the Council on, L, Cardiometabolic, H., Council on Cardiovascular Disease in the, Y, Council on Clinical, C, & Stroke, C. (2016). Sleep duration and quality: Impact on lifestyle behaviors and Cardiometabolic health: A scientific statement from the American Heart Association. *Circulation*, *134*(18), e367–e386. https://doi.org/10.1161/CIR.0000000000000444

Tabar, Y. R., Mikkelsen, K. B., Rank, M. L., Hemmsen, M. C., Otto, M., & Kidmose, P. (2021). Ear-EEG for sleep assessment: A comparison with actigraphy and PSG. *Sleep & Breathing*, *25*(3), 1693–1705. https://doi.org/10.1007/s11325-020-02248-1

Tabar, Y. R., Mikkelsen, K. B., Shenton, N., Kappel, S. L., Bertelsen, A. R., Nikbakht, R., Toft, H. O., Henriksen, C. H., Hemmsen, M. C., Rank, M. L., Otto, M., & Kidmose, P. (2023). At-home sleep monitoring using generic ear-EEG. *Frontiers in Neuroscience*, *17*, 987578. https://doi.org/10.3389/fnins.2023.987578

Thompson, K. I., Chau, M., Lorenzetti, M. S., Hill, L. D., Fins, A. I., & Tartar, J. L. (2022). Acute sleep deprivation disrupts emotion, cognition, inflammation, and cortisol in young healthy adults. *Frontiers in Behavioral Neuroscience*, *16*, 945661. https://doi.org/10.3389/fnbeh.2022.945661

Troester, M. M., Brooks, R., Gamaldo, C. E., Harding, S. M., Lloyd, R. M., Quan, S. F., Vaughn, B. V., Berry, R. B., … for the American Academy of Sleep Medicine. (2023). The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Version 3.

Uccella, S., Cordani, R., Salfi, F., Gorgoni, M., Scarpelli, S., Gemignani, A., Geoffroy, P. A., De Gennaro, L., Palagini, L., Ferrara, M., & Nobili, L. (2023). Sleep deprivation and insomnia in adolescence: Implications for mental health. *Brain Sciences*, *13*(4), 569. https://doi.org/10.3390/brainsci13040569

Vanacore, A., & Pellegrino, M. S. (2022). Robustness of κ -type coefficients for clinical agreement. *Statistics in Medicine*, *41*(11), 1986–2004. https://doi.org/10.1002/sim.9341

Verma, R. K., Dhillon, G., Grewal, H., Prasad, V., Munjal, R. S., Sharma, P., Buddhavarapu, V., Devadoss, R., Kashyap, R., & Surani, S. (2023). Artificial intelligence in sleep medicine: Present and future. *World Journal of Clinical Cases*, *11*(34), 8106–8110. https://doi.org/10.12998/wjcc.v11.i34.8106

Wild, C. J., Nichols, E. S., Battista, M. E., Stojanoski, B., & Owen, A. M. (2018). Dissociable effects of self-reported daily sleep duration on high-level cognitive abilities. *Sleep*, *41*(12), zsy182. https://doi.org/10.1093/sleep/zsy182

Xie, L., Kang, H., Xu, Q., Chen, M. J., Liao, Y., Thiyagarajan, M., O'Donnell, J., Christensen, D. J., Nicholson, C., Iliff, J. J., Takano, T., Deane, R., & Nedergaard, M. (2013). Sleep drives metabolite clearance from the adult brain. *Science*, *342*(6156), 373–377. https://doi.org/10.1126/science.1241224

You, Y., Zhong, X., Liu, G., & Yang, Z. (2022). Automatic sleep stage classification: A light and efficient deep neural network model based on time,

frequency and fractional Fourier transform domain features. *Artificial Intelligence in Medicine*, *127*, 102279. https://doi.org/10.1016/j.artmed.2022.102279

Zibrandtsen, I., Kidmose, P., Otto, M., Ibsen, J., & Kjaer, T. W. (2016). Case comparison of sleep features from ear-EEG and scalp-EEG. *Sleep Science*, *9*(2), 69–72. https://doi.org/10.1016/j.slsci.2016.05.006

Zibrandtsen, I. C., Kidmose, P., Christensen, C. B., & Kjaer, T. W. (2017). Ear-EEG detects ictal and interictal abnormalities in focal and generalized epilepsy – a comparison with scalp EEG monitoring. *Clinical Neurophysiology*, *128*(12), 2454–2461. https://doi.org/10.1016/j.clinph.2017.09.115

How to cite this article: Borges, D. F., Soares, J. I., Silva, H., Felgueiras, J., Batista, C., Ferreira, S., Rocha, N. B., & Leal, A. (2024). A custom-built single-channel in-ear electroencephalography sensor for sleep phase detection: an interdependent solution for at-home sleep studies. *Journal of Sleep Research*, e14368. https://doi.org/10.1111/jsr.14368