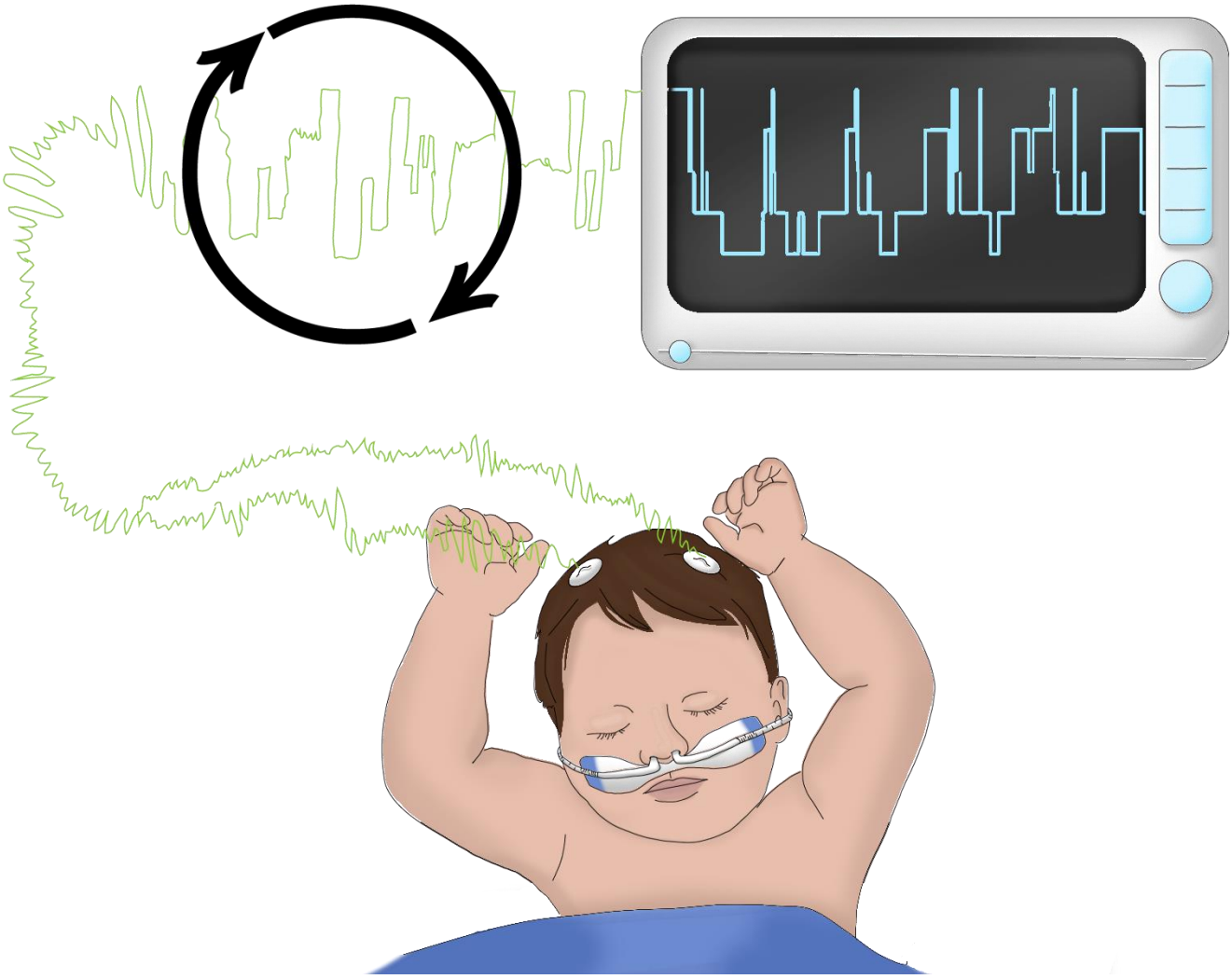
MASTER'S THESIS

Automated EEG-based sleep monitoring in critically ill children

F.W. (Floor) Hiemstra
MSc Technical Medicine, track: Sensing & Stimulation
September 2021



Universiteit Leiden
The Netherlands

TU Delft Delft University of Technology

Erasmus University Rotterdam

Cover illustration by Daniela Estrada

MASTER'S THESIS

Automated EEG-based sleep monitoring in critically ill children

F.W. (Floor Willemijn) Hiemstra
Student number: 4395468
September 13, 2021

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University | Delft University of Technology | Erasmus University Rotterdam

Technical Medicine, Track: Sensing & Stimulation
Faculty of Mechanical, Maritime and Materials Engineering (3mE)
Mekelweg 2
2628 CD Delft
The Netherlands

Master thesis project (TM30004 - 35 ECTS)
Intensive Care Unit, Department of Pediatric Surgery
Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands
Dec 2020 – Sept 2021

Supervisors

Dr. R.C.J. de Jonge	Erasmus MC	Medical supervisor 1
Dr. J.W. Kuiper	Erasmus MC	Medical supervisor 2
Dr. ir. D.M.J. Tax	TU Delft	Technical supervisor
Drs. A.B.G. Cramer	Erasmus MC	Daily supervisor

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Assessment committee

Chairman	Prof. dr. K.F.M. (Koen) Joosten Professor of Pediatrics, pediatrician-intensivist Intensive Care Unit, Department of Pediatric Surgery Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands
Medical supervisor 1	Dr. R.C.J. (Rogier) de Jonge Pediatrician-intensivist, neonatologist Intensive Care Unit, Department of Pediatric Surgery Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands
Medical supervisor 2	Dr. J.W. (Jan Willem) Kuiper Pediatrician-intensivist Intensive Care Unit, Department of Pediatric Surgery Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands
Technical supervisor	Dr. D.M.J. (David) Tax Assistant professor Pattern Recognition Laboratory, Faculty of Electrical Engineering, Mathematics and Computer Science Delft University of Technology, Delft, The Netherlands
Daily supervisor	Drs. A.B.G. (Arnout) Cramer PhD candidate Intensive Care Unit, Department of Pediatric Surgery Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands

Preface

With this thesis, my journey at the TU Delft will be finished. I started this journey as the first cohort of the bachelor program Clinical Technology, without having any idea where it would bring me. During my bridging program to Medicine after completing my bachelor, I realized that it was not only the clinical practice that excited me, but it was especially the combination of technology and medicine. I started the Technical Medicine master track Sensing & Stimulation, together with 9 enthusiastic students. After the first year with inspiring teachers and interesting topics, I performed my second-year internships at the department of Transplantation Surgery, the Intensive Care Unit and the Neonatal Intensive Care Unit. I was immediately triggered by the intensive care environment and its complex patient population, multidisciplinary team and advanced monitoring and technology offers many opportunities and challenges for a technical physician to be. During my industry internship at Philips Research, I had the opportunity to improve my skills and increase my interest in signal processing and patient monitoring. It was therefore that this present thesis project, in which I could combine my passion for the ICU, patient monitoring and signal processing, sparked my interest. The subject of sleep and circadian rhythms was very appealing, perhaps also because I conducted my high school research project on the effects of circadian disturbances on expression of brown fat in mice. Without having in depth knowledge of machine learning, I kicked off this internship by following a TU Delft machine learning course. Whereas I thought I had already finished my last exam by ending my first master's year, there was one more to go: this machine learning exam, topped off by the experience of digital home exams during the COVID-19 pandemic. During this thesis project, I learned a lot regarding clinical, technical and scientific aspects, as I performed an explorative study and developed machine learning models on physiological data sets. I really enjoyed working at the PICU environment, where I learned so much about the fascinating congenital disorders, technical challenges encountered in the PICU and relevance of sleep and its disturbances in PICU patients.

I am really excited to work in multidisciplinary environments and combine my enthusiasm for the human body and technology in clinical practice. Therefore, I am looking forward to start my career as a technical physician and to discover what the future has in store!

Summary

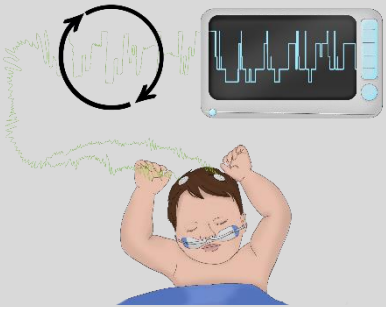
Sleep deprivation is commonly encountered in critically ill children admitted to the pediatric intensive care unit (PICU) and is associated with poor clinical outcome. Automated electroencephalography (EEG)-based monitoring of sleep enables the study of sleep without the need for visual assessment of the EEG signals, the golden standard. The real-time and continuous assessment of a patient's sleep state that could be established by automated EEG-based sleep monitoring is a prerequisite for individual optimization of sleep.

This thesis aimed to explore the potentials and pitfalls of automated sleep monitoring in PICU patients, that has – to our knowledge – not been developed yet. First, a literature study was performed to summarize the effects of various PICU factors on the sleep EEG and discuss proposed sleep monitoring methods in adult ICU and neonatal ICU patients. It was concluded that many medical conditions, sedatives and analgesics cause slowing of the EEG that challenge PICU sleep monitoring. In healthy adults, many classification methods have been proposed, varying from simple threshold-based methods to complex neural networks. The study of automated sleep monitoring in adult ICU patients has been limited, although more studies have investigated automated sleep monitoring in NICU patients. Next, an explorative study was performed in which various classification methods for automated EEG-based sleep monitoring in both non-critically as well as critically ill children were developed and evaluated. It was concluded that a simple index measure is a promising method to monitor sleep in PICU patients. Machine learning models developed in non-critically ill patients cannot easily be applied to PICU patients in whom the sleep EEG is frequently deviant. Future efforts should focus on further tuning, training and validating the classification models with more PICU data. In Part I of this thesis, the literature study can be found. Part II consists of the research report of the explorative study. Supplementary materials are provided at the end of this thesis report.

Although the results do not encourage immediate implementation in clinical practice, they do warrant further development and testing. With this thesis, a first step towards automated sleep monitoring in the PICU has been made.

Table of contents

Part I – Literature study



The potential and challenges of automated EEG-based sleep monitoring in critically ill children: a literature review

F.W. (Floor) Hiemstra, MSc student Technical Medicine

Abstract

Sleep deprivation is commonly encountered in critically ill children admitted to the paediatric intensive care unit (PICU) and is associated with poor clinical outcome. Automated electroencephalography (EEG)-based depth of sleep monitoring enables real-time continuous study of sleep in PICU patients without the need for visual assessment of the EEG signals, the gold standard. Real-time and continuous knowledge of a patient's sleep state is an essential prerequisite for individual optimization of sleep. A variety of PICU factors interfere with the EEG, including the patient's age and developmental state, underlying illness and medication. The heterogeneous PICU population challenges the development of an automated sleep monitoring method, which has, to our knowledge, not been developed yet. This review aids in this development by providing an overview of the potential and barriers of an automated EEG-based sleep monitoring method in critically ill children by summarizing the effects of various PICU factors on the sleep EEG and discussing proposed sleep monitoring methods in adult ICU and neonatal ICU patients.

1. Introduction

Sleep is a vital state of the human body that is essential to life. Although its function is not fully understood, sleep is thought to have a restorative and memory consolidative function[1]. Children admitted to the pediatric intensive care unit (PICU) are exposed to various risk factors for sleep deprivation including environmental factors, medication, morbidity and discomfort[2]. Sleep studies in PICU patients demonstrated frequent occurrence of sleep deprivation, characterized by fragmentation of sleep, reduced total sleep time, disrupted sleep architecture and a disproportional amount of sleep occurring during daytime[3–7]. Sleep deprivation affects the homeostatic processes of the body[8] and is associated with immune dysfunction[9,10] and development of delirium[11], all potentially leading to poor outcomes and prolonged PICU stay. Besides, in children, sleep also has an important role in brain maturation[12]. It must be clear that the importance of sleep during (recovery from) critical illness deserves attention, particularly for children who are undergoing active neurologic maturation. Monitoring of sleep enables the revelation of links between negative outcomes and sleep deprivation, and optimization of sleep.

Sleep can be measured using a combination of electroencephalography (EEG), electromyography (EMG) and electrooculography (EOG). Based on the spectral composition and features in the EEG, EMG and EOG signals, sleep can be divided into four different stages that indicate the depth of sleep: rapid eye movement (REM) sleep and non-REM (NREM) sleep, subdivided into NREM stage 1 (N1), NREM stage 2 (N2) and NREM stage 3 (N3). In healthy adults, N1, N2, N3 and REM alternate in a cyclical fashion defining the normal sleep architecture. Sleep staging is traditionally done by visual analysis of these signals according to the American Association of Sleep Medicine (AASM) criteria[13]. However, the use of the AASM criteria in characterizing sleep of critically ill patients is often debated due to confounding of the EEG signals by effects of the underlying illness and medication[14–17]. This confounding and the subjective nature of visual sleep scoring leads to a high inter-observer variability. Ambrogio et al. showed that interobserver variability of visually scored sleep EEGs from sedated and mechanically ventilated critically ill adult patients ($\kappa = 0.52 \pm 0.23$) is high compared to that for the healthy control patients ($\kappa = 0.89 \pm 0.13$; $p = 0.03$)[18]. Furthermore, visual scoring of the EEG signals is time-consuming and requires skilled personnel. Automated monitoring of sleep has the potential to continuously and in real time indicate depth of sleep, which could be used to directly detect disturbed sleep patterns and optimize sleep.

Sleep staging could be automated by computerized processing of EEG signals. Next to the EEG signals, other physiological signals could also provide information about a patient's sleep stage, including ECG, EMG, EOG and respiratory signals, or a combination of multiple physiological signals. However, the intensive support of heart and lung function and administration of muscle relaxants limits the use of ECG, EMG and respiratory signals in sleep staging. Actigraphy has also been suggested to assess sleep by measuring motion with an accelerometer, typically placed on the wrist or ankle[19,20]. By counting the movements within an epoch, the device can determine if the patients is probably asleep or awake. Clearly, this method could not provide any information related to the stage of patient's sleep. The primary weakness of actigraphy in the assessment of sleep is that it is solely based on the quantification of movements, while most ICU patients have reduced movement due to sedation, muscle relaxants and critical illness. Therefore, EEG remains the golden standard in the assessment of sleep.

Several attempts have been made to develop an automated sleep staging algorithm to monitor sleep of patients admitted to the adult ICU or neonatal intensive care unit (NICU), using EEG signals. These algorithms determine features in the EEG signal that correlate with the depth of sleep, which is related to the previously described sleep stages, in critically ill patients. Since maturation of the brain during childhood is associated with changes in the sleep EEG[21], these algorithms developed for ICU or NICU may not be suitable for depth of sleep monitoring in all critically ill children admitted to the PICU. To our knowledge, an algorithm that is specifically suitable for depth of sleep monitoring in critically ill children has not been developed yet.

The heterogeneity in age and critical illness encountered in the PICU introduces several challenges in the development of an automated sleep monitoring method. The objective of this literature review is to a) summarize the effects of PICU factors on the sleep EEG that might interfere with the sleep staging process, and b) discuss the methods used to automatically monitor depth of sleep based on EEG signals in critically ill adults and neonates. This review could aid in the development of an automated sleep monitoring method in the PICU by summarizing the potentials and challenges of sleep EEG assessment. In section 2, background information on the normal sleep EEG is provided. Section 3 discusses the changes in the sleep EEG occurring with the progression of age. The effects of critical illness on the EEG, including underlying illness and medication, are discussed in section 4. In section 5, methods used for automated sleep monitoring in healthy adults, critically ill adults and neonates are discussed. The review ends with a discussion and conclusion in section 6 and 7.

Table 1. EEG frequencies

Frequency band	Frequency range
Gamma	30 – 48 Hz
Beta	13 – 30 Hz
Alpha	8 – 13 Hz
Theta	4 – 8 Hz
Delta	0.5 – 4 Hz

2. Normal sleep EEG

The EEG measures the electrical activity of the brain with multiple electrodes placed along the scalp. The measured voltage fluctuations over time are a result of the ionic current within the neurons of the brain. Oscillations in the EEG represent synchronized activity over a network of neurons. Sleep is characterized by slowing of the neuronal firing patterns, which could be observed as the low-frequency waves in the EEG during sleep[22]. Together with EMG on the chin, which measures the submental muscle tone by measuring the electric potential of the muscle cells, and EOG, which detects the eye movements by measuring the retinal potential, EEG is used to distinguish the four different sleep stages by their spectral composition and specific features. This staging is traditionally conducted by visual analysis of the EEG, EMG and EOG signals on a 30 second epoch basis according to the AASM criteria[13]. Sleep stages were originally defined by Rechtschaffen and Kales (R&K) in 1968 based on polysomnographic (EEG, EOG, EMG) riteria[23]. These criteria were used from 1968 to 2007 until the AASM updated the criteria and introduced the current international guidelines for the assessment of sleep. The definition of the various

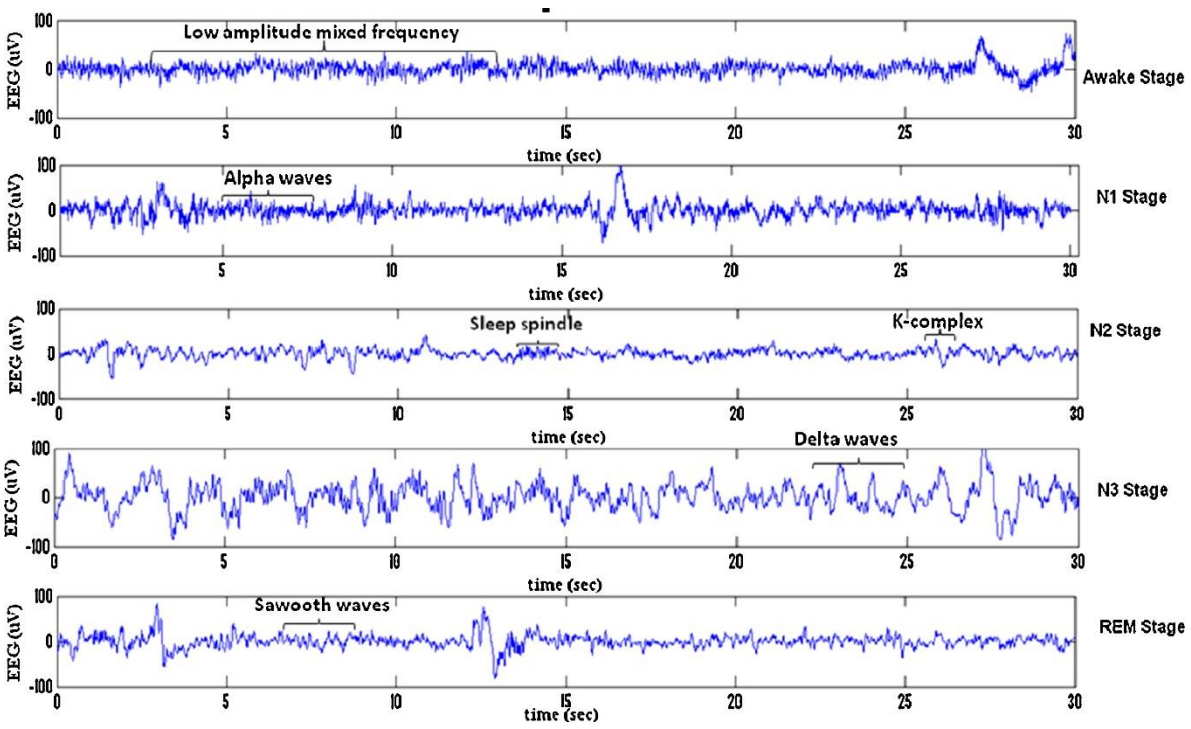


Figure 1. EEG for different stages of sleep: awake, N1, N2, N3 and REM sleep. The background frequencies (mixed frequency, alpha and delta), K-complexes, sleep spindles and sawtooth waves are shown in the figure. Retrieved from: Fraiwan et al. (2012)[135]

sleep stages is originally based on the EEG observations rather than physiological observations. In the EEG, the background activity and transients are important characteristics for each stage. To describe the background activity, the following EEG frequency bands are defined: gamma activity as 30 to 48 Hz, beta activity is 13 to 30 Hz, alpha is 8 and 13 Hz, theta is 4 and 8 Hz, and delta is the slowest activity, 0.5 of 4 Hz (Table 1). The EEG frequencies are thought to represent oscillatory communications between systemic interconnections of neurons[24]. An important rhythm in the assessment of the EEG is the dominant posterior rhythm (DPR). In healthy adults, this rhythm has a frequency in the alpha range (8.5 – 12 Hz) and is seen in the posterior head regions during relaxed wakefulness with the eyes closed. The DPR typically attenuates with opening of the eyes as a result of the activation of the visual cortex. Another characteristic to describe the EEG are transients, i.e., short-lasting EEG features, like typical waves or burst patterns. The muscle tone is assessed by the amplitude of the chin EMG. In the EOG, the speed of eye movements and eye blinks can be observed.

The wake stage is characterized by high frequency EEG activity (> 50% of the epoch is alpha activity) and high muscle tone that decreases when the eyes are closed[13]. After falling asleep, stage N1 is typically first entered. N1 is the transitional stage and lightest stage of sleep with a very short duration in the sleep cycle. N1 is characterized by low voltage, fast EEG activity. Stage N1 sleep is scored when more than 15 seconds (≥50%) of the epoch is made up of theta activity, sometimes intermixed with low-amplitude beta activity. Toward the end of stage N1, vertex sharp waves may occur. Vertex sharp waves are sharply contoured, negative-going bursts that stand out from the background activity. In N1, muscle tone decreases and only slow eye movements are present. In stage N2, the EEG activity slows down and amplitudes are increased. Muscle tone further decreases and eye movements disappear. N2 is marked by predominant theta activity with bursts of faster activity. In this stage, K-complexes and sleep spindles occur for the first time. K-complexes are represented by a sharp negative wave followed by a slower positive component. Sleep spindles are phasic bursts of 11 to 16 Hz activity of short duration. N1 and N2 are together referred to as "non-slow wave sleep"

(NSWS). Stage N3 is the deepest stage of sleep with a high arousal threshold. Characterized by its high amplitude and low frequency EEG waves, this stage is referred to as "slow wave sleep" (SWS). K-complexes and sleep spindles may be seen in N3. If >20% of the epoch is delta activity, the epoch is scored as N3. The time spent in N3 decreases with age. REM sleep is a paradoxical sleep stage in which the EEG resembles wakefulness and physiological activity is high, while muscle tone is very low. Dreaming most often occurs during this stage. The EEG of REM sleep is characterized by relatively low-amplitude, mixed frequency EEG theta waves, intermixed with some alpha waves. In this stage, sawtooth waves can be present. These are 2-6 Hz, sharply contoured triangular EEG patterns that occur in series for a few seconds. As the name suggests, REM sleep is marked by its rapid eye movements[13].

3. Sleep EEG in children

Age is a crucial factor to take into account when evaluating sleep. Where neonates typically spend 16 to 18 hours per day sleeping, sleep time in adults is decreased to one-third of the day. Also, significant changes in sleep architecture occur with age. These changes are most significant in the first months of life and correspond well with critical periods of brain maturation[25]. For example, the amount of REM sleep decreases considerably. A term infant spends 50% of sleep time in REM sleep, falling to 30% by 6 months of age and to 20-25% by 5 year of age, equal to adult levels[26]. Beside these developmental changes in sleep architecture, the sleep EEG also varies considerably as children mature. Remarkable development in continuity of EEG activity, dominant frequency, presence of typical patterns and waveforms and differentiation in sleep stages are observed in the sleep EEG. The EEG sleep patterns mature as the child matures. A delay in the development of the sleep EEG is associated with a global developmental delay[27]. As the sleep EEG has reached its adult-like version after 6 months, the DPR keeps increasing in frequency with age until adolescence[28]. Due to these changes, the AASM provided separate criteria for scoring sleep/wake states for infants aged 0 to 2 months and children aged between 2 months and 18 years[13]. In this section, the most significant developmental changes in the sleep EEG are described for different age categories. Since the PICU admits critically ill infants and children aged 0 – 18 years with the exception of preterm neonates, developmental changes in the sleep EEG are discussed for term neonates to 18-year adolescents. Term neonates are defined as neonates born > 37 weeks gestational age (GA). The GA is defined as the time elapsed between the first day of the mother's last menstrual period and the day of birth.

As the sleep EEG develops most significantly in the neonatal period, understanding of the exact age of the infant is essential in this period to correctly interpret the sleep EEG. It is important to note that the sleep EEG of the neonate does not reflect the postnatal age of the brain (i.e. the number of days following birth), but rather the postmenstrual age (PMA), which refers to the time from the last menstrual period to the date of assessment[29]. Under low-risk conditions, the brain, and thereby the EEG, develops independent of whether the infant is in utero or post-delivery. This implies that the EEG of a 3-week-old term neonate born at gestational age (GA) 40 weeks, is comparable to that of a 11-week-old low-risk premature infant born at GA 32 weeks. Up to the postnatal age of 3 months, the age expressed in PMA best reflects the EEG[30]. In preterm neonates born prior to 32 weeks GA, sleep and wake could not be identified electrographically[31]. The EEG is asynchronous between both hemispheres and discontinuous with periods of suppression alternated by intermittent bursts of activity characterized by sharply contoured EEG waves. With progression of the GA, the EEG becomes more continuous and the duration of periods of suppression and bursts diminishes[32].

0 – 2 months

In the first weeks, sleep onset is typically REM sleep, which changes to N1 sleep onset after 3 months. A neonate typically has a sleep-wake cycle of 2 to 4 hours, independent from day or night. This sleeping pattern gradually changes during the weeks following birth. Sleep continuity increases and sleep becomes more dominant during the night and wakefulness during the daytime. By the age of 2 months, this circadian rhythm is established[25]. In

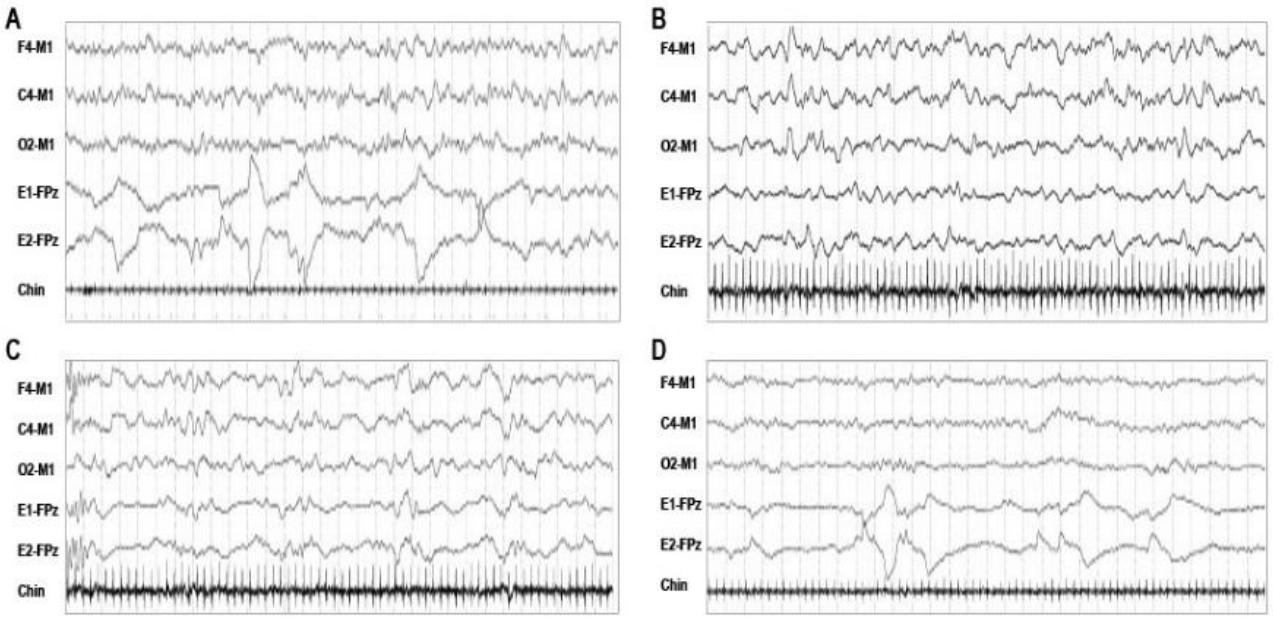


Figure 2. EEG patterns seen in infants 0 – 2 months. A) Active sleep with a mixed pattern and low EMG chin tone, B) Quiet sleep with high voltage slow (HVS) pattern and higher chin EMG tone, C) Quiet sleep with tracé alternant (TA) and higher chin EMG tone, D) Active sleep with low voltage irregular (LVI) pattern and low chin EMG tone. Retrieved from: Grigg-Damberger et al. (2007)[28].

term infants aged 0 to 2 months, three sleep stages are distinguished: active (or REM) sleep, quiet (or NREM) sleep or transitional sleep. EEG patterns during active sleep are continuous. Similar to adult sleep, rapid eye movement and low muscle tone are seen. Two types of active sleep could be observed in neonates: AS1 and AS2/LVI. AS1 typically occurs following wake and contains intermixed high amplitude delta activity with lower voltage theta activity. This pattern is termed 'mixed' (Figure 2D). AS2/LVI typically occurs following a period of quiet sleep and consists of primarily continuous irregular low voltage mixed frequency with low voltage theta activity intermixed with occasional low voltage delta activity, referred to as "low voltage irregular" (LVI) (Figure 2B). The same EEG patterns could be observed during wakefulness, making them hard to distinguish. During quiet sleep, the EEG is characterized by an alternating background pattern with high voltage bursts of delta activity alternating with periods of lower voltage theta activity. This pattern is referred to as "tracé alternant" (TA) (Figure 2C). Another typically seen pattern during quiet sleep is continuous high voltage delta activity, referred to as "high voltage slow" (HVS) (Figure 2B). In normal brain maturation, the proportion of HVS increases while the proportion of TA decreases during quiet sleep with progression of age[33]. By 4 weeks of age, TA usually disappears, fully replaced by HVS. Muscle tone during quiet sleep is higher than muscle tone during active sleep. The wake stage is therefore most reliably scored by behavioural observation because many of the distinctive EEG features of the wake stage are not seen at this age. Sleep spindles may be seen as early as 6 weeks age. If present, the epoch is scored as quiet sleep (and not N2 at this age). In contrary to adult sleep spindles, the waveforms may occur asynchronously over the brain. Sleep spindles could therefore only be seen in some of the EEG leads, typically over in the central brain region. With progression of age, the spindles become more synchronized. Vertex waves and K-complexes are not present yet. An epoch is scored as transitional sleep if both active and quiet sleep characteristics are present. As the name suggest, this stage most often occurs in transitions from wake to active or from active of quiet sleep. The addition of this stage in infants of this age was needed because it is commonly present in infants and is seen as a marker of development and maturation[30]. Neonatal transients that are often seen in drowsiness and all sleep stages are short-lasting runs of sharply contoured triangular theta or delta waves over the frontal regions. These so-called "encoches frontales" typically disappear after 2 months[28].

2 - 12 months

From the age of 2 months, the sleep EEG increasingly reaches its adult differentiation as part of the transition from neonatal to infantile sleep. The neonatal transients (encoches frontales) disappear and the neonatal EEG background patterns (LVI, TA, HVS) are gradually replaced by more rhythmical theta waves. At 3 to 4 months, the DPR is first seen. In contrast to the alpha frequency range in adults, this DPR is in the 3-4 frequency range at 3 months, increasing to 5 Hz at 5 months and 6-7 Hz at 12 months[28]. From the age of 6 – 8 months on, high voltage bursts of 4 Hz, increasing to 5-6 Hz over the following months, are seen during drowsiness, in the transition from wake to sleep. This pattern is known as "hypnagogic hypersynchrony"[28]. With the progression of age, the sleep stage differentiation becomes clearer by the appearance of the adult sleep EEG transients. K-complexes and rare broad vertex waves are first seen 5-6 months post term and gradually acquire adult appearance with progression of age. The presence of these transients enables the differentiation of NREM into N1, N2 and N3. At 5 months age, the sawtooth waves, characteristic for REM sleep, appear for the first time[28].

1 - 18 years

In early childhood, the DPR increases with age, from 6-7 Hz in 2-year-olds to 7-9 Hz in 3-year-olds. Hypnagogic hypersynchrony progressively diminishes and vertex waves become very prominent. The sleep spindles become more synchronous between both hemispheres. By the age of 1 year, 70% of the sleep spindles is synchronous, 100% by the age of 2 years[34]. The REM sleep is characterized by medium-voltage theta waves. During SWS, diffuse waves can be observed. In older children, the DPR keeps gradually increasing in amplitude and frequency to those in adults. The REM sleep EEG now shows low-voltage theta activity. During SWS, the delta activity increases in amplitude. The hypnagogic hypersynchrony disappears and is rarely seen in children after age 6 years. Vertex waves are usually of high amplitude and occur in bursts. K-complexes are often seen with spindle activity. During adolescence, the sleep EEG reaches its adult differentiation. The DPR is in the alpha range, equal to adults, with a lower amplitude compared to younger children[28].

4. Sleep EEG of critically ill patients

The sleep EEG of critically ill patients is often confounded by derangements induced by the underlying critical illness and by the central effect of various medications that are typically used for sedation or analgesia. Furthermore, the EEG is prone to artifacts arising from the noisy ICU environment and 50-Hz electrical inference with electrical equipment simultaneously used in the ICU. These effects on the EEG signal challenge the application of the AASM criteria to score sleep in critically ill patients admitted to the ICU and introduce high interobserver variability in scoring sleep of critically ill patients[18]. This highlights the need to identify the atypical EEG findings in critically ill patients.

Unfortunately, studies characterizing the sleep EEG in critically ill children in the PICU are scarce. One study of Kudchadkar et al. was found, in which they studied the sleep EEG power spectrum during sleep in mechanically ventilated critically ill children who were admitted to the PICU with respiratory failure[35]. At the time of monitoring, each patient received a continuous infusion of opioids and benzodiazepines. When comparing to healthy age- and gender-matched children, they found that the average power over the night in all frequency bands (beta, alpha, theta and delta) was lower in the critically ill children. Furthermore, unlike the healthy children, PICU patients did not demonstrate expected temporal variability in delta and theta power during the night, in which the power declines gradually over the night. Also, no day-night organization was observed in the PICU patients.

More is known about sleep in critically ill adults. In ICU studies, both a disrupted sleep architecture characterized by increased N1 and N2 sleep and decreased N3 and REM sleep as well as atypical sleep EEG findings were observed[14,15,36]. Drouot et al. reported that 28% of the non-sedated ICU patients showed atypical sleep EEG findings[16],

compared to 85% in sedated ICU patients, as reported by Watson et al.[14]. Therefore, the addition of new sleep stages as various subcategories of atypical sleep, for scoring sleep in critically ill adults was proposed by these authors. Atypical sleep was characterized by high amplitude, polymorphic delta activity without superimposed fast frequencies and rapid eye movements with a low chin muscle tone. Concordant to this, Ambrogio et al. showed that the delta/alpha and delta/beta power ratio is significantly higher in a 24-hour period in sedated and ventilated critically ill adult patients compared to healthy controls[18]. Beside the polymorphic and increased delta activity, other atypical EEG findings, reported by Watson et al., are presence of burst suppression or isoelectric EEG[14]. Burst suppression is an EEG pattern characterized by periods of high voltage activity alternating by periods of suppression (activity $<$ 10 mV). Typically, the episodes of suppression are longer (usually 5-10 seconds) than the periods of high voltage activity (usually 1-3 seconds). Isoelectric EEG could be observed as a flat line EEG. Both patterns are typically found in patients with inactivated brain states, such as in coma or deep sedation[37]. Furthermore, K-complexes and sleep spindles were observed to be absent in critically ill adults. The absence of these N2 markers was present in 20-44% of the ICU patients[14–16,18], challenging the scoring of N2. In line with this, Ambrogio et al. reported a poor interobserver reliability for scoring N1 and N2 ($\kappa = 0.19$)[18]. Also, dissociation of EEG findings and behaviour was observed in critically ill adults. Drouot et al. observed that relative powers of delta and theta were significantly higher in some non-sedated patients than one would expect during normal wakefulness[16]. On the opposite, alpha and beta relative powers in these patients were significantly lower. This observation was referred to as "pathological wakefulness". Pathological wakefulness is defined as EEG epochs in which behavioural characteristics of wakefulness (opening eyes to verbal stimuli, making eye contact, following simple commands) occur with EEG features of N3 (high-amplitude, low-frequency waves), which are not seen in normal wakefulness. Remarkably, the same shift in EEG frequency is observed in ICU patients with delirium[38,39], implicating that pathological wakefulness is a potential marker for subclinical or hypoactive delirium. Concordant to this, Drouot et al. found a significantly higher occurrence of delirium in the days following the EEG recording in patients experiencing atypical sleep with pathological wakefulness than in patients with usual sleep[16]. Conversely to the EEG dissociation in pathological wakefulness, unresponsive comatose patients were noted to have alpha or beta activity[14].

These studies show that atypical EEG findings like polymorphic delta activity, absence of sleep spindles and K-complexes, burst suppression, isoelectric activity and EEG dissociation are often encountered in critically ill patients. However, these studies do not take into account the independent effect of the patient-related comorbidities and risk factors for atypical EEG findings, such as sedatives and analgesics. Atypical sleep is more often observed in sedated patients compared to non-sedated patients[14,16]. Furthermore, as mentioned before, the shift in EEG frequency observed in pathological wakefulness is also observed in ICU patients with delirium[38,39]. Cooper et al. mentioned that patients identified as experiencing atypical sleep had a lower Glasgow coma scale (GCS), indicating a lower consciousness level. This is concordant with the burst suppression and isoelectric activity that is typically seen in patients with an inactivated brain state[37]. Thus, the EEG abnormalities observed in these studies could therefore be drug-induced or an effect of the underlying illness. Interestingly, the changes in background EEG activity during wakefulness reported in atypical sleep is similar to those reported in healthy individuals subjected to sleep deprivation for 24 hours[40,41]. Likewise, a decreased density of sleep spindles and K-complexes was observed during recovery sleep after sleep deprivation[42,43]. Since ICU patients are exposed to sleep deprivation with sleep fragmentation and decreased REM and SWS sleep, this suggests that the observed atypical sleep EEG could also be a result of the experienced sleep deprivation. Identification of the cause of atypical EEG characteristics remains challenging in the heterogeneous population of the critically-ill. In the following sections, evidenced influences on the sleep EEG of frequently used medications and commonly encountered morbidities at the PICU are identified.

4.1. Influence of medication on the sleep EEG

Numerous types of medication are routinely administered at the PICU for therapeutic purposes or to provide comfort to the patient. Some of them interact with the brain and could thereby alter EEG characteristics or disrupt sleep architecture. Analgesics and sedatives relieve pain and reduce the level of consciousness, respectively, by interacting with the central nervous system. The effects on the EEG of commonly used sedatives, analgesics and some other drugs in the PICU are identified in this section. An overview of the EEG effects of the discussed agents is provided in Table 2.

Sedatives

Sedatives might promote quality of sleep in critically ill patients by their anxiolytic and calming properties. However, many frequently used sedatives have been shown to disrupt sleep architecture, leading to sleep deprivation. The effects of this disruption could be observed after withdrawal of sedative medication. Similar to in natural sleep deprivation, the characteristics of recovery sleep after withdrawal of medication are dependent on the characteristics of the lost sleep. If, for example, a patient is withdrawn of a REM-suppressive sedative, its recovery sleep is characterized by a disproportionally high percentage of REM sleep[44]. Thus, sedation in critically ill patients is paradoxical, because it is both a potential treatment and cause for sleep disruption. When a patient is sedated, they look comfortable and even asleep. Although sedatives intersect with the same neural pathways involved in the process of sleep, the sedative state is not thought to have the same restorative effect that natural sleep has[45]. Whereas natural sleep is a biological process with a cyclic behaviour, influenced by circadian rhythmicity and reversed by external stimuli, sedative sleep is rather continuous and dependent on the type of sedative agent and dose administered. Also, the EEG findings that define the sleep stages in natural sleep are atypical in sedative sleep.

The depth of sedation has been correlated with the number of burst suppressions seen on the EEG[46]. This pattern is typically seen in inactivated brain states, in this case induced by deep sedation. The presence of burst suppression patterns indicates that the cerebral metabolic rate is reduced[47] and could be used to titrate sedation individually. With deepening sedation, the number of burst suppression patterns increases until eventually the EEG becomes completely isoelectric[46]. Commonly used sedative medication at the PICU include benzodiazepines, propofol, ketamine, and clonidine. Where most of them could be leading to burst suppression patterns and isoelectric EEG in high dosages, their effects on the EEG in lower dosages are variable.

Many sedative agents work by activation of the gamma-aminobutyric acid type A ($GABA_A$) receptor. The same receptor is involved in the initiation and maintenance of physiological NREM sleep and the generation of sleep spindles. One of the most commonly used sedative agents in the PICU are the GABA-agonistic benzodiazepines (e.g. midazolam, lorazepam, diazepam). Benzodiazepines are shown to increase total sleep time and N2, while decreasing REM and N3 in adults[48]. Veselis et al. showed that in a critically ill patient population the EEG power spectrum varied consistently with the depth of sedation induced by the benzodiazepine midazolam with a significant decrease in median frequency, spectral edge (i.e., the frequency below which 95% of the spectral power is located) and log absolute power in the beta bandwidth was found in the EEG of deeper sedated patients[49]. Jennekens et al. found contrasting EEG changes in full-term neonates with ischemic stroke after administration of midazolam[50]. In contrast to to EEG changes in adults, an increase in theta power was observed while alpha and beta powers remained constant. A decrease in total power and delta power was observed, in accordance with findings in adult studies. This discrepancy is hypothesized to be the result of immaturity of the neonatal brain, influencing neuronal signal transmission. Propofol is also believed to bind to the GABA-receptor at a site different from the benzodiazepine binding site, thereby activating the receptor[51]. Similar to benzodiazepines, propofol suppresses N3 and REM sleep[52]. Propofol affects the EEG by diffuse slowing of the EEG frequency and depression of amplitudes[53]. Blood concentration of propofol is negatively correlated with EEG frequency and amplitude. Also in children, increases in total EEG power and slow waves during sedation with propofol

could be observed from 0 to 6 years old[54]. A next GABA-agonist is the short-acting sedative etomidate. In low dosage, used during induction of anaesthesia, epileptic activity on the EEG causing involuntary myoclonic movements is common[55]. This epileptic activity disappears in higher dosages at which burst suppression could be observed. Barbiturates also work on the GABA-receptor. They typically reduce REM sleep[56]. In dosages used for general anaesthesia, the EEG slows down and irregular theta and delta activity is present[57]. The effect on the EEG for dosages used for anticonvulsive purposes is discussed later in this section.

Another class of sedative agents are the α-2-receptor agonists, that bind to the α-2-receptor in the locus coeruleus in the brainstem to decrease activity of noradrenergic innervation and include clonidine and dexmedetomidine[58]. Sedation under α-2-receptor agonists resembles natural sleep more closely than sedation under GABA agonists, clinically observed by easy arousals by external stimuli and better cognitive functioning when aroused[59]. This observation could be explained by the fact that the α-2-receptor agonists interact with the natural sleep pathway at a site farther upstream than the GABA agonists do. However, sleep architecture is still disturbed as marked by a decrease in REM sleep and increase in N3 and N2 sleep[60]. The increase in N2 sleep under α-2-receptor agonists sedation is partly based on an increase in sleep spindles. These sleep spindles resemble sleep spindles present during natural sleep in density, amplitude and frequency content, although a significant increase in spindle duration was found in spindles during dexmedetodine (1.11 ± 0.14 seconds) compared to spindles in normal sleep (0.88 ± 0.14 seconds; $p<0.01$)[61]. Another effect of α-2-receptor agonists on the EEG is a dose-dependent slowing. Mason et al. studied the effects of dexmedetodine sedation on the EEG in children with epilepsy[62]. They observed that the EEG during sedation resembled N2 sleep, although modest significant increases in theta, alpha and beta power were found, suggesting that dexmedetodine does not hinder EEG interpretation for sleep assessment. Similar to during natural sleep, a significant increase in spike activity during dexmedetodine sedation was observed (15 vs. 22 per minute; $p=0.01$).

Table 2. Overview of the EEG effects of medication.

Drug class or individual drug	EEG effect
Sedatives	
All	Burst suppression, isoelectric
Benzodiazepines	*Low dosage (anticonvulsive):* ↑ Beta *High dosage (sedative):* Slowing
Propofol	Slowing, amplitude depression
Etomidate	*Low dosage (induction):* Epileptic activity *High dosage (sedation):* Burst suppression
Barbiturates	*Low dosage (anticonvulsive):* ↑ Beta *High dosage (sedative):* EEG slowing
α-2-receptor agonists	↑ Sleep spindles, slowing
Ketamine	↑ Beta + theta
Analgesics	
Opioids	Slowing
Acetylsalicylic acid and azapropazone	Slowing
Other	
Antipsychotics	Slowing
Muscle relaxants	Absence of muscle artifacts
Cocaine, amphetamines, methylphenidate'	↑ Beta (low voltage)

Ketamine is a fast-acting sedative agent that also has analgesic effects. Its action relies on blockage of the N-methyl-D-aspartate (NMDA) receptor, although its exact mechanism appears to be more complex[63]. Ketamine is known to increase total sleep time with increases of N3 and REM sleep[64]. In the EEG, effects of ketamine administration could be observed as increased beta activity and dominant high amplitude theta activity[65].

Analgesics

Pain is a common cause of sleep disruption in critically ill patients[66]. Poor sleep may also increase

a patient's perception of pain[67]. Analgesics are therefore often used in the PICU to provide adequate pain relief. However, similar to sedatives, many analgesics are known to disturb sleep architecture[56]. Opioids are often used with sedatives in mechanically ventilated PICU patients and include morphine, remifentanil and methadone. They exert their analgesic effect by binding to the opioid receptor principally found in the central and peripheral nervous system and gastrointestinal tract, causing constipation as one of the main adverse effects. Opioids have a dose-dependent suppressive effect on REM-sleep and N3 sleep and increase N2 sleep[68]. On the EEG, they cause a dose-dependent slowing of the EEG with increased delta activity[69]. In high doses, they decrease not only the median frequency, but also the spectral power[70]. A systematic review performed by Malver et al. did not identify studies evaluating the effect of paracetamol on the EEG[71]. In the non-steroidal anti-inflammatory drugs (NSAIDs) group, no studies were identified evaluating the effect of diclofenac, ibuprofen, ketoprofen and phenazone. However, the NSAIDs acetylsalicylic acid and azapropazone were found to cause slowing of the EEG[72,73].

Other

Beside the discussed sedatives and analgesics, numerous other medications routinely administered to critically ill children affect sleep architecture or sleep EEG. Inotropes, vasopressors, steroids, bronchodilators and antiarrhythmic agents are all known to disrupt sleep by interaction with the sympathetic nervous system or hormone secretion[19]. Many of these medications typically do not greatly affect the sleep EEG, although some central nervous system stimulants such as cocaine, amphetamines and methylphenidate could increase beta activity at low voltage[57]. Muscle relaxants, sometimes used in mechanically ventilated patients, do not have an effect on the brain waves. However, their effect on the EEG could be observed as the absence of muscle artifacts in the EEG, Thus, from the EEG quality perspective, they have a beneficial effect. Antipsychotic drugs, such as haloperidol and risperidone are used at the PICU in delirious patients. A study by Armour et al. in paediatric burn patients showed that haloperidol increased total sleep time and N2 sleep compared with control nights[74]. Antipsychotics are known to slow down the EEG with a decrease in beta activity and increase in theta and delta activity[75]. The alpha activity increases in amplitude while decreasing somewhat in frequency. In high dosages, they could induce epileptic activity. Antiepileptic, or anticonvulsive, drugs influence the EEG significantly, beside their suppression of epileptic activity. Carbamazepine, valproate and phenytoin cause slowing of the EEG frequency[76]. In low dosage, some sedative agents, such as benzodiazepines or barbiturates have anticonvulsive properties. Contrary to its effect in dosages used for sedative purposes, benzodiazepines will induce a decrease in alpha activity and a diffuse increase in beta activity[57]. The barbiturate phenobarbital affects the EEG by increased beta activity correlated to the plasma concentration, contrary to its slowing effect seen in higher dosages used for sedative purposes[76].

4.2. Influence of medical conditions on the sleep EEG

Critical illness in children admitted to the PICU includes a broad range of medical and surgical diagnoses, such as trauma, multi-organ failure, or postoperative care requiring close monitoring and life support. Morbidities that affect the brain, directly or indirectly, could also influence the EEG. In this section, some medical conditions that are frequently encountered in the PICU with known effects on the EEG are discussed. An overview of the EEG effects of the discussed medical conditions is provided in Table 3.

Coma

Coma is the state of prolonged unconsciousness from which a person cannot be awakened and fails to respond to external stimuli. Also, there is a lack of a normal sleep-wake cycle with a complete absence of wakefulness. Coma can be caused by a variety of problems, either by natural causes or medically induced. The EEG in medically-induced comas is discussed in section 4.1. Where the comatose patients may show the same clinical features, the EEG can reveal diagnostic and prognostic features. EEG features observed in comatose patients include slowing of the EEG, burst suppression, epileptiform activity, triphasic waves, amplitude suppression, or even isoelectric activity[77].

Triphasic waves are high-amplitude, positive sharp transients that is preceded and followed by negative waves of relatively lower amplitude. Isoelectric activity is used to determine brain death. A patient is declared brain dead if electrical cerebral activity is absent in a 30-minute registration[78]. A normal EEG in a comatose-looking patient suggests a locked-in syndrome, or pseudo-coma, caused by a brainstem lesion[77].

Some specific types of coma may be discerned based on EEG findings. In some comatose patients, alpha or theta activity could be observed. These tracings contain frequencies found in normal wakefulness and depart from typical EEG findings in comatose patients of slowing and suppression. These so-called "alpha-coma" and "theta-coma" patterns, or if they coexist "alpha-theta-coma" differ from EEG activity during wakefulness in its non-reactivity to eye opening and more diffuse presence[77]. The prognostic value of these patterns depends on the timing and aetiology[79]. The presence of alpha coma patterns, for example, carry an extremely poor prognosis in hypoxic encephalopathies, while prognosis is quite good in toxic encephalopathies[80]. In spindle coma, the EEG of the comatose patient contains sleep-like activity with sleep spindles, vertex waves and K-complexes. Spindle coma implies the functional preservation of the cerebral hemispheres which can be associated with a good prognosis[81]. Alpha, theta, alpha-theta and spindle coma have also been observed in children[82]. The youngest reported patient with alpha coma is a 2-month old infant with phenobarbital intoxication after a period of hypoxia[83]. The alpha activity in the infant, at whose age alpha activity is not even present during normal wakefulness, was similar to the alpha coma pattern observed in adult patients.

The presence and prognostic value of EEG features in coma depend on the cause of the coma. For example, triphasic waves are typically present in metabolic encephalopathies, while they will rarely be seen in hypoxic encephalopathy[84]. EEG findings in various causes of encephalopathies, potentially leading to coma, will be pointed out in the following sections.

Epilepsy and status epilepticus

The EEG is routinely used in the evaluation of patients with seizures or at risk of developing seizures, such as in hypoxic-ischemic encephalopathy or patients with status epilepticus. Epileptic activity on the EEG could be recognized as a pattern that represents a clear change from the background frequencies, containing epileptiform transients that occur isolated or in bursts. These epileptiform transients include spikes and sharp waves, that stand out to the background EEG by their wave morphology and high amplitude. Spikes and sharp waves differ in their duration; spikes usually have a duration of 20-70 milliseconds, while sharp waves last 70-200 milliseconds. In contrast to spikes, sharp waves often have a multiphasic behaviour. Most spikes and sharp waves have a negative polarity and are often followed by slow waves, then referred to as "spike-and-slow-wave complex". The postictal (i.e., after an epileptic seizure) EEG often shows slowing with or without attenuation of the background activity, usually lasting for several minutes[85]. A large inter- and intrasubject variability in the presence, morphology and clinical value of the epileptic EEG activity exists that is beyond the scope of this review.

In many epileptic syndromes, NREM sleep activates epileptic discharges, increasing with depth of sleep NREM sleep[86], while epileptic activity is relatively suppressed during REM sleep[87]. However, epileptic EEG activity could also be present in non-epileptic patients during sleep. A study in nine hundred otherwise healthy children, found a prevalence of epileptiform activity during sleep in 1.45%, most prominent during NREM sleep[88]. The epileptic activity might complicate waveform interpretation during visual sleep analysis. Marzec et al. studied the features in EEGs of patient with epilepsy interfering with sleep staging[89]. 48% of the EEG recordings of 43 epileptic patients contained epileptic features to the extent that sleep scoring was interfered. In some cases, the epileptic activity obliterated the sleep waveforms. Besides, spikes and sharp waves may resemble vertex waves or K-complexes, confounding the scoring of N1 and N2. The postictal slowing might resemble the delta activity during N3. Also, the general slowing of the EEG background due to pathologic brain activity

and the effects of anticonvulsive drugs (see section 4.1) could interfere with sleep staging. In some patients, abnormally low frequency background activity was observed during visually observed wakefulness, that might be defined as N1 sleep following the sleep staging criteria.

Central nervous system infections

Central nervous system infections are a frequent cause of paediatric encephalopathy. They generally result from blood-born spread of bacteria, fungi and viruses. Based on the structure involved, central nervous system infections are broadly divided into three categories: meningitis, encephalitis and abscesses. Although their effect on the EEG could vary with the pathogen involved and the location of the lesion, common EEG findings are discussed here. In meningitis, the EEG could be normal or show diffuse, irregular delta activity when the infection is limited to the meninges[84]. Remarkable is the rapid and parallel decrease in EEG abnormalities with clinical improvement[90]. If the infection penetrates the brain parenchyma (encephalitis, meningo-encephalitis), the EEG is always and more strongly abnormal. Encephalitis or meningo-encephalitis also produces delta activity as the most dominant EEG abnormality, but epileptiform activity may also occur[84]. The same EEG abnormalities were observed in children and neonates[91,92]. On the EEG of patient with a brain abscess, focal arrhythmic delta waves and epileptic activity could be observed in the EEG leads covering the abscess location[93]. These features could be used to localize the abscess. In the EEG leads of the affected hemisphere, suppression and slowing of the background activity could be observed.

Traumatic brain injury

In traumatic brain injury (TBI), the effects on the EEG depend on the exact mechanism and severity of the injury. In general, the EEG in TBI shows slowing of the background activity, decreasing with the severity of the injury[94]. Sleep architecture is commonly disturbed in patients with TBI, characterized by increased N3 sleep and disturbed order of the sleep cycle[95]. Cerebral contusion can be associated with multiple microhaemorrhages, producing focal theta or delta activity with high amplitude[94]. Severe TBI can cause epi- or subdural haemorrhages. In the acute phase, these haemorrhages could cause increased generalized slowing of the background EEG or focal abnormalities with high amplitude polymorphic delta activity[96]. Hematoma, contusion or oedema are mass lesions that could increase the intracranial pressure (ICP). Increased ICP causes slowing of the EEG as a result of reduced cerebral blood flow and has been negatively correlated with the median frequency and power in delta activity[97]. Changes in ICP dynamics are linked to burst activity in the EEG[98]. Interestingly, ICP has been shown to increase during REM, N1 and N2 sleep in TBI patients, thereby also affecting the EEG by introducing slow waves[99]. In skull defects, either traumatic or postoperative, locally increased alpha and beta activity with high amplitude could be observed, referred to as "activité de brêche" or "breach rhythm"[100]. This activity is caused by changes in the conduction of electrical potentials from the brain tissue to the EEG electrode by the absence of the skull that normally acts as a high frequency and amplitude filter.

Hypoxic-ischemic encephalopathy

In hypoxic-ischemic encephalopathy, brain damage is caused by oxygen deprivation and limited blood flow. As reviewed by Bauer et al., EEG findings in hypoxic-ischemic encephalopathy are epileptic activity, background suppression, burst suppression patterns, increased theta and delta activity with preserved alpha rhythm and alpha or alpha-theta coma[101]. During resuscitation, the EEG could be flat, which will persist several hours after circulation is restored. In neonates, the EEG rhythm will take longer to recover than in older patients[102]. Similar to observations in older children, the EEG in neonatal hypoxic-ischemic encephalopathy is characterized by epileptic activity, background suppression and/or burst suppression[103].

Metabolic encephalopathy

Metabolic encephalopathy is caused by a broad range of aetiologies that impair cerebral metabolism, either by shortage of nutrients, hormonal or electrolyte imbalances or the presence of toxic agents. In many metabolic encephalopathies, the EEG is characterized by the presence of triphasic waves[84].. However, triphasic waves are infrequently

observed in children[104]. As reviewed by Kaplan et al.[105], electrolyte imbalances could produce diffuse slowing of the EEG and in more severe cases epileptic discharges. Diffuse slowing and epileptic discharges are also observed in hypoglycaemia, although increased delta activity could also be asymmetric. EEG changes in hyperglycaemia are less prominent. The hormonal imbalances in hyperadrenalism and hyperthyroidism cause increases in alpha frequency with prominent beta activity or slow background activity with superimposed fast frequency activity. On the other hand, hypothyroidism and hypoadrenalism both cause slowing of the EEG background activity and decreased reactivity. Hepatic encephalopathy results from the inability of the liver to remove toxins from the blood in severe liver disease. In adults, the EEG has been shown to demonstrate hepatic encephalopathy before clinical presentation[106]. Most common EEG findings in children with acute liver failure are diffuse slowing and epileptiform discharges[104]. Whereas triphasic waves are frequently seen in adults suffering from liver failure, they are rarely present in children. Similar EEG findings are observed in uremic encephalopathy in renal failure patients. Slowing of the EEG and presence of triphasic waves has been correlated to blood urea nitrogen levels[107]. Also, epileptiform discharges or bursts of theta activity could be observed. In patients with respiratory failure, hypercapnia could also induce various abnormalities in brain function leading to metabolic encephalopathy. In healthy awake adults, hypercapnia has been shown to produce slowing of the EEG[108].

Cerebrovascular accidents

Ischemic cerebral vascular accidents (CVA) affect the EEG in a similar fashion as hypoxic-ischemic encephalopathy. The EEG background activity is slowed down, either bilaterally or unilaterally in the affected hemisphere[84]. In the region that is supplied by the affected cerebral artery, polymorphic theta and delta activity could be observed, often mixed with sharp waves. Similarly, haemorrhagic CVA causes localized polymorphic delta activity in the affected brain region or diffuse slowing of the background EEG. A study by Hirose et al. showed that intracerebral hematoma only shows diffuse

Table 3. Overview of the EEG effects of medical conditions

Medical conditions	EEG effect
Coma (general)	Slowing, burst suppression, epileptic activity, triphasic waves, amplitude suppression, isoelectric activity, alpha, theta or spindle patterns
Epilepsy, status epilepticus	Epileptic activity, slowing
Central nervous system infections	*Meningitis*: normal, irregular delta activity *Encephatlis*: slowing, epileptic activity, background suppression *Abscess*: focal polymorphic delta activity
Traumatic brain injury	*General*: Slowing, burst activity *Local (contusion/haemorrhages):* polymorphic delta activity *Local (skull defects):* breach rhythm
Hypoxic-ischemic encephalopathy	Slowing, epileptic discharges, background suppression, burst suppression
Metabolic encephalopathy	Slowing, epileptic discharges, triphasic waves (less frequent in children)
Cerebral vascular accident	*General*: Slowing *Local*: polymorphic delta activity *Neonates*: rolandic sharp waves
Delirium	Slowing

polymorphic delta activity in patients with hematomas larger than 30 mL, that cause a shift of the midline structures[109]. In smaller hematomas, the polymorphic activity is restricted more locally to the affected region. In subarachnoid haemorrhages, increases in delta activity, often mixed with sharp waves, could indicate the presence of vasospasm[110]. Neonatal intraventricular haemorrhage is characterized by the presence of positive rolandic sharp waves, defined as sharp transients or positive polarity appearing in the rolandic regions, around the central sulcus, in the brain[111].

Delirium

Delirium is an organically caused disturbance in attention, cognition and consciousness. While sleep deprivation is regarded to be a risk factor for the

development of delirium, it is also likely that delirium itself contributes to sleep disturbances[11]. A study by Trompeo et al. in adult ICU patients demonstrated an association between delirium and severe REM sleep disruption[112]. However, a cause-and-effect relationship has been hard to establish. The EEG of delirious patients shows slowing of the peak and median frequency, with a reduction of alpha power and an increase of theta and delta power[38,39]. In children with febrile delirium, similar spectral changes were observed, with the most significant change being the increase in relative delta power[113]. Decrease of the relative delta power was associated with clinical improvement. As mentioned in section 4, similar spectral changes were found in ICU patients experiencing pathological wakefulness[16], implicating that this finding is a marker for subclinical or hypoactive delirium. Van der Kooi et al. used the characteristic spectral changes in delirious patients to develop an EEG-based tool for delirium detection[114]. By using the relative delta power in EEG in the frontal-parietal electrode derivation, delirium could be detected with a sensitivity of 100% and specificity of 96%.

4.3. Other factors influencing the sleep EEG

Beside the discussed influences of medication and medical conditions on the EEG, various extracerebral sources could introduce artifacts that interfere with the EEG signal[115]. The electrical field caused by eye movements introduces high amplitude slow wave artifacts in the frontal EEG leads, that could be confused with delta and theta activity. Muscle activity of the facial, neck or shoulder muscles introduces high frequency (15-35 Hz) activity. Furthermore, electrical activity from the heart interferes with the EEG, observed as a rhythmic pattern corresponding with the heart rate. The electrolyte content of sweat drops changes the electrical baseline in the EEG electrodes and thereby introduce a low frequency artifact (usually <0.5 Hz). Patient movement is also reflected in the EEG. Movement due to respiration could be observed as a low frequency artifact in the delta or theta range. Pathological conditions such as scalp oedema or a caput succedaneum could affect the EEG by damping the electrical signal from the brain, causing lower EEG amplitudes[116].

On top of these physiological sources, artifacts could arise from technical aspects anywhere in the recordings system. Electrical interference artifacts at 50 Hz could be introduced by electromagnetic fields from surrounding electrical devices. Changes in electrode impedance, for example caused by touching the electrodes or loose electrodes, introduces abrupt and high amplitude transients. During impedance checks, total absence of EEG activity is observed. The noisy PICU environment introduces additional artifact sources[117]. Electrical interference artifacts are more common due to the numerous electrical devices that are simultaneously used in the PICU. The frequent nursing activities introduce movement artifacts. Interestingly, motion artifacts were also observed in patients connected to a hemoperfusion machine[118]. The motion artifacts, observed as saw-tooth waveforms, were correlated to the rotary pump action of the device.

A last factor that might influence the EEG is gender. A recent study by Markovic et al. studied the sleep EEG in early adolescence (9-14 years old) and found significant differences between males and females[119]. Girls were found to have greater spindle activity and more power in the higher frequency bands (16.2-44 Hz) during all sleep stages.

5. Automated sleep monitoring methods

To our knowledge, a method to automatically monitor depth of sleep in critically ill children using computerized EEG signal processing has not been developed yet. Although Kudchadkar et al. used power spectral analysis to assess spectral changes in the sleep EEG of PICU patients, they did not use the obtained power values in the various frequency bands to automatically classify sleep or compare it to the hypnogram obtained by visual sleep scoring[35]. However, some attempts have been made to develop an automated depth of sleep monitoring method in critically ill adults or neonates. Meanwhile, methods to stage sleep in healthy adults are numerous, varying from simple single feature with threshold-based algorithms to advanced deep learning methods. A complete overview of all methods is

beyond the scope of this review. A selection of the methods will be discussed in this section.

General methods for sleep stage classification

Literature on automated sleep staging methodologies in healthy adults is numerous. Currently, a literature search on EEG-based methods in healthy adults in the PubMed database yields already more than 2000 results, with a rapid increase in publications in the past decade. These algorithms generally include the following series of steps: signal pre-processing, feature extraction and classification. The methods used for each of these steps vary greatly among the various sleep staging methods, with varying degrees of classification accuracy as a result. Although some of these methods will achieve an acceptable accuracy in a subpopulation of PICU patients, many will not due to the different EEG patterns in young children and confounding with effects of medication and underlying illness. However, the knowledge on the accuracy and efficiency of the various signal pre-processing algorithms, feature extraction methods and classification approaches obtained from the studies in healthy adults, could aid in the development of a PICU method. Based on recent reviews by Zhao et al. (2019)[120], Faust et al. (2019)[121] and Chriskos et al. (2021)[122], a short overview of these methods is provided in this section. An overview of various proposed algorithms and their classification results can be found in Table 1 in Faust et al.[121] and Table 6 in Zhao et al.[120].

Signal pre-processing

Signal pre-processing is used to remove unwanted content from the EEG signal. This unwanted content is a result of extracerebral sources that interfere with the EEG signal and lead to spurious results when processing the EEG data. As reviewed by Chriskos et al., most of the works use a signal bandpass filter or a combination of high pass and low pass filters, with lower cut-off frequencies varying from 0.3 to 0.5 Hz and upper cut-off frequencies from 30 to 380 Hz to remove unwanted frequencies[122]. Notch filters are often used to filter the electrical interference artifacts near 50 or 60 Hz. Although frequency filtering removes a major portion of unwanted spectral content, the noise within the frequency range of interest will not be removed. One method that is often used to remove this noise is independent component analysis (ICA). ICA is based on the assumption that artifacts and brain activities in the EEG are generated by independent processes and decomposes the signal into independent components. The components that are likely to arise from noise sources can then be rejected and a clean signal could be reconstructed. For the general noise in EEG signals, wavelet denoising has been shown to be a good way to improve the signal to noise ratio[120]. Wavelet denoising is based on decomposition of the signal into multiple lower resolution levels by controlling scaling and shifting factors of a single wavelet function, and shrink or remove the resolution levels that most likely contain noise. Other methods of artifact reduction in EEG processing include principal component analysis (PCA), adaptive filtering techniques or artifact identification algorithms. The latter one detects artifacts based on artifact characteristics and removes the artifact segment from the signal. It must be noted that many signal pre-processing methods tend to remove a portion of the EEG signal that might contain valuable information. Furthermore, advanced pre-processing methods increase the required computational power. Therefore, many authors choose to not use advanced artifact filters[121].

Feature extraction

After the pre-processing steps, features can be extracted from the EEG signal. The variety of used features in EEG-based sleep staging algorithms is extensive. From the time domain, the following statistical features are often used: mean, median, interquartile ranges, variance, skewness and kurtosis. Also, energy, zero-crossing rate, peak-to-peak amplitude entropy and the Hjorth parameters are frequently extracted from the EEG signal. The Hjorth parameters were introduced in 1970 by Hjorth as indicators of statistical properties of the EEG signal in time domain[123]. Nowadays, they are still commonly used in EEG feature extraction. The parameters are Activity, Mobility and Complexity (equation 1, 2 and 3).

1) $Activity = var\big(y(t)\big)$

2) $Mobility = \sqrt{\frac{var\left(\frac{dy(t)}{dt}\right)}{var(y(t))}}$

3) $Complexity = \frac{Mobility\ (\frac{dy(t)}{dt})}{Mobility(y(t))}$

The Activity parameter is the variance of the signal in time domain and represents the signal power. The Mobility is defined as the square root of the variance of the first derivative of the signal divided by the variance of the signal, thereby representing the mean frequency or the proportion of standard deviation of the power spectrum. From the Mobility parameter, the Complexity parameter can be calculated. The Complexity represents the change in frequency and is derived as the ratio between the Mobility of the first derivative and the Mobility of the EEG itself. Spectral features that are used include the spectral power in the various frequency bands (beta, alpha, theta, delta) and ratios between them, mean frequency, spectral edge frequencies, spectral roll-off and spectral entropy. Also, features from wavelet and empirical mode decomposition are used. Features based on presence of EEG transients, for example the number of sleep spindles per epoch, are rarely used, although some have been mentioned.

Classification approaches

Many classification methods are based on the visual sleep staging methods, by the AASM criteria or the Rechtschaffen and Kales (R&K) criteria, that were used until the introduction of the AASM criteria in 2007. By thresholding spectral power features according to the visual sleep staging criteria, an epoch can be classified as one of the sleep stages. Several other classification methods have been used, including decision trees, k-means, support vector machines, random forests, bootstrap aggregating and neural networks. As reviewed by Chriskos et al., methods using the k-nearest neighbour classifier achieved accuracies up to 89%, decision trees up to 97%, support vector machines up to 94%, random forests up to 95% and neural networks including multilayer perceptron and recurrent neural networks up to 91%[122]. However, one must note that the accuracies do not depend on the classification method solely, but also on the pre-processing and feature extraction methods used.

5.1. Methods for sleep stage classification in adult ICU patients

Fewer methods have been studied for the evaluation of sleep in adult ICU patients, although some interesting attempts have been made. An overview of the discussed methods in adult ICU patients is provided in Table S1. In 2014, Reinke et al. introduced a novel method for ICU depth of sleep analysis, the ICU depth of sleep index (IDOS index)[124]. The IDOS index is defined as the ratio between the gamma (30 to 48 Hz) and delta (0.5 to 4 Hz) band power in each epoch using a single channel EEG. By manual selection of thresholds for each patient, the IDOS score was used to classify the EEG data as wake, NSWS (REM, N1 and N2) or SWS (N3). In their proof-of-concept study, the IDOS index was compared to visual sleep scoring in 5 non-sedated ICU patients and 15 healthy subjects. Average agreement for the three-stage classification defined by Cohen's kappa was 0.83 (standard deviation, SD: ± 0.06) for the healthy subject recordings. In the ICU patients, Cohen's kappa statistic had a high variance, ranging from 0.46 to 0.90. The lowest Cohen's kappa was found in a patient with the highest APACHE (Acute Physiology and Chronic Health Evaluation) score (APACHE II: 35, APACHE IV: 98) and a severely disturbed sleep pattern. Average sensitivity and specificity (± SD) in the ICU patients were respectively 0.88 (0.10) and 0.87 (0.09) for the wake epochs, 0.66 (0.15) and 0.69 (0.14) for non-SWS epoch and 0.68 (0.22) and 0.59 (0.27) for SWS epochs.

This study by Reinke et al. has, to our knowledge, been the only one using an EEG feature score with thresholding to classify sleep stages in adult ICU patients. However, other studies have studied features in the sleep EEG of ICU patients that could be used to assess sleep. Although these studies were not directly intended to develop an automated sleep scoring method, their results could provide valuable information in the development of such a method. Ambrogio et al. compared spectral features per epoch of sleep EEGs recorded in mechanically ventilated and sedated ICU patients and age-matched healthy controls[18]. Spectral features included relative powers in the beta, alpha, theta and delta frequency bands and the delta/alpha and

delta/beta ratio. Although significant changes were found between the critically ill patients and healthy controls, unfortunately, no results of the spectral feature changes over time were presented. Four years later, Gehlbach et al. also studied the spectral changes in sleep EEGs of mechanically ventilated ICU patients receiving intravenous sedation, and visualized the changes over time, although they did not use the calculated spectral features to classify sleep[125]. In addition to relative delta power, they also calculated the spectral edge frequency 95% (SEF95) for each epoch. SEF95 is defined as the frequency below which 95% of the spectral power resides and is normally higher during wakefulness than during sleep[126]. In line with previously discussed findings, SEF95 and delta power profiles generally lacked the circadian and ultradian rhythmicity characteristic of normal sleep. Unfortunately, the correlation between the spectral feature changes and sleep stages was not studied. However, effects of changes in medication or responsiveness in individual temporal profiles of delta power and SEF95 were discussed. Furthermore, they showed that SEF95 is more robust to high frequency artifacts, suggesting this might be a valuable feature to use in an automated sleep monitoring method.

A number of processed EEG-based brain function measures have been introduced that were originally developed to monitor depth of sedation during anaesthesia. Some have undergone limited testing as potential method to monitor sleep in ICU patients. One of these measures is the bispectral index (BIS), calculated by a nonlinear function of several EEG-based subparameters in time and frequency domain[127]. Its exact algorithm is proprietary from which only a portion has been identified[128]. BIS values near 100 represent an 'awake' clinical state, whereas 0 equals EEG silence. Studies of sleep in healthy patients demonstrate that the BIS value falls during physiological sleep and rises during arousals, but that there is a significant overlap of values for a given sleep stage[63,126]. Nicholson et al. used the BIS and submental EMG to investigate sleep in ICU patients[129]. They used the following BIS and EMG values to classify patients: awake (BIS > 85), non-SWS (BIS 60-85), SWS (BIS < 60) or REM (BIS > 60 or decrease in EMG power > 30% or the presence of REM-like waves on the frontal EEG). These thresholds were based on a pilot study, from which no results were presented. They confirmed the clinical observation and results of previous studies that almost no ICU patient shows normal sleep and aimed to highlight the fact that the traditional classification criteria of EEG sleep staging are deficient to use in critical care patients. They did not correlate the BIS-based sleep stages to the visually scored sleep stages. Vacas et al. tested the sleep staging performance in critically ill adults by an EEG-based brain monitor, the SedLine® Brain Function Monitor (Masimo Corp., Irvine, CA)[130]. The SedLine calculates the patient state index (PSI) by using a proprietary algorithm incorporating the EEG power, frequency and phase information from anterior-posterior relationships of the brain as well as coherence between bilateral brain regions[131]. In the ICU patients, the observation of disturbed sleep architectures was confirmed. Although not in ICU patients but in three healthy subjects, the sleep stages defined by the SedLine monitor were compared to visually scored hypnograms. The percentage agreements were 67% for the wake stage, 77% for the non-REM (N1 = 29%, N2 = 88%, and N3 = 6%), and 89% for the REM stage, with an overall agreement of 75%.

5.2. Methods for sleep stage classification in NICU patients

Clearly, the previously discussed methods for ICU depth of sleep monitoring are not suitable for neonatal sleep monitoring due to the significant differences in sleep EEG in neonates, as discussed in section 3. Sleep state analysis in neonates is challenging during this period of rapid brain development, introducing a large variation in EEG patterns. Yet, automated sleep staging algorithms have gained much attention in NICU research when compared to PICU and adult ICU research. The important role of neonatal sleep in the brain maturation makes the assessment of sleep a valuable measure that provides crucial insight into brain function integrity[33]. In contrast to adults, sleep staging in neonates is generally considered as a three-class problem: wake, active (REM) sleep, and quiet (NREM) sleep. Various EEG characteristics have been used to develop an automated sleep staging algorithm, specifically for preterm or term neonates, or both. The NICU patient population

mainly consists of preterm neonates. Since intrauterine brain development is similar to extrauterine brain development, methods evaluated on neonates > 38 PMA could be interesting for PICU patients. A selection of the methods developed for neonatal sleep monitoring, that have potential to be applied in the neonatal patients admitted to the PICU, is discussed in this section. An overview of the discussed methods in NICU sleep staging is provided in Table S2.

Koolen et al. developed an EEG-based sleep state classifier that performed consistently well over a wide range of ages (24-45 weeks PMA)[132]. They extracted 57 features from the time and frequency domains and features representing spatial connectivity for each epoch and used a forward feature selection algorithm to define a reduced feature subset. The features that were most distinctive for quiet and active sleep classification were relative delta power and the 5th percentile of the peak-to-peak EEG amplitude. The reduced feature subset, consisting of 7 features, was used to train a nonlinear support vector machine classifier with radial basis kernel function. Performance tests showed that the algorithm was able to classify quiet and active sleep epochs with 85% accuracy, 87% specificity and 83% sensitivity. The classifier performance was slightly better in neonates > 32 weeks PMA compared to the more premature neonates (<32 weeks), although these differences were not statistically significant. Piryatinska et al. used the frequency content of the EEG as input for a classification method in preterm and term infants with the same PMA of 40 weeks[27]. The following features were significantly different between active and quiet sleep: logarithm (log) of alpha power, log of beta power, log of theta power, log of spectral moment (i.e., a measure of how the power spectrum is distributed), spectral entropy (i.e., measure of the peakiness of the power spectrum), log of SEF90 (i.e., the frequency under which 90% of total power is accumulated) and finally the loglog of fractional dimension (i.e., a measure of smoothness of the power spectrum surface). A change point detection algorithm was used for each of the features and the created clusters where classified using a k-means clustering algorithm. Rates of agreement were evaluated for 4 combinations of features in the full-term cohort and 5 combinations of features for the preterm cohort. Agreement rates varied significantly among EEG recordings and per run, but were in the 80-90% range when averaged over 100 runs. EEG spectral features were also used by Scher et al.[27]. In addition, cardiorespiratory, EMG and EOG parameters were used. 13 features were used in a linear discriminant model, obtaining an overall accuracy of 93.8% in term and preterm neonates measured at term age (38 – 43 weeks).

Beside the spectral characteristics, continuity of the EEG background could be used to assess neonatal sleep patterns. Palmu et al. developed an EEG index based on the presence of activity bouts, referred to as "spontaneous activity transients" (SAT), to detect sleep wake cycles in early preterm infants (<34 weeks GA)[133]. The proportion of the EEG covered by SATs fluctuates with the sleep cycle, with more SATs during wakefulness. SATs were detected using an automated SAT detection algorithm that was presented in an earlier study and is based on classification by thresholding the output of a nonlinear energy operator reflecting both the amplitude and frequency of the EEG signal[134]. The percentage of time covered by SATs (SAT%) was found to show temporal behaviour that compared well with the hypnogram obtained by visual sleep scoring, with significant differences in SAT% between deep and REM sleep.

Recent studies used more advanced machine learning algorithms for neonatal sleep stage classification. Fraiwan et al. used time-frequency analysis for automated sleep stage identification[135]. Wigner–Ville distribution (WVD), Hilbert–Hough spectrum (HHS) and continuous wavelet transform (CWT) time frequency distributions were used to represent the EEG signal in time-frequency domain. Features were extracted from WVD, HHS and CWT by entropy values. Classification was done by an artificial neural network (ANN). The system was trained and tested using data taken from neonates of 40 weeks PMA for both preterm and full-term neonates. The data set was the same as was used by Piryatinska et al.[27]. Classification based on WVD outperformed the approaches based on CWT and HHS and achieved a kappa coefficient of 0.84 and 0.65 in full-term neonate recordings and 0.74 and

0.50 in preterm neonate recordings. In their later study, Fraiwan et al. used a long short-term memory (LSTM) neural network[136]. A LSTM is used in sequence learning that has the ability to learn long-term dependencies between data. The method achieved high accuracy levels (96.81%) in three-class sleep staging on the same dataset as used in their previous study[135] and by Piryatinska et al.[27].

An active research group in the field of machine learning based neonatal sleep state analysis methods from the University of Leuven introduced various methods[137–140]. In 2017, they presented the class-based adaptive sleep staging (CLASS) algorithm to detect quiet sleep[139]. The CLASS algorithm relies on the more discontinuous background pattern observed during quiet sleep. High power artifacts could be easily confused with periods of discontinuity. Therefore, a nonstationary artifact reduction method that applies principal component analysis over a sliding window to separate the artifact from the EEG signal was used. Next, an adaptive segmentation method was used that defines segment boundaries where large changes in amplitude and frequency behaviour occur. This allows the application in a larger range of PMA since it is based on the relative discontinuity at each PMA between quiet sleep and other states. The CLASS performance proved optimal across recordings from neonates at 31-38 weeks PMA, compared to neonates <31 weeks or > 38 weeks, with a median sensitivity and specificity of 94% and 83%, respectively. After 38 weeks, the EEG becomes more continuous and relative changes in discontinuity become less distinguishable, resulting in a lower CLASS performance. In a study published in 2018 by Pillay et al., they trained both a Hidden Markov Model (HMM) as well as a Gaussian Mixture Model (GMM) for sleep state classification in neonates 27-41 weeks PMA admitted to the NICU without sedative or anticonvulsive medication or cerebral lesions[137]. They evaluated their method both for two state (quiet and active sleep) as well as four state classification. In four state classification, active sleep was subdivided into active sleep I and LVI, and quiet sleep is subdivided into tracé alternant and HVS. The use of an HMM enables the incorporation of prior knowledge of the sleep state transition probabilities. 112 features were extracted from time and frequency domain and wavelet and empirical mode decompositions from which the best features were selected for each method by minimum redundancy maximum relevance. This paper also suggests a patient-wise rescaling of the features before feeding them into the classifier. For both two-state as well as four state classification, the HMM performed better than the GMM model (two state: accuracy is 95%, in HMM and 92% in GMM, four-state: accuracy Is 86% in HMM and 82% in GMM), suggesting that the introduction of the transition probability is favourable. Feature scaling improved the classification performance. Ansari et al. have used a convolutional neural network for sleep stage classification in preterm and term neonates[138]. The neural network has been directly trained on a multichannel EEG after downsampling the EEG signal to 30 Hz. For two-state classification, this method achieved a mean Cohen's kappa of 0.76. For the four-state classification, this Cohen's kappa was 0.66. In the most recent study of this Belgian research group, published by Ghimatgar et al. in 2020, they used a multichannel approach based on a LSTM network and HMM[140]. Sequential forward feature selection was used for feature and channel selection to identify the features in the various brain areas that are most relevant for sleep staging in neonates. A LSTM then classifies the epoch as sleep stages. Finally, an HMM-based postprocessing stage was used to reduce false positives by incorporating the knowledge of transition probabilities between stages into the classification process. The final classifier achieved an overall accuracy 78.9-82.4% on a dataset with 16 neonates PMA 38-42 weeks with six-bipolar EEG channels.

6. Discussion

This review provides a broad overview of the potential factors that influence the sleep EEG in critically ill children and discusses the available methods for automated sleep monitoring. The presented overview of the most important findings could contribute to the required knowledge for the development of an EEG-based sleep monitoring method for PICU patients. The findings demonstrate that sleep monitoring in the PICU is challenged by a variety of PICU factors that interfere with the sleep EEG. It is important to take

these factors into account during the development of a PICU sleep monitoring method and during the assessment of sleep.

First, the significant changes in the sleep EEG with age must be taken into account, with the most rapid development in EEG patterns during the first 2 months of life. The sleep EEG patterns observed in neonates differ greatly from those observed in older children and adults, thereby challenging the use of a single sleep monitoring algorithm for all PICU patients. In the first months of life, the discontinuous and asynchronous EEG during sleep becomes more continuous and synchronous. Neonatal EEG patterns disappear and adult sleep transients (sleep spindles, K-complexes, vertex waves) appear, enabling the differentiation into N1, N2 and N3 sleep. In later childhood, the EEG changes are less drastic. With progression of age, the frequency and amplitude of the DPR keeps gradually increasing until it reaches its adult characteristics during adolescence. Although the sleep EEG changes with age have been extensively studied, it must be realized that the developmental status of a child determines the EEG, rather than the age. On top of the age-dependent changes, the critical illness introduces additional changes in the EEG. Unfortunately, sleep EEG studies in PICU patients are limited. Adult ICU studies show that the atypical sleep EEG is characterized by polymorphic delta activity, the absence of N2 markers, burst suppression and isoelectric activity. In general, all medical conditions in which the brain is involved, either directly or indirectly, could affect the EEG. Whereas most cerebral pathologies cause diffuse slowing of the EEG, focal changes could be observed in pathologies with local lesions such as brain abscesses, local contusion or CVA. Epileptic discharges could be present in various neurological conditions, including status epilepticus, central nervous infections or hypoxic-ischemic, renal, hepatic or other metabolic encephalopathies. Furthermore, many sedatives and analgesics are known to cause slowing of the EEG. The effect of medication on the sleep EEG in critically ill children is especially complex due to various factors. The critical illness introduces unpredictable pharmacodynamics and pharmacokinetics caused by hemodynamic instability, altered protein binding and impaired organ function. On top of that, the age and physiological characteristics of the child influence the drug interaction. The effect of drugs on the brain, and thereby the EEG, is dependent on the brain maturation state. Another challenge in the prediction of EEG effects caused by medication is the drug-drug interaction. Medication is often used in combination with other classes of medication that both interact with the EEG and could introduce complex drug-drug interactions. For some medications, a temporal discrepancy between EEG changes and blood concentration of the drug is known, which could be altered by the unpredictable pharmacodynamics and pharmacokinetics present in PICU patients. Although for many medications its dose-dependent effect on the sleep EEG is widely studied in healthy individuals, less is known about their effects in critically ill children. Lastly, it is important to mention that for many drugs and medical conditions the influence on the wake EEG has been widely studied, while its influence in the various sleep stages remains unclear. During normal sleep, slowing of the EEG is physiological. The EEG deviations introduced by drugs and medical conditions that have a slowing effect could be less clear during sleep. Finally, the EEG is – often unavoidably - confounded with artifacts from various physiological or non-physiological sources. The noisy PICU environment introduces additional artifact sources, such as from surrounding electrical devices or frequent nursing activities. During the development of a sleep monitoring method, attention must be paid to these artifacts and efforts should be made to remove the artifacts from the EEG signal and to evaluate the effect of artifact removal.

The discussed automated sleep monitoring methods used in adult ICU patients, NICU patient and healthy patients provide knowledge on potential methods and EEG features to use in a PICU specific method. Some of the methods used and validated in adult patients will perform quite well in older children in the PICU, while its performance will heavily decrease when applied to the neonates in the PICU. The same goes the other way; a neonatal sleep staging algorithm will probably not perform well in the older PICU population. During evaluation of these algorithms, the patient-related factors

should be taken into account, including age, medication and medical condition. From studies in healthy adults, knowledge on the accuracy and efficiency of the various signal pre-processing algorithms, feature extraction methods and classification is obtained. Signal pre-processing is often minimal, with only the use of a single bandpass, low-pass or high-pass filter to remove unwanted frequencies. Most often used features are derived from the frequency domain, although time, wavelet and empirical mode domain features have also been shown to achieve good results. Several classification methods have been used in sleep stage classification. Complex classifiers, for example using neural networks, do not always achieve better performance results than simpler classifiers such as decision trees or support vector machines do.

The IDOS index introduced by Reinke et al. has to our knowledge been the only method to automatically asses sleep in critically ill adult patients[124]. This simple and intuitive method, based on the gamma/delta ratio, seems to be a promising method for sleep assessment in the critically ill and is therefore more discussed in depth here. However, the use of the gamma frequency band in the analysis of sleep is controversial. The electrical activity in the gamma range is contaminated with muscle activity as a result of its overlap with the EMG frequency band. Although increased muscle activity, and thus an increased gamma power, is a useful marker in sleep state analysis, its use in patients receiving muscle relaxants and sedatives is therefore debatable. Furthermore, noise and artifacts often introduce high frequency activity, thereby increasing the gamma power, and thus interfering with the IDOS index. Also, the general slower EEG activity observed in children might reduce its usefulness for the application in PICU patients. Further efforts on improving the index score and validation in a heterogeneous ICU, and finally PICU, population are needed before clinical implementation. Other EEG features that are shown to reflect sleep in critically ill patients should be evaluated for their implementation as a variable in an EEG-based index score. These EEG features include the relative band powers, ratios of other band powers and SEF95. Next, processed EEG-based brain function measures, such as the BIS value or PSI, have potential as measures of sleep in ICU, but additional validation studies in ICU patients are needed to correlate the processed EEG measure with the various sleep stages defined by visual sleep scoring. One disadvantage of using these measures is the nondisclosure of their calculation algorithms. This complicates the interpretation of unexpected values and the identification of factors influencing the measure. Besides, it eliminates the ability to calculate the measure from raw EEG data retrieved from brain monitors other than the proprietary ones. Also, it must be highlighted that the development of the BIS and PSI values have primarily been based on depth of sedation, which is a different process than natural sleep.

In neonatal sleep staging methods, additional EEG characteristics are used, such as measures to define the EEG discontinuity. Clearly, methods based on EEG discontinuity are only beneficial in neonates up to several weeks postnatal age, thereby limiting the application in a broader PICU population. However, used features in time, frequency or time-scale domain are similar to those used in adult sleep state analysis. Most discriminating features for active and quiet sleep discrimination appeared to be in the frequency domain, similar to observations in adult ICU and healthy adult sleep. However, amplitude-based features, for example using the range EEG, were also frequently selected by feature selection algorithms. In contrast to the discussed adult ICU methods, many neonatal methods rely on more complex algorithms. The use of the complex machine learning methods potentially enables the extraction of patterns in the complex EEG signal that are not easily observed by eye or simple computerized methods. Koolen et al. introduced an intuitive method with the best performance in neonates > 32 weeks PMA[132]. The use of multiple EEG channels enabled the incorporation of asynchrony measures, typically observed in neonates. Discontinuity of the EEG was quantified by burst intervals, as a valuable feature in neonatal sleep stage classification. Adaptive approaches, as the one proposed by Dereymaeker et al., have the advantage to detect relative changes and adjust to the individual EEG[139]. Palmu et al. studied a simple index based on the presence of SATs in the sleep EEG as a marker for neonatal sleep stages[133]. SATs

are typically present in preterm neonates. A method solely based on the presence of SATs presence might therefore be of limited value in the PICU. Scher et al. introduced a simple method with a high accuracy using EOG, EMG, ECG and respiratory channels in addition to the EEG that is worth considering during the development of a PICU method[141]. The incorporation of knowledge on the temporal structure of the sleep cycle, used in the methods of Ghimatgar et al.[140] and Pillay et al.[137], both using HMM, could efficiently increase the algorithm accuracy. The sleep transition rules embedded in the models incorporate the information from neighbouring segments and use this information mainly to rule out infrequent stage transitions, such as from wake to deep sleep. Ghimatgar et al. showed that postprocessing with HMM could improve the overall accuracy up to 3% by reduction of false positive sleep stages[140]. Thus, the use of an HMM seems a promising method for sleep stage classification.

Some NICU methods choose a four-state classification approach (AS1, LVI, HVI, TA) since the proportion of the various neonatal EEG patterns observed during quiet and active sleep has prognostic value on developmental and clinical outcomes in the NICU. However, this four-state classification approach is not advised by the AASM criteria and might therefore not be beneficial for the assessment of sleep in the PICU. Also, it must be noted that in the discussed NICU studies little or no attention was paid to the neurophysiological changes caused by medication and neurological conditions that influence the EEG. Most studies did not include neonates receiving sedative or analgesic medication or with neurological disorders.

The broad subject of this review and extensive literature on each subtopic limited the completeness of this review. For most subtopics, no systematic search was performed, thereby risking incompleteness of the provided information. However, systematic searches were used for the automated sleep methods in adult ICU and NICU. To stay within the scope of this review, the results of these studies were only briefly discussed and not systematically analysed. They serve as basic knowledge on classification approaches for ICU sleep. To bound this review's scope, a selection of commonly used medication and frequently administered medical conditions in the PICU was made. EEG effects were described in general, without specifying frequency and amplitude changes. However, for most factors, EEG effects are highly variable per individual and are therefore irrelevant to quantify. Also, for many factors, only EEG effects on healthy adults were identified, not taken into account the potentially different effects during sleep, critical illness or in children.
Not discussed in this review is the EEG data acquisition, which still has to be evaluated for the development of a PICU sleep monitoring method. Next to the technical aspects of the EEG data acquisition, this implies the channel selection, or the EEG electrodes. The use of fewer channels, or even a single one, would be patient-friendlier, quicker to apply, easier to interpret and require less computational power. However, a single channel is less robust to disconnection of the electrode due to patient movement. Reinke et al. used the C3/C4 electrodes, placed centrally on the left and right hemisphere, to calculate the IDOS index[124]. This channel has been shown to be most representative for the classification of sleep stages in healthy individuals with minimal EMG interference[142]. However, the frontal channel (F3/F4) are known to better measure K-complexes and delta waves than other channels do[143]. When selecting a channel, the asynchronous behaviour and topographical differentiation of electrical activity in children should be taken into account.

It must be emphasized that the EEG is just a biomarker of the underlying sleep state that might not always reflect the sleep. The EEG resembles electrical activity arising from various processes in the brain, influenced by underlying factors present in PICU patients that also interact with each other. Unfortunately, neurosciences have not progressed to the point where it is possible to exactly understand the neurophysiological activity that causes the observed EEG pattern. The neurophysiology of sleep is a complex process that has not fully been elucidated. Although the AASM criteria form a good support in the assessment of sleep, it must be realized that these criteria, and thus definitions of sleep stages, were originally defined based on visual

analysis of the EEG signal rather than physiological substrates that reflect the true underlying sleep stage. This is especially relevant in sedated patients or patient with significant neurological disorders, where features observed in the EEG signal are not always related to the physiological process of sleep. The slowing of the EEG activity observed in encephalopathy or under sedatives, could potentially lead to the, perhaps incorrect, classification into deeper sleep stages following the AASM criteria. The same applies to the increase in N2 features induced by α-2-receptor agonists leading to increased N2 staging. However, it should be questioned whether the increase in N2 markers really means that there is an increase in N2 sleep, and thus whether the EEG findings reflect the physiological processes that occur during N2 sleep. The effectiveness of EEG-based sleep staging based on the AASM criteria should therefore be discussed in the critically ill patient population.

Despite these challenges introduced by various interfering factors, EEG remains the golden standard in the assessment of sleep, especially in ICU patients where heart rate, respiratory rate and EMG are artificially influenced. However, it must be noted that sleep is multi-dimensional and is more than sleep staging only. Other dimensions of sleep include total sleep time, movement, awakenings, perception of sleep, tiredness upon awakening, daytime energy and functioning. Various methods measure various dimensions of sleep. Bourne et al.[19] and Richards et al.[20] both reviewed various methods to assess sleep in critically ill patients. Next to objective methods based on EEG or actigraphy, they discussed subjective method including patient and clinical assessment. In patient assessment, a patient's perceptions of sleep quality and tiredness could be evaluated via questionnaires. However, this method requires patients that are alert, orientated and able to respond and provide feedback, which limits its use in PICU patients. Clinician assessment is useful in patients who are unable to self-report. However, this method is time-consuming, introduces interrater variability and is intermittent.

This review can aid in the development of an EEG-based automated sleep monitoring method in the PICU by summarizing the barriers and EEG deviations in the PICU. Despite the challenges introduced by the various PICU factors, the detrimental effects on clinical outcome of sleep deprivation in critically ill children emphasize the need for the development of a PICU sleep monitoring method. An automated method eliminates the need for human intervention in the sleep scoring process, resulting in a cost-effective, objective method that could be applied in real-time and continuously. The first step in this development is the study of the sleep EEG in PICU patients. The addition of the EOG signal in the algorithm might be beneficial in the detection of REM sleep. Future efforts in the development are the optimization of signal pre-processing, feature selection and classification approaches. The challenge remains to develop a method that is suitable for patients over the whole PICU age range, from 0 to 18 years. Due to the significant EEG difference in this age range, an easily adjustable method could be a solution. The method could be adjusted to the patient age by predefined and validated settings for various age subclasses. These settings could include the selection of a different feature subset, or classifier parameters. For example, in an EEG index-based method, the threshold value could be adjusted for various age categories. Besides, the three-stage classification in neonates (wake, quiet and active sleep) versus the four (or more)-stage classification in older children (wake, REM, non-SWS (N1, N2), SWS (N3)) should be taken into account during the development of a PICU depth of sleep monitoring method. Finally, the developed sleep monitoring method should be simple, intuitive and robust to artifacts, and able to assess sleep real-time. Large scale validation in critically ill patients is required before clinical implementation is justified. The algorithm performance should be correlated to sedation scores and illness severity. Identification of patients in which sleep is difficult and ineffective to assess could be useful. In addition to the comparison with visually scored sleep stages, the monitoring outcomes could be compared to other automated sleep staging methods and patient and nurse perception of sleep. Finally, the sleep assessment should be correlated to clinical outcomes.

7. Conclusion

In conclusion, the PICU environment provides unique challenges in the development of an automated sleep monitoring methods. The broad ranges of age, critical illness and medication encountered in the PICU introduce a variety of factors that affect the sleep EEG. The sleep EEG patterns observed in neonates differ greatly from those observed in older children and adults, as they are discontinuous, asynchronous and lack adult sleep transients. This challenges the use of a single sleep monitoring algorithm for all PICU patients. Many medical conditions, sedatives and analgesics cause slowing of the EEG. These EEG influences should be taken into account when evaluating the performance of the sleep monitoring algorithm and during assessment of sleep. The discussed sleep monitoring methods provide knowledge on the pre-processing, feature extraction and classification methods. The most discriminating features lie in the frequency domain, where the simple band powers appear to be discriminative between sleep stages. Also, time domain features indicating amplitude ranges seem to be valuable measures. Classification methods vary from simple threshold-based methods to complex neural networks. Methods incorporating the sequential characteristics of the sleep data seems to be promising, but also simple threshold-based methods have potential. To conclude, based on these literature findings, the development of an EEG-based sleep monitoring method for PICU patients is challenging but seems to be achievable.

List of abbreviations

AASM	–	American Academy of Sleep Medicine
BIS	–	Bispectral index
CLASS	–	Class-based adaptive sleep staging
CVA	–	Cerebral Vascular Accident
CWT	–	Continuous wavelet transform
DPR	–	Dominant posterior rhythm
ECG	–	Electrocardiogram
EEG	–	Electroencephalogram
EMG	–	Electromyogram
EOG	–	Electrooculogram
GA	–	Gestational age
GABA	–	Gamma-aminobutyric acid
GMM	–	Gaussian mixture model
HHM	–	Hidden Markov model
HHS	–	Hilbert-Hough spectrum
HVS	–	High voltage slow
ICP	–	Intracranial pressure
ICU	–	Intensive care unit
IDOS	–	Intensive Care Unit Depth of Sleep
ICA	–	Independent component analysis
LSTM	–	Long short-term memory
LVI	–	Low voltage irregular
N1	–	Non slow wave sleep stage 1
N2	–	Non slow wave sleep stage 2
N3	–	Non slow wave sleep stage 3
NICU	–	Neonatal intensive care unit
NMDA	–	N-methyl-D-aspartate
NREM	–	Non rapid eye movement
NSWS	–	Non slow wave sleep
PCA	–	Principal component analysis
PICU	–	Paediatric intensive care unit
PMA	–	Postmenstrual age
REM	–	Rapid eye movement
R&K	–	Rechtschaffen & Kales
SD	–	Standard deviation
SAT	–	Spontaneous activity transient
SEF95	–	Spectral edge frequency 95%
SWS	–	Slow wave sleep
TA	–	Tracé alternant
TBI	–	Traumatic brain injury
WVD	–	Wigner-Ville distribution

References

1. Schwartz, J. & Roth, T. Neurophysiology of Sleep and Wakefulness: Basic Science and Clinical Implications. *Curr. Neuropharmacol.* 6, 367–378 (2009).
2. Kudchadkar, S. R., Aljohani, O. A. & Punjabi, N. M. Sleep of critically ill children in the pediatric intensive care unit: A systematic review. *Sleep Medicine Reviews* vol. 18 103–110 (2014).
3. Cureton-Lane, R. A. & Fontaine, D. K. Sleep in the pediatric ICU: An empirical investigation. *Am. J. Crit. Care* 6, 56–63 (1997).
4. Corser, N. C. Sleep of 1- And 2-Year-old Children in Intensive Care. *Compr. Child Adolesc. Nurs.* 19, 17–31 (1996).
5. Carno, M. A. & Connolly, H. V. Sleep and sedation in the pediatric intensive care unit. *Critical Care Nursing Clinics of North America* vol. 17 239–244 (2005).
6. Armour, A. D., Khoury, J. C., Kagan, R. J. & Gottschlich, M. M. Clinical assessment of sleep among pediatric burn patients does not correlate with polysomnography. *J. Burn Care Res.* 32, 529–534 (2011).
7. Al-Samsam, R. H. & Cullen, P. Sleep and adverse environmental factors in sedated mechanically ventilated pediatric intensive care patients. *Pediatr. Crit. Care Med.* 6, 562–567 (2005).
8. Kamdar, B. B., Needham, D. M. & Collop, N. A. Sleep deprivation in critical illness: Its role in physical and psychological recovery. *Journal of Intensive Care Medicine* vol. 27 97–111 (2012).
9. Palmblad, J., Petrini, B., Wasserman, J. & Akerstedt,

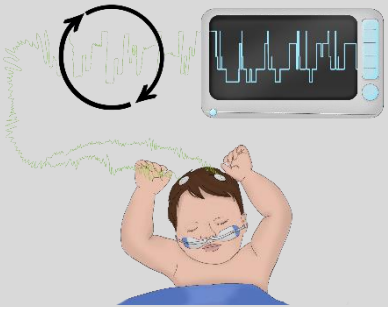
T. Lymphocyte and granulocyte reactions dueing sleep deprivation. *Psychosom. Med.* 41, 273–278 (1979).
10. Besedovsky, L., Lange, T. & Born, J. Sleep and immune function. *Pflugers Archiv European Journal of Physiology* vol. 463 121–137 (2012).
11. Weinhouse, G. L. *et al.* Bench-to-bedside review: Delirium in ICU patients - importance of sleep deprivation. *Critical Care* vol. 13 (2009).
12. Peirano, P. D. & Algarín, C. R. Sleep in brain development. in *Biological Research* vol. 40 471–478 (Society of Biology of Chile, 2007).
13. Iber, A. L. C. C., Ancoli-Israel, S., Chesson, A., Quan, S. F. & Medicine., for the A. A. of S. The AASM manual for the scoring of sleep and associated events: Rules, Terminology and Technical Specification: Version 2.6 (2020). *Am. Acad. Sleep Med.* (2007).
14. Watson, P. L. *et al.* Atypical Sleep in Ventilated Patients. *Crit. Care Med.* 41, 1958–1967 (2013).
15. Cooper, A. B. *et al.* Sleep in critically ill patients requiring mechanical ventilation. *Chest* 117, 809–818 (2000).
16. Drouot, X. *et al.* A new classification for sleep analysis in critically ill patients. *Sleep Med.* 13, 7–14 (2012).
17. Danker-Hopfe, H. *et al.* Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *J. Sleep Res.* 13, 63–69 (2004).
18. Ambrogio, C., Koebnick, J., Quan, S. F., Ranieri, V. M. & Parthasarathy, S. Assessment of Sleep in Ventilator-Supported Critically Ill Patients. *Sleep* 31, 1559–1568 (2008).
19. Bourne, R. S., Minelli, C., Mills, G. H. & Kandler, R. Clinical review: Sleep measurement in critical care patients: Research and clinical implications. *Critical Care* vol. 11 (2007).
20. Richards, K. C., Wang, Y. Y., Jun, J. & Ye, L. A Systematic Review of Sleep Measurement in Critically Ill Patients. *Frontiers in Neurology* vol. 11 (2020).
21. Olbrich, E., Rusterholz, T., LeBourgeois, M. K. & Achermann, P. Developmental Changes in Sleep Oscillations during Early Childhood. *Neural Plast.* 2017, (2017).
22. Amzica, F. & Steriade, M. Electrophysiological correlates of sleep delta waves. *Electroencephalogr. Clin. Neurophysiol.* 107, 69–83 (1998).
23. Rechtschaffen, A. & Kales, A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects.* (1968).
24. Louis, E. K. St. *et al.* Appendix 1. The Scientific Basis of EEG: Neurophysiology of EEG Generation in the Brain. (2016).
25. Mirmiran, M., Maas, Y. G. H. & Ariagno, R. L. Development of fetal and neonatal sleep and circadian rhythms. *Sleep Medicine Reviews* vol. 7 321–334 (2003).
26. Horne, J. Why REM sleep? Clues beyond the laboratory in a more challenging world. *Biological Psychology* vol. 92 152–168 (2013).
27. Piryatinska, A. *et al.* Automated detection of neonate EEG sleep stages. *Comput. Methods Programs Biomed.* 95, 31–46 (2009).
28. Grigg-Damberger, M. *et al.* The visual scoring of sleep and arousal in infants and children. *Journal of Clinical Sleep Medicine* vol. 3 201–240 (2007).
29. Nunes, M. L., Da Costa, J. C. & Moura-Ribeiro, M. V. L. Polysomnographic quantification of bioelectrical maturation in preterm and fullterm newborns at matched conceptional ages. *Electroencephalogr. Clin. Neurophysiol.* 102, 186–191 (1997).
30. De Weerd, A. W. & Van den Bossche, R. A. S. The development of sleep during the first months of life. *Sleep Med. Rev.* 7, 179–191 (2003).
31. Selton, D. *et al.* EEG at 6 weeks of life in very premature neonates. *Clin. Neurophysiol.* 121, 818–822 (2010).
32. Hughes, J. R., Fino, J. & Gagnon, L. Periods of activity and quiescence in the premature EEG. *Neuropediatrics* 14, 66–72 (1983).
33. Dereymaeker, A. *et al.* Review of sleep-EEG in preterm and term neonates. *Early Hum. Dev.* 113, 87–103 (2017).
34. Lenard, H. G. The development of sleep spindles in the EEG during the first two years of life. *Neuropadiatrie* 1, 264–276 (1970).
35. Kudchadkar, S. R. *et al.* Temporal characteristics of the sleep EEG power spectrum in critically ill children. *J. Clin. Sleep Med.* 11, 1449–1454 (2015).
36. Freedman, N. S., Gazendam, J., Levan, L., Pack, A. I. & Schwab, R. J. Abnormal sleep/wake cycles and the effect of environmental noise on sleep disruption in the intensive care unit. *Am. J. Respir. Crit. Care Med.* 163, 451–457 (2001).
37. Amzica, F. What does burst suppression really mean? *Epilepsy Behav.* 49, 234–237 (2015).
38. Plaschke, K. *et al.* EEG changes and serum anticholinergic activity measured in patients with delirium in the intensive care unit. *Anaesthesia* 62, 1217–1223 (2007).
39. Koponen, H., Partanen, J., Pääkkönen, A., Mattila, E. & Riekkinen, P. J. EEG spectral analysis in delirium. *J. Neurol. Neurosurg. Psychiatry* 52, 980–985 (1989).
40. Rodin, E. A., Luby, E. D. & Gottlieb, J. S. The electroencephalogram during prolonged experimental sleep deprivation. *Electroencephalogr. Clin. Neurophysiol.* 14, 544–551 (1962).
41. Naitoh, P., Kales, A., Kollar, E. J., Smith, J. C. & Jacobson, A. Electroencephalographic activity after prolonged sleep loss. *Electroencephalogr. Clin. Neurophysiol.* 27, 2–11 (1969).
42. Sforza, E., Chapotot, F., Pigeau, R., Paul, P. N. & Buguet, A. Effects of sleep deprivation on spontaneous arousals in humans. *Sleep* 27, 1068–1075 (2004).
43. De Gennaro, L., Ferrara, M. & Bertini, M. Effect of slow-wave sleep deprivation on topographical distribution of spindles. *Behav. Brain Res.* 116, 55–59 (2000).
44. Novak, M. & Shapiro, C. M. Drug-induced sleep disturbances. Focus on nonpsychotropic medications. *Drug Safety* vol. 16 133–149 (1997).

45. Saper, C. B., Scammell, T. E. & Lu, J. Hypothalamic regulation of sleep and circadian rhythms. *Nature* vol. 437 1257–1263 (2005).
46. Witte, H. *et al.* Interrelations between EEG frequency components in sedated intensive care patients during burst-suppression period. *Neurosci. Lett.* 260, 53–56 (1999).
47. Kassell, N. F., Hitchon, P. W., Gerk, M. K., Sokoll, M. D. & Hill, T. R. Alterations in cerebral blood flow, oxygen metabolism, and electrical activity produced by high dose sodium thiopental. *Neurosurgery* 7, 598–603 (1980).
48. Lancel, M. Role of GABA(A) receptors in the regulation of sleep: Initial sleep responses to peripherally administered modulators and agonists. *Sleep* 22, 33–42 (1999).
49. Veselis, R. A., Reinsel, R., Marino, P., Sommer, S. & Carlon, G. C. The effects of midazolam on the EEG during sedation of critically ill patients. *Anaesthesia* 48, 463–470 (1993).
50. Jennekens, W. *et al.* Effects of midazolam and lidocaine on spectral properties of the EEG in full-term neonates with stroke. *Eur. J. Paediatr. Neurol.* 16, 642–652 (2012).
51. Hales, T. G. & Lambert, J. J. The actions of propofol on inhibitory amino acid receptors of bovine adrenomedullary chromaffin cells and rodent central neurones. *Br. J. Pharmacol.* 104, 619–628 (1991).
52. Kondili, E., Alexopoulou, C., Xirouchaki, N. & Georgopoulos, D. Effects of propofol on sleep quality in mechanically ventilated critically ill patients: A physiological study. *Intensive Care Med.* 38, 1640–1646 (2012).
53. Herregods, L., Rolly, G., Mortier, E., Bogaert, M. & Mergaert, C. EEG and SEMG monitoring during induction and maintenance of anesthesia with propofol. *Int. J. Clin. Monit. Comput.* 6, 67–73 (1989).
54. Walsh, E. C. *et al.* Age-Dependent Changes in the Propofol-Induced Electroencephalogram in Children With Autism Spectrum Disorder. *Front. Syst. Neurosci.* 12, 23 (2018).
55. Modica, P. A., Tempelhoff, R. & White, P. F. *Pro-and Anticonvulsant Effects of Anesthetics (Part 11).*
56. Pandharipande, P. & Ely, E. W. Sedative and Analgesic Medications: Risk Factors for Delirium and Sleep Disturbances in the Critically Ill. *Crit. Care Clin.* 22, 313–327 (2006).
57. Bauer, G. & Bauer, R. EEG, drug effects, and central nervous poisening. in *E. Niedermeyer and F. Lopes da Sila (Eds,), Electroencephalography: Basic Principles, Clinical Applications and Related Fields* 701–724 (Urband and Schwarzenberg, 2005).
58. Nelson, L. E. *et al.* The α2-adrenoceptor agonist dexmedetomidine converges on an endogenous sleep-promoting pathway to exert its sedative effects. *Anesthesiology* 98, 428–436 (2003).
59. Hall, J. E., Uhrich, T. D. & Ebert, T. J. Sedative, analgesic and cognitive effects of clonidine infusions in humans. *Br. J. Anaesth.* 86, 5–11 (2001).
60. Weinhouse, G. L. & Watson, P. L. Sedation and Sleep Disturbances in the ICU. *Critical Care Clinics* vol. 25 539–549 (2009).
61. Huupponen, E. *et al.* Electroencephalogram spindle activity during dexmedetomidine sedation and physiological sleep. *Acta Anaesthesiol. Scand.* 52, 289–294 (2008).
62. Mason, K. P., O'Mahony, E., Zurakowski, D. & Libenson, M. H. Effects of dexmedetomidine sedation on the EEG in children. *Paediatr. Anaesth.* 19, 1175–1183 (2009).
63. Sleigh, J., Harvey, M., Voss, L. & Denny, B. Ketamine - more mechanisms of action than just NMDA blockade. *Trends in Anaesthesia and Critical Care* vol. 4 76–81 (2014).
64. Duncan, W. C. *et al.* Concomitant BDNF and sleep slow wave changes indicate ketamine-induced plasticity in major depressive disorder. *Int. J. Neuropsychopharmacol.* 16, 301–311 (2013).
65. Schwartz, M. S., Virden, S. & Scott, D. F. Effects of ketamine on the electroencephalograph. *Anaesthesia* 29, 135–140 (1974).
66. Bihari, S. *et al.* Factors affecting sleep quality of patients in intensive care unit. *J. Clin. Sleep Med.* 8, 301–307 (2012).
67. Pilowsky, I., Crettenden, I. & Townley, M. Sleep disturbance in pain clinic patients. *Pain* 23, 27–33 (1985).
68. Dimsdale, J. E., Norman, D., Dejardin, D. & Wallace, M. S. *The Effect of Opioids on Sleep Architecture. Journal of Clinical Sleep Medicine* vol. 3 (2007).
69. Matejcek, M., Pokorny, R., Ferber, G. & Klee, H. Effect of morphine on the electroencephalogram and other physiological and behavioral parameters. *Neuropsychobiology* 19, 202–211 (1988).
70. Smith, N. T., Dec-Silver, H. & Sanford, T. J. EEGs during high-dose fentanyl-, sufentanil-, or morphine-oxygen anesthesia. *Anesth. Analg.* 63, 386–393 (1984).
71. Malver, L. P. *et al.* Electroencephalography and analgesics. *Br. J. Clin. Pharmacol.* 77, 72–95 (2014).
72. Fink, M. & Irwin, P. Central nervous system effects of aspirin. *Clin. Pharmacol. Ther.* 32, 362–365 (1982).
73. Lotsch, J. *et al.* Effects of azapropazone on pain-related brain activity in human subjects. *Br. J. Clin. Pharmacol.* 40, 545–552 (1995).
74. Armour, A., Gottschlich, M. M., Khoury, J., Warden, G. D. & Kagan, R. J. A randomized, controlled prospective trial of zolpidem and haloperidol for use as sleeping agents in pediatric burn patients. *J. Burn Care Res.* 29, 238–247 (2008).
75. Centorrino, F. *et al. EEG Abnormalities During Treatment With Typical and Atypical Antipsychotics. Am J Psychiatry* vol. 159 (2002).
76. Drake, M. E., Huber, S. J., Pakalnis, A. & Denio, L. Electroencephalographic effects of antiepileptic drug therapy. *J. Epilepsy* 3, 75–79 (1990).
77. Young, G. B., McLachlan, R. S., Kreefi, J. H. & Demelo, J. D. An electroencephalographic classification for coma. *Can. J. Neurol. Sci.* 24, 320–325 (1997).

78. *Besluit vaststelling van de dood bij postmortale orgaandonatie*. (wetten.overheid.nl, 2020).
79. Iragui, V. J. & McCutchen, C. B. Physiologic and prognostic significance of 'alpha coma'. *J. Neurol. Neurosurg. Psychiatry* 46, 632–638 (1983).
80. Kaplan, P. W., Genoud, D., Ho, T. W. & Jallon, P. Etiology, neurologic correlations, and prognosis in alpha coma. *Clin. Neurophysiol.* 110, 205–213 (1999).
81. Emidio, A. C. *et al.* Spindle Coma in the Intensive Care Unit: Different Aetiologies – Different Outcomes. *Eur. J. Case Reports Intern. Med.* 6, 1 (2019).
82. Horton, E. J., Goldie, W. D. & Baram, T. Z. Rhythmic Coma in Children. *J. Child Neurol.* 5, 242–247 (1990).
83. Homan, R. W. & Jones, M. G. Alpha-pattern coma in a 2-month-old child. *Ann. Neurol.* 9, 611–613 (1981).
84. Young, G. B. The EEG in coma. *J. Clin. Neurophysiol.* 17, 473–485 (2000).
85. Kaibara, M. & Blume, W. T. The postictal electroencephalogram. *Electroencephalogr. Clin. Neurophysiol.* 70, 99–104 (1988).
86. Malow, B. A., Kushwaha, R., Lin, X., Morton, K. J. & Aldrich, M. S. Relationship of interictal epileptiform discharges to sleep depth in partial epilepsy. *Electroencephalogr. Clin. Neurophysiol.* 102, 20–26 (1997).
87. Wieser, H. G. Temporal lobe epilepsy, sleep and arousal: stereo-EEG findings. *Epilepsy research. Supplement* vol. 2 97–119 (1991).
88. Capdevila, O. S., Dayyat, E., Kheirandish-Gozal, L. & Gozal, D. Prevalence of epileptiform activity in healthy children during sleep. *Sleep Med.* 9, 303–309 (2008).
89. Marzec, M. L. & Malow, B. A. Approaches to staging sleep in polysomnographic studies with epileptic activity. *Sleep Med.* 4, 409–417 (2003).
90. Turrell, R. C. & Roseman, E. Electroencephalographic studies of the encephalopathies: Iv. serial studies in meningococcic meningitis. *Arch. Neurol. Psychiatry* 73, 141–148 (1955).
91. Wang, I. J. *et al.* The correlation between neurological evaluations and neurological outcome in acute encephalitis: A hospital-based study. *Eur. J. Paediatr. Neurol.* 11, 63–69 (2007).
92. Sainio, K., Grandström, M. L., Pettay, O. & Donner, M. EEG in neonatal herpes simplex encephalitis. *Electroencephalogr. Clin. Neurophysiol.* 56, 556–561 (1983).
93. Vignaendra, V., Ghee, L. T. & Chawla, J. EEG in brain abscess: Its value in localization compared to other diagnostic tests. *Electroencephalogr. Clin. Neurophysiol.* 38, 611–622 (1975).
94. Nuwer, M. R., Hovda, D. A., Schrader, L. M. & Vespa, P. M. Routine and quantitative EEG in mild traumatic brain injury. *Clinical Neurophysiology* vol. 116 2001–2025 (2005).
95. Mantua, J. *et al.* A systematic review and meta-analysis of sleep architecture and chronic traumatic brain injury. *Sleep Medicine Reviews* vol. 41 61–77 (2018).
96. Ianof, J. N. & Anghinah, R. Traumatic brain injury: An EEG point of view. *Dement. e Neuropsychol.* 11, 3–5 (2017).
97. Chen, H., Wang, J., Mao, S., Dong, W. & Yang, H. A new method of intracranial pressure monitoring by EEG power spectrum analysis. *Can. J. Neurol. Sci.* 39, 483–487 (2012).
98. Connolly, M., Vespa, P., Pouratian, N., Gonzalez, N. R. & Hu, X. Characterization of the Relationship Between Intracranial Pressure and Electroencephalographic Monitoring in Burst-Suppressed Patients. *Neurocrit. Care* 22, 212–220 (2015).
99. Munari, C. & Calbucci, F. Correlations between intracranial pressure and EEG during coma and sleep. *Electroencephalogr. Clin. Neurophysiol.* 51, 170–176 (1981).
100. Cobb, W. ., Guiloff, R. . & Cast, J. Breach rhythm: The EEG related to skull defects. *Electroencephalogr. Clin. Neurophysiol.* 47, 251–271 (1979).
101. Bauer, G., Trinka, E. & Kaplan, P. W. EEG patterns in hypoxic encephalopathies (post-cardiac arrest syndrome): Fluctuations, transitions, and reactions. *J. Clin. Neurophysiol.* 30, 477–489 (2013).
102. Jørgensen, E. O. & Holm, S. The natural course of neurological recovery following cardiopulmonary resuscitation. *Resuscitation* 36, 111–122 (1998).
103. Murray, D. M., Boylan, G. B., Ryan, C. A. & Connolly, S. Early EEG findings in hypoxic-ischemic encephalopathy predict outcomes at 2 years. *Pediatrics* 124, e459–e467 (2009).
104. Hussain, E. & Nordli, D. EEG Patterns in Acute Pediatric Encephalopathies. *J Clin Neurophysiol* 30, 539–544 (2013).
105. Kaplan, P. W. The EEG in metabolic encephalopathy and coma. *J. Clin. Neurophysiol.* 21, 307–318 (2004).
106. Van der Rijt, C. C. D., Schalm, S. W., de Groot, G. H. & de Vlieger, M. Objective measurement of hepatic encephalopathy by means of automated EEG analysis. *Electroencephalogr. Clin. Neurophysiol.* 57, 423–426 (1984).
107. Hughes, J. R. Correlations between EEG and chemical changes in uremia. *Electroencephalogr. Clin. Neurophysiol.* 48, 583–594 (1980).
108. Wang, D. *et al.* Comparing the effect of hypercapnia and hypoxia on the electroencephalogram during wakefulness. *Clin. Neurophysiol.* 126, 103–109 (2015).
109. Hirose, G. *et al.* Delta Waves in the EEGs of Patients With Intracerebral Hemorrhage. *Arch. Neurol.* 38, 170–175 (1981).
110. Claassen, J., Mayer, S. A. & Hirsch, L. J. Continuous EEG monitoring in patients with subarachnoid hemorrhage. *Journal of Clinical Neurophysiology* vol. 22 92–98 (2005).
111. Clancy, R. R., Tharp, B. R. & Enzman, D. Eeg in premature infants with intraventricular hemorrhage. *Neurology* 34, 583–590 (1984).
112. Trompeo, A. C. *et al.* Sleep disturbances in the critically ill patients: role of delirium and sedative agents. 77, 604–612 (2011).
113. Onoe, S. & Nishigaki, T. EEG spectral analysis in

children with febrile delirium. *Brain Dev.* 26, 513–518 (2004).
114. Van Der Kooi, A. W. *et al.* Delirium detection using EEG: What and how to measure. *Chest* 147, 94–101 (2015).
115. Sazgar, M., Young, M. G., Sazgar, M. & Young, M. G. EEG Artifacts. in *Absolute Epilepsy and EEG Rotation Review* 149–162 (Springer International Publishing, 2019). doi:10.1007/978-3-030-03511-2_8.
116. Gupta, N., Pappas, A., Thomas, R. & Shankaran, S. Reference values for three channels of amplitude-integrated EEG using the Brainz BRM3 cerebral function monitor in normal term neonates: A pilot study. *Pediatr. Neurol.* 52, 344–348 (2015).7. White, D. M. & Van Cott, A. C. EEG artifacts in the intensive care unit setting. *Neurodiagnostic Journal* vol. 50 8–25 (2010).
118. Young, B., Osvath, L., Jones, D. & Socha, E. A novel EEG artifact in the intensive care unit. *J. Clin. Neurophysiol.* 19, 484–486 (2002).
119. Markovic, A., Kaess, M. & Tarokh, L. Gender differences in adolescent sleep neurophysiology: a high-density sleep EEG study. *Sci. Rep.* 10, 1–13 (2020).
120. Zhao, D. Comparative analysis of different characteristics of automatic sleep stages. *Comput. Methods Programs Biomed.* 175, 53–72 (2019).
121. Faust, O., Razaghi, H., Barika, R., Ciaccio, E. J. & Acharya, U. R. A review of automated sleep stage scoring based on physiological signals for the new millennia. *Computer Methods and Programs in Biomedicine* vol. 176 81–91 (2019).
122. Chriskos, P. *et al.* A review on current trends in automatic sleep staging through bio-signal recordings and future challenges. *Sleep Med. Rev.* 55, 101377 (2021).
123. Hjorth, B. EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* 29, 306–310 (1970).
124. Reinke, L., van der Hoeven, J. H., van Putten, M. J. A. M., Dieperink, W. & Tulleken, J. E. Intensive care unit depth of sleep: Proof of concept of a simple electroencephalography index in the non-sedated. *Crit. Care* 18, R66 (2014).
125. Gehlbach, B. K. *et al.* Temporal Disorganization of Circadian Rhythmicity and Sleep-Wake Regulation in Mechanically Ventilated Patients Receiving Continuous Intravenous Sedation. *Sleep* 35, 1105–1114 (2012).
126. Nieuwenhuijs, D., Coleman, E. L., Douglas, N. J., Drummond, G. B. & Dahan, A. Bispectral Index Values and Spectral Edge Frequency at Different Stages of Physiologic Sleep. *Anesth. Analg.* 94, 125–129 (2002).
127. Sigl, J. C. & Chamoun, N. G. An introduction to bispectral analysis for the electroencephalogram. *J. Clin. Monit.* 10, 392–404 (1994).
128. Lee, H. C. *et al.* Data Driven Investigation of Bispectral Index Algorithm. *Sci. Rep.* 9, 1–8 (2019).
129. Nicholson, T., Patel, J. & Sleigh, J. W. Sleep Patterns in Intensive Care Unit Patients: A Study Using the Bispectral Index. *Crit. Care Resusc.* 3, 86–91 (2001).
130. Vacas, S. *et al.* The Feasibility and Utility of Continuous Sleep Monitoring in Critically Ill Patients Using a Portable Electroencephalography Monitor. *Anesth. Analg.* 123, 206–212 (2016).
131. Drover, D. & Ortega, H. R. R. Patient state index. *Best Practice and Research: Clinical Anaesthesiology* vol. 20 121–128 (2006).
132. Koolen, N. *et al.* Automated classification of neonatal sleep states using EEG. *Clin. Neurophysiol.* 128, 1100–1108 (2017).

Part II – Research report



Automated EEG-based sleep monitoring in critically ill children

F.W. (Floor) Hiemstra, MSc student Technical Medicine

Abstract

Introduction: Sleep deprivation is commonly encountered in critically ill children admitted to the pediatric intensive care unit (PICU) and is associated with poor clinical outcome. Automated electroencephalography (EEG)-based depth of sleep monitoring enables real-time continuous study of sleep in PICU patients without the need for visual assessment of the EEG signals, the gold standard. This study aims to evaluate the classification performance of various index measures and machine learning models for sleep monitoring in critically ill children.

Method: Two EEG-index-based approaches, calculated as the ratio gamma/delta and of gamma/(theta+delta) spectral powers, as well as three machine learning models - decision tree (DT), support vector machine (SVM) and extreme gradient boosting (XGBoost) - were trained and evaluated. The classification into three as well as four sleep states was evaluated. Polysomnography (PSG) recordings of 120 non-critically ill patients were used for model optimization, training and internal validation. As a proof-of-concept, the models were tested on the PSG data of 10 PICU patients.

Results: Whereas the machine learning models outperformed the index-measures in both three- as well as four-state classification in PSG recordings of non-critically ill children, the opposite was true for the PICU PSG data. Best results for PSG data of non-critically ill patients were obtained with the XGBoost model, with a 5-fold cross-validation accuracy of 0.79 ($\pm$ 0.01) for three-state classification. Performances for PICU PSG data were remarkably worse for all models. The best results for PICU data were obtained with the index-based approach (accuracy = 0.60) and the gamma/delta and gamma/(theta+delta) performed equally. The individual assessment of model performances per PICU patient revealed large variation between them.

Conclusion: A simple index measure is a promising method to monitor sleep in PICU patients. Machine learning models developed in non-critically ill patients cannot easily be applied to PICU patients in whom the sleep EEG is frequently deviant. Future efforts should focus on further tuning, training and validating the classification models with more PICU data.

1. Introduction

Sleep is a vital state of the human body that is essential to life. Although its function is not fully understood, sleep is thought to have a restorative and memory consolidative function[1]. Children admitted to the pediatric intensive care unit (PICU) are exposed to various risk factors for sleep deprivation including the noisy intensive care unit (ICU) environment, medication, morbidity and discomfort[2]. Sleep studies in PICU patients demonstrated frequent occurrence of sleep deprivation, characterized by fragmentation of sleep, reduced total sleep time, disrupted sleep architecture and a disproportional amount of sleep occurring during daytime[3–6]. Deprivation of sleep is associated with various physiological and psychological disturbances and might lead to prolonged ICU stay, increased mortality and the development of delirium[7,8].

Monitoring of sleep has the potential to reveal links between negative outcomes and sleep deprivation and may ultimately aid clinicians to optimize sleep. Sleep can be measured using polysomnography (PSG), which is a multi-parametric sleep measurement used to diagnose sleep disorders. PSG records physiological changes during sleep using the electroencephalography (EEG), electromyography (EMG), electrooculography (EOG), electrocardiography (ECG) and pulse oximetry.

Table 1. EEG frequencies.

Frequency band	Frequency range
Gamma	30-48 Hz
Beta	13-30 Hz
Alpha	8-13 Hz
Theta	4-8 Hz
Delta	0.5-4 Hz

Based on the spectral composition and features in the EEG, EMG and EOG signals, sleep is divided into rapid eye movement (REM) and non-REM sleep (NREM). In normal human sleep, NREM and REM alternate in a cyclical fashion, defining the normal sleep architecture. During REM sleep, the EEG resembles wakefulness but muscle activity is typically greatly reduced. NREM sleep is divided into N1, N2 and N3. N1 is the lightest stage of sleep, followed by N2. N1 and N2 are characterized by slowing of the EEG frequency. N3 is a deeper, more restful stage of sleep with a high arousal threshold. Characterized by its high amplitude and low frequency EEG waves, this stage is referred to as "slow wave sleep" (SWS). N1 and N2 are together referred to as "non slow wave sleep" (NSWS).

Sleep staging is traditionally done by visual analysis of these signals according to the American Association of Sleep Medicine (AASM) criteria. However, the use of the AASM criteria in characterizing sleep of critically ill patients is often debated due to confounding of the EEG signals by effects of the underlying illness and medication[9–11]. Furthermore, visual scoring of the EEG signals is time-consuming, requires skilled personnel and the subjective nature leads to considerable inter-observer variability[12]. Several attempts have been made to develop an automated sleep staging algorithm to monitor sleep of adult or neonatal patients admitted to the ICU or neonatal intensive care unit (NICU), respectively[13–19]. These algorithms use PSG signals to determine features that correlate with the depth of sleep, which is related to the previously described sleep stages, in critically ill patients. On top of the elimination of the human effort in the sleep staging process, the potential of these algorithms lies in their ability to continuously and real-time indicate depth of sleep, which is an essential prerequisite for individual optimization of sleep. Machine learning models are frequently used in sleep staging algorithms and could enable the extraction of patterns in the EEG signal that are not easily observed by eye[20]. However, they are often complex to interpret and their added value in sleep staging in critically ill patients has not been validated. A simpler method is the use of index measures that correlate with the depth of sleep. Index measures could be used for sleep stage classification by determining thresholds for the index value to distinguish the various sleep stages. The use of a simple EEG-based index based on the ratio of the gamma (30-48 Hz) to delta (0.5-4 Hz) spectral power in a single channel EEG signal to assess depth of sleep in adult ICU patients was proposed by Reinke et al.: the ICU depth of sleep (IDOS) index[15]. The dynamic properties of spectral powers in the EEG frequency ranges (Table 1) during sleep are well-known: low frequency activity (theta, delta) increases as sleep deepens, while high frequency activity (gamma, beta, alpha) increases during wakefulness[21]. Therefore, using spectral power ratio in a sleep index introduces an interpretable and simple sleep measure.

To our knowledge, automated EEG-based sleep monitoring methods to assess the sleep quantity and quality in critically ill children has not been investigated in the PICU setting. In the heterogenous population of PICU patients, not only the influences of sedative and analgesic medication and critical illness on the sleep EEG[22], but also the changes in the sleep EEG that are associated with the maturation of the brain during childhood, should be taken into account[23]. The overall aim of this study was to develop a single channel EEG-based sleep monitoring method that is able to detect the changes of depth of sleep over time in critically ill children. To achieve this, we aimed to gain insight in the potential and pitfalls of various sleep stage classification models for PICU sleep monitoring by developing and evaluating both index-measures as well as various machine learning models. Therefore, we developed and tested the various classification models on PSG data obtained from non-critically ill children, and further tested the validity of these models in PSG data from PICU patients. Although the index-based approach to classify sleep by the use of thresholds does not really meet de definition of a

'model', they are for the sake of the consistency further referred to as 'index-based models'.

2. Methods

Patients

PSG recordings of non-critically ill children were obtained from the outpatient clinical database in the period between January 2017 to June 2021. These patients were referred to the Erasmus Medical Center (MC), Sophia Children's Hospital, Rotterdam, the Netherlands, to receive a PSG for suspected or follow-up of sleep problems. Due to its retrospective nature, formal informed consent of subjects was not required. This was confirmed by the Medical Research Ethics Committee (MREC) of the Erasmus MC after review of the study protocol. To take into account the developmental changes in the sleep EEG, eight age categories were defined that globally correspond with the EEG changes during maturation[24]: 0-2 months, 2-6 months, 6-12 months, 1-3 years, 3-5 years, 5-9 years, 9-13 years, 13-18 years. For patients born preterm (<37 weeks gestational age), age was corrected until the postnatal age of 2 years. Fifteen recordings were collected for each age category, resulting in a total of 120 recordings. PSG recordings were included if the PSG showed normal physiological sleep with presence of all sleep stages and without atypical EEG findings. PSG recordings obtained from patients with severe neurological illness or who received sedative or analgesic medication were excluded. Also, PSG recordings were excluded if the hypnogram or PSG recording was incomplete or the data quality was low due to the presence of many artifacts. The PSG data obtained from the recordings of non-critically ill patients are further referred to as 'reference PSG data'.

PSG recordings of critically ill children were obtained prospectively from patients admitted to the PICU of the Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands. These patients participated in an ongoing trial in which the effect of continuous versus intermittent nutrition in PICU patients is investigated (ContInNuPIC trial, approved by the Erasmus MC MREC). As a part of this study, PSG recordings were done to investigate the effects of nutrition on circadian rhythm. Informed consent was obtained from each patient. Patients were randomized to receive either continuous or intermittent nutrition. Besides the nutrition protocol, they received normal care according to the standardized hospital protocols. All available PSG recordings were used. From all patients, age, gender, PSG or PICU indication, medical history and sedative or analgesic medication that was administered during the PSG were collected.

Data acquisition

The PSGs were performed with a standard device (Brain RT, OSG, Rumst, Belgium or Morpheus, Micromed Sp.A., Treviso, Italy) using an eight-channel EEG, two-channel EOG and EMG. EEG electrodes included the frontal (F3, F4), central (C3, C4), occipital (O1, O2) and auricular (A1, A2) electrodes and were placed according to the international 10-20 system with Ag/AgCl electrodes, sharing the same grounded electrode as reference (Fz), resulting in 8 unipolar EEG signals. Classification performance was assessed across various EEG channels to obtain the best performing EEG channel for final model development. Therefore, the following bipolar EEG channels were derived by subtraction of the unipolar electrode pairs: F3-C3, F3-C4, F3-O2, F3-A2, C3-C4, C3-O2, C3-A2, O1-O2, O1-A2. These EEG channels were chosen based on recommendations of the AASM[25] and good classification results in previous sleep staging studies in adults and neonates[22]. It was assumed that brain activity during sleep was synchronous between both hemispheres and therefore, only one of the bilaterally paired electrodes was used. The EMG electrode was placed on the submental muscle and the EOG electrodes were placed on the right and left outer canthus (ROC and LOC) of the eye, with the ROC electrode one centimeter superior and LOC one centimeter inferior of the outer canthus. The ROC-LOC channel was derived from the two EOG signals. EEG, EOG and EMG signals were sampled at 250 Hz or 256 Hz, dependent on the PSG device used. All recordings were visually scored by PSG technicians on a 30-second epoch basis according to the AASM criteria[26]. The reference PSG recordings were scored by different PSG technicians, the PICU recordings were scored by a single PSG technician. The

recordings containing the raw PSG signals together with the visually scored hypnogram were manually exported from the PSG software environment BrainRT (OSG, Rumst, Belgium) in European Data Format (EDF). Further signal analysis was performed in Python (3.9.1) using EEGlib (0.4), XGBoost (1.4.2), PyEDFlib (0.1.20), Skicit-learn (0.24.0), Scipy packages (1.6.1).

Three- and four-state classification

All models were developed for both three- as well as four-state classification (wake/NSWS/SWS or wake/REM/NSWS/SWS, respectively). Three-state classification labels were obtained from the visually scored hypnogram by combining N1, N2 and REM sleep to form NSWS, while SWS consists of N3. In four-state classification, REM is considered as separate sleep stage. For neonatal sleep, in which only two sleep stages (REM and NREM) are distinguished, NREM sleep is defined as SWS and REM as NSWS. Due to this absence of stage N1 and N2 in neonates and young infants, four-state labels were not retrieved from patients < 6 months of age. The N stage was used by the PSG technicians in PICU patients for epochs that have characteristics of NREM sleep, but could not be classified as either N1, N2 or N3 due to atypical or absent EEG characteristics. This stage is considered as SWS.

Preprocessing

All PSG signals were divided into 30-second epochs. A simple artifact detection algorithm was used to identify and label epochs than contain significant artifacts in the PSG signals. Epochs with signal amplitude exceeding a predefined threshold (most often movement or 50-Hz electrical interference artifacts) or zero activity (impedance measurement artifact) were detected and removed from the dataset (see Supplementary Methods 1 for more details). Next, a 16th order Butterworth band-pass filter was used for each PSG signal for additional artifact reduction by removing irrelevant frequencies. All EEG signals were filtered between 0.5-48 Hz, the EOG signal between 0-30 Hz and the EMG signal between 5-40 Hz[20].

Feature extraction

For each PSG signal, the signal's characteristics per epoch were mathematically described by the calculation of various features. A set of features widely used in sleep studies in adults and neonates was extracted from each epoch as potential candidates for the classification algorithm. An overview of the features used is presented in Table 2, including references to existing applications and definitions. Details and formulas of the feature calculation are provided in the Supplementary Methods 2. Fifty-one EEG features were calculated per channel for all unipolar channels of the left hemisphere (n=4) and derived bipolar channels (n=9). Time-domain features consisted of statistical features, as well as measures of signal complexity and self-similarity, such as the Hjorth parameters, fractal dimension and detrended fluctuation analysis. Twenty-five features were extracted from the frequency domain after discrete short-term Fourier transform by using a 2-second Hanning window with 50% overlap. Spectral descriptors, bandpowers and ratio of bandpowers were extracted from each epoch using the five EEG frequency bands (Table 1). To characterize non-stationary properties of EEG signals, a 5-level discrete wavelet transform (DWT) was used using a fourth-order Daubechies wavelet (see Supplementary Methods 3 for more details). Before DWT, the signals were down-sampled to 100 Hz to obtain frequency bands that are associated with the sleep EEG frequency bands. Statistical features were calculated for each of the level coefficients. EOG features were based on the presence of rapid eye movements (REMs) or slow eye movements (SEMs) in the ROC-LOC channel. Spectral bandpowers were determined from the power spectral density estimated by discrete short-term Fourier transform using a 10-second Hanning window with 50% overlap, to obtain a frequency resolution of 0.1 Hz. The characteristic changes in muscle power, and thus EMG signal amplitude, in the various sleep stages were used as EMG features. All features were standard scaled by removing the mean and scaling to unit-variance. Patient age was used as feature by using one-hot encoding for the age categories.

Data sampling and splitting strategy

Clear definitions and splitting of the data into optimization, training, validation and test sets are necessary to ensure unbiased validation and test results and correct interpretation of the results. The methods for model development and computational

Table 2. Overview of the calculated features for each epoch. EEG = electroencephalogram, EMG = electromyogram, EOG = electrooculogram, REM = rapid eye movements, SEM = slow eye movements.

Feature category	Feature description	Reference
EEG features (p=51)		
• ***Time domain (p=14)***	Statistical features: Mean of absolute amplitude, variance, zero-crossing-rate, interquartile range (25^{th}-75^{th}), signal sum, energy, kurtosis, skewness, Shannon entropy	20,27,28
	Hjorth parameters: Activity, Mobility, Complexity	29
	Higuchi fractal dimension	18,30–32
	Detrended fluctuation analysis	32–34
• ***Frequency domain (p=25)***	Spectral bandpowers: total signal power, delta, theta, alpha, beta, gamma (relative and absolute)	20,27,28
	Spectral bandpower ratio: gamma/delta, gamma/theta, beta/delta, beta/theta, alpha/delta, alpha/theta	20,27,28
	Sleep spindles: spectral bandpower 11-15 Hz (sigma)	35
	Spectral descriptors: spectral edge 95%, median and mean frequency, spectral kurtosis, spectral skewness, spectral entropy	20,27,28
• ***Time-frequency domain (p=12)***	Mean absolute value and standard deviation of coefficient amplitudes in D1, D2, D3, D4, D5 and A5 bands	30,36–38
EOG features (p=4)	Absolute spectral bandpower 0.35-0.5 Hz (REMs), 0.35-2 Hz (REMs) and 0.1-0.35 Hz (SEMs)	39, 40
	Variance	41
EMG features (p=2)	Mean absolute amplitude and energy	39
Age (p=8)	Age categories: 0-2 months, 2-6 months, 6-12 months, 1-3 years, 3-5 years, 5-9 years, 9-13 years, 13-18 years (one-hot encoding)	

costs differ between the index-based and machine learning models. Therefore, different data sampling and splitting strategies that were used are described separately below. For both, the reference PSG data was used for model optimization, training and internal validation. The test score on the PICU PSG data set is referred to as 'external validation', for which all available PICU data was used. During each cross-validation (CV), folds were made on a patient-level, meaning that the PSG data obtained from one patient will not appear in two different folds. Sampling of the data to create subsets was done via random selection of epochs. For model development and evaluation in four-state classification, subjects $<$ 6 months of age were removed from the data. The data sampling and splitting strategies are visualized in Figure 1 for index-based models and in Figure 2 for machine learning models.

Index-based models

An exploration data set was created for exploration of potential index measures by sampling 10,000 epochs from the reference PSG data. The whole reference PSG dataset was used for channel evaluation, final model training and internal validation. Channel evaluation and internal validation was done via 5-fold CV.

Machine learning models

Model exploration was done by creating learning curves using the whole reference PSG data. A learning curve is a plot of the model's classification performance for the training and validation set against the number of training samples. Next, for development of the machine learning models, the reference PSG data was split into an optimization set (two-third) and a training set (one-third). Model optimization, which comprises forward feature selection and hyperparameter tuning, was done via 3-fold CV. Within this CV, another 3-fold CV was

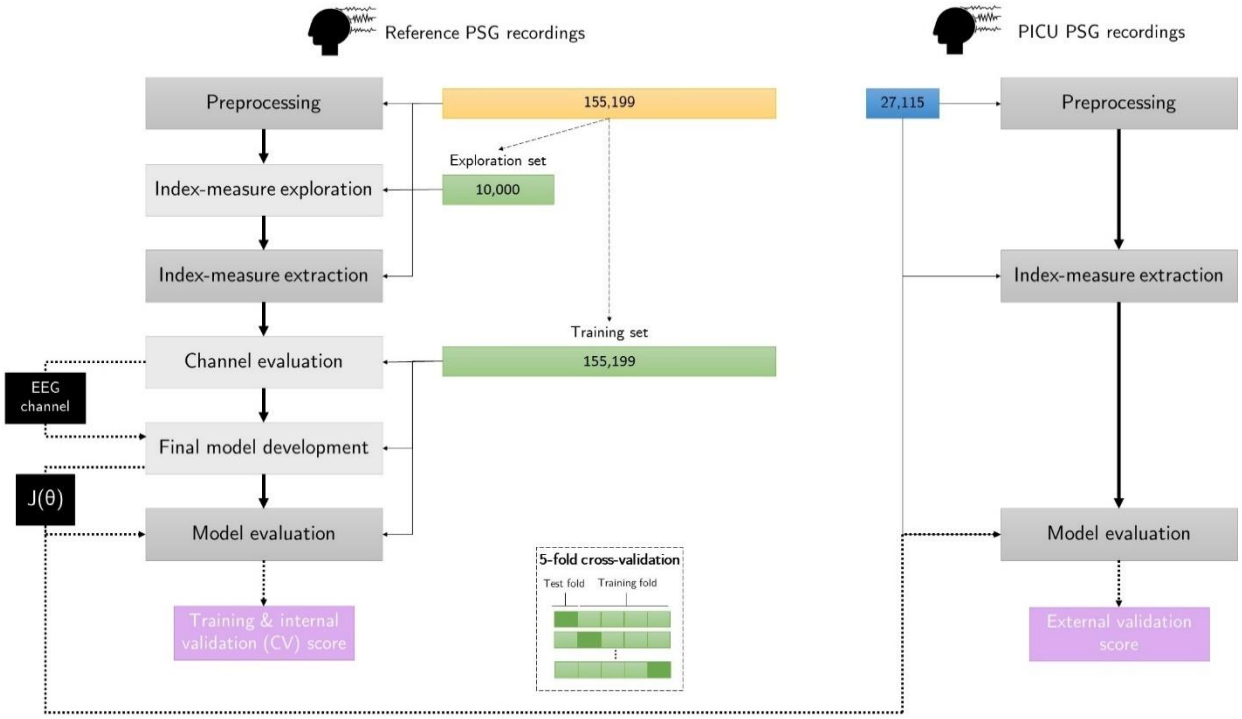


Figure 1. Method flowchart and data sampling and splitting strategy for development and evaluation of index-based models. J(θ) represents the trained model (i.e., the determined thresholds). The number in the boxes are the number of epochs in each data subset. The data splitting into various sets was done by randomly selecting epochs. CV was used during channel evaluation and internal validation. In each CV, folds were made on a patient-level. EEG = electroencephalogram, PICU = pediatric intensive care unit, PSG = polysomnography

used for forward feature selection. Channel evaluation and internal validation was done via 5-fold CV. Due to high computational costs, it was not feasible to use all data in each step in the model development. For channel evaluation, 10,000 epochs were randomly selected from the optimization set and from the training set. During final model development, optimization and training set sizes consisted of 50,000 epochs. Model optimization of the SVM and XGBoost models required significantly more run time than for the other models. In order to stay within feasible run times, 50% of the optimization sets was used during channel evaluation and final model development of SVM and XGBoost. The training set sizes were equal for all models.

Model development

Index-based models

Spectral power features in the EEG frequency ranges were used to develop the index measure. To assess which spectral power ratios have potential to be used in an index measure, Spearman correlations of the absolute spectral powers with the ordinal sleep stage labels were obtained using the C3-C4 channel from the exploration data set. Multiple combinations of positively correlated frequency bands and negatively correlated frequency bands were used as ratio and its classification performances for across all EEG channels were evaluated on the exploration data set. Optimal thresholds for sleep stage classification were determined by maximizing the classification accuracy over a threshold range. The best performing indices were selected for further development. For this, training and 5-fold CV scores were obtained across all EEG channels for channel evaluation. This was done for three- and four state classification on the training data set, which comprises all reference PSG data. The thresholds obtained from the best performing channel were selected as the final model.

Machine learning models

Learning curves for various machine learning models were created for model exploration. This graphical representation of the model's learning behavior gives insight into the model performance, computational

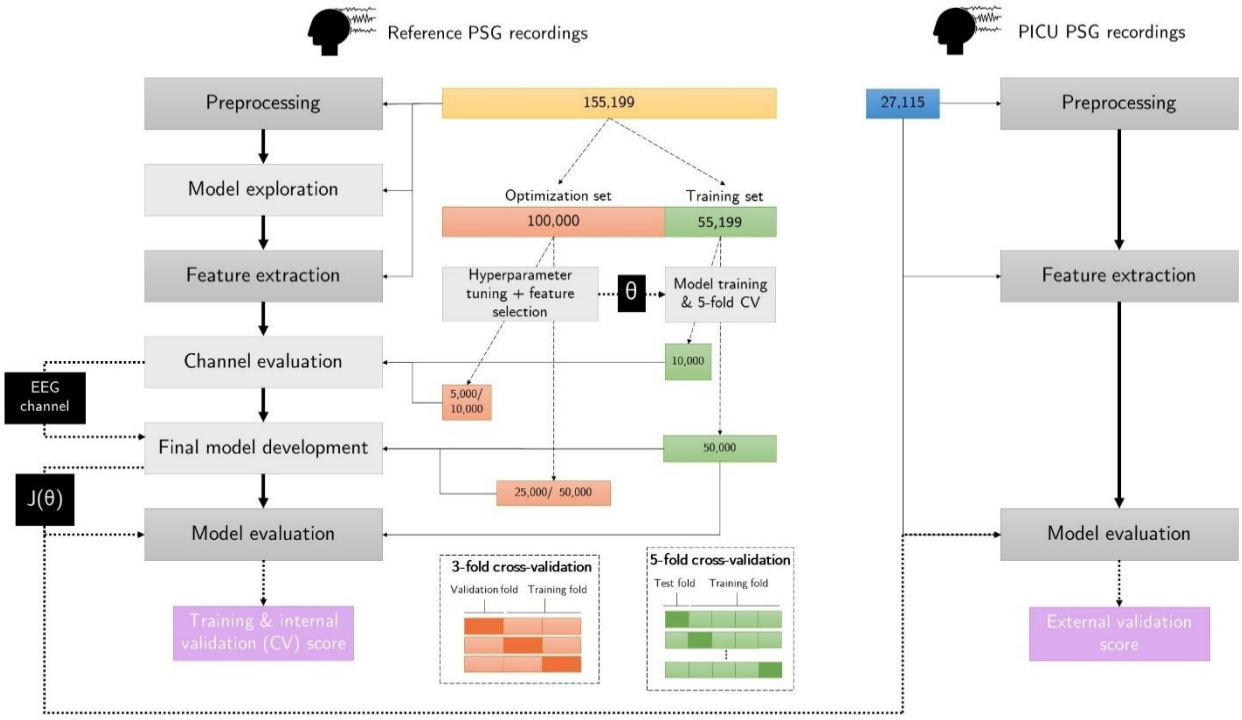


Figure 2. Method flowchart and data sampling and splitting strategy for development and evaluation of machine learning models. θ represents the hyperparameters, J(θ) represents the trained model. The number in the boxes are the number of epochs in each data subset. During model optimization (hyperparameter tuning + feature selection), 5,000 and 25,000 epochs were used for channel evaluation and final model development, respectively, for SVM and XGBoost. For DT, 10,000 and 50,000 epochs were used. The data splitting into various sets was done by randomly selecting epochs. CV was used during model optimization (hyperparameter tuning + feature selection), channel evaluation and internal validation. In each CV, folds were made on a patient-level. Not visualized in this figure is the 3-fold CV that was used for forward feature selection during model optimization. CV = cross-validation, DT = decision tree, EEG = electroencephalogram, PICU = pediatric intensive care unit, PSG = polysomnography, SVM = support vector machine, XGBoost = extreme gradient boosting

costs and convergence point, where the model will not benefit from any more training samples. Learning curves were determined for 7 machine learning models: decision trees (DT), logistic regression (LR), linear discriminant analysis (LDA), k-nearest neighbors (KNN), support vector machine (SVM), random forest (RF) and XGBoost. Age, EOG, EMG and EEG features from the C3-C4 channel were used. With a large set of features, high correlations between some features are inevitable. This results in unnecessarily large feature data sets, thereby risking the 'curse of dimensionality'. Therefore, the minimum redundancy maximum relevance (mRMR) method was used[42]. This feature selection approach tends to select features that have a high correlation with the sleep stage labels and a low correlation between themselves by using the mutual information between them. 20 features were selected from the 57 PSG features. Since age features will not likely to be selected by the mRMR method as they are not correlated with the sleep stage labels, they were added to the feature data set after mRMR feature selection. The models were trained using empirically chosen hyperparameter settings, i.e., the internal settings of the machine learning model whose value can control the learning process of the model (Table S2). The validation score was determined using 10-fold CV. Training and validation scores were expressed in accuracy. Model selection was done based on a trade-off between model performance, computational costs and convergence point.

We then compared the performance across EEG channels for channel evaluation of the three best-performing machine learning models. First, the models were optimized when using features extracted from one EEG channel at a time in combination with the EOG, EMG and age features. Again, mRMR was used to reduce the number of

features by selecting 20 from the 57 features, after which the age features were added to the data set. Next, the reduced feature subset was fed into a pipeline to find the optimal feature subset in combination with the optimal set of hyperparameters. Sequential forward feature selection (SFS) with 3-fold CV was used to further reduce the feature subset. In this feature selection method, features are sequentially added to an empty feature set until the addition of further features does not further improve classification performance, in this study expressed as accuracy. A hyperparameter space for hyperparameter tuning was defined by specifying the range and step size for each hyperparameter to be tuned (Table S3). To reduce the computational costs, only 10% of the total hyperparameter space was selected by a randomized grid search algorithm for evaluation. Each combination of hyperparameters and feature subset was trained and tested during 3-fold CV with classification accuracy as scoring metric to select the best combination. The optimal hyperparameters and feature subsets were used to train the models per channel and obtain training and 5-fold CV scores. The model was retuned on the best performing channel on a larger optimization and retrained on another training data set to develop the final model. Overviews of the methods for index-based and machine learning model development and evaluation is provided in Figure 1 and 2, respectively.

Model evaluation

Area under the receiver operating characteristic (ROC) curve (AUC), accuracy and Cohen's kappa were used as final model performance metrics for training scores and internal and external validation. In this multiclass classification problem, the AUCs were computed as the unweighted average AUCs of each class versus the rest. For the index-based models, class probabilities for AUC calculation were estimated by multiplication of the kernel density estimate for the index-value with the prior probability, for which class proportions were used. In the kernel density estimate, a Gaussian kernel was used with a bandwidth equal to 10% of the 5-95th percentile range of index values. Class probabilities from the machine learning models were retrieved using standard methods (Skicit-learn package[43]).

Additional internal validation was done by evaluation of the classification performance across different age categories for assessment of differences between age groups. In contrast to the machine learning models, the index-based models did not have age as input feature. Therefore, it was also tested whether the index-based models would provide a better fit when trained per age category separately. The same was tested with the PICU PSG data for individually trained index-based models. Next to external validation on the whole PICU data set, performance scores were also obtained per PICU patient, to be able to study the differences in performance scores between PICU patients.

3. Results

Polysomnography data and patient characteristics

From the 120 non-critically ill patients, a total of 1293.3 hours of recording was obtained, yielding 155,199 epochs (Table 3). The non-critically ill patients had a median age of 3.1 years and most of them were referred to the sleep laboratory for suspected or follow-up of airway obstruction (N = 61) or neuromuscular diseases (N = 32) (Table 4). Mean recording time per patient was 10.8 hours (standard deviation (SD): ± 1.3). 0.3% of the epochs was labelled as impedance artifact and high amplitude artifacts were present in 13.8% of all unipolar signals (Table 3). The presence of high amplitude artifacts varied across the channels, but were most frequent in the occipital and auricular electrodes (Table S7).

Ten critically ill patients admitted to our PICU between May 2020 to May 2021 were enrolled in this study. The PICU patients had a median age of 0.6 years and were admitted to the PICU for postoperative care (N = 3), cardiac failure (N = 3), sepsis (N=2), or exacerbation of neurological (N=1) or oncological disease (N = 1) (Table 4). Patient characteristics per patient are summarized in Table 5. The PSG recordings of the ten PICU patients yielded a total of 225.9 hours of PSG data (27,115 epochs), with a mean recording time of 22.6 hours (± 4.4) per patient. Impedance artifacts were present in 0.04% of the epochs and high amplitude artifacts in 31.7% (Table 3). Due to this large

Table 3. PSG data characteristics from non-critically ill (reference) and PICU patients, obtained from the visually scored hypnogram. Total sleep time is the time spent in any of the sleep stages during the PSG recording. PICU = pediatric intensive care unit, PSG = polysomnography, N1/N2/N3/N = non rapid eye movement sleep stage 1, 2, 3 or quiet sleep, respectively, REM = rapid eye movement sleep, SD = standard deviation

	Non-critically ill patients (N=120)	PICU patients (N=10)
Average length of PSG recording, hours (± SD)	10.8 (± 1.2)	22.6 (± 4.4)
Total sleep time, hours (± SD)	8.2 (± 1.4)	13.9 (± 4.2)
Impedance artifacts, % of total number of epochs (number of epochs)	0.3 (n=403)	0.04 (n=12)
High amplitude artifacts, % of total number of all unipolar epochs (number of epochs)	13.8 (n=236,463)	31.7 (n=94,482)
Mean time spent in each stage, % of total sleep time (± SD) (number of epochs)		
• REM	24.5 (± 9.0) (n=29,478)	9.9 (± 10.7) (n=1,674)
• N1	9.8 (± 7.8) n=11,308)	11.1 (± 14.8) (n=1,665)
• N2	25.3 (± 15.2) (n=29,439)	22.1 (± 23.4) (n=3,367)
• N3	30.5 (± 16.3) n=35,591)	23.5 (± 23.6) (n=3,325)
• N	9.9 (± 22.5) (n=11,851)	33.4 (± 37.9) (n=6,706)

amount of high amplitude artifacts and limited amount of PICU data, the PICU PSG data was manually checked to review the artifact labels. 50-Hz electrical interference was the main cause of the artifacts and was frequently present in one channel for the entire recording. In one of the PICU recordings (patient D), a baseline drift in some of the EEG signals (F3, F4, A1, A2 and O1) was observed. This baseline drift resulted in falsely detected artifacts, as a consequence of the increased mean absolute amplitude. However, it was observed that the baseline drift was filtered out after applying the 0.5-48 Hz bandpass filter and the remaining EEG signals were of high quality. Therefore, the artifact labels from the drifted EEG signals were removed. Detailed patient, PSG and artifact characteristic per age category for the reference group and per patient for the PICU patients are provided in Table S4-S8.

Model exploration and selection

During exploration of potential index-measures, Spearman correlations between the spectral powers and the sleep stages were highest for relative gamma (ρ=0.61), beta (ρ=0.43) and delta power (ρ=-0.43) or absolute delta (ρ=-0.53) and theta (ρ=-0.44) power (Table S9). Therefore, these spectral powers were used in various combinations to construct an index measure and their classification performance was tested on the exploration data set for three-state classification. Best results were obtained with the gamma to delta ratio (CV accuracy = 0.71) and gamma to theta+delta ratio (CV accuracy = 0.72) (Table S10). These ratios were therefore selected for further model development. For the machine learning models, the learning curves showed similar classification performances for the RF, LDA and LR model and superior classification performances of the XGBoost, SVM, KNN and DT models (Figure S3). Although the convergence points of the XGBoost and DT models did not seem to be fully reached, a training set size of 50,000 samples was considered to be both reasonable and feasible. The KNN and SVM models both have high computational costs, and therefore only the SVM model was selected for further model development, together with XGBoost and DT.

Channel evaluation

Across the EEG channels, the variation in classification performances of the index-based models was within a range of 0.05 accuracy for three-state classification (Table S11). Performance

Table 4. Patient characteristics of the non-critically ill (reference) and PICU patients. PSG = polysomnography, PICU = pediatric intensive care unit, SD = standard deviation

Patient characteristic	Non-critically ill patients (N = 120)	PICU patients (N=10)
Median age (years)	3.1	0.6
Males/females (number of patients)	62/58	6/4
PSG/PICU indication, % of total patients (number of patients)	Airway obstruction: 50.8 (N=61) Neuromuscular disease: 26.7 (N=32) Pulmonary disease: 7.5 (N=9) Central sleep apnea: 7.5 (N=9) Unknown: 7.5 (N=9)	Cardiac failure: 30 (N=3) Cardiothoracic surgery: 20 (N=2) Sepsis: 20 (N=2) Abdominal surgery: 10 (N=1) Exacerbation of neurological disease (N=1) or oncological disease (N=1)
Intubated during PSG, % of total patients (number of patients)	n/a	80 (N=8)
Days prior on PICU at time of PSG (mean (± SD))	n/a	6.4 (± 4.1)
Medication during PSG (number of patients)	None	Paracetamol (N=6), benzodiazepines (N=7), remifentanil (N=2), levetiracetam (N=1), morphine (N=6), (es)ketamine (N=4)

Table 5. Detailed patient characteristics of the PICU patients, per patient. PICU = pediatric intensive care unit, PSG = polysomnography

	Patient A	Patient B	Patient C	Patient D	Patient E	Patient F	Patient G	Patient H	Patient I	Patient J
Age category	13-18 years	6-12 months	2-6 months	2-6 months	2-6 months	0-2 months	6-12 months	13-18 years	5-9 years	13-18 years
Gender	f	f	f	f	m	f	m	m	f	m
PICU indication	Neuromuscular disease	Post-resuscitation	Cardiothoracic surgery	Cardiothoracic surgery	Heart failure	Abdominal surgery	Heart failure	Sepsis	Respiratory obstruction	Sepsis
Intubated, during PSG	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Days prior on PICU, at time of PSG	3	6	7	7	14	2	1	3	10	11
Medication during PSG	Remifentanil	Lorazepam, levetiracetam	Morphine, midazolam, ketamine	Morphine, midazolam, ketamine, paracetamol	Midazolam, paracetamol, remifentanil	Paracetamol	Morphine, midazolam, paracetamol	Morphine, midazolam	Morphine, midazolam, ketamine, paracetamol	Morphine, ketamine, paracetamol

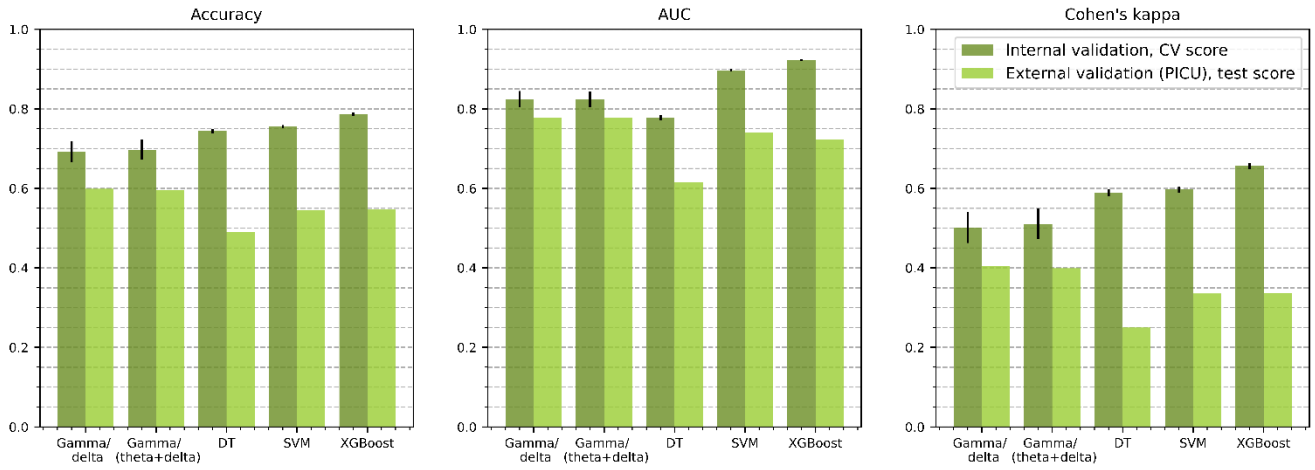


Figure 3a. Classification performances for three-state classification.

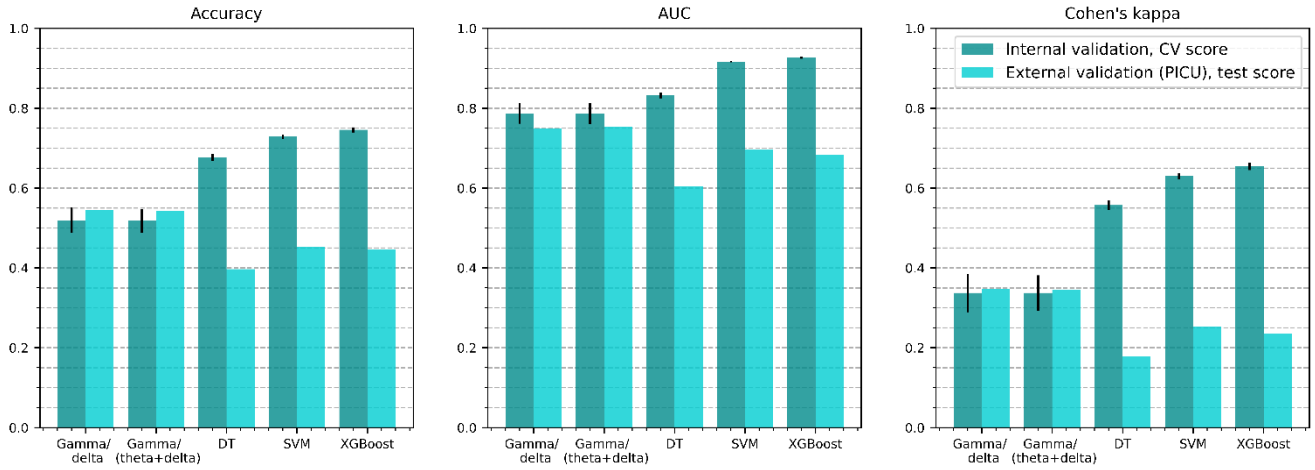


Figure 3b. Classification performances for four-state classification.
Internal validation scores were determined on the training data set (reference PSG data sets) with 5-fold CV. External validation scores were determined on the PICU PSG data set. The black bars represent the standard deviation around the mean of the CV scores. AUC = area under the receiver operating characteristics (ROC) curve, CV = cross-validation, DT = decision tree, PICU = pediatric intensive care unit, SVM = support vector machine, XGBoost = extreme gradient boosting

variation between all EEG channels of the DT and XGBoost models was even smaller and stayed within a range of 0.04 and 0.02, respectively (Table S12). In all, the frontal and central channels and their derivations performed slightly better than the occipital and auricular channels. Interestingly, the classification performance of the SVM model showed more variation between the channels and varied from an accuracy of 0.52 to 0.75 for three-state classification. The lowest performances with the SVM model were obtained in the F3-O2 and A1, C3 channels. For all models, the variation in performances between channels slightly increased in the four-state classification models. However, the overall link between individual EEG channels and performance did not differ between classification into three or four states. The F3-C3 channel was chosen as EEG channel for final model development, as this channel performed consistently superior to other channels in all models.

Internal validation on reference PSG data

Classification performance of the gamma/delta and gamma/(theta+delta) final models was very close for both three- as well as four-state classification (Figure 3, Table S13). With internal validation, the final gamma/delta model achieved an accuracy of 0.69 (± 0.03) for three-state classification. Performance of the final gamma/(theta+delta) model was marginally better (0.70 (± 0.03)). Four-state classification index

-based models performed significantly worse (0.52 (± 0.03) for both index-based models). Interestingly, training and CV scores were very close. With internal validation, all machine learning models outperformed the index-based models, most significant in four-state classification. The final XGBoost model performed best for both three-state and four-state classification, followed by the SVM model. Three-state classification accuracies for DT, SVM and XGBoost were 0.74 (± 0.00), 0.76 (± 0.00) and 0.79 (± 0.00), respectively. Four-state classification performances were slightly worse, although its difference with three-state classification performance was significantly smaller than for the index-based models (for DT, SVM and XGBoost respectively: 0.68, 0.73 and 0.75). While training and CV scores of the SVM model were very close, the gap between them was larger for the DT and XGBoost models, characterizing their overfitting risk. For all models, Accuracy, AUCs and Cohen's kappa scores for internal validation are provided in Table S13.

External validation on PICU data

From the 10 PICU PSG recordings, two contained 50-Hz electrical interference artifacts in 100% of the EEG signal in either the F3 or C3 channel, rendering them unusable for analysis (Patient G and I, Table S8). Therefore, the F4-C4 channel was used in these recordings, based on the assumption that brain activity is synchronous, and thus the EEG signal in the F3-C3 channel is similar to the EEG signal in the F4-C4 channel. Patient C, D, E and F were aged under 6 months, and therefore, four-state classification might not be applicable in these patients. However, in patient C and E, more than three sleep stages were distinguished. Hence, patient C and E were considered to be applicable for four-state classification.

In all models, classification accuracy for the pooled PICU data decreased relative to the internal validation scores (Figure 3, Table 13). In contrast to during internal validation, the index-based models now outperformed the machine learning models in both three- as well as four-state classification. Gamma/delta and gamma/(theta+delta) models performed similar (0.60 for three-state classification and 0.55 for four-state classification). The DT models performed worse of all models with an accuracy of 0.49 in three-state and 0.40 in four-state classification. Whereas the XGBoost model had outperformed the SVM model during internal validation, performances in external validation are comparable (0.54 and 0.55 for XGBoost and SVM respectively in three-state and 0.45 for both in four-state classification).

With individual model assessment, it was observed that the classification performance had a strong variation between PICU patients (Figure 4, Table S14 & S15). The DT model performed consistently worse across all patients, concordant to the pooled results. Between the index-based, SVM and XGBoost models, the best performing models varied per PICU patient and per three- or four-state classification task. Whereas the accuracy, AUC and Cohen's kappa provide inside into the overall model performance, visual assessment of the agreement between the predicted hypnogram and visually scored hypnogram is essential to gain insight into the performance per sleep stage and stability of the models. In Figure 5, the agreement of gamma/delta ratio and the XGBoost predicted hypnogram with the visually scored hypnograms for three-state classification in PICU patient A are provided. This figure shows good agreement of the gamma/delta ratio and XGBoost model for all stages, although both show instability of the model, observed as the noisy index signal and frequent state transitions in the predicted hypnogram. The importance of visual assessment of the model results is emphasized when assessing the agreement of gamma/delta with the hypnogram in patient B (Figure 6). The wake stage was never predicted by the gamma/delta model as the index signal never exceeds the wake-NSWS thresholds. This finding could not have been noticed when only looking into the accuracy, AUC and Cohen's kappa. In 6 patients, the thresholds – especially the wake-NSWS thresholds - were too high relative to the index signal amplitudes (patient A, B, D, E, F, J), suggesting decreased gamma/delta and gamma/(theta+delta) ratios in these patients (Figure S5, S6, S10, S11, S20, S21, S25, S26, S30, S31, S50, S51). In patient H, only wake and N (or undefined/atypical) sleep was distinguished. The agreement of the index measures and predicted hypnograms with the visually scored sleep labels were extremely poor is this patient (Figure S40-S44).

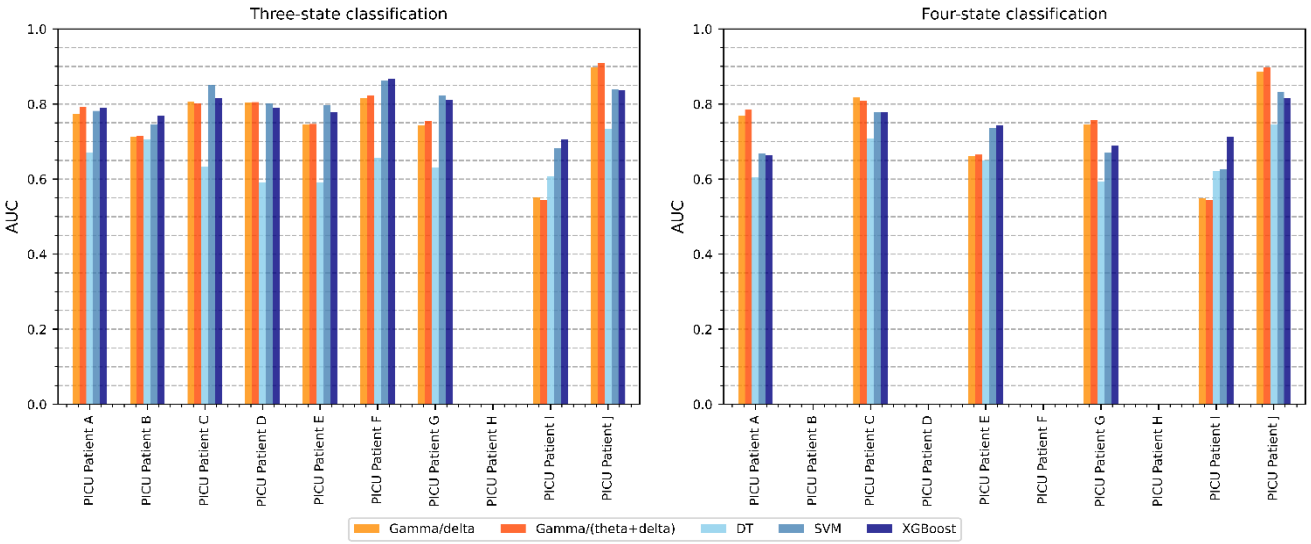


Figure 4. Classification performance (AUC) per PICU patient. AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, SVM = support vector machine, XGBoost = extreme gradient boosting.

Individualizing the thresholds per PICU patient improved the classification performance (Figure S4, Table S16), although its effects were minor. The mean accuracy improved from 0.59 to 0.62 (min-max range: 0.30 - 0.87)) for both indices when using individualized thresholds instead of the final model thresholds, the AUC's improved from 0.76 to 0.79 (0.70 – 0.96) for gamma/delta and to 0.80 (0.69 – 0.92) for gamma/(theta+delta) for three-state classification. However, the improvement in performance when using individualized thresholds differed per patient. Plots of the agreement between the visually scored hypnogram and the index-measures with final model thresholds and with individualized thresholds, and the by the machine learning models predicted hypnograms are provided in Figure S5-S54.

Validation per age category

Between age categories, the classification performance in the training data set showed strong variation (Table S17 and S18). This variation was most stand out in the index-based models, where accuracy ranged from 0.58 to 0.74 (gamma/delta, three-state classification). Performance was consistently worse in the recordings obtained from the youngest patients, 0 – 2 months and 2-6 months. Remarkably, the DT model seemed to handle the variation between ages better than the other machine learning models, with only a 0.02 range in accuracy. Model fits of the index-based models were trained per age category separately for three-state classification. However, accuracies did not remarkably improve when using the thresholds determined per age category separately compared to when using the final model thresholds. The finding that the per age category determined thresholds did not improve performance was concordant with the nonlinear relation and small variation in thresholds between age categories (Figure S55), although the thresholds for patients < 6 months seemed to be slightly deviating from the other thresholds. Also, the high inter-subject variation could have contributed to the insensitivity to the per age category determined thresholds. Variation between absolute gamma, theta and delta power between age categories across sleep stages is visualized in Figure S56.

Feature evaluation

High correlations between features were found, thereby highlighting the importance of the feature selection (Figure S57). The features that were selected by the mRMR and sequential forward feature selection and used in the final models did not show high variation between the models (Figure 7, Table 6). DWT features from all decomposition levels were frequently selected in all models. Also, several age features were always used in the final models, with

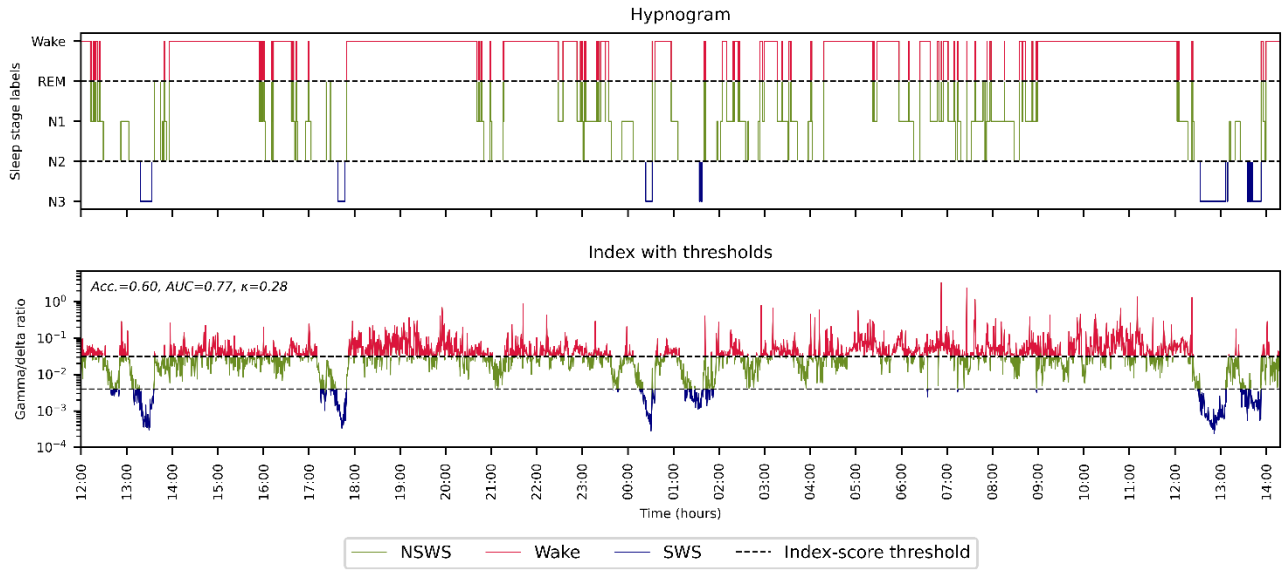


Figure 5a. Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model thresholds) for three-state classification in PICU patient A.

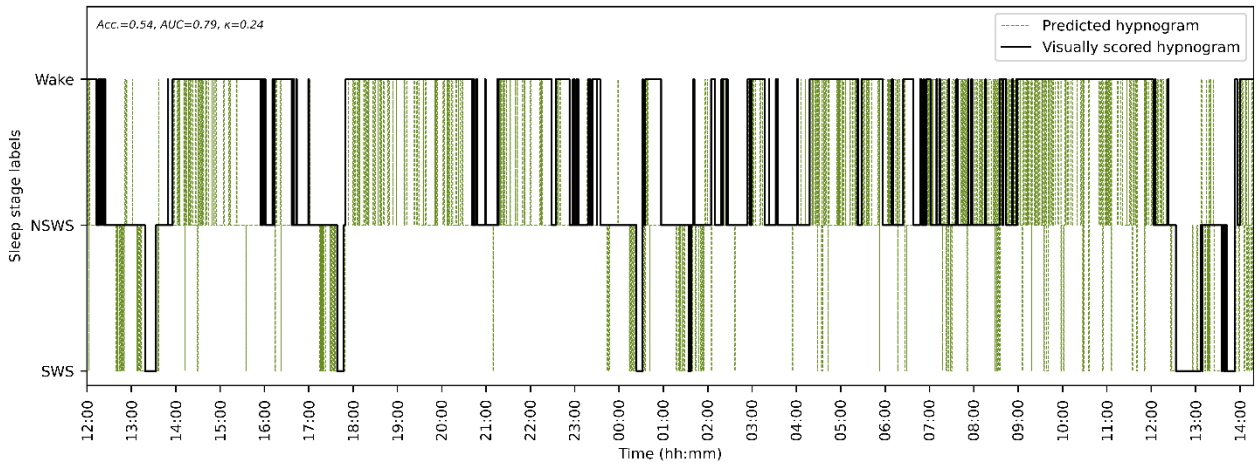


Figure 5b. Agreement of with the visually scored hypnogram with the hypnogram predicted by the XGBoost model for three-state classification in PICU patient A.

Visual assessment of the agreement of the predicted hypnogram and visually scored hypnogram provides information on the performance per sleep stage and stability of the model. The gamma/delta shows good agreement with the hypnogram for all stages. It could be observed that the index variation is high, resulting in a noisy signal. The agreement of the XGBoost predicted hypnogram with the visually scored hypnogram is reasonable. The instability and inaccuracy of the model could be observed by the frequent transition between stages that are not concordant with the visually scored hypnogram. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, PICU = pediatric intensive care unit, REM = rapid eye movement sleep, SWS = slow wave sleep

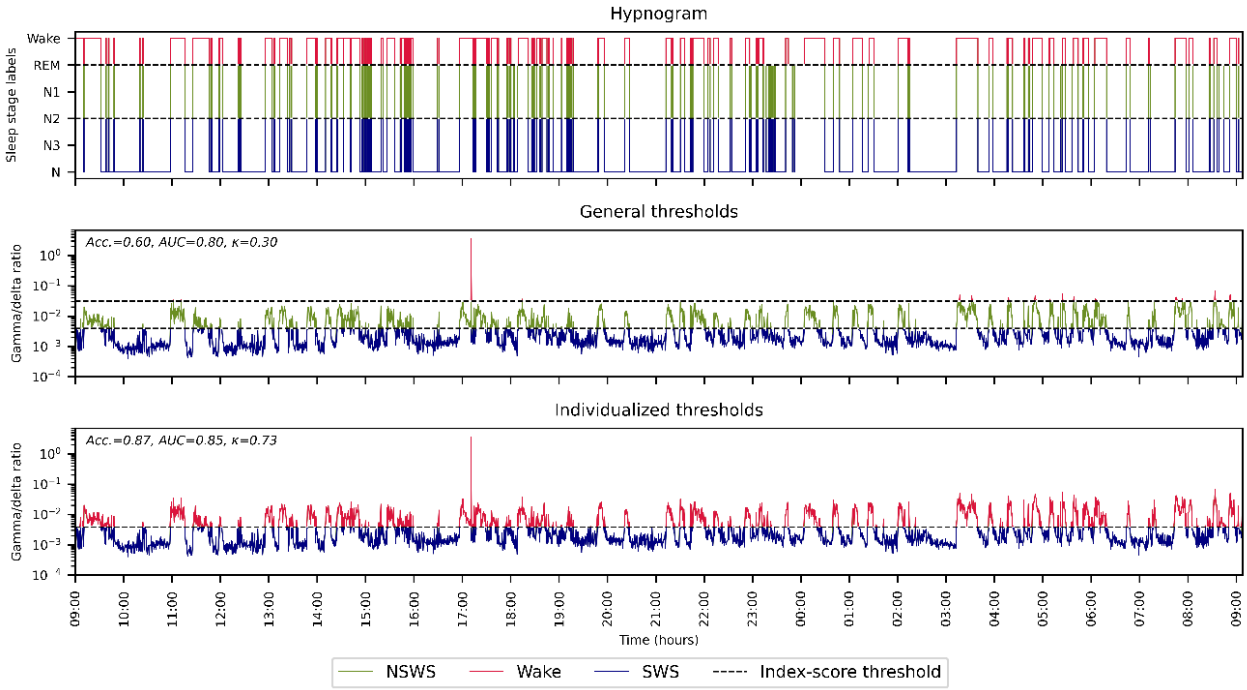


Figure 6. Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification in PICU patient D. This figure shows the importance of visual assessment and evaluation of performance per sleep stage. Although the accuracy of classification by the general thresholds is still reasonable, it could be observed that the wake stage is seldom predicted as the index signal seldom exceeds the above threshold. The AUC and Cohen's kappa might therefore provide a better impression of the model performance. Individualizing the thresholds significantly improves the classification performance. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, PICU = pediatric intensive care unit, REM = rapid eye movement sleep, SWS = slow wave sleep

Table 6. Selected features for the SVM model. abs. = absolute, DWT = discrete wavelet transform, EMG = electromyogram, rel. = relative, SD = standard deviation, SVM = support vector machine

Three-state classification		Four-state classification	
Abs. mean amplitude	DWT cA SD	Age 1-3 years	DWT cD1 SD
Age 0-2 months	DWT cD4 mean	Age 13-18 years	DWT cD4 mean
Age 1-3 years	DWT cD4 SD	Age 6-12 months	DWT cD5 mean
Age 13-18 years	DWT cD5 mean	Age 9-13 years	EMG Chin: Abs. mean amplitude
Age 2-6 months	DWT cD5 SD	Abs. Delta power	Hjorth mobility
Age 3-5 years	Gamma/delta ratio	Abs. mean amplitude	Interquartile range
Age 5-9 years	Gamma/theta ratio	Abs. Sigma power	Rel. Gamma power
Age 6-12 months	Interquartile range	DWT cA mean	Spectral edge
Age 9-13 years	Rel. Gamma power	DWT cA SD	Spectral entropy
DWT cA mean	Signal sum	DWT cD1 mean	Zero crossing rate

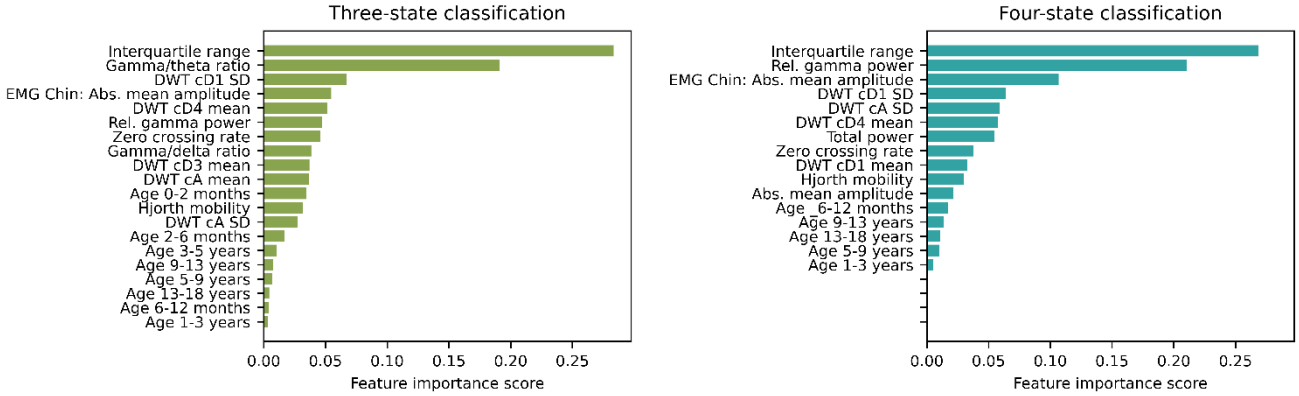


Figure 7a. Feature importances for DT models.

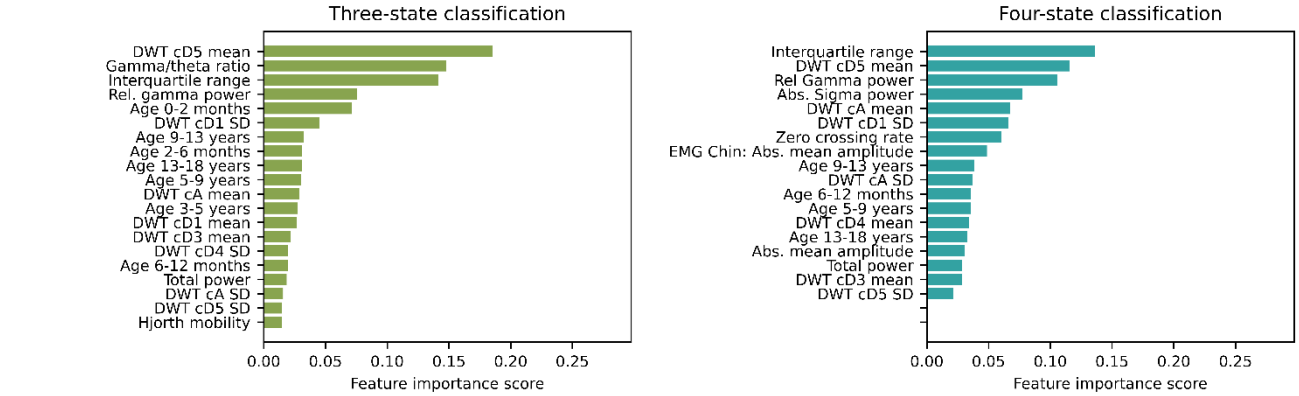


Figure 7b. Feature importances for XGBoost models.

Feature importance scores were calculated as the total reduction of the Gini criterion brought by that feature, weighted by the number of samples the node is responsible for. For the XGBoost model, the feature importances are averaged across all of the trees within in the model. Feature importances for SVM models with radial basis function kernels are more complicated to extract due to the transformation to an infinite-dimensional feature space and are therefore not provided. The corresponding frequency ranges for each DWT level (cD1-cD5, cA) is provided in Table S1. abs. = absolute, DWT = discrete wavelet transform, EMG = electromyogram, rel. = relative, SD = standard deviation, SVM = support vector machine, XGBoost = extreme gradient boosting

the whole range of age categories being selected. Interquartile range and absolute mean amplitude were popular features in the time-domain and gamma, theta and delta powers in the frequency domain. Measures of complexity or EOG features were never selected. In all four-state classification models, the beta or sigma power and EMG absolute mean amplitude were selected, which were seldom selected in the three-state classification models. Feature importance scores for the DT and XGBoost model are provided in Figure 7.

4. Discussion

This study demonstrated the potentials and challenges of automated sleep monitoring methods in the PICU by proposing various classification methods and evaluating them on both non-critically ill as well critically ill children. Although the results are explorative, they seem to be promising and should stimulate further investigation in classification methods to study PICU sleep. Automated monitoring of sleep in the PICU eliminates the need for human intervention in the sleep scoring process, resulting in a cost-effective, objective method. Next to facilitating future large scale PICU sleep research, the real-time and continuous application opens up possibilities to individually monitor and optimize sleep in clinical

practice. Despite the challenges introduced by the complexity of sleep in PICU patients, the detrimental effects on clinical outcome of sleep deprivation emphasize the need for future developments in PICU sleep monitoring to enable individual optimization of sleep.

Our classification results are in line with the reported inter-rater variability for the visual assessment of sleep for sleep classification in critically ill adults, with Cohen's kappa statistics ranging from 0.52[14] to as low as 0.19[44]. To our knowledge, this was the first attempt to develop an automated sleep monitoring method for PICU patients. Therefore, results of this study could not be compared with results on similar datasets. In general, classification results reported in automated sleep staging studies have a very large variation as they strongly depend on the data set, the classification task (i.e., number of stages to distinguish) and validation methods and measures used. Previous studies reported classification accuracy of sleep staging algorithms in healthy adults up 89% with DT and 94% with SVM models[17,4520,28]. Our classification results are inferior to the ones reported in neonatal sleep staging studies, that obtained classification accuracies of 75[17,45] up to 93%[46,47]. Attempts to study the automated monitoring of sleep in adult ICU patients have been limited. As far as we know, no studies were published that mention the results of machine learning methods to classify sleep in critically ill adult patients. The use of index-based models for ICU sleep staging was proposed by Reinke et al[15]. Their IDOS method, based on the gamma to delta spectral power, achieved average agreement, expressed as Cohen's kappa, in non-critically ill outpatient recordings of 0.82. In ICU patients, Cohen's kappa's ranged from 0.46 to 0.90. Although our gamma/delta model was based on the same spectral power ratio as the IDOS index, there are some import differences that might declare our inferior results. First, the IDOS method uses individually determined thresholds. As our results show, the use of individually determined threshold remarkably improved the classification performances of the index-based models. Another importances difference between our gamma/delta model and the IDOS method are the additional smoothing filter that was used to smooth the IDOS signal before classification. Although the influence of the smoothing is hypothesized to be minor, the noisy gamma/delta ratio (Figure 5a, 6, S5-S55), that was calculated on a 30-second epoch basis might have contributed to the inferior results.

Whereas in this study the machine learning models outperformed the index-based models in both three- as well as four-state classification in PSG recordings of non-critically ill children, the opposite was true for the PICU PSG data. This implicates that the machine learning models generalize not well to the PICU data as they have a higher risk of overfitting on the training data. The higher risk of overfitting was also observed when comparing the gap between the training and internal CV scores of the index-based and machine learning models. This was most dominant in the DT models, which had consistently worse external validation performances. The SVM and XGBoost models had similar performances on the PICU data. A practical disadvantage of the SVM is the high computational power that is required to train the models. The XGBoost models is very fast to train, but requires extensive hyperparameter tuning as there are many hyperparameters to tune. In this study, only a few of them were tuned. In the exploration phase of this study, several combinations of index-measures were tested with superior results of the gamma/delta and gamma/(theta+delta) ratios. It was hypothesized that the addition of theta power might increase the classification performance in children due to the increased theta powers in the pediatric EEG[48]. However, the classification results of the gamma/delta and gamma/(theta+delta) were very similar. Also, when looking to the index over time (Figure S5-S54), the signals are almost perfectly correlated. Figure S56 shows that the theta power is relatively small relative to the delta power, which might declare its limited contribution. Further finetuning of the bandpower ranges and smoothing could improve the classification performance of the index measures.

The physiology of sleep in ICU patients is complex which challenges the measurement of ICU sleep. The EEG is just a biomarker of the underlying sleep state that might not always reflect the underlying sleep state. The EEG resembles electrical activity arising from various processes in the brain. Whereas the analgesic and sedative medication often cause slowing

of the EEG, the critical illness itself might also interfere with the EEG and introduces confounding derangements[22]. This results in frequent atypical findings in the EEG of critically ill patients, including polymorphic and increased delta activity, burst suppression, isoelectric EEG and absence of K-complexes and sleep spindles[11,14,49]. Clearly, these EEG atypical findings complicate the sleep staging process. Despite these challenges, PSG remains the golden standard in the assessment of sleep. However, the high reported inter-rater variability – and thus low Cohen's kappas – could question whether the comparison of our proposed classification models with the visually scored hypnogram is reliable.

The variation in classification performance among different PICU patients highlights the importance of individual model assessment in PICU sleep studies rather than combining PSG data of all PICU patients together. The heterogenous PICU population introduces a large variation in sleep EEGs that are affected by various factors. In this study, the severity of illness and dose of administered drugs were not taken into account. Future studies should link classification performance on patient characteristics to ultimately be able to identify patients for who sleep monitoring might be effective. PICU patients whose EEGs exhibit normal sleep EEG characteristics might benefit more from machine learning models, while for other PICU patients the use of a coarser approach as the index-measure is more sufficient to measure sleep. An individually trained model would provide optimal performance results, which was observed when using the individualized thresholds for the index measures. However, this approach might not be realistic for implementation in clinical practice. Next to the identification of patients for who sleep monitoring might not be effective, the exploration of calibration tools to initialise the monitoring method and evaluate its effectiveness is worth considering.

The effectiveness of classification of sleep in ICU patients with severe critical illness or high doses of sedative or analgesic medication into discrete classes has often been discussed[11,14,15,49]. The use of an additional sleep stage, 'atypical sleep', has therefore been proposed by various authors[11,14,49]. This atypical sleep stage could be compared with the 'N' label that was used in our study, although this 'N' label lacked clear definitions. This label was used in sleep of PICU patients for epochs that have characteristics of NREM sleep, but could not be classified as either N1, N2 or N3 due to atypical or absent EEG characteristics. The decision to combine the 'N' label with N3 to form SWS in our three-state and four-state classification approach is debatable. Therefore, future effort should focus on clear definitions of the 'N' label and the efficiency to use the conventional sleep stage definitions for sleep monitoring in PICU patients. As discussed before, in some of the PICU patients, the visual assessment of the index-measure over time might be more suitable to indicate depth of sleep. In previous studies, the bispectral index (BSI) and patient state index (PSI) have been proposed to assess depth of sleep over time in critically ill adults[13,50]. Both indices are calculated by commercially available algorithm based on several EEG features[51,52]. Although their correlation with depth of sleep in critically ill adults seems promising, the nondisclosure of their calculating algorithm is a big drawback as it complicates the interpretation of unexpected values or trends and identification of factors influencing the measure. The added value of the index-measures proposed by Reinke et al.[15] and in this study lies in its easy interpretability and computational simplicity.

In this study, multiple performance metrics were used to enable more insight in the classification model performances. Accuracy was used during channel evaluation and model development, where threshold optimization, feature selection and hyperparameter tuning was performed by maximizing the accuracy. However, accuracy is sensitive to imbalanced classes and the possibility of agreement occurring by chance. The pitfalls of using the accuracy for model assessment could be observed in patient D (Figure 6). The wake-NSWS threshold of the final model was too low for the specific patient. However, accuracy was still relatively high as the presence of wake epochs was high (Figure S15). This example also shows the effect of maximizing the thresholds on the accuracy with imbalanced classes. Since the NSWS stage are seldom present, the wake-NSWS and NSWS-SWS thresholds are very close. The use of Cohen's kappa as performance metric to maximize during model optimization and development could have resulted in different models. Cohen's kappa corrects for the

possibility of agreement by chance and is thus more suitable for class-imbalanced problems. Also, obtaining the AUC has additional value next to the accuracy and Cohen's kappa as it is a measure of the overall performance of the classifier rather than the performance at a fixed threshold. The AUC represents the trade-off between sensitivity and specificity for different thresholds on class probabilities that are predicted by the classifier and is insensitive to imbalanced classes. The AUC is more suitable for imbalanced classes as it is the unweighted average of the AUC of each class versus the rest. Although the contingency table and performance per sleep stage separately per patient provides valuable information, this approach is undesirable as it results in overloads of result data. Therefore, during classification model development, it is important to carefully choose and interpret the model performance metrics and take into account how the metric is influenced by imbalances in predicted labels.

The goal of this study was to use a single channel EEG for sleep stage classification. Single channel EEG has the benefit to have low complexity for both practical as well as computational considerations. The PSG recordings that were performed using a multichannel EEG setup enabled the extensive channel evaluation. In general, the variation in classification performance of all models between the various EEG channels was very small. The frontal and central channels and their derivations performed consistently well, while the auricular and occipital channels lagged behind. Whereas across all channels, performance in three-state classification was better than with four-state classification, the overall link between individual EEG channels and performance did not differ between classification into three or four states. This suggests that the discriminative performance of features that are characteristic for REM sleep did not differ among EEG channels, and similar channels could be used for three- and four-state classification problems. Next to the classification performance of the EEG channels, the position of the EEG electrodes and their relative risk to introduce artifacts should be considered. Most artifacts were present in the occipital and auricular electrodes, which is not surprising as their location increases the risk of loosening of the EEG electrodes or movement artifacts. The all or nothing functionality that comes with the single channel approach highlights the importance of this practical consideration.

The results of this study show that age does not seem to influence the model performance remarkably. In the machine learning models, the age feature was frequently selected. This suggests that age has a significant influence, although this was not consistent with the findings in performance variation between age categories in the index-based models. The influence of age should be further investigated, as the sleep EEG varies considerably as children mature with a remarkable development in continuity of EEG activity, dominant frequency, presence of typical patterns and waveforms and differentiation in sleep stages occur, most dominant in the first months of life[24].

EEG recordings in ICU patients are more prone to artifacts arising from the noisy ICU environment and 50-Hz electrical interference with electrical equipment simultaneously used in the ICU[53]. This was also observed in our PICU data, which contained significantly more artifacts than the reference PSG signals. Most of these artifacts were caused by 50-Hz electrical interference. This is in contrast to the artifacts in reference PSG signals, which are mostly caused by movements. Also, ECG artifacts were often present. Both ECG and 50-Hz artifacts are associated with increased electrode impedances. In our study, the PSG electrodes were – in order to reduce patient discomfort - often not attached to the skin with clay in the PICU patients, which was standard practice in the reference PSG recordings. This might have contributed to increased impedances, and thus artifacts, in combination with the long duration of the recordings. It was also observed that impedance artifacts were less frequently present in the PICU PSG recordings compared to the reference PSG recordings, suggesting that impedances were less frequently checked. In future studies, attention should be paid to the execution of the PSG recordings in the PICU in order to increase the quality of the data. A limitation in the artifact handling was the simplicity of the artifact detection algorithm that was used. This could have led to – potentially avoidable – missed artifacts or removal of lots of data, as was the case in PICU patient D. The use of more

sophisticated artifact removal methods could have improved the accuracy of the artifact removal[20,28]. Another important note on the artifact handling method is that the presence of artifacts per sleep stage was not taken into account. Since movement artifact are typically present during wakefulness, this might have introduced a bias.

A strong limitation of this study is the limited number of PSG recordings obtained from PICU patients to study the external validity of the classification models. The heterogenous PICU population requires a large amount of PICU patients to validate the models in wide ranges of patient characteristics. On top of that, the tuning range and amount of reference PSG data for model optimization and training was suboptimal due to computational limitations. A remarkable variation was found in optimal hyperparameter settings and feature subsets after repeated optimization on different optimization sets. This implies that further attention should be paid to the effect of extending the optimization data set, hyperparameter space and grid search on the final model performances. However, classification performances obtained during exploration and channel evaluation, for which smaller subsets of the optimization and training data sets were used, were comparable with the final model performances. This suggests that adding more data for model optimization and training would not drastically change the obtained classification performances.

This explorative study shows many possibilities for future efforts to improve methods for monitoring of sleep in critically ill children. In future studies, the identification of patients for who sleep monitoring is both effective as well as clinically relevant is an essential step. This could aid in specifying the requirements and focus of the monitoring methods, such as the number of states to classify and sensitivity versus specificity trade-offs. Also, the choice to discriminate into discrete stages or rather to focus on a continuous indicator of depth of sleep is important to make. Next, future efforts should focus on improving the tuning, optimization and training of the models. Increasing the computational resources to reduce run time could enable the extending of the optimization and training process. Whereas the models strongly rely on the quality of features that are calculated, attention should be paid to extend and improve the feature engineering. In addition to EEG, EOG and EMG, the use of vital parameters should be explored. The heart rate variability, for example, has been shown to reveal significant differences between NREM and REM sleep[54,55]. Furthermore, the use of other machine learning models should be studied. Models that take into account the sequential nature of sleep stages, such as convolutional neural networks and Hidden Markov models, could increase accuracy. However, the added value of the incorporation of knowledge on sleep stage transition for its application in PICU patients, whose sleep is known to have disturbed temporal structures, should be discussed. Finally, the availability of more data obtained from critically ill children could increase the external validity of the models. A larger amount of data could also enable a PICU sleep approach for the development of a monitoring method. Whereas in this study, we started with data obtained from non-critically ill patients to develop a model and evaluate the model on the PICU population, it might also be interesting to develop a model on the PICU data, either with supervised or unsupervised methods, such as clustering. In all future approaches, it is essential to also take into account the practical consideration of the proposed methods, such as the computational power and the artifact robustness. Altogether, it might be clear that considerable efforts in further investigation are required before clinical implementation. However, with this study, a first step towards automated sleep monitoring in the PICU has been made.

5. Conclusion

This study shows the potentials and pitfalls of various classification methods for automated EEG-based sleep monitoring in both non-critically as well as critically ill children. Whereas the machine learning models outperformed the index-measures in both three- as well as four-state classification in PSG recordings of non-critically ill children, the opposite was true for the PICU PSG data. A simple index measure seems to be a promising method to monitor sleep in PICU patients. However, machine learning models developed in non-critically ill patients cannot easily be applied to PICU patients in whom the sleep

EEG is frequently deviant. The variation of classification performance between PICU patients emphasizes the importance of individual assessment and identification of patient characteristics that challenge the measurement of sleep. Although the results do not encourage immediate implementation in clinical practice, they do warrant further development and testing. Future efforts should focus on further tuning, training and validating the classification models with more PICU data.

List of supplementary materials

Supplementary methods:

1. Artefact detection algorithm
   - Figure S1: EEG artifacts
2. Additional feature information
   - Figure S2: Structure of 5-level DWT transform
   - Table S1: Frequency ranges of five levels of decomposition in DWT
3. Discrete wavelet transform

Supplementary figures and tables:

- Table S2: Hyperparameter settings during learning curve determination
- Table S3: Hyperparameter space and final tuning configuration
- Table S4: Detailed patient characteristics of the non-critically ill (reference) patients, per age category
- Table S5: Detailed PSG data characteristics from the non-critically ill (reference) patients, obtained from the visually scored hypnogram
- Table S6: Detailed PSG data characteristics from the PICU patients, obtained from the visually scored hypnogram.
- Table S7: Results of the artifact detection algorithm in the non-critically ill (reference) patient data per age category.
- Table S8: Results of the artifact detection algorithm in the PICU patient data per age category.
- Table S9: Spearman correlation of spectral powers with sleep stage labels.
- Table S10: Classification performance (train scores, accuracy) of various potential index-measures constructed from the ratio of spectral powers.
- Figure S3: Learning curve for three-state classification.
- Table S11: Classification performance (CV scores, accuracy) of the index-based models across various EEG channels for three- and four-state classification.
- Table S12: Classification performance (CV scores, accuracy) of the machine learning models across various EEG channels for three- and four-state classification.
- Table S13: Classification performances of the index-based and machine learning models.
- Table S14: Three-state classification performances of the index-based and machine learning models per PICU patient.
- Table S15: Four-state classification performances of the index-based and machine learning models per PICU patient.
- Table S16: Classification performance (training score, accuracy) per PCIU patient for the individually trained models, three-state classification
- Figure S4: Classification performance (AUC) with individualized versus general thresholds per PICU patient for three-state classification.
- Figure S5: PICU patient A - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S6: PICU patient A - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S7: PICU patient A - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S8: PICU patient A - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S9: PICU patient A - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Figure S10: PICU patient B - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S11: PICU patient B - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S12: PICU patient B - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S13: PICU patient B - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S14: PICU patient B - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Figure S15: PICU patient C - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S16: PICU patient C - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S17: PICU patient C - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S18: PICU patient C - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S19: PICU patient C - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Figure S20: PICU patient D - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the

final model, or general, and individualized thresholds) for three-state classification.
- Figure S21: PICU patient D - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S22: PICU patient D - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S23: PICU patient D - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S24: PICU patient D - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Figure S25: PICU patient E - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S26: PICU patient E - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S27: PICU patient E - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S28: PICU patient E - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S29: PICU patient E - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Figure S30: PICU patient F - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S31: PICU patient F - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S32: PICU patient F - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S33: PICU patient F - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S34: PICU patient F - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Figure S35: PICU patient G - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S36: PICU patient G - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S37: PICU patient G - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S38: PICU patient G - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S39: PICU patient G - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Figure S40: PICU patient H - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S41: PICU patient H - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S42: PICU patient H - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S43: PICU patient H - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S44: PICU patient H - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Figure S45: PICU patient I - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S46: PICU patient I - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S47: PICU patient I - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S48: PICU patient I - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S49: PICU patient I - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Figure S50: PICU patient J - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S51: PICU patient J - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification.
- Figure S52: PICU patient J - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification.
- Figure S53: PICU patient J - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification.
- Figure S54: PICU patient J - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification.
- Table S17: Classification performance (training subgroup scores, accuracy) of the final index-based models per age category.

- Table S18: Classification performance (training subgroup scores, accuracy) of the final machine learning models per age category.
- Table S19: Classification performance (accuracy) of the final index-based models versus the per age category trained models, three-state classification.
- Figure S55: Variation in thresholds between age categories for the index-based models, three-state classification.
- Figure S56: Absolute spectral powers across sleep stages per age category.
- Figure S57: EEG feature correlation heatmap (F3-C3 channel).

List of abbreviations

A1/A2	- Auricular electrodes
AUC	- Area under the receiver operating characteristic (ROC) curve (AUC),
AASM	- American Academy of Sleep Medicine
C3/C4	- Central electrodes
CV	- Cross-validation
DT	- Decision tree
EEG	- Electroencephalogram
ECG	- Electrocardiogram
EMG	- Electromyogram
EOG	- Electrooculogram
F3/F4	- Frontal electrodes
ICU	- Intensive care unit
IDOS	- Intensive care unit depth of sleep
KNN	- k-nearest neighbours
LDA	- Linear discriminant analysis
LOC	- Left outher canthus
LR	- Logistic regression
MREC	- Medical Research Ethics Committee
mRMR	- Minimum redundancy maximum relevance
N1/N2/N3/N	- Non rapid eye movement sleep stages
NICU	- Neonatal intensive care unit
NREM	- Non rapid eye movement sleep
NSWS	- Non slow wave sleep
O1/O2	- Occipital electrodes
PICU	- Pediatric intensive care unit
PSG	- Polysomnography
REM	- Rapid eye movement sleep
ROC	- Right outer canthus
RF	- Random forest
SD	- Standard deviation
SFS	- Sequential forward feature selection
SVM	- Support vector machine
SWS	- Slow wave sleep
XGBoost	- Extreme gradient boosting

References

1. Schwartz, J. & Roth, T. Neurophysiology of Sleep and Wakefulness: Basic Science and Clinical Implications. *Curr. Neuropharmacol.* 6, 367–378 (2009).
2. Kudchadkar, S. R., Aljohani, O. A. & Punjabi, N. M. Sleep of critically ill children in the pediatric intensive care unit: A systematic review. *Sleep Medicine Reviews* vol. 18 103–110 (2014).
3. Cureton-Lane, R. A. & Fontaine, D. K. Sleep in the pediatric ICU: An empirical investigation. *Am. J. Crit. Care* 6, 56–63 (1997).
4. Corser, N. C. Sleep of 1- And 2-Year-old Children in Intensive Care. *Compr. Child Adolesc. Nurs.* 19, 17–31 (1996).
5. Carno, M. A. & Connolly, H. V. Sleep and sedation in the pediatric intensive care unit. *Critical Care Nursing Clinics of North America* vol. 17 239–244 (2005).
6. Al-Samsam, R. H. & Cullen, P. Sleep and adverse environmental factors in sedated mechanically ventilated pediatric intensive care patients. *Pediatr. Crit. Care Med.* 6, 562–567 (2005).
7. Ely, E. W. *et al.* Delirium as a Predictor of Mortality in Mechanically Ventilated Patients in the Intensive Care Unit. *J. Am. Med. Assoc.* 291, 1753–1762 (2004).
8. Weinhouse, G. L. *et al.* Bench-to-bedside review: Delirium in ICU patients - importance of sleep deprivation. *Critical Care* vol. 13 (2009).
9. Watson, P. L. *et al.* Atypical Sleep in Ventilated Patients. *Crit. Care Med.* 41, 1958–1967 (2013).
10. Cooper, A. B. *et al.* Sleep in critically ill patients requiring mechanical ventilation. *Chest* 117, 809–818 (2000).
11. Drouot, X. *et al.* A new classification for sleep analysis in critically ill patients. *Sleep Med.* 13, 7–14 (2012).
12. Danker-Hopfe, H. *et al.* Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *J. Sleep Res.* 13, 63–69 (2004).
13. Vacas, S. *et al.* The Feasibility and Utility of Continuous Sleep Monitoring in Critically Ill Patients Using a Portable Electroencephalography Monitor. *Anesth. Analg.* 123, 206–212 (2016).
14. Ambrogio, C., Koebnick, J., Quan, S. F., Ranieri, V. M. & Parthasarathy, S. Assessment of Sleep in Ventilator-Supported Critically Ill Patients. *Sleep* 31, 1559–1568 (2008).
15. Reinke, L., van der Hoeven, J. H., van Putten, M. J. A. M., Dieperink, W. & Tulleken, J. E. Intensive care unit depth of sleep: Proof of concept of a simple electroencephalography index in the non-sedated. *Crit. Care* 18, R66 (2014).
16. Ansari, A. H. *et al.* A convolutional neural network outperforming state-of-the-art sleep staging algorithms for both preterm and term infants. in *Journal of Neural Engineering* vol. 17 016028 (Institute of Physics Publishing, 2020).
17. Piryatinska, A. *et al.* Automated detection of neonate EEG sleep stages. *Comput. Methods Programs Biomed.* 95, 31–46 (2009).
18. Ghimatgar, H. *et al.* Neonatal EEG sleep stage classification based on deep learning and HMM. *J. Neural Eng.* 17, 036031 (2020).
19. Estévez, P. A. *et al.* Polysomnographic pattern recognition for automated classification of sleep-waking

states in infants. *Med. Biol. Eng. Comput.* 40, 105–113 (2002).
20. Chriskos, P. *et al.* A review on current trends in automatic sleep staging through bio-signal recordings and future challenges. *Sleep Med. Rev.* 55, 101377 (2021).
21. Mann, K., Bäcker, P. & Röschke, & J. Dynamical properties of the sleep eeg in different frequency bands. *Int. J. Neurosci.* 73, 161–169 (1993).
22. Hiemstra, F. W. The potential and challenges of automated EEG-based sleep monitoring in critically ill children : a literature review. *Not Publ.* (2021).
23. Olbrich, E., Rusterholz, T., LeBourgeois, M. K. & Achermann, P. Developmental Changes in Sleep Oscillations during Early Childhood. *Neural Plast.* 2017, (2017).
24. Dereymaeker, A. *et al.* Review of sleep-EEG in preterm and term neonates. *Early Hum. Dev.* 113, 87–103 (2017).
25. Duce, B., Rego, C., Milosavljevic, J. & Hukins, C. The AASM Recommended and Acceptable EEG Montages are Comparable for the Staging of Sleep and Scoring of EEG Arousals. *J. Clin. Sleep Med.* 10, 803–809 (2014).
26. Iber, A. L. C. C., Ancoli-Israel, S., Chesson, A., Quan, S. F. & Medicine., for the A. A. of S. The AASM manual for the scoring of sleep and associated events: Rules, Terminology and Technical Specification: Version 2.6 (2020). *Am. Acad. Sleep Med.* (2007).
27. Zhao, D. Comparative analysis of different characteristics of automatic sleep stages. *Comput. Methods Programs Biomed.* 175, 53–72 (2019).
28. Faust, O., Razaghi, H., Barika, R., Ciaccio, E. J. & Acharya, U. R. A review of automated sleep stage scoring based on physiological signals for the new millennia. *Computer Methods and Programs in Biomedicine* vol. 176 81–91 (2019).
29. Hjorth, B. EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* 29, 306–310 (1970).
30. Pillay, K. *et al.* Automated EEG sleep staging in the term-age baby using a generative modelling approach. *J. Neural Eng.* 15, (2018).
31. Piryatinska, A., Woyczynski, W. A., Scher, M. S. & Loparo, K. A. Optimal channel selection for analysis of EEG-sleep patterns of neonates. *Comput. Methods Programs Biomed.* 106, 14–26 (2012).
32. Cabañero-Gomez, L., Hervas, R., Gonzalez, I. & Rodriguez-Benitez, L. eeglib: A Python module for EEG feature extraction. *SoftwareX* 15, 100745 (2021).
33. Lee, J., Kim, D., Kim, I., Suk Park, K. & Kim, S. Nonlinear-analysis of human sleep EEG using detrended fluctuation analysis. *Med. Eng. Phys.* 26, 773–776 (2004).
34. Farag, A. F., El-Metwally, S. M. & Morsy, A. A. A Sleep Scoring System Using EEG Combined Spectral and Detrended Fluctuation Analysis Features. *J. Biomed. Sci. Eng.* 07, 584–592 (2014).
35. Purcell, S. M. *et al.* Characterizing sleep spindles in 11,630 individuals from the National Sleep Research Resource. *Nat. Commun. 2017 81* 8, 1–16 (2017).
36. Silveira, T. L. T. da, Kozakevicius, A. J. & Rodrigues, C. R. Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain. *Med. Biol. Eng. Comput. 2016 552* 55, 343–352 (2016).
37. Şen, B., Peker, M., Çavuşoğlu, A. & Çelebi, F. V. A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms. *J. Med. Syst. 2014 383* 38, 1–21 (2014).
38. Fraiwan, L. *et al.* Time frequency analysis for automated sleep stage identification in fullterm and preterm neonates. *J. Med. Syst.* 35, 693–702 (2011).
39. Malafeev, A. *et al.* Automatic Human Sleep Stage Scoring Using Deep Neural Networks. *Front. Neurosci.* 12, 781 (2018).
40. Zhang, Y. *et al.* Automatic sleep staging using multi-dimensional feature extraction and multi-kernel fuzzy support vector machine. *J. Healthc. Eng.* 5, 505–520 (2014).
41. Estrada, E. *et al.* EOG and EMG: Two important switches in automatic sleep stage classification. *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.* 2458–2461 (2006) doi:10.1109/IEMBS.2006.260075.
42. Peng, H., Long, F. & Dong, C. Feature selection based on mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1–6 (2005).
43. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830 (2011).
44. Elliott, R., McKinley, S., Cistulli, P. & Fien, M. Characterisation of sleep in intensive care using 24-hour polysomnography: anobservational study. *Crit. Care 2013 172* 17, 1–10 (2013).
45. Fraiwan, L. & Alkhodari, M. Neonatal sleep stage identification using long short-term memory learning system. *Med. Biol. Eng. Comput.* 58, 1383–1391 (2020).
46. Dereymaeker, A. *et al.* An automated quiet sleep detection approach in preterm infants as a gateway to assess brain maturation. *Int. J. Neural Syst.* 27, (2017).
47. Löfhede, J. *et al.* Automatic classification of background EEG activity in healthy and sick neonates Related content Automatic classification of background EEG activity in healthy and sick neonates. *J. Neural Eng* 7, 12 (2010).
48. Grigg-Damberger, M. *et al.* The visual scoring of sleep and arousal in infants and children. *Journal of Clinical Sleep Medicine* vol. 3 201–240 (2007).
49. Watson, P. L. Measuring sleep in critically ill patients: Beware the pitfalls. *Critical Care* vol. 11 (2007).
50. Nicholson, T., Patel, J. & Sleigh, J. W. Sleep Patterns in Intensive Care Unit Patients: A Study Using the Bispectral Index. *Crit. Care Resusc.* 3, 86–91 (2001).
51. Sigl, J. C. & Chamoun, N. G. An introduction to bispectral analysis for the electroencephalogram. *J. Clin. Monit.* 10, 392–404 (1994).
52. Drover, D. & Ortega, H. R. R. Patient state index. *Best Practice and Research: Clinical Anaesthesiology* vol. 20

121–128 (2006).
53. White, D. M. & Van Cott, A. C. EEG artifacts in the intensive care unit setting. *Neurodiagnostic Journal* vol. 50 8–25 (2010).
54. Long, X., Fonseca, P., Haakma, R., Foussier, J. & Aarts, R. M. Automatic detection of overnight deep sleep based on heart rate variability : a preliminary study. *2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC 2014* 50–53 (2014) doi:10.1109/EMBC.2014.6943526.
55. Uçar, M. K., Bozkurt, M. R., Bilgin, C. & Polat, K. Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques. *Neural Comput. Appl. 2016 298* 29, 1–16 (2016).

Acknowledgements

I would like to express my sincere appreciation to my supervisors for their support and effort throughout this project. Jan Willem and Rogier, thanks for your involvement and clinical and academic perspectives during my final internship. I really appreciated your positive attitudes and our discussions about technology and data science in healthcare. You both gave me the opportunity to learn and see as much as possible in the Sophia Children's Hospital, which I really enjoyed. David, even though – due to COVID-19 related restrictions - we were not able to physically meet more than once during this project, you were able to transfer your enthusiasm through my computer screen. Thanks for your help and advice on how to use data science and machine learning in this medical application. Arnout, thanks for your endless enthusiasm about sleep monitoring and data. You really took time to help me, teach me and discuss the potentials and challenges of PICU sleep monitoring. It was really fun working with you, and hopefully we will collaborate someday in the future. Next, I would like to thank Koen. You introduced me to the pediatric sleep EEG by letting me browse your collection of books. I really appreciate your clinical perspective on my project and that you are willing to be part of my assessment committee, despite the fact that you held your inaugural lecture in the same period of time. Karla and Els, your knowledge and experience on sleep staging in children and PICU patients helped me a lot, thanks! Thanks to Laurens (UMCG) for our interesting discussions on how to measure sleep in ICU patients. Thanks to all staff doctors, residents, nurses and nurse practitioners for your supervision and giving me the opportunity to learn and develop my clinical skills. Thanks to the PhD students for the fun times in room 3508. I really appreciated the relaxed and positive atmosphere in the department.

At last, I would like to thank my friends, roommates and family for being there. Daniela, it was really fun to reactivate my creative mind together during the making of the illustration of the cover design. Thanks to you and your creativity, it is the most beautiful illustration of a sleeping PICU patient one will ever see. And the final thanks to my big brother Daan, who helped me to see the possibilities of Python. I might reconsider what my favorite programming language will be.

Supplementary materials
Part I – Literature study

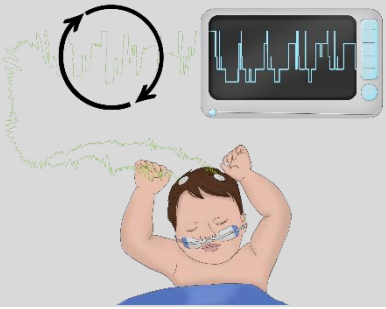


Table S1. Overview of methods used in adult ICU for automated sleep monitoring. BIS = bispectral index, EMG = electromyogram, ICU = Intensive Care Unit, MV = mechanically ventilated, N1 = non rapid eye movement stage 1, N2 = non rapid eye movement stage 2, N3 = non rapid eye movement stage 3, non-sed. = non-sedated, NSWS = non slow wave sleep, PSI = patient state index, sed. = sedated, REM = rapid eye movement, SEF95 = spectral edge frequency 95%, SWS = slow wave sleep, W = wake.

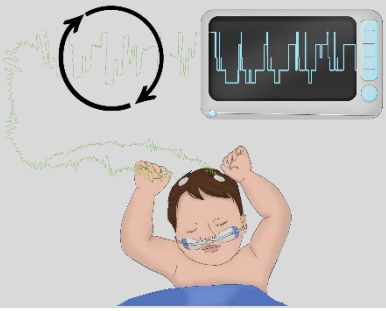
Study	Patients	Classification into	EEG channels	Features	Classification method	Results
Reinke et al. (2014)[124]	Adult ICU, MV, non-sed. (n=5) Healthy controls (n=15)	W, SWS, NSWS	C3-C4	IDOS (=gamma/delta power ratio)	Thresholding (manually selected)	ICU patients: κ=0.46-0.90 Healthy controls: κ=0.82
Ambrogio et al. (2008)[18]	Adult ICU, MV, sed. (n=17) Healthy controls (n=17)	n/a	C3-C4	Delta, theta, alpha and beta power, delta/alpha power ratio, delta/beta power ratio	n/a	*-No classification or comparison with golden standard, only feature evaluation-*
Gehlbach et al. (2012)[125]	Adult ICU, MV, sed. (n=12)	n/a	C3-A1/A2, C4-A1/A2, O1-A1/A2, O2-A1/A2	Absolute delta power + SEF95	n/a	*-No classification or comparison with golden standard, only feature evaluation-*
Nicholson et al. (2001)[129]	Adult ICU, in recovery phase, minimally sed. (n=29)	W, REM, SWS, NSWS	F7-Fpz, F8-Fpz	BIS + chin EMG power	Thresholding	*-Not compared with golden standard-*
Vacas et al. (2016)[130]	Adult ICU (n=23) Healthy controls (n=3)	W, REM, N1, N2, N3	Fp1-F8, Fp2-F7	PSI	Unknown (SedLine monitor)	Healthy controls: κ=0.61, accuracy=75%

Table S2. Overview of methods used in NICU for automated sleep monitoring. a: Similar datasets, b: Similar datasets. Abbreviations: ANN = artificial neural network, AS = active sleep, AS1 = active sleep 1, ASI = activation synchrony index, CNN = convolutional neural network, CWT = continuous wavelet transform, DWT = discrete wavelet transform, ECG = electrocardiogram, EMG = electromyogram, EMD = empirical mode decomposition, EOG = electrooculogram, GMM = Gaussian Mixture Model, HHS = Hilbert-Hough spectrum, HI = histogram index, HMM = Hidden Markov Model, HR = heart rate, HVS = high voltage slow wave, IBI = inter-burst interval, LDA = linear discriminant analysis, LSTM = long short term memory, LVI = low voltage irregular, max. = maximum, min.= minimum, NICU = neonatal intensive care unit, PMA = postmenstrual age, prc. = percentile, QS = quiet sleep, RBF = radial basis function, rEEG = range electroencephalogram, ref. = reference electrode, REM = rapid eye movement, RR = respiratory rate, SAT% = spontaneous activity transient index, SVM = support vector machine, sqrt = square root, TA = tracé alternant, W = wake, WVD = Wigner-Ville distribution.

Study	Patients	Classification into	EEG channels	Features	Classification method	Results
Koolen et al. (2017)[132]	NICU, PMA 24 – 45 weeks, preterm (n=67)	W, AS, QS	FP1, FP2, C3, C4, T3, T4, O1, O2, to ref. G2	p=7 (from p=57); delta1 power, 5th prc. rEEG, ASI C3O1-C4O2, ASI Fp1C3-Fp2C4, HI median, rEEG median, IBI median	SVM + RBF kernel	Accuracy=85% (best performance PMA>32 weeks)
Piryatinska et al. (2009)[27]	NICU, PMA 40 weeks, preterm (n=21) + term (n=16)[a]	W, AS, QS	Fp1-C3	Preterm: p=4 (from p=8); delta power, spectral entropy, log log of fractional dimension Term: p=4 (from p=8); delta power, spectral entropy, log log of fractional dimension, spectral moment	Change point detection + k-means clustering	Accuracy=80-90%
Palmu et al. (2013)[133]	NICU, PMA < 34 weeks, preterm (n=15)	W, AS, QS	C3-O2, O2-A1	SAT%	n/a	*-No classification or comparison with golden standard, only feature evaluation-*
Scher et al. (1996)[141]	NICU, PMA 29-43 weeks, preterm (n=26) + term (n=28)	W, AS, QS	FP1, FP2, C3, C4, T3, T4, 01, O2, Fz, Cz, Pz + EMG, EOG, ECG and 3 respiratory channels	p=13 (from p=32); EEG: delta, beta, alpha, theta and total power, EOG: REMs, EMG: energy, respiratory channels: RR, RR ratio, RR variance, ECG: HR, HR ratio, HR variance	Linear discriminant analysis	Accuracy=93%
Fraiwan et al. (2012)[135]	NICU, PMA 40 weeks, preterm (n=14) + term (n=15)[a]	W, AS, QS	Fp1-C3	Entropy in delta, theta, alpha, beta1, beta2, sleep spindle and K-complex frequency bands in WVD, CWT and HHS.	ANN	WVD term: accuracy=84%, κ=0.65, WVD preterm: accuracy=74%, κ=0.51 HHS term: accuracy=72%, κ=0.41, HHS preterm: accuracy=68%, κ=0.39

						CWT term: accuracy=75%, κ=0.39, CWT preterm: accuracy=64%, κ=0.28
Fraiwan et al. (2020)[136]	NICU, PMA 40 weeks, preterm (n=21) + term (n=16)[a]	W, AS, QS	Fp1-C3	*-By LSTM-*	LSTM network	Accuracy=96.81%, κ=0.91
Dereymaeker et al. (2017)[139]	NICU, PMA 27-42 weeks, preterm (n=26)	QS or non-QS	Fp1, Fp2, C3, C4, T3, T4, O1, O2, to ref. Cz	p=9; amplitude std., amplitude min.-max., max. of 1st derivative, max. of 2nd derivative, mean frequency, sqrt delta power, sqrt theta power, sqrt alpha power, sqrt beta power	Adaptive segmentation + k-means clustering	Sensitivity=93%, specificity=80% (optimal performance PMA 31-38 weeks)
Pillay et al. (2018)[137]	NICU, PMA 27-41 weeks, n=16[b]	QS or non QS or AS1, HVS, TA, LVI	Fp1, Fp2, C3, C4, T3, T4, O1, O2, to ref. Cz	p=17 (from p=112); 6 from EMD, 5 from time domain, 3 from frequency domain, 3 from DWT (for more details, see Table 3 in Pillay et al.[137])	HMM or GMM	Four-state performance: HMM: accuracy=86%, κ=0.62, GMM: accuracy=82%, κ=0.55 Two-state performance: HMM: accuracy=95%, κ=0.89, GMM: accuracy=92%, κ=0.85
Ghimatgar et al. (2020)[140]	NICU, PMA 38-42 weeks, n=16 [b]	QS or non QS or AS1, HVS, TA, LVI	Fp1-C3, T3-O1, C4-T4, O1-O2, Fp2-T4, T3-T4	p=57 (from p=160); 15 from time domain, 18 from frequency domain, 15 from DWT, 1 from cepstral domain, 11 from nonlinear domain (for more detail, see Table 2 in Ghitmatgar et al.[140])	LSTM + HMM	Two-state performance: Accuracy=94%, κ=0.88 Four-state performance: Accuracy=80%, κ=0.76
Ansari et al. (2020)[138]	NICU, PMA 38-42 weeks, preterm (n=16[b]), term (n=16)	QS or non QS or AS1, HVS, TA, LVI	Fp1, Fp2, C3, C4, Cz, T3, T4, O1, O2, to ref. Cz.	*-By CNN-*	CNN	Two-state performance: κ=0.76 Four-state performance: κ=0.64 (better performance in preterm)

Supplementary materials
Part II – Research report



Supplementary methods

1. Artifact detection algorithm

Polysomnography (PSG) signals are known to contain a lot of artifacts that might influence the algorithm's training and test performance[1]. Therefore, a simple artifact detection algorithm was used to identify and label epochs containing significant artifacts. The most common artifacts observed in the electroencephalogram (EEG) include impedance measurement artifacts (Figure S1a), movement artifacts (Figure S1b), 50-Hz electrical interference (Figure S1c) or low-frequency artifacts from other physiological sources such as breathing or sweating. During impedance measurements, the impedance of all PSG signals is checked at the same time, resulting in a flat line with zero amplitude. Impedance measurement artifacts were therefore detected if one of the PSG signals sampled at 2 Hz is equal to zero for more than 4 subsequent samples. Many other artifacts, such as from movements or 50-Hz interference, are characterized by high amplitudes. These artifacts were therefore detected when the mean of the absolute signal amplitude in the epoch exceeds a predefined threshold, which were empirically chosen and optimized for EEG, electrooculogram (EOG) and electromyogram (EMG) separately. The following thresholds used for the EEG, EOG and EMG signals, respectively: 68, 152 and 152 uV. The artifact detection algorithm was applied to all unipolar PSG signals (EEG F3, F4, C3, C4, O1, O2, A1 and A2, EOG ROC, EOG LOC and EMG chin) and labelled for each signal separately. Epochs from the unipolar PSG signal that were labelled as 'artifact', or from the bipolar PSG signal derived from the unipolar PSG signal in question, were removed from the dataset. After the artifact detection algorithm, additional artifacts were reduced by bandpass filtering the signals to remove irrelevant frequencies. These irrelevant frequencies often consist of 50-Hz electrical interference noise or low-frequency artifacts from movement, sweating or breathing, that did not exceed the amplitude thresholds and were therefore not removed by the simple artifact detection algorithm.

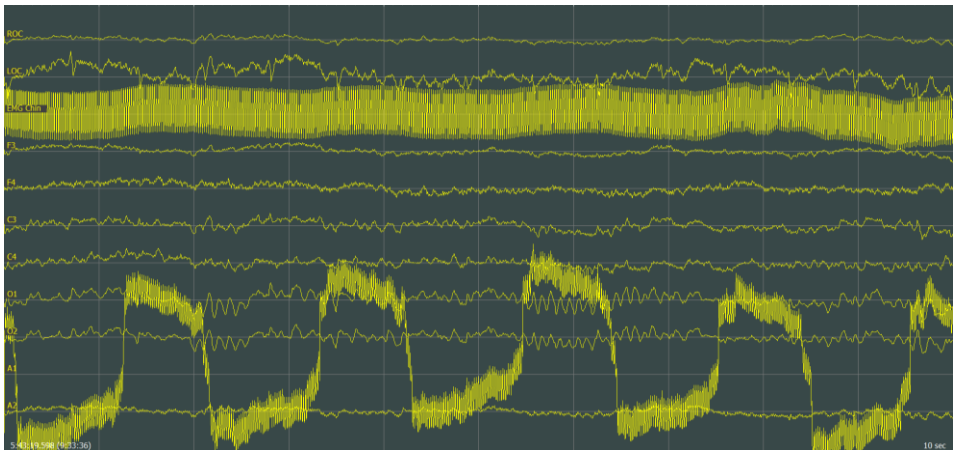


Figure S1a. 50-Hz electrical interference artifact (A2, EMG).

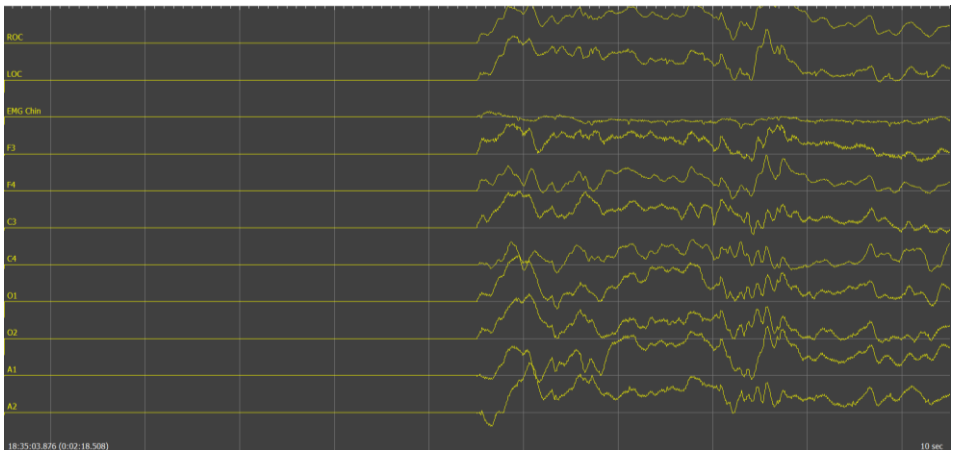


Figure S1b. Impedance measurement artifact.

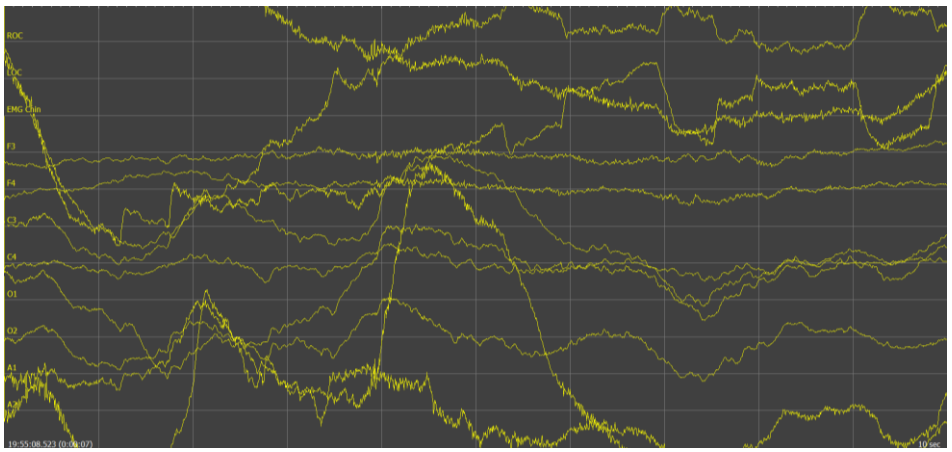


Figure S1c. Movement artifact.

2. Additional feature information

The features were calculated using the following formulas, where:

f = frequency
i = sample
μ = mean
n = number of samples in epoch
k = frequency bin
$P(f)$ = spectral power
y = epoch signal

Bandpower

$$\int_{f_{lower}}^{f_{upper}} P(f)\, df$$

Detrended fluctuation analysis

First, y is integrated into a new series $y = [y(1), \ldots., y(N)]$ where $y(k) = \sum_{i=1}^{k}(x - \mu)$.
The integrated series is then sliced into boxes of equal length n. In each box of length n, a least-squares line is fit to the data, representing the trend in that box. The y-coordinate of the straight-line segments is denoted by $y_n(k)$.
The root-mean-square fluctuation of the integrated series is calculated by:
$F(n) = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y(k) - y_n(k))^2}$.
The fluctuation can be defined as the slope of the line relating $F(n)$ to $\log(n)$.

Energy

$$\frac{1}{n}\sum_{i=0}^{n}(y_i)^2$$

Higuchi fractal dimension

First, k new series are constructed from y by:
$y_m, y_{m+k}, y_{m+2k}, \ldots, y_{m+\left(\frac{n-m}{k}\right)k}$, where $m = 1, 2, \ldots, k$.
For each time series constructed from the previous decomposition, the length $L(m,k)$ is computed by:

$$L(m,k) = \frac{\sum_{i=2}^{(n-m)/k} | x_{m+ik^{-x}_{m+(i-1)k}}|(n-1)}{((n-m)/k)k}$$

The average length is computed as $L(k) = (\sum_{i=1} L(i,k))/k)$

This procedure repeats k_{max}times for each k from 1 to k_{max}, and then uses a least-square method to determine the slope of the line that best fits the curve of $\ln(L(k))$ versus $\ln(1/k)$. The slope is the Higuchi Fractal Dimension.

Hjorth Activity

$$\frac{1}{n}\sum_{i=0}^{n}(y_i - \mu)$$

Hjorth Mobility	$\sqrt{\frac{variance(\frac{dy_i}{i})}{variance(y_i)}}$	
Hjorth Complexity	$\frac{Mobility(\frac{dy_i}{di})}{Mobility(y_i)}$	
Interquartile range (25th-75th)	$Q3 - Q1$	where $Q1$ = 25th percentile and $Q3$ = 75th percentile
Kurtosis	$\frac{\sum_{i=0}^{n}(y_i - \mu)^4 / n}{\sigma^4}$	where σ = standard deviation
Mean frequency	$\frac{\sum_{i}^{n} P(f_k) * f_k}{\sum_{i=0}^{n} P(f_k)}$	
Mean absolute amplitude	$\frac{1}{n}\sum_{i=0}^{n} \lvert y_i \rvert$	
Median frequency	$\frac{\sum_{k=0}^{n} P(f_k) * f_k}{\sum_{k=0}^{n} P(f_k)}$	
Shannon entropy	$\frac{1}{n}\sum_{i=0}^{n} p(y_i) * \log(p(y_i))$	where $p(y_i)$ = probability of signal value y_i
Signal sum	$\sum_{i=0}^{n} \lvert y_i \rvert$	
Skewness	$\frac{\sum_{i=0}^{n}(y_i - \mu)^3 / n}{\sigma^3}$	where σ = standard deviation
Standard deviation	$\frac{1}{n}\sqrt{\sum_{i=0}^{n}(y_i - \mu)}$	
Spectral edge 95%	f_k where $\sum_{k=0}^{f_{max}} P(f_k) > 0.95 * P(f)$	
Spectral entropy	$\frac{-\sum_{k=0}^{f_{max}} P(f_k) * \log(P(f_k))}{\log(f_{max})}$	
Spectral kurtosis	$\frac{\sum_{k=0}^{f_{max}}(f_k - \mu_f)^4 * P(f_k)}{{\sigma_f}^4 \sum_{k=0}^{f_{max}} P(f_k)}$	where μ_f = mean frequency and σ_f = standard deviation of spectral power

Spectral skewness	$\frac{\sum_{k=0}^{f_{max}}(f_k - \mu_1)^3 * P(f_k)}{{\mu_2}^3 \sum_{k=0}^{f_{max}} P(f_k)}$	where μ_f = mean frequency and σ_f = standard deviation of spectral power
Variance	$\frac{1}{n}\sum_{i=0}^{n}(y_i - \mu)$	
Zero-crossing-rate	$\frac{1}{n}\sum_{i=0}^{n}\lvert s_i - s_{i-1}\rvert$	where s_i = polarity, $s_i = 1$ if $y_i > 0$

3. Discrete wavelet transform

The discrete wavelet transform is a time-frequency analysis method that is regularly used in EEG signal analysis and feature extraction. Because of its flexible way to represent the time-frequency domain of a signal, it has the benefit to account for the non-stationarity of the signal. The EEG signal is considered a non-stationary signal as its properties change during each sleep stage by transient waves and alternating frequencies. The DWT is obtained by convolution with wavelets with varying scales and breaks down the input signal into multiple frequency subbands. A wavelet is characterized by its zero mean and finite energy, which enables localization in time and thus temporal resolution. In DWT, the time window size varies with the frequency. Lower frequencies have larger windows and higher frequencies have smaller windows, resulting in better frequency resolution in lower frequencies and better time resolution in higher frequencies. The DWT is calculated by passing it to a series of low- and high-pass filters. Each filter output is down-sampled by factor 2 before further processing so account for the removed frequencies after filtering. Convolution with the wavelet with varying scale results in detail (high-pass filtered signal) and approximation coefficients (lowpass filtered signals). In this study, a debauchies-4 (db-4) wavelet was used, as it has previously been shown to perform well in other EEG analyses[2]. A block diagram of a 5-level DWT process with the frequency range per sub-band is shown in Figure S2. The frequency ranges of the five levels of decomposition and their associated EEG frequency range are provided in Table S1.

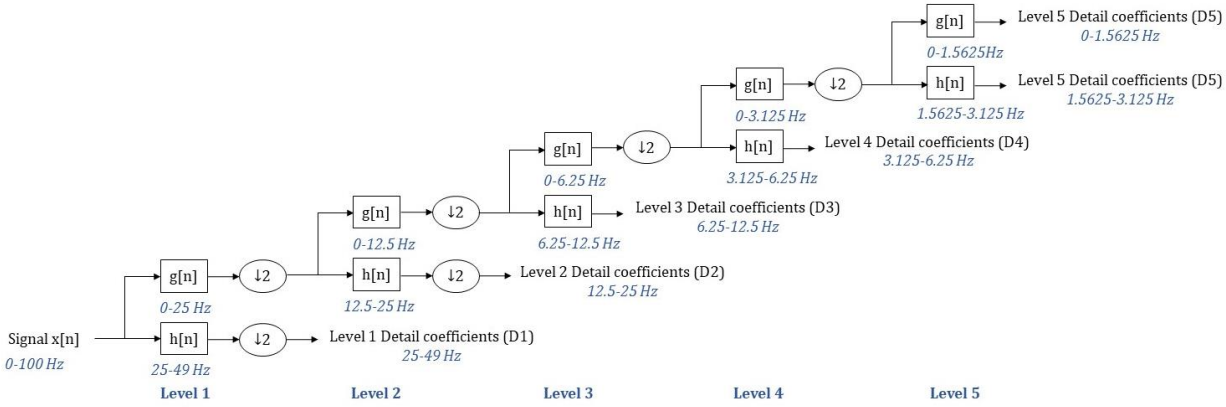


Figure S2. Structure of a 5-level discrete wavelet transform with sample frequency 100 Hz. h[n] represents high pass filters, g[n] represents low pass filters, ↓ represents the subsampling by factor 2.

Table S1. Frequency ranges of five levels of decomposition in discrete wavelet transform with sampling frequency 100 Hz.

Coefficient set	Frequency range	Associated EEG frequency band
D1	25-49 Hz	Gamma
D2	12.5-25 Hz	Beta
D3	6.25-12.5 Hz	Alpha
D4	3.125-6.25 Hz	Theta
D5	1.5625-3.125 Hz	Delta
A5	0-1.5625 Hz	Delta

References:

1. Devuyst, S. *et al.* Automatic Processing of EEG-EOG-EMG Artifacts in Sleep Stage Classification. *IFMBE Proc.* 23, 146–150 (2009).
2. Pillay, K. *et al.* Automated EEG sleep staging in the term-age baby using a generative modelling approach. *J. Neural Eng.* 15, (2018).

Supplementary tables and figures

Table S2. Hyperparameter settings during learning curve determination. DT = decision tree, LR = logistic regression, LDA = linear discriminant analysis, KNN = k-nearest neighbours, SVM = support vector machine, RF = random forest, XGBoost = gradient boosting algorithm

Machine learning model	Hyperparameter settings for learning curve determination
DT	Splitting criterion: Gini impurity, maximum tree depth: 10
LR	Penalty: L2, regularization term (C): 1.0
LDA	Solver: singular value decomposition, priors: class proportions
KNN	Number of neighbours: 40, distance metric: Minkowski, weights: uniform
SVM	Kernel: radial basis function, regularization term (C): 10, kernel coefficient (gamma): scaled (=1 / (n_features * $variance_{features}$))
RF	Number of estimators: 100, splitting criterion: Gini impurity, maximum tree depth = 6, bootstrapping: on
XGBoost	Booster: decision tree, objective: linear regression, learning rate: 0.1, minimum child weight = 1, number of estimators: 100, splitting criterion: Gini impurity, maximum tree depth = 6

Table S3. Hyperparameter space and final tuning configuration. Other hyperparameters are equal to the ones used to create the learning curves (Table S2). DT = decision tree, SVM = support vector machine, XGBoost = extreme gradient boosting

Machine learning model	Hyperparameter	Hyperparameter space	Final configuration three-state classification	Final configuration four-state classification
DT	Number of features	[2, 4, 6, 8, 10, 12, 14, 16, 18, 20]	20	16
	Max depth	[2, 4, 6, 8, 10, 12, 14, 16, 18]	18	12
SVM	Number of features	[2, 4, 6, 8, 10, 12, 14, 16, 18, 20]	20	20
	Regularization term, C	[0.01, 0.1, 1, 10, 100]	10	1
	Gamma	[0.001, 0.01, 0.1, 1, 10]	0.01	0.01
XGBoost	Number of features	[2, 4, 6, 8, 10, 12, 14, 16, 18, 20]	20	18
	Learning rate	[0.05, 0.1, 0.15, 0.2]	0.05	0.05
	Min. child weight	[1, 3, 5, 7]	3	1
	Subsample ratio of columns when constructing each tree	[0.5, 0.75, 1]	1	0.5

Table S4. Detailed patient characteristics of the non-critically ill (reference) patients, per age category. PSG = polysomnography

	0-2 months (n=15)	2-6 months (n=15)	6-12 months (n=15)	1-3 years (n=15)	3-5 years (n=15)	5-9 years (n=15)	9-13 years (n=15)	13-18 years (n=15)
Median age (years)	0.1	0.3	0.7	2.3	4.1	6.3	11.4	16.0
Males/females	7/8	8/7	8/7	8/7	9/6	8/7	7/8	7/8
PSG indication (% of total)								
• Airway obstruction	73.3	33.3	33.3	73.3	73.3	46.7	33.3	40
• Neuromuscular disease	6.7	26.7	26.7	6.7	13.1	40	53.3	40
• Pulmonary disease	0	13.3	26.7	6.7	0	0	6.7	6.7
• Central sleep apnea	13.1	13.3	13.3	6.7	0	0	6.7	6.7
• Unknown	6.7	13.3	0	6.7	13.3	13.1	0	6.7

Table S5. Detailed PSG data characteristics from the non-critically ill (reference) patients, obtained from the visually scored hypnogram. Total sleep time is the time spent in any of the sleep stages during the PSG recording. PICU = pediatric intensive care unit, PSG = polysomnography, N1/N2/N3/N = non rapid eye movement sleep stage 1, 2, 3 or quiet sleep, respectively, REM = rapid eye movement sleep, SD = standard deviation

	0-2 months	2-6 months	6-12 months	1-3 years	3-5 years	5-9 years	9-13 years	13-18 years
Total recording length (hours)	175,8 (n=21,100)	170.2 (n=20,428)	168.4 (n=20,204)	163.9 (n=19,667)	161.1	163.2	151.1	142.1
Average length of PSG recording, hours (± SD)	11.7 (± 1.5)	11.3 (± 1.3)	11.2 (± 0.8)	10.9 (± 1.1)	10.7 (± 0.7)	10.9 (± 0.85)	10.1 (± 1.1)	9.5 (± 0.7)
Total sleep time, hours (± SD)	8.3 (± 1.4)	8.8 (± 0.8)	8.9 (± 1.5)	8.2 (± 1.5)	9.0 (± 0.6)	8.6 (± 0.7)	6.8 (± 1.5)	6.9 (± 1.20
Mean time spent in each stage, % of total sleep time (± SD) (number of epochs)								
• REM	34.4 (± 9.0) (n=5,295)	32.6 (± 8.1) (n=5,120)	28.02 (± 3.4) (n=4,482)	22.9 (± 5.6) (n=3,355)	21.7 (±4.0) (n=3,539)	22.4 (± 4.4) (n=3,458)	16.2 (± 6.0) (n=2,086)	17.9 (± 9.3) (n=2,143)
• N1	0.5 (± 1.7) (n=87)	5.5 (± 6.6) (n=855)	12.1 (± 4.9) (n=1,950)	15.2 (± 8.4) (n=2,294)	9.4 (± 5.4) (n=1,1514)	8.2 (± 4.3) (n=1,279)	14.9 (± 7.4) (n=1,716)	13.2 (± 8.4) (n=1,613)
• N2	0	14.2 (± 13.7) (n=2,203)	28.1 (± 8.7) (n=4,499)	29.1 (± 10.9) (n=4,318)	27.3 (±10.2) (n=4,461)	32.7 (± 10.2) (n=5,135)	32.7 (± 7.6) (n=4,005)	38.0 (± 11.3) (n=4,818)
• N3	5.6 (±14.3) (n=838)	28.0 (± 16.5) (n=4,365)	31.8 (± 12.5) (n=5,070)	32.9 (±10.5) (n=4,743)	41.5 (±13.7) (n=6,671)	36.8 (± 12.1) (n=5,646)	36.3 (± 10.0) (n=4,456)	30.8 (± 10.2) (n=3,802)
• N	0.006 (± 0.06) (n=5)	20.0 (± 26.5) (n=8,736)	n/a	n/a	n/a	n/a	n/a	n/a

Table S6. Detailed PSG data characteristics from the PICU patients, obtained from the visually scored hypnogram. Total sleep time is the time spent in any of the sleep stages during the PSG recording. a: From the PSG of patient H, no sleep stages could be distinguished and all sleep was staged as sleep stage N, b: Sleep was staged as stage N when no clear characteristics were present to discriminate between stage N1, N2 or N3. PICU = pediatric intensive care unit, PSG = polysomnography, N1/N2/N3/N = non rapid eye movement sleep stage 1, 2, 3 or quiet sleep, respectively, REM = rapid eye movement sleep, SD = standard deviation

	Patient A	Patient B	Patient C	Patient D	Patient E	Patient F	Patient G	Patient H[a]	Patient I	Patient J
Length of PSG recording, hours	26.3	10.2	24.4	24.1	24.3	24.8	23.6	24.3	20.0	23.9
Total sleep time, hours	12.2	6.2	10.9	15.6	18.6	16.0	12.8	22.3	13.1	11.9
Mean time spent in each stage, % of total sleep time (number of epochs)										
• REM	2.2 (n=32)	11.0 (n=82)	10.6 (n=138)	4.7 (n=87)	13.6 (n=303)	39.6 (n=758)	4.1 (=63)	0 (n=0)	9.2 (n=145)	4.6 (n=66)
• N1	41.3 (n=603)	0 (n=0)	4.8 (n=63)	0 (n=0)	0 (n=0)	0 (n=0)	33.6 (n=516)	0 (n=0)	21.5 (n=338)	10.2 (n=145)
• N2	45.1 (n=648)	0 (n=0)	11.3 (n=147)	0 (n=0)	10.0 (n=224)	0 (n=0)	45.9 (n=704)	0 (n=0)	61.1 (n=958)	47.3 (n=676)
• N3	11.4 (n=166)	47.7 (n=335)	70.9 (n=925)	0 (n=0)	43.9 (n=980)	0 (n=0)	15.0 (n=230)	0 (n=0)	8.2 (n=128)	37.9 (n=541)
• N[b]	0 (n=0)	41.3 (n=208)	2.5 (n=32)	95.3 (n=1,1779)	32.6 (n=728)	60.5 (n=1,159)	1.4 (n=21)	100 (n=2679)	0 (n=0)	0 (n=0)

Table S7. Results of the artifact detection algorithm in the non-critically ill (reference) patient data per age category. EEG = electroencephalogram, EMG = electromyogram, EOG = electrooculogram, LOC = left outer canthus, ROC = right other canthus

	0-2 months (n=15)	2-6 months (n=15)	6-12 months (n=15)	1-3 years (n=15)	3-5 years (n=15)	5-9 years (n=15)	9-13 years (n=15)	13-18 years (n=15)
Impedance artifacts, % of total epochs (number of epochs)	0.6 (n=119)	0.4 (n=80)	0.2 (n=35)	0.2 (n=34)	0.1 (n=24)	0.2 (n=42)	0.2 (n=44)	0.1 (n=25)
Total high amplitude artifacts, % of total epochs (number of epochs)	18.6% (n=43,361)	18.6 (n=41,213)	10.8 (n=24,025)	13.4 (n=29,092)	9.5 (n=20,268)	13.9 (n=29,918)	16.6 (n=33,208)	8.2 (n=15,378)
EOG ROC high amplitude artifacts, % of total epochs (number of epochs)	9.4 (n=1,977)	11.2 (n=2,258)	3.4 (n=691)	1.6 (n=318)	1.3 (n=245)	4.2 (n=829)	8.2 (n=1,484)	8.5 (n=1,442)
EOG LOC high amplitude artifacts, % of total epochs (number of epochs)	13.8 (n=2,910)	9.5 (n=1,913)	9.1 (n=1,832)	3.9 (n=775)	3.8 (n=741)	11.8 (n=2,317)	13.8 (n=2,505)	4.1 (n=695)
EMG Chin high amplitude artifacts, % of total epochs (number of epochs)	15.0 (n=3,161)	17.5 (n=3,525)	8.1 (n=1,639)	21.1 (n=4,142)	7.4 (n=1,421)	17.0 (n=3,332)	23.8 (n=4,318)	24.6 (n=4,200)
EEG F3 high amplitude artifacts, % of total epochs (number of epochs)	12.7 (n=2,682)	12.3 (n=2,477)	7.6 (n=1,538)	13.3 (n=2,616)	2.0 (n=392)	7.1 (n=1,392)	7.8 (n=1,421)	0.9 (n=154)
EEG F4 high amplitude artifacts, % of total epochs (number of epochs)	17.7 (n=3,732)	15.1 (n=3,043)	13.2 (n=2,665)	9.6 (n=1,882)	5.3 (n=1,034)	10.2 (n=1,994)	10.9 (n=1,980)	2.0 (n=336)
EEG C3 high amplitude artifacts, % of total epochs (number of epochs)	18.0 (n=3,795)	17.0 (n=3,411)	7.9 (n=1,593)	9.1 (n=1,784)	5.0 (n=957)	8.1 (n=1,583)	11.7 (n=2,114)	2.2 (n=383)
EEG C4 high amplitude artifacts, % of total epochs (number of epochs)	11.7 (n=2,479)	12.1 (n=2,429)	4.1 (n=829)	4.0 (n=779)	2.9 (n=558)	8.5 (n=1,663)	10.0 (n=1,813)	0.8 (n=129)
EEG O1 high amplitude artifacts, % of total epochs (number of epochs)	28.8 (n=6,071)	29.4 (n=5,922)	22.9 (n=4,619)	22.6 (n=4,445)	19.7 (n=3,802)	24.3 (n=4,761)	25.7 (n=4,664)	8.1 (n=1,383)
EEG O2 high amplitude artifacts, % of total epochs (number of epochs)	26.3 (n=5,548)	28.1 (n=5,663)	18.7 (n=3,787)	28.5 (n=5,604)	19.9 (n=3,850)	19.1 (n=3,740)	25.7 (n=4,664)	9.3 (n=1,581)

EEG A1 high amplitude artifacts, % of total epochs (number of epochs)	25.9 (n=5,473)	28.7 (n=5,768)	12.1 (n=2,454)	18.9 (n=3,718)	19.6 (n=3,785)	26,2 (n=5,124)	20.5 (n=3,717)	13.2 (n=2,247)
EEG A2 high amplitude artifacts, % of total epochs (number of epochs)	25.7 (n= 5,414)	23.5 (n=4,724)	11.6 (n=2,343)	15.2 (n=2,995)	17.9 (n=3,459)	16.0 (n=3,141)	24.8 (n=4,484)	16.4 (n=2,803)

Table S8. Results of the artifact detection algorithm in the PICU patient data per age category. a: Artifact labels were manually removed since they were detected as artifacts due to the baseline drift in the EEG signal. This baseline drift could be removed with the bandpass filter. EEG = electroencephalogram, EMG = electromyogram, EOG = electrooculogram, LOC = left outer canthus, ROC = right other canthus

	Patient A	Patient B	Patient C	Patient D[a]	Patient E	Patient F	Patient G	Patient H	Patient I	Patient J
Impedance artifacts, % of total epochs (number of epochs)	0.03 (n=1)	0 (n=0)	0.03 (n=1)	0.03 (n=1)	0.2 (n=7)	0.03 (n=1)	0 (n=0)	0 (n=0)	0.04 (=1)	0 (n=0)
Total high amplitude artifacts, % of total epochs (number of epochs)	2.5 (n=882)	23.0 (n=3,110)	3.6 (n=1,175)	46.2 (n=14,725)	33.2 (n=10,637)	6.7 (n=2,187)	21.5 (n=6,725)	30.9 (n=9.907)	35.9 (n=9,468)	27.1 (n=8,539)
EOG ROC high amplitude artifacts, % of total epochs (number of epochs)	8.7 (n=276)	16.2 (n=199)	0.1 (n=4)	0.03 (n=1)	0 (n=0)	13.7 (n=407)	0.5 (n=14)	0.5 (n=16)	0.3 (n=6)	0.9 (n=26)
EOG LOC high amplitude artifacts, % of total epochs (number of epochs)	7.7 (n=243)	29.6 (n=364)	1.2 (n=35)	7.0 (n=202)	1.6 (n=7)	0.2 (n=7)	0.03 (n=1)	0.4 (n=11)	2.3 (n=54)	0.2 (n=7)
EMG Chin high amplitude artifacts, % of total epochs (number of epochs)	0.09 (n=3)	0 (n=0)	35.1 (n=1029)	4.2 (n=121)	12.0 (n=348)	31.5 (n=936)	14.3 (n=406)	0.4 (n=12)	1.9 (n=45)	0.8 (n=23)
EEG F3 high amplitude artifacts, % of total epochs (number of epochs)	0.03 (n=1)	1.4 (n=17)	0.3 (n=8)	98.0 (n=2,837)	0.2 (n=5)	1.2 (n=36)	100 (n=2,837)	0.3 (n=8)	100 (n=2,399)	0.3 (n=8)
EEG F4 high amplitude artifacts, % of total	0.06 (n=2)	19.9 (n=245)	0.4 (n=13)	98.0 (n=2,837)	99.5 (n=2,896)	1.2 (n=36)	2.6 (n=73)	100 (n=2,916)	0.5 (n=13)	99.0 (n=2,839)

epochs (number of epochs)										
EEG C3 high amplitude artifacts, % of total epochs (number of epochs)	1.0 (n=32)	15.1 (n=185)	0.2 (n=7)	1.6 (n=45)	0.4 (n=13)	1.2 (n=13)	23.4 (n=664)	0.4 (n=12)	100 (n=2,399)	0.6 (n=19)
EEG C4 high amplitude artifacts, % of total epochs (number of epochs)	0.06 (n=2)	17.0 (n=209)	0.3 (n=10)	0.1 (n=3)	99.1 (n=2,883)	1.0 (n=31)	7.8 (n=221)	100 (n=2,2916)	4.0 (n=97)	97.6 (n=2800)
EEG O1 high amplitude artifacts, % of total epochs (number of epochs)	4.5 (n=144)	74.7 (n=918)	0.4 (n=13)	100 (n=2,896)	28.2 (n=820)	5.6 (n=166)	23.1 (n=656)	1.8 (n=52)	100 (n=2,399)	0 (n=0)
EEG O2 high amplitude artifacts, % of total epochs (number of epochs)	3.8 (n=120)	17,3 (n=213)	1.1 (n=31)	0.2 (n=6)	98.6 (n=2,870)	4.6 (n=137)	33.9 (n=961)	100 (n=2,916)	60.2 (n=1,444)	97.4 (n=2,794)
EEG A1 high amplitude artifacts, % of total epochs (number of epochs)	0.5 (n=15)	30.2 (n=371)	0.5 (n=15)	99.5 (n=2,881)	13.0 (n=377)	3.6 (n=106)	15.9 (n=451)	6.6 (n=191)	12.3 (n=296)	0.3 (n=10)
EEG A2 high amplitude artifacts, % of total epochs (number of epochs)	1.4 (n=44)	31.7 (n=389)	0.3 (n=10)	100 (n=2,896)	13.0 (n=378)	9.8 (n=290)	15.5 (n=441)	29.4 (n=857)	13.2 (n=316)	0.5 (n=13)

Table S9. Spearman correlation of spectral powers with sleep stage labels. Correlations were determined for three-state classification using spectral powers obtained from the C3-C4 channel from the exploration data set (10,000 samples). The categorical sleep stage labels were converted to linear sleep stage labels as follows: Wake = 2, NSWS = 1, SWS = 0. NSWS = non slow wave sleep, SWS = slow wave sleep

	Absolute	Relative
Total power	-0.49	n/a
Gamma power	0.26	0.61
Beta power	-0.08	0.43
Alpha power	-0.24	0.39
Theta power	-0.44	0.18
Delta power	-0.53	-0.43

Table S10. Classification performance (train scores, accuracy) of various potential index-measures constructed from the ratio of spectral powers. Performance results were obtained from the exploration data set (10,000 samples) for three-state classification. EEG = electroencephalogram

EEG channel	Gamma/delta	Gamma/theta	(Gamma+beta) / (theta+delta)	Gamma / (theta+delta)
EEG F3	0.71	0.70	0.66	0.72
EEG C3	0.70	0.64	0.60	0.69
EEG O1	0.68	0.60	0.61	0.68
EEG A1	0.71	0.63	0.64	0.71
EEG F3-C3	0.70	0.66	0.64	0.70
EEG F3-C4	0.70	0.68	0.66	0.71
EEG F3-O2	0.71	0.66	0.66	0.71
EEG F3-A2	0.71	0.65	0.68	0.71
EEG C3-C4	0.67	0.63	0.63	0.66
EEG C3-O2	0.68	0.61	0.63	0.68
EEG C3-A2	0.70	0.61	0.65	0.69
EEG O1-O2	0.64	0.59	0.62	0.64
EEG O1-A2	0.68	0.62	0.65	0.67

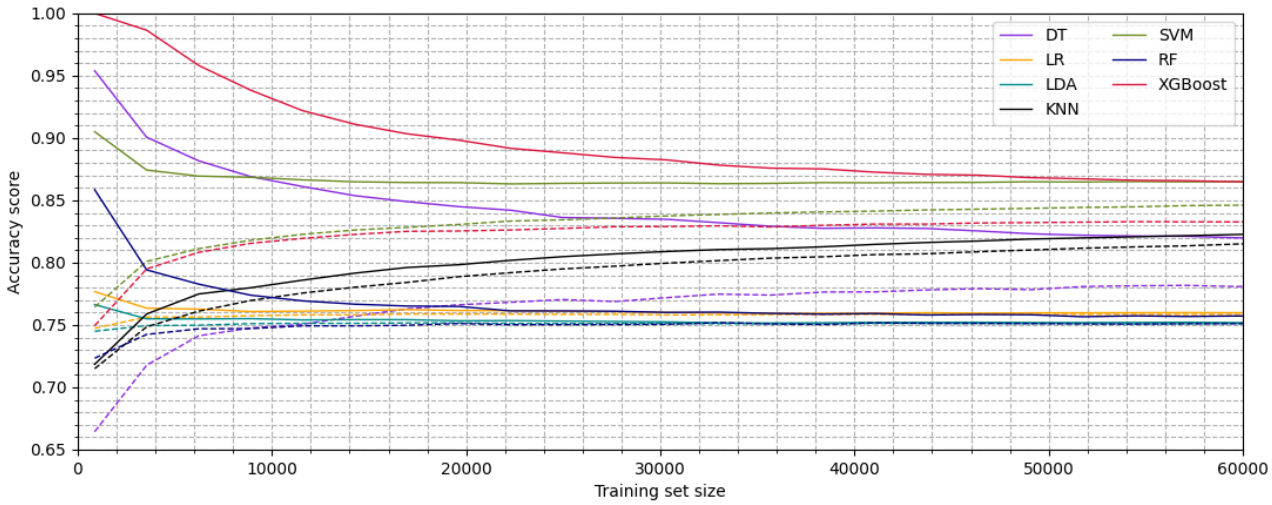


Figure S3. Learning curve for three-state classification. The learning curves shows the classification performance of 7 machine learning models across different training set sizes. The learning curves were created using 20 features, selected by the mRMR method, in combination with the age features on the C3-C4 channel, using all reference PSG data. Validation scores were obtained via 10-fold CV. The used hyperparameter settings can be found in Table S2. The plot shows that best classification results were obtained with the XGBoost, SVM, DT and KNN models. Performance of the RF, LDA and LR models was similar after they reached the convergence point. While the RF, KNN, LR and LDA models did not seem to benefit from more than, approximately, 10,000 samples used for the training set, the DT, SVM and XGBoost models did benefit from more training data. Although the convergence points of the XGBoost and DT models did not seem to be fully reached, a training set size of 50,000 samples was considered to both reasonable and feasible. No overfitting was observed with increasing training set sizes. Against expectations, the performance scores obtained during the determination of the learning curve were higher than the performance scores of the final models. It is hypothesized that this is a result of the difference in the CV strategy that is used. Whereas during learning curve determination the CV folds were made on an epoch-level, the CV folds during internal validation of the final models were made rather on a patient-level. Also, the feature selection that was performed on the same data set as was used to obtain the learning curve validation scores could have contributed to the high performances.
CV = cross-validation, DT = decision tree, KNN = k-nearest neighbour, LDA = linear discriminant analysis, LR = logistic regression, mRMR = minimum redundancy maximum relevance, PSG = polysomnography, RF = random forest, SVM = support vector machine, XGBoost = extreme gradient boosting

Table S11. Classification performance (CV scores, accuracy) of the index-based models across various EEG channels for three- and four-state classification. Performance results were obtained from the training data set (i.e. all reference PSG data). CV results were obtained using 5-fold CV. CV = cross-validation, EEG = electroencephalogram, PSG = polysomnography, SD = standard deviation

	Gamma/delta		Gamma/(theta+delta)	
EEG channel	**Three-state**	**Four-state**	**Three-state**	**Four-state**
EEG F3	0.70 (± 0.03)	0.54 (± 0.01)	0.70 (± 0.02)	0.53 (± 0.01)
EEG C3	0.70 (± 0.05)	0.52 (± 0.01)	0.69 (± 0.02)	0.52 (± 0.01)
EEG O1	0.68 (± 0.03)	0.47 (± 0.01)	0.68 (± 0.04)	0.47 (± 0.02)
EEG A1	0.69 (± 0.01)	0.51 (± 0.02)	0.68 (± 0.00)	0.50 (± 0.02)
EEG F3-C3	0.70 (± 0.03)	0.52 (± 0.03)	0.70 (± 0.03)	0.52 (± 0.03)
EEG F3-C4	0.71 (± 0.03)	0.53 (± 0.03)	0.71 (± 0.03)	0.52 (± 0.03)
EEG F3-O2	0.71 (± 0.02)	0.51 (± 0.03)	0.71 (± 0.02)	0.51 (± 0.02)
EEG F3-A2	0.70 (± 0.01)	0.49 (± 0.02)	0.70 (± 0.00)	0.50 (± 0.01)
EEG C3-C4	0.69 (± 0.03)	0.51 (± 0.02)	0.69 (± 0.03)	0.51 (± 0.02)
EEG C3-O2	0.68 (± 0.03)	0.47 (± 0.02)	0.69 (± 0.03)	0.47 (± 0.03)
EEG C3-A2	0.70 (± 0.03)	0.49 (± 0.03)	0.70 (± 0.02)	0.49 (± 0.02)
EEG O1-O2	0.66 (± 0.01)	0.40 (± 0.01)	0.66 (± 0.01)	0.40 (± 0.01)
EEG O1-A2	0.66 (± 0.03)	0.44 (± 0.02)	0.67 (± 0.03)	0.46 (± 0.03)

Table S12. Classification performance (CV scores, accuracy) of the machine learning models across various EEG channels for three- and four-state classification. Performance results were obtained from the training data set for channel evaluation (i.e, 10,000 epochs of the reference PSG data). CV results were obtained using 5-fold CV. a: Not tested due to run time problems. CV = cross-validation, DT = decision tree, EEG = electroencephalogram, PSG = polysomnography, SD = standard deviation. SVM = support vector machine, XGBoost = extreme gradient boosting

	DT		SVM		XGBoost	
EEG channel	**Three-state**	**Four-state**	**Three-state**	**Four-state**	**Three-state**	**Four-state**
EEG F3	0.75 (± 0.00)	0.65 (± 0.01)	0.72 (± 0.01)	0.62 (± 0.01)	0.78 (± 0.01)	0.70 ((± 0.02)
EEG C3	0.73 (± 0.01)	0.64 (± 0.01)	0.51 (± 0.01)	0.34 (± 0.00)	0.77 (± 0.00)	0.71 (± 0.01)
EEG O1	0.74 (± 0.01)	0.61 (± 0.00)	0.63 (± 0.01)	0.62 (± 0.02)	0.78 (± 0.01)	[a]
EEG A1	0.76 (± 0.00)	0.62 (± 0.01)	0.52 (± 0.00)	0.62 (± 0.01)	0.78 (± 0.01)	[a]
EEG F3-C3	0.73 (± 0.00)	0.66 (± 0.01)	0.75 (± 0.00)	0.69 (± 0.01)	0.79 (± 0.01)	0.70 (± 0.01)
EEG F3-C4	0.75 (± 0.00)	0.64 (± 0.01)	0.66 (± 0.00)	0.68 (± 0.01)	0.78 (± 0.01)	0.71 (± 0.01)
EEG F3-O2	0.74 (± 0.00)	0.62 (± 0.00)	0.54 (± 0.00)	0.37 (± 0.00)	0.79 (± 0.01)	0.66 (± 0.01)
EEG F3-A2	0.77 (± 0.00)	0.67 (± 0.00)	0.66 (± 0.00)	0.49 (± 0.01)	0.79 (± 0.01)	0.70 (± 0.01)
EEG C3-C4	0.72 (± 0.00)	0.66 (± 0.01)	0.72 (± 0.00)	0.67 (± 0.02)	0.77 (± 0.01)	0.71 (± 0.00)
EEG C3-O2	0.71 (± 0.01)	0.64 (± 0.01)	0.63 (± 0.01)	0.65 (± 0.02)	0.78 (± 0.01)	0.69 (± 0.01)
EEG C3-A2	0.72 (± 0.00)	0.67 (± 0.01)	0.62 (± 0.01)	0.51 (± 0.02)	0.79 (± 0.01)	0.70 (± 0.01)
EEG O1-O2	0.74 (± 0.00)	0.61 (± 0.01)	0.61 (± 0.00)	0.54 (± 0.01)	0.77 (± 0.01)	[a]
EEG O1-A2	0.72 (± 0.00)	0.64 (± 0.00)	0.67 (± 0.01)	0.65 (± 0.01)	0.79 (± 0.01)	[a]

Table S13. Classification performances of the index-based and machine learning models. AUC = area under the (receiver operating characteristic) cuve, CV = cross-validation, PICU = pediatric intensive care unit, PSG = polysomnography, XGBoost = extreme gradient boosting, DT = decision tree, SVM = support vector machine

	EEG channel	Performance metric	Training scores	Internal validation score (5-fold CV score, reference data)	External validation scores (test score, PICU data)
Three-state classification					
Gamma/delta	**F3-C3**	**Accuracy**	0.69	0.69 (± 0.03)	0.60
		AUC	0.83	0.82 (± 0.02)	0.78
		Cohen's kappa	0.50	0.50 (± 0.04)	0.40
Gamma/(theta+delta)	**F3-C3**	**Accuracy**	0.70	0.70 (± 0.03)	0.60
		AUC	0.83	0.82 (± 0.02)	0.78
		Cohen's kappa	0.51	0.51 (± 0.04)	0.40
DT	**F3-C3**	**Accuracy**	0.96	0.74 (± 0.00)	0.49
		AUC	0.99	0.78 (± 0.01)	0.62
		Cohen's kappa	0.92	0.59 (± 0.01)	0.25
SVM	**F3-C3**	**Accuracy**	0.76	0.76 (± 0.00)	0.54
		AUC	0.90	0.90 (± 0.00)	0.74
		Cohen's kappa	0.60	0.60 (± 0.01)	0.34
XGBoost	**F3-C3**	**Accuracy**	0.81	0.79 (± 0.00)	0.55
		AUC	0.94	0.93 (± 0.00)	0.72
		Cohen's kappa	0.70	0.66 (± 0.01)	0.34
Four-state classification					
Gamma/delta	**F3-C3**	**Accuracy**	0.57	0.52 (± 0.03)	0.55
		AUC	0.79	0.79 (± 0.03)	0.76
		Cohen's kappa	0.40	0.34 (± 0.04)	0.36
Gamma/(theta+delta)	**F3-C3**	**Accuracy**	0.57	0.52 (± 0.03)	0.55
		AUC	0.79	0.79 (± 0.03)	0.76
		Cohen's kappa	0.40	0.34 (± 0.04)	0.35
DT	**F3-C3**	**Accuracy**	0.81	0.68 (± 0.01)	0.40
		AUC	0.96	0.83 (± 0.01)	0.60
		Cohen's kappa	0.75	0.56 (± 0.01)	0.18
SVM	**F3-C3**	**Accuracy**	0.74	0.73 (± 0.01)	0.45
		AUC	0.93	0.92 (± 0.00)	0.70
		Cohen's kappa	0.65	0.63 (± 0.01)	0.25
XGBoost	**F3-C3**	**Accuracy**	0.80	0.75 (± 0.01)	0.45
		AUC	0.95	0.93 (± 0.00)	0.68
		Cohen's kappa	0.73	0.65 (± 0.01)	0.23

Table S14. Three-state classification performances of the index-based and machine learning models per PICU patient. In some patients, only three sleep stages were distinguished, thereby hampering the individual assessment of four-state classification performance (marked by '-'). In patient H, only two sleep stages were distinguished. Therefore, classification performance could not be individually assessed for this patient. a: Artifact labels were manually removed since they were detected as artifacts due to the baseline drift in the EEG signal. This baseline drift could be removed with the bandpass filter. b: The F4-C4 channel was used instead of the F3-C3 channel. AUC = area under the (receiver operating characteristic) curve, DT = decision tree, SVM = support vector machine, XGBoost = extreme gradient boosting

	Performance metric	Patient A	Patient B	Patient C	Patient D[a]	Patient E	Patient F	Patient G[b]	Patient H	Patient I[b]	Patient J
Three-state classification											
Gamma/delta ratio	**Accuracy**	0.60	0.64	0.72	0.60	0.48	0.56	0.60	-	0.41	0.71
	AUC	0.77	0.71	0.81	0.80	0.75	0.82	0.74	-	0.55	0.90
	Cohen's kappa	0.28	0.19	0.52	0.30	0.19	0.37	0.28	-	-0.12	0.56
Gamma/(theta+delta) ratio	**Accuracy**	0.62	0.63	0.71	0.60	0.44	0.54	0.59	-	0.37	0.77
	AUC	0.79	0.72	0.80	0.80	0.75	0.82	0.75	-	0.54	0.91
	Cohen's kappa	0.32	0.20	0.49	0.32	0.17	0.35	0.26	-	-0.14	0.64
DT	**Accuracy**	0.54	0.41	0.52	0.50	0.47	0.60	0.63	-	0.46	0.55
	AUC	0.67	0.71	0.63	0.59	0.59	0.66	0.63	-	0.61	0.73
	Cohen's kappa	0.24	0.23	0.26	0.23	0.19	0.37	0.33	-	0.11	0.35
SVM	**Accuracy**	0.51	0.34	0.63	0.67	0.58	0.68	0.67	-	0.56	0.56
	AUC	0.78	0.75	0.85	0.80	0.80	0.86	0.82	-	0.68	0.84
	Cohen's kappa	0.21	0.17	0.46	0.45	0.35	0.52	0.42	-	0.23	0.36
XGBoost	**Accuracy**	0.54	0.39	0.58	0.66	0.58	0.69	0.68	-	0.53	0.60
	AUC	0.79	0.77	0.82	0.79	0.78	0.87	0.81	-	0.71	0.84
	Cohen's kappa	0.24	0.22	0.40	0.45	0.36	0.52	0.45	-	0.20	0.41

Table S15. Four-state classification performances of the index-based and machine learning models per PICU patient. In some patients, only three sleep stages were distinguished, thereby hampering the individual assessment of four-state classification performance (marked by '-'). In patient H, only two sleep stages were distinguished. Therefore, classification performance could not be individually assessed for this patient. a: Artifact labels were manually removed since they were detected as artifacts due to the baseline drift in the EEG signal. This baseline drift could be removed with the bandpass filter. b: The F4-C4 channel was used instead of the F3-C3 channel. AUC = area under the (receiver operating characteristic) curve, DT = decision tree, SVM = support vector machine, XGBoost = extreme gradient boosting

	Performance metric	Patient A	Patient B	Patient C	Patient D[a]	Patient E	Patient F	Patient G[b]	Patient H	Patient I[b]	Patient J
Four-state classification											
Gamma/delta ratio	**Accuracy**	0.59	-	0.71	-	0.38	-	0.59	-	0.37	0.70
	AUC	0.77	-	0.82	-	0.66	-	0.74	-	0.55	0.89
	Cohen's kappa	0.27	-	0.50	-	0.12	-	0.28	-	-0.11	0.54
Gamma/(theta+delta) ratio	**Accuracy**	0.62	-	0.70	-	0.35	-	0.57	-	0.34	0.76
	AUC	0.78	-	0.81	-	0.66	-	0.76	-	0.54	0.90
	Cohen's kappa	0.31	-	0.48	-	0.11	-	0.26	-	-0.14	0.62
DT	**Accuracy**	0.36	-	0.50	-	0.44	-	0.36	-	0.38	0.49
	AUC	0.60	-	0.71	-	0.65	-	0.59	-	0.62	0.75
	Cohen's kappa	0.08	-	0.32	-	0.20	-	0.12	-	0.10	0.08
SVM	**Accuracy**	0.46	-	0.60	-	0.48	-	0.32	-	0.42	0.61
	AUC	0.67	-	0.78	-	0.74	-	0.67	-	0.63	0.83
	Cohen's kappa	0.17	-	0.42	-	0.28	-	0.14	-	0.13	0.43
XGBoost	**Accuracy**	0.48	-	0.60	-	0.47	-	0.33	-	0.46	0.52
	AUC	0.66	-	0.78	-	0.74	-	0.69	-	0.71	0.82
	Cohen's kappa	0.19	-	0.40	-	0.25	-	0.11	-	0.17	0.32

Table S16. Classification performance (training score, accuracy) per PCIU patient for the individually trained models, three-state classification. Individualized thresholds were not determined for patient H since there were only two stages distinguished in this patient. a: Artifact labels were manually removed since they were detected as artifacts due to the baseline drift in the EEG signal. This baseline drift could be removed with the bandpass filter. b: The F4-C4 channel was used instead of the F3-C3 channel. AUC = area under the receiver operating characteristic (ROC) curve, CV = cross-validation

	Performance metric	Patient A	Patient B	Patient C	Patient D[a]	Patient E	Patient F	Patient G[b]	Patient H	Patient I[b]	Patient J
Gamma/delta ratio – individual thresholds	**Accuracy**	0.67	0.68	0.79	0.87	0.67	0.66	0.63	-	0.30	0.81
	AUC	0.86	0.76	0.94	0.94	0.81	0.85	0.78	-	0.70	0.96
	Cohen's kappa	0.34	0.29	0.62	0.73	0.38	0.49	0.32	-	-0.11	0.69
Gamma/(theta+delta) ratio – individual thresholds	**Accuracy**	0.67	0.70	0.79	0.86	0.68	0.66	0.65	-	0.30	0.81
	AUC	0.81	0.75	0.83	0.86	0.75	0.83	0.77	-	0.69	0.92
	Cohen's kappa	0.35	0.34	0.62	0.70	0.39	0.50	0.35	-	-0.11	0.69

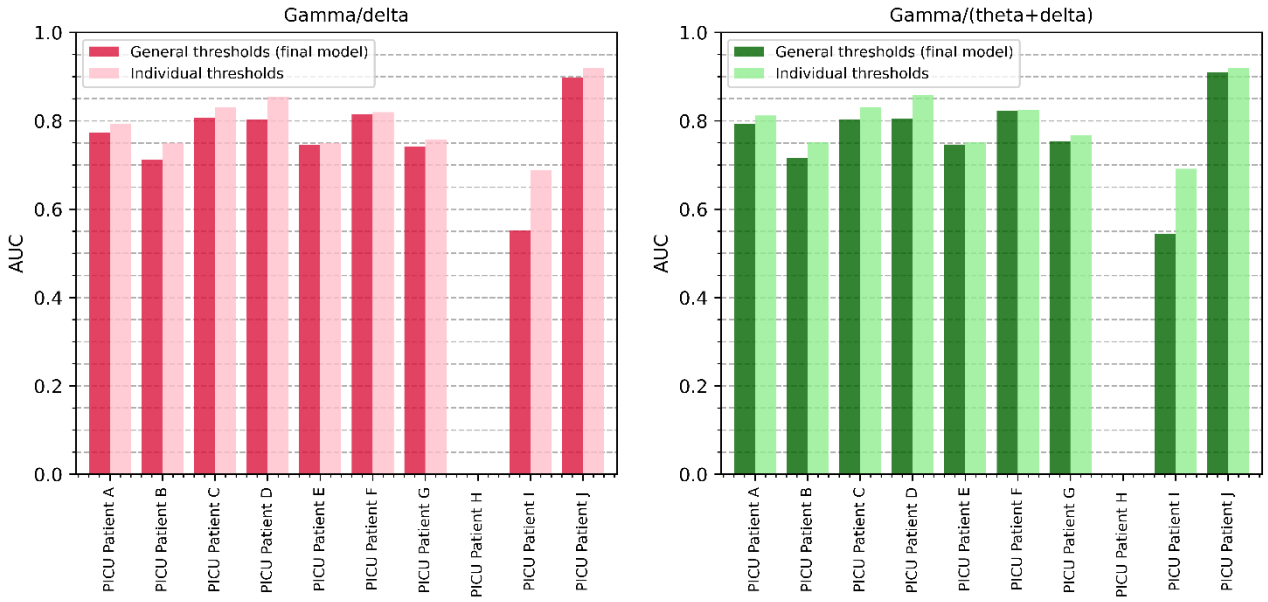


Figure S4. Classification performance (AUC) with individualized versus general thresholds per PICU patient for three-state classification. AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit.

Patient A

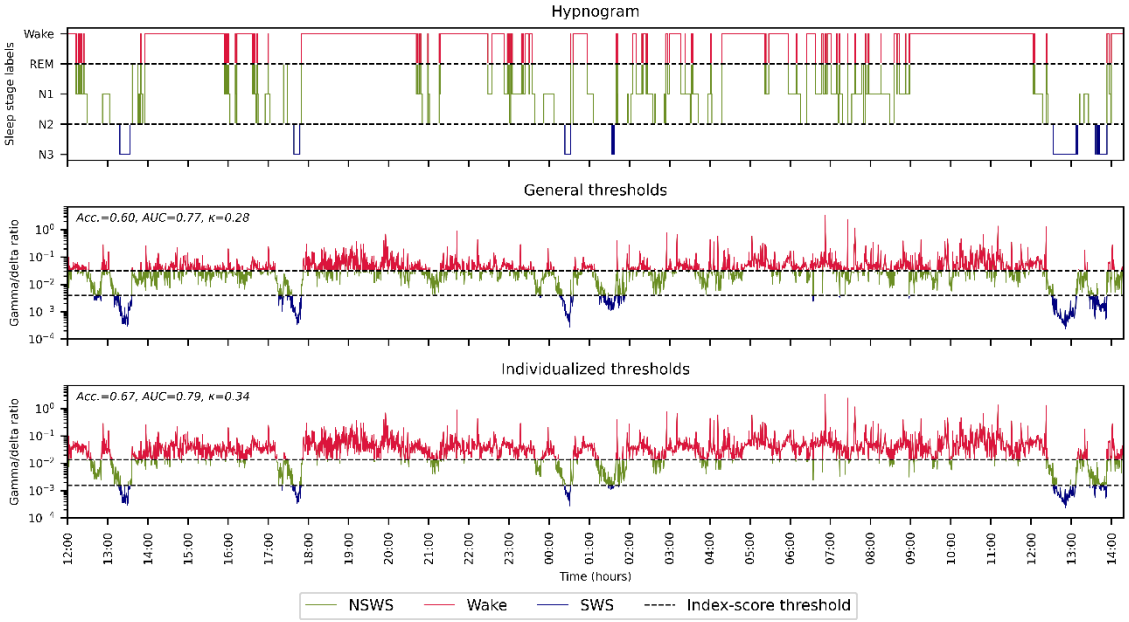


Figure S5. PICU patient A - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

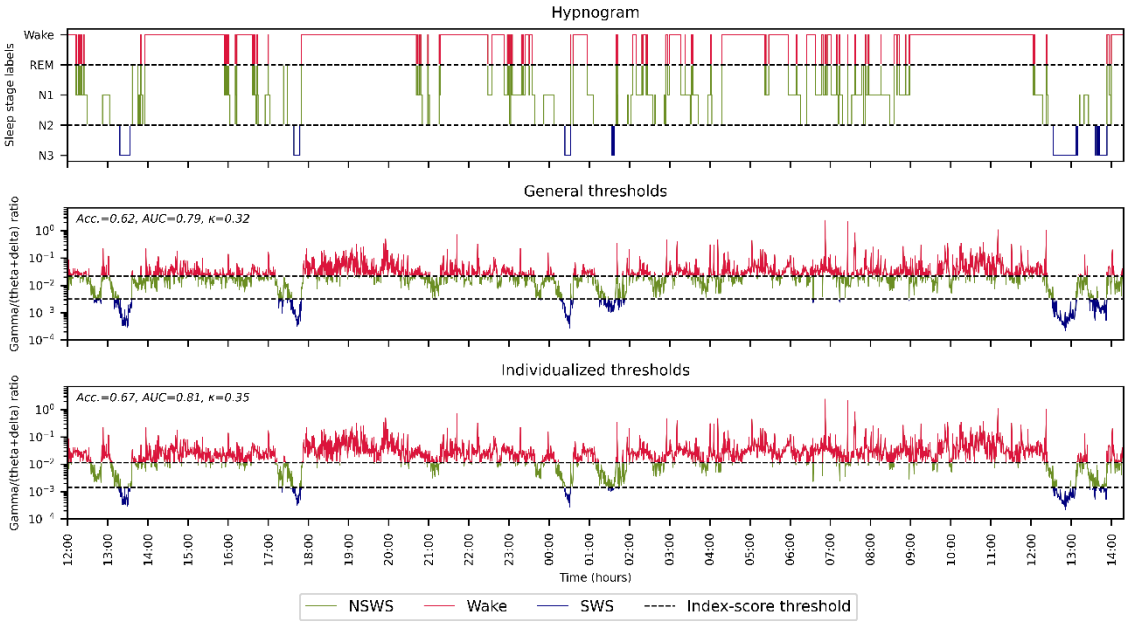


Figure S6. PICU patient A - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

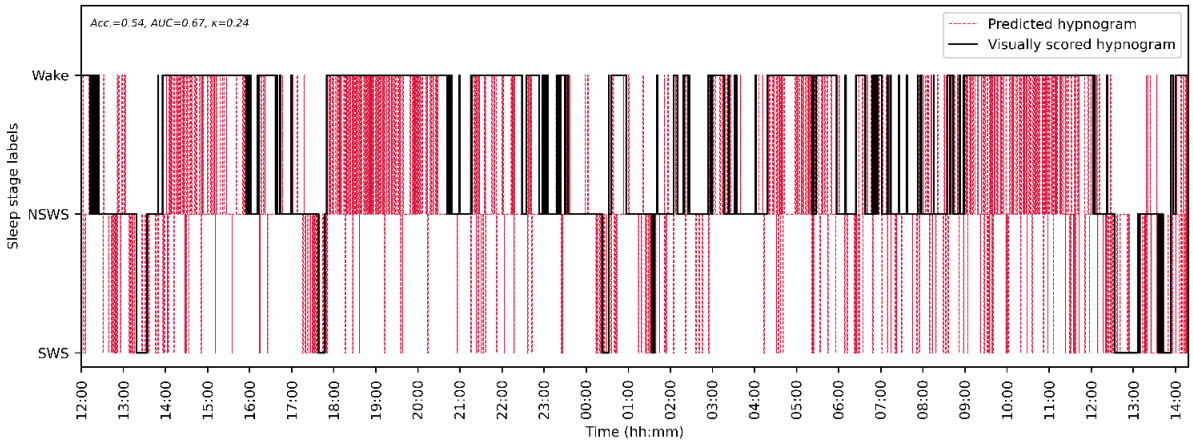


Figure S7. PICU patient A - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

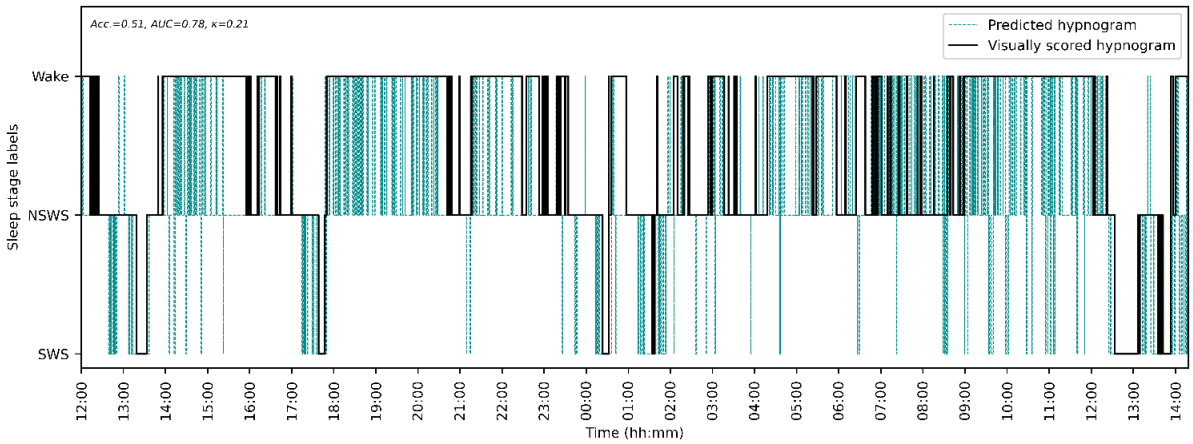


Figure S8. PICU patient A - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

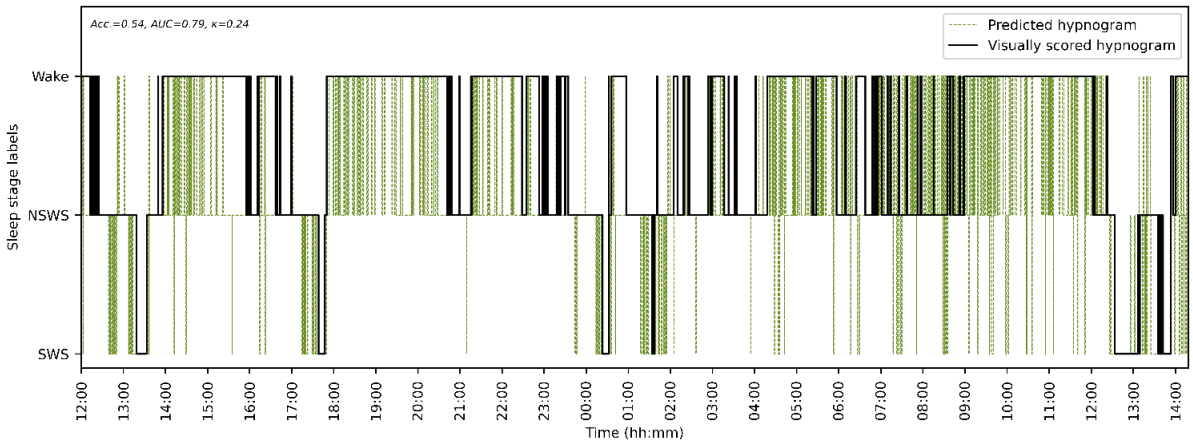


Figure S9. PICU patient A - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Patient B

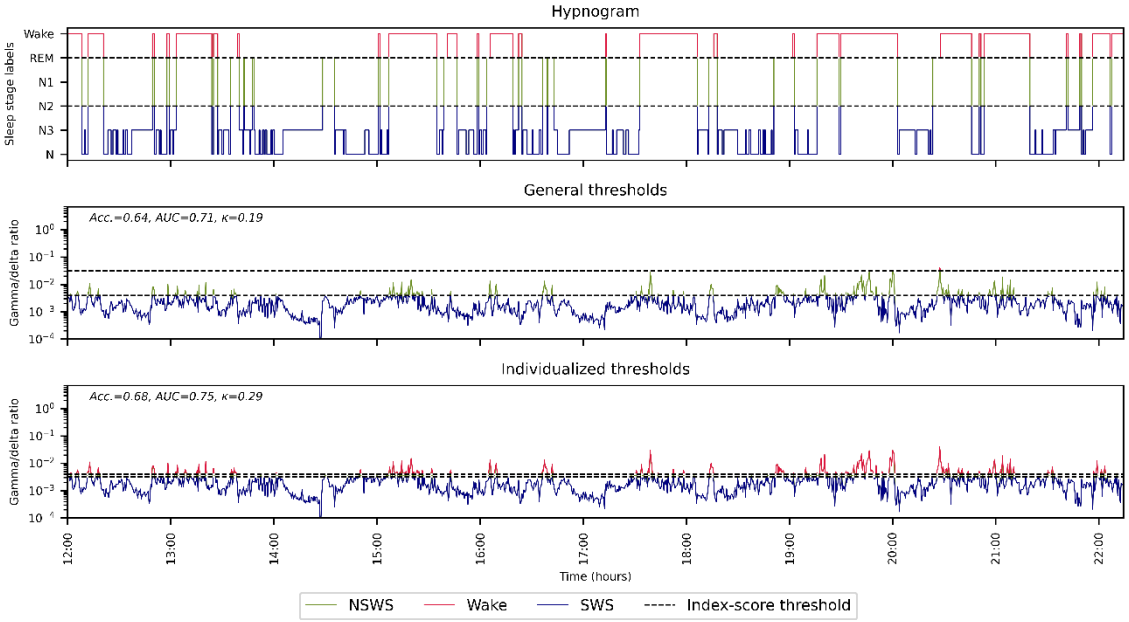


Figure S10. PICU patient B - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

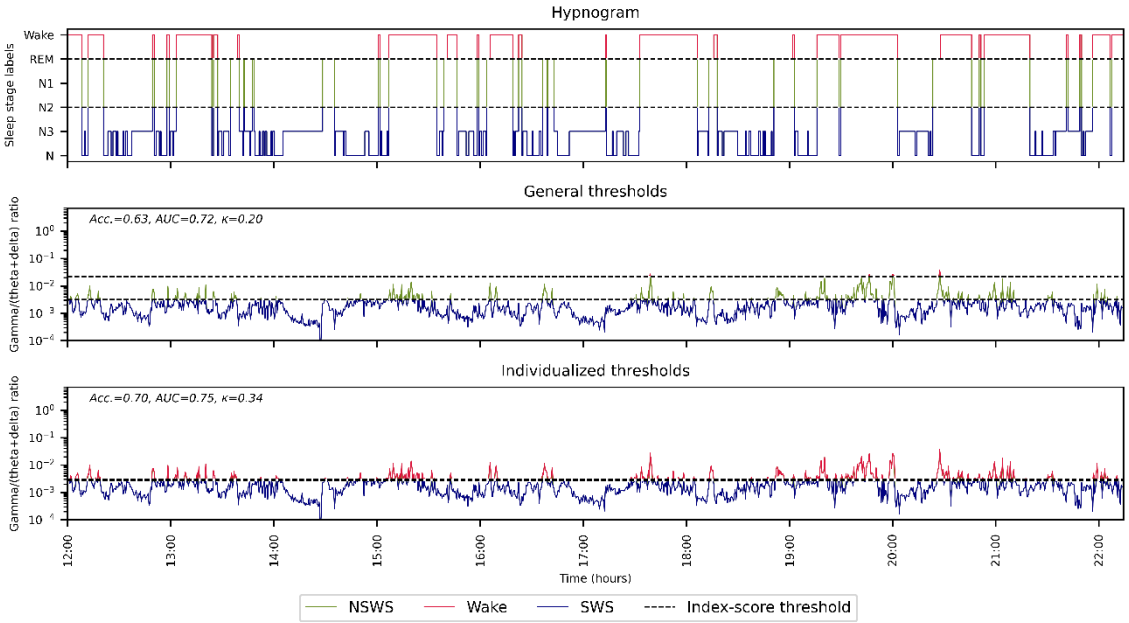


Figure S11. PICU patient B - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

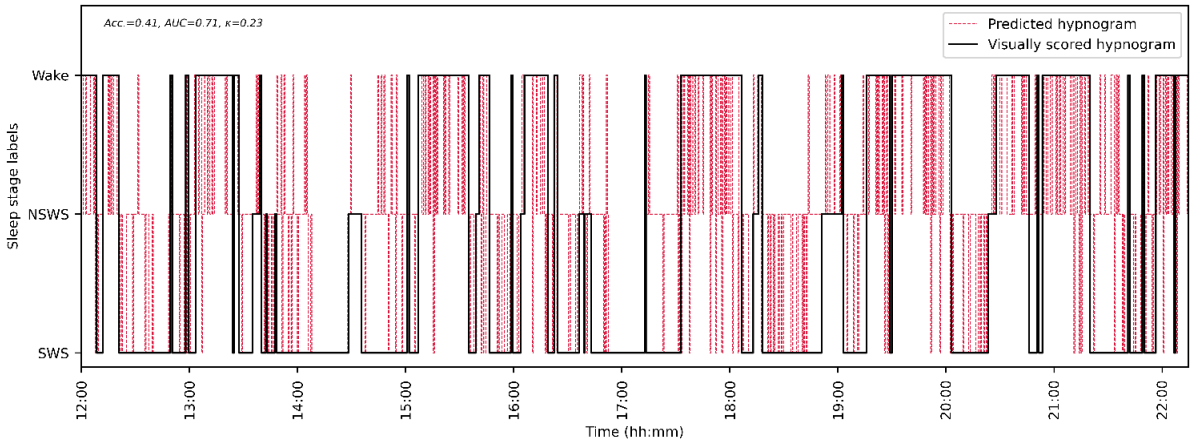


Figure S12. PICU patient B - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

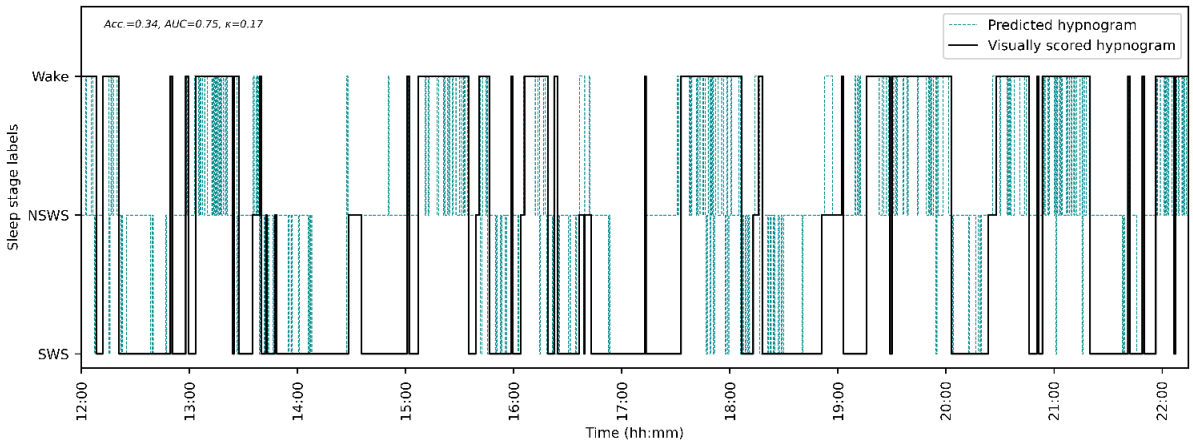


Figure S13. PICU patient B - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

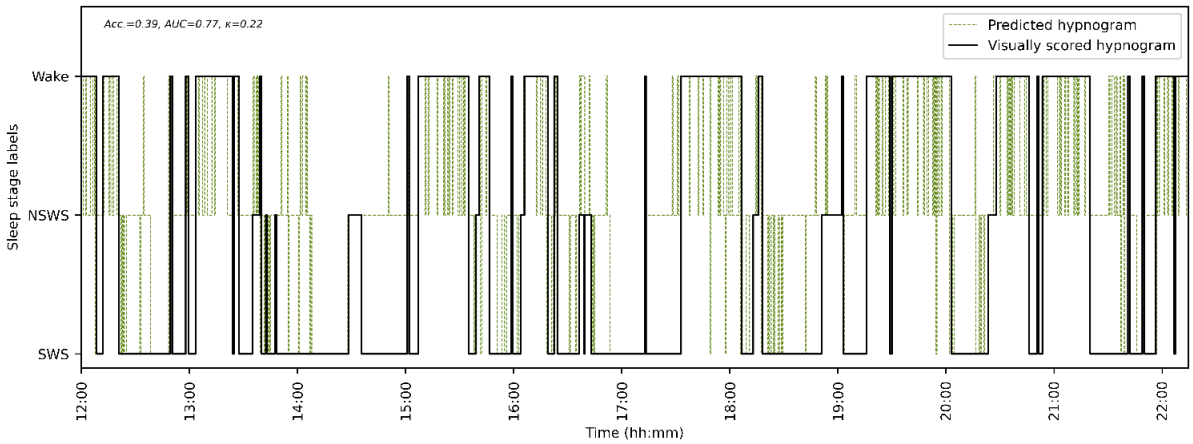


Figure S14. PICU patient C - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Patient C

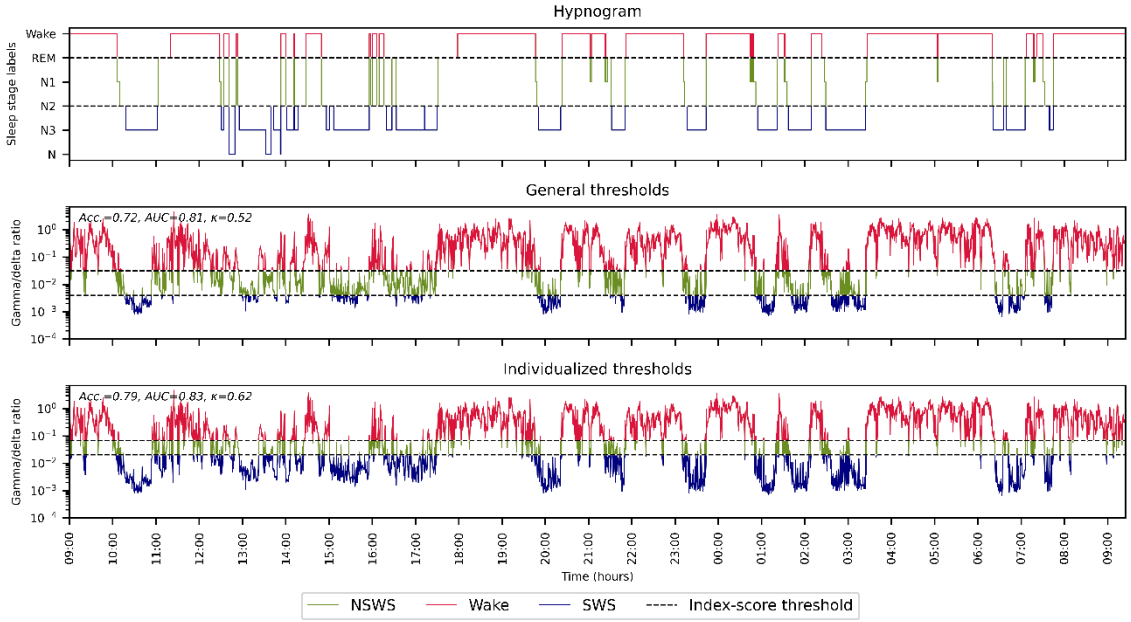


Figure S15. PICU patient C - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

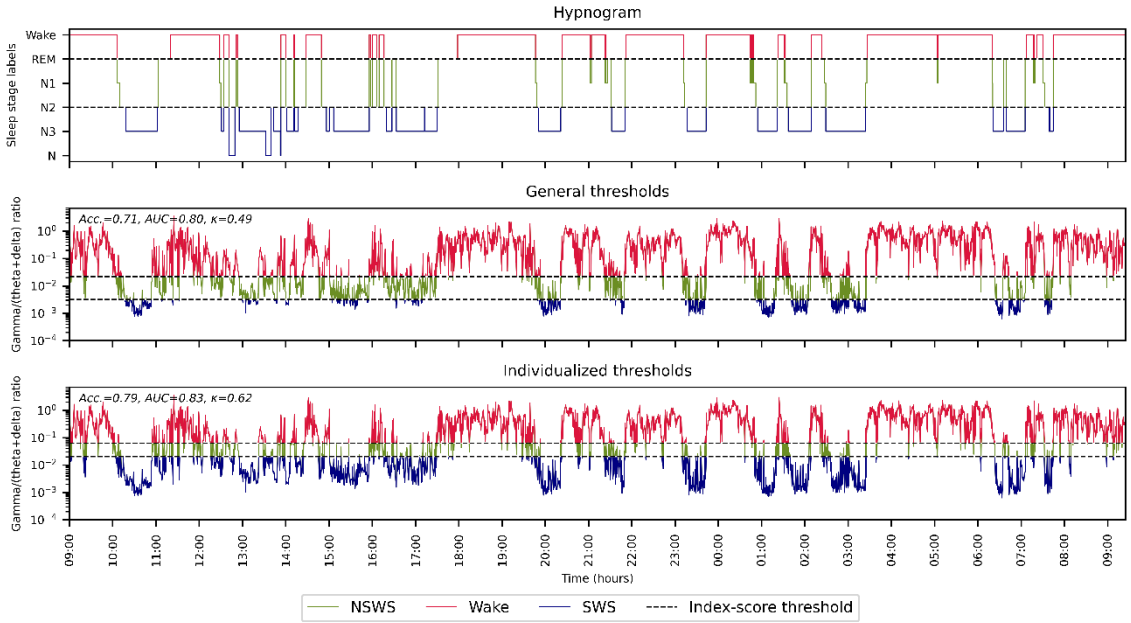


Figure S16. PICU patient C - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

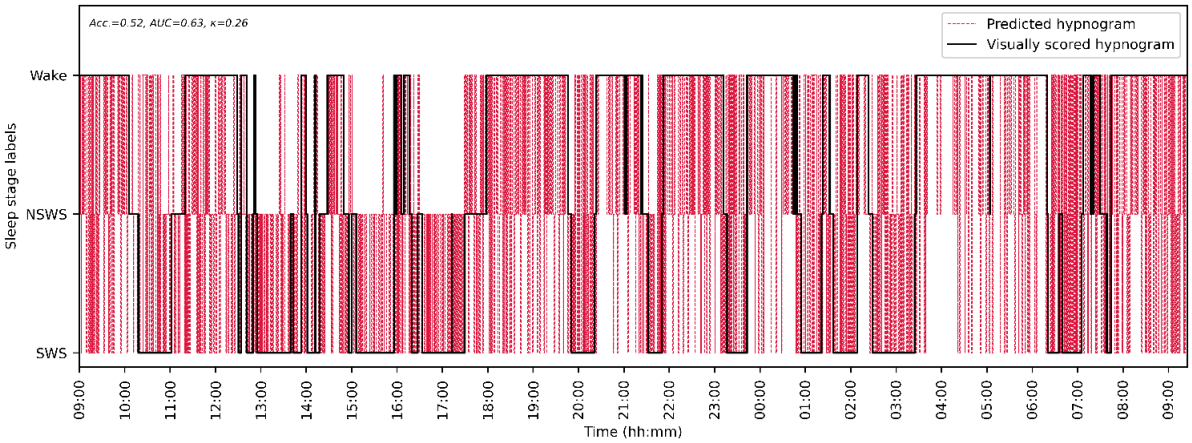


Figure S17. PICU patient C - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

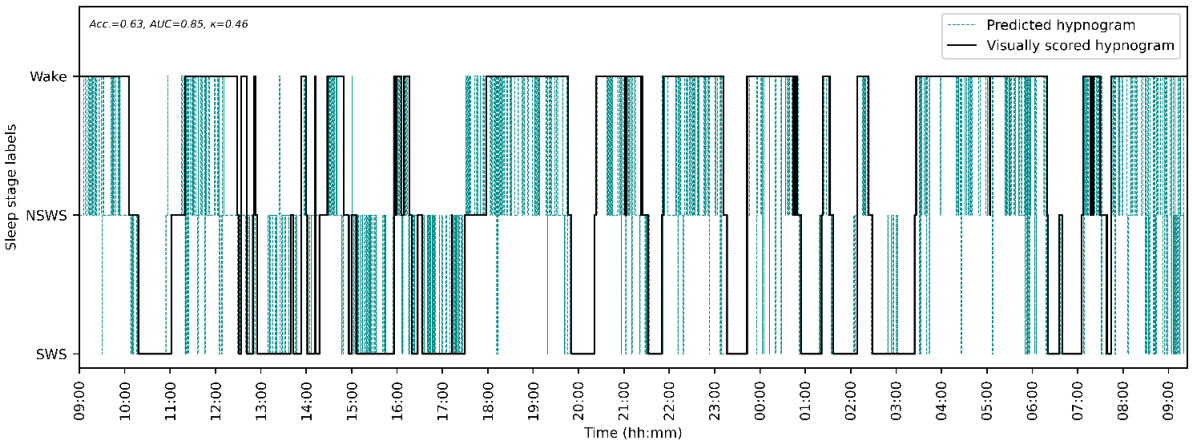


Figure S18. PICU patient C - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

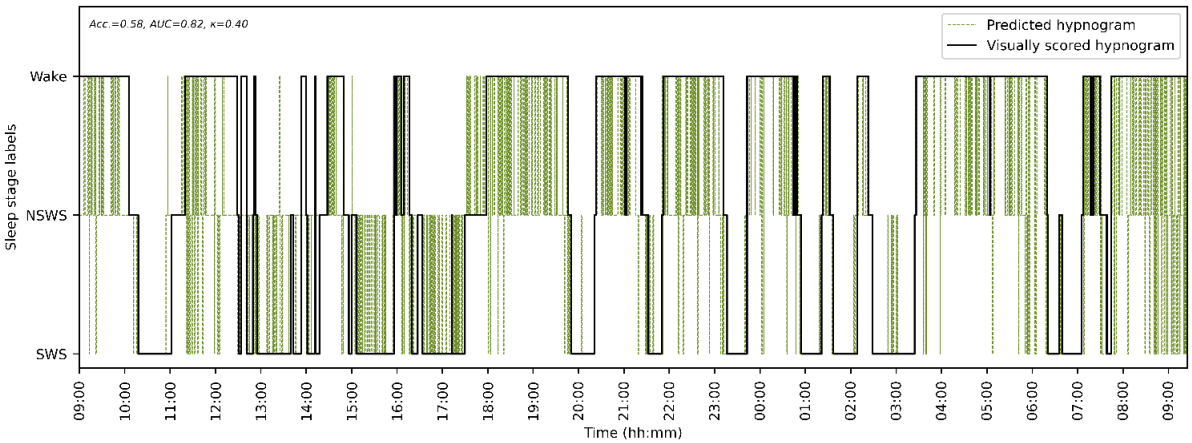


Figure S19. PICU patient C - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Patient D

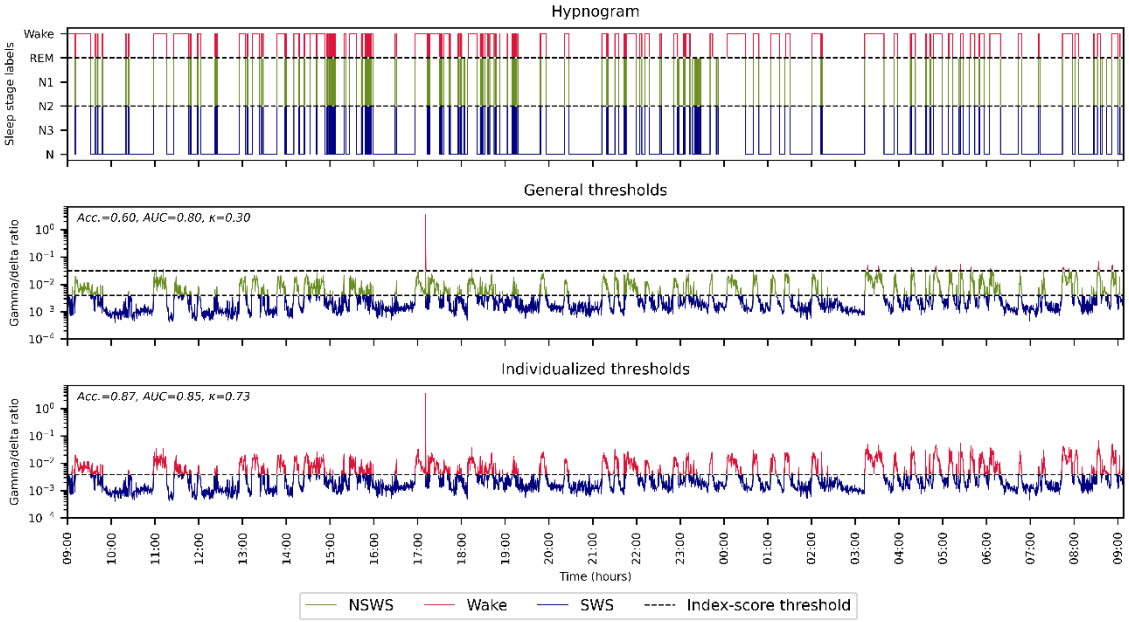


Figure S20. PICU patient D - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

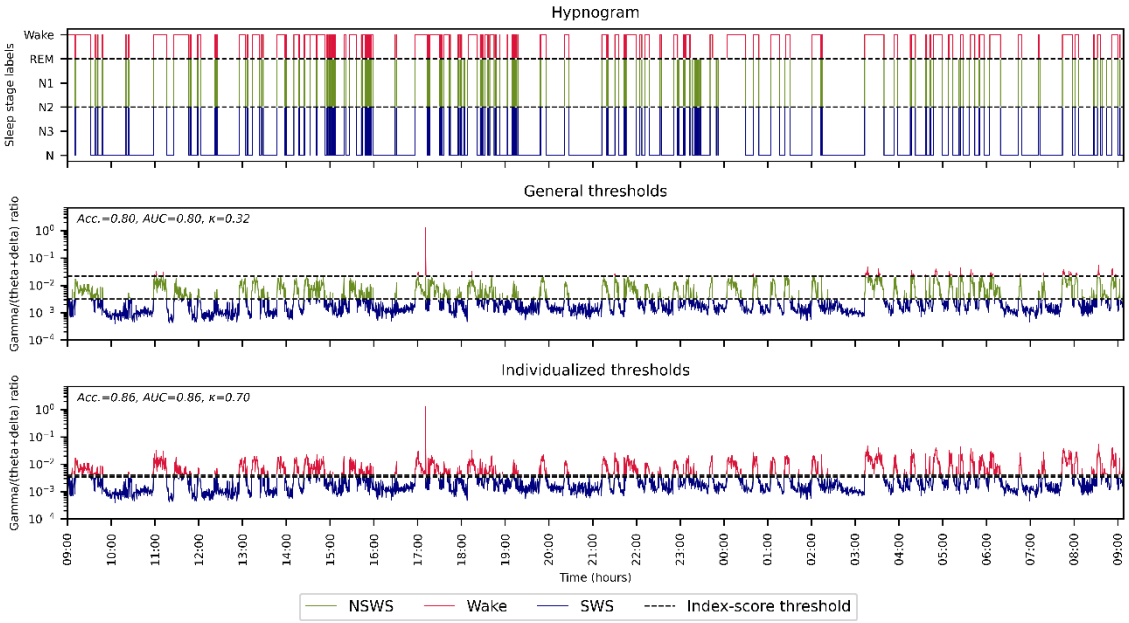


Figure S21. PICU patient D - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

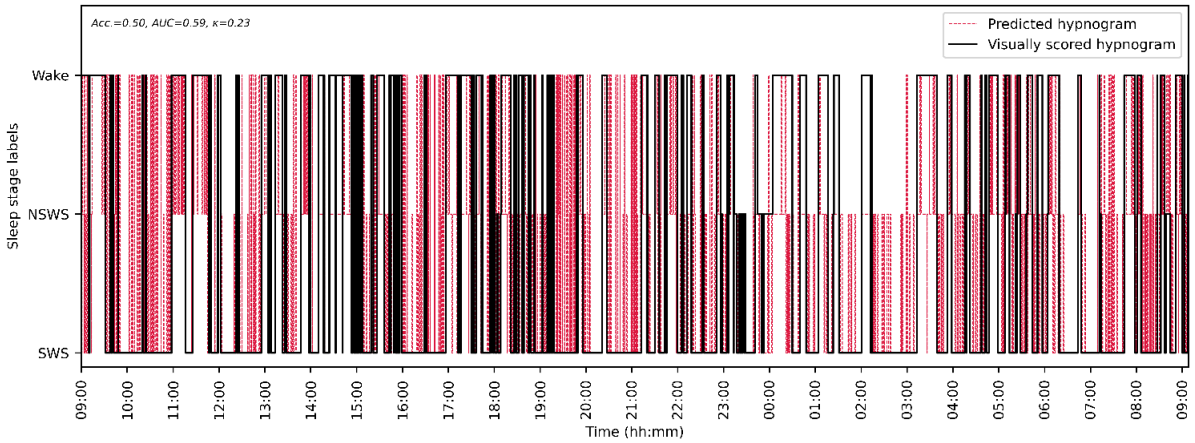


Figure S22. PICU patient D - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

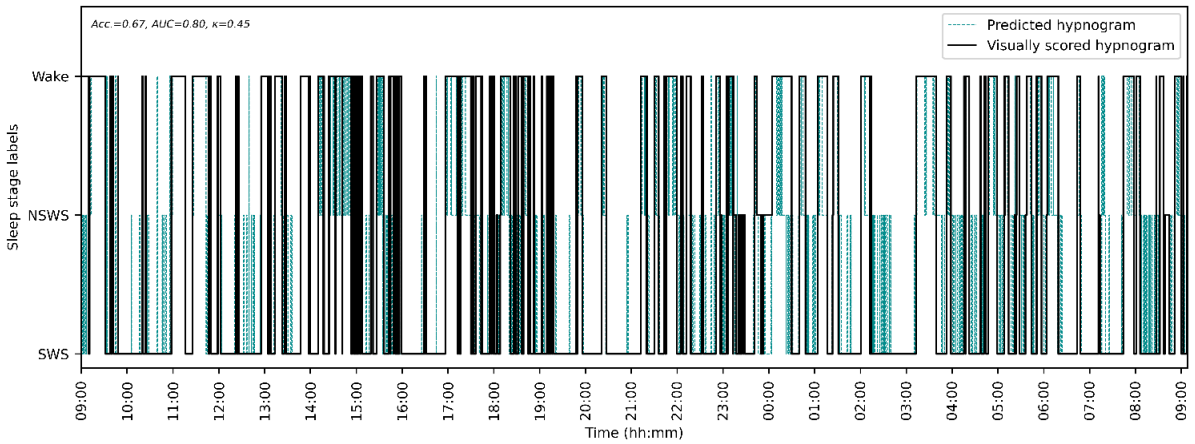


Figure S23. PICU patient D - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

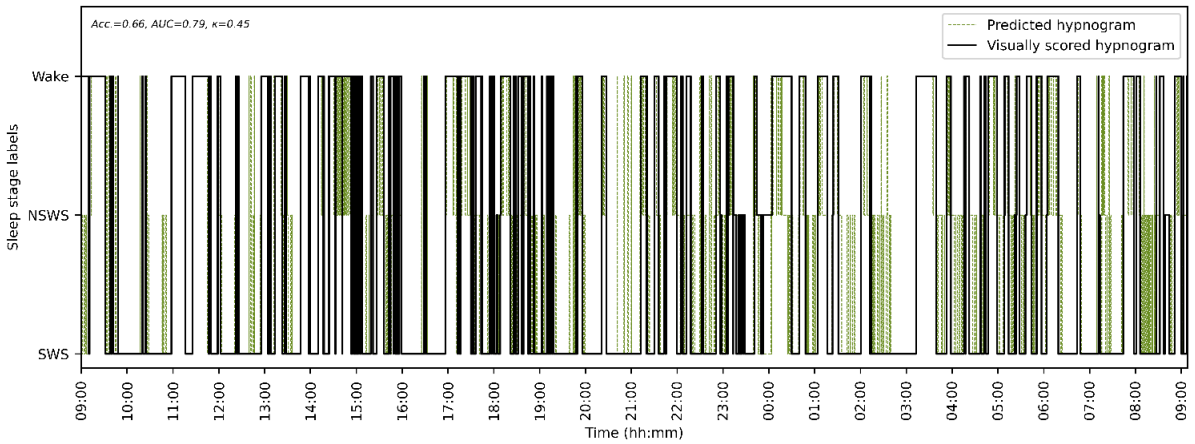


Figure S24. PICU patient D - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Patient E

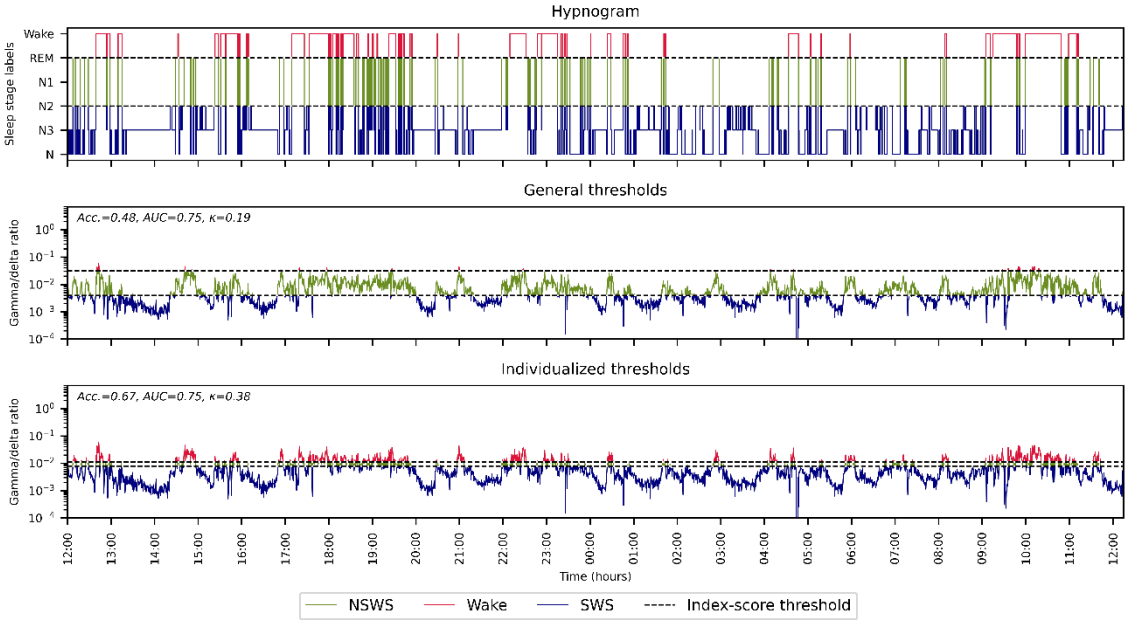


Figure S25. PICU patient E - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

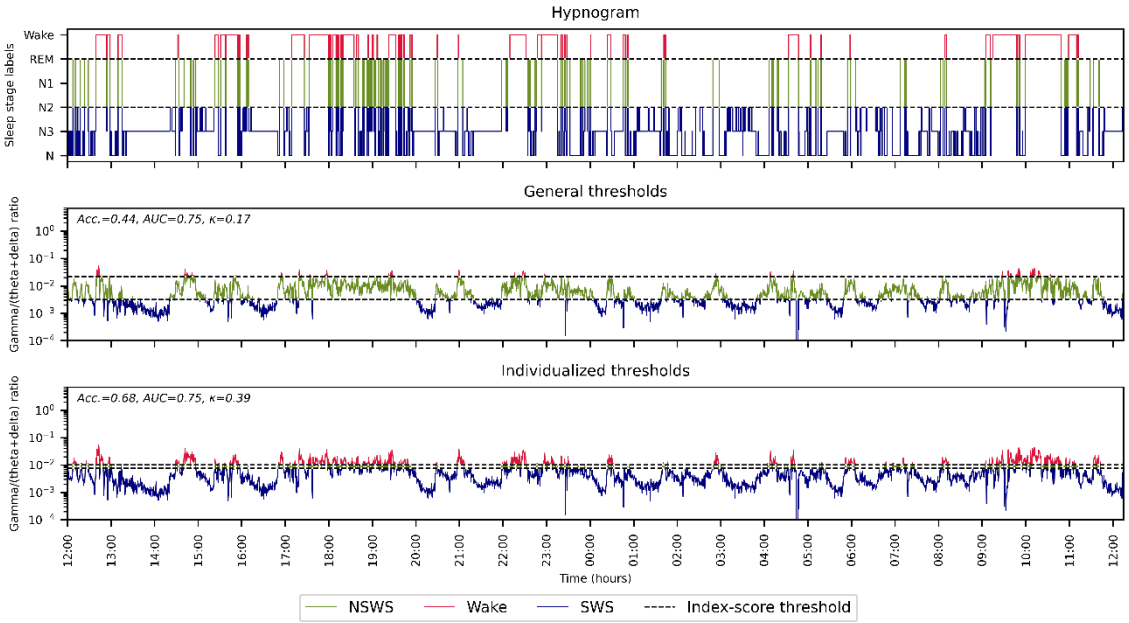


Figure S26. PICU patient E - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

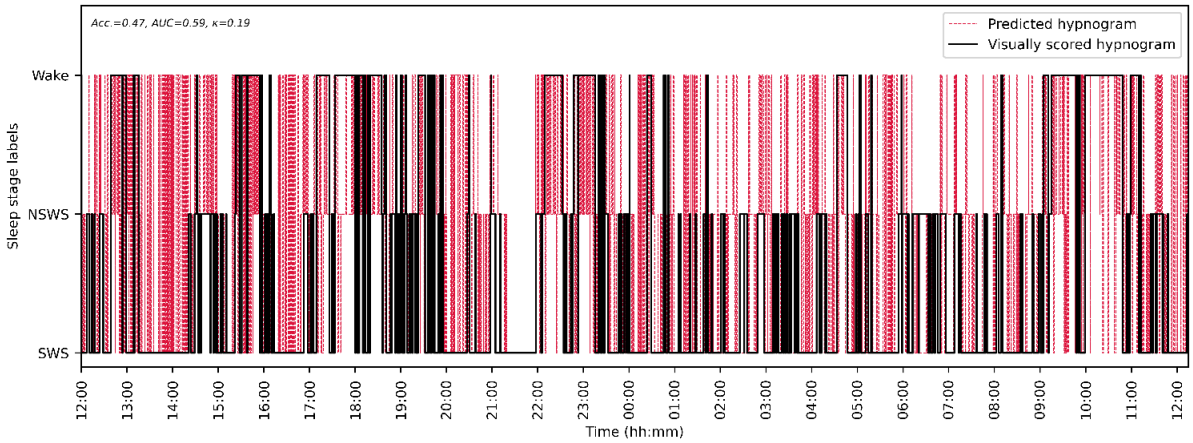


Figure S27. PICU patient E - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

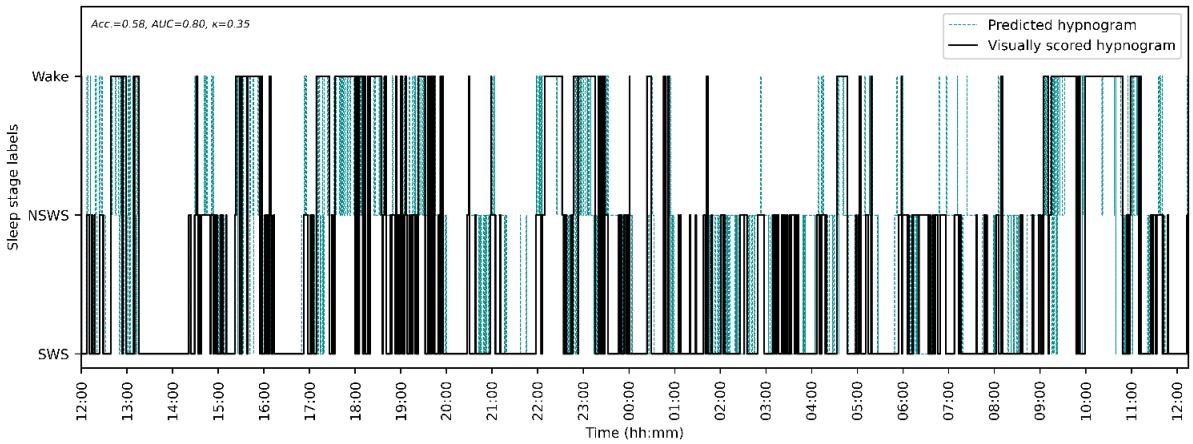


Figure S28. PICU patient E - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

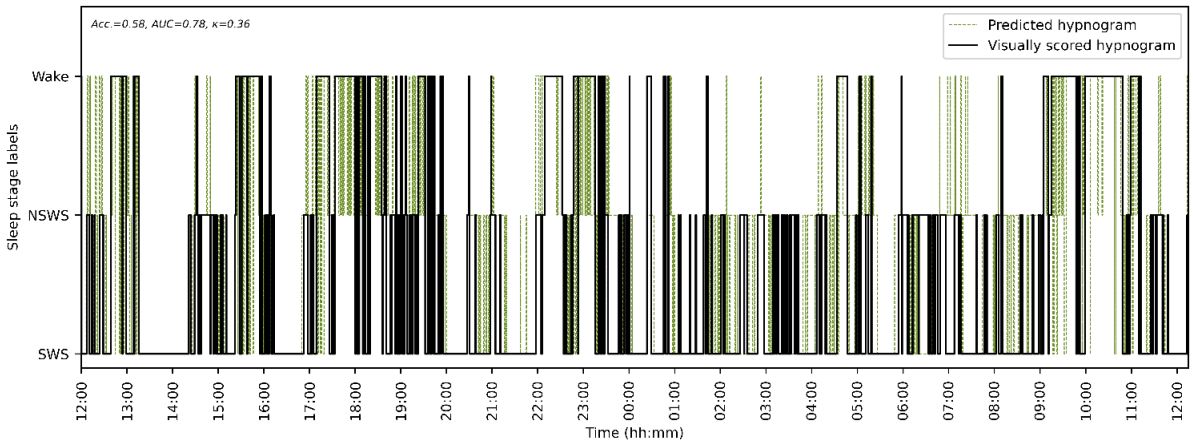


Figure S29. PICU patient E - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Patient F

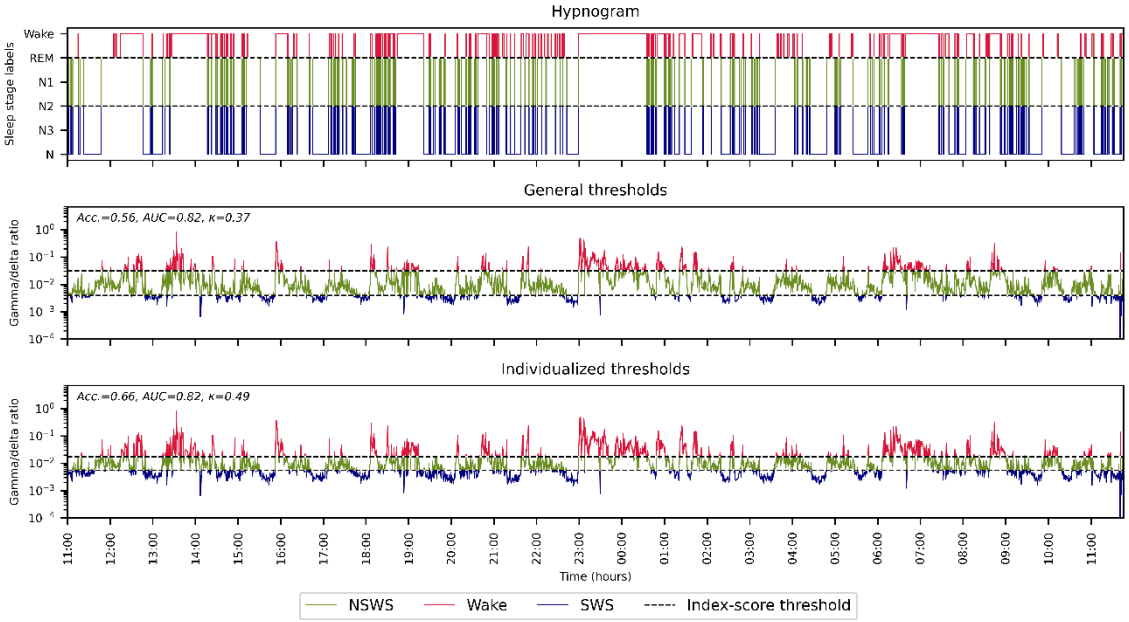


Figure S30. PICU patient F - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

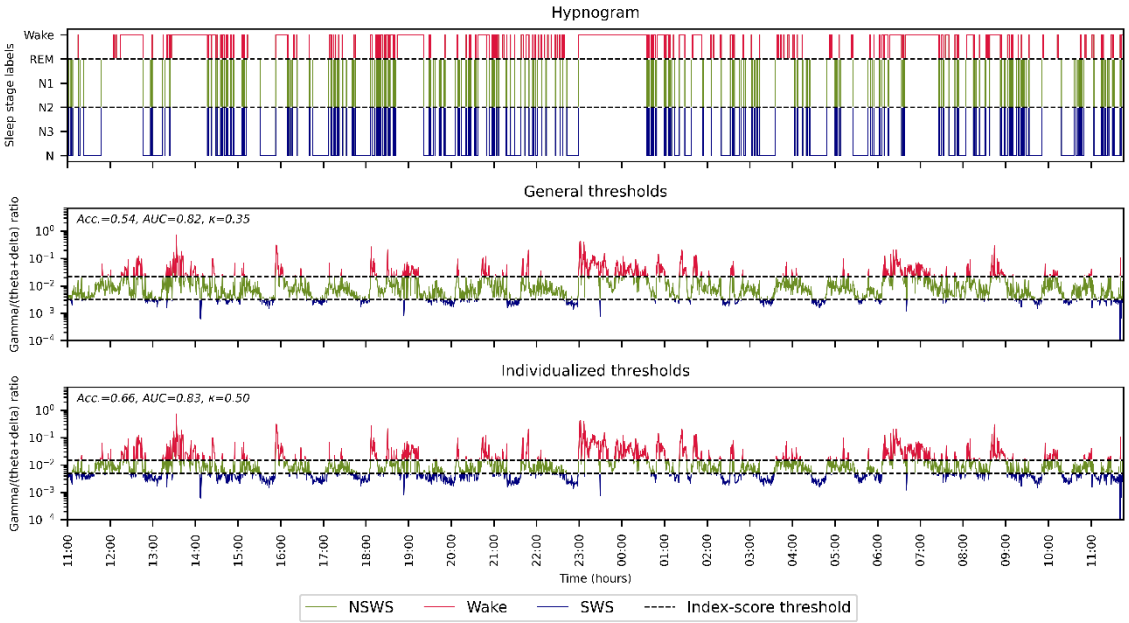


Figure S31. PICU patient F - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

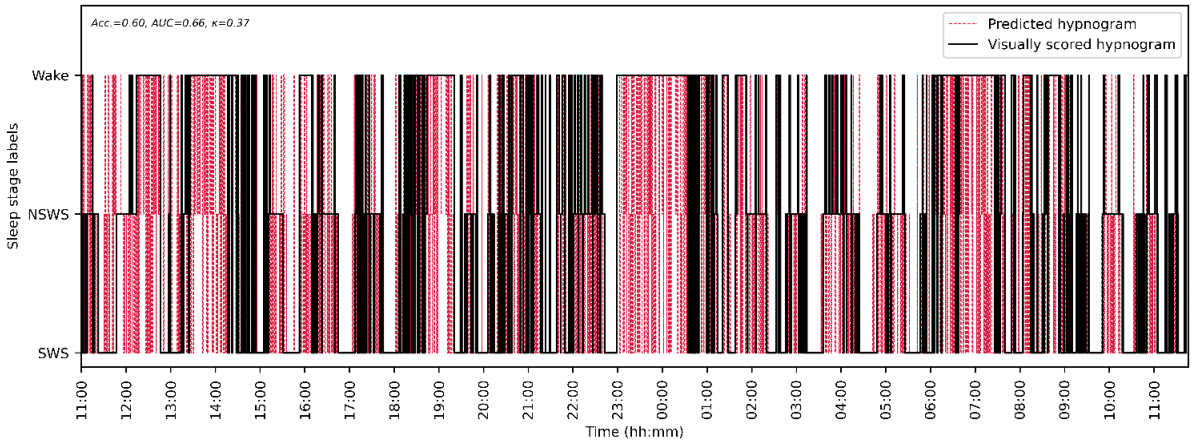


Figure S32. PICU patient F - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

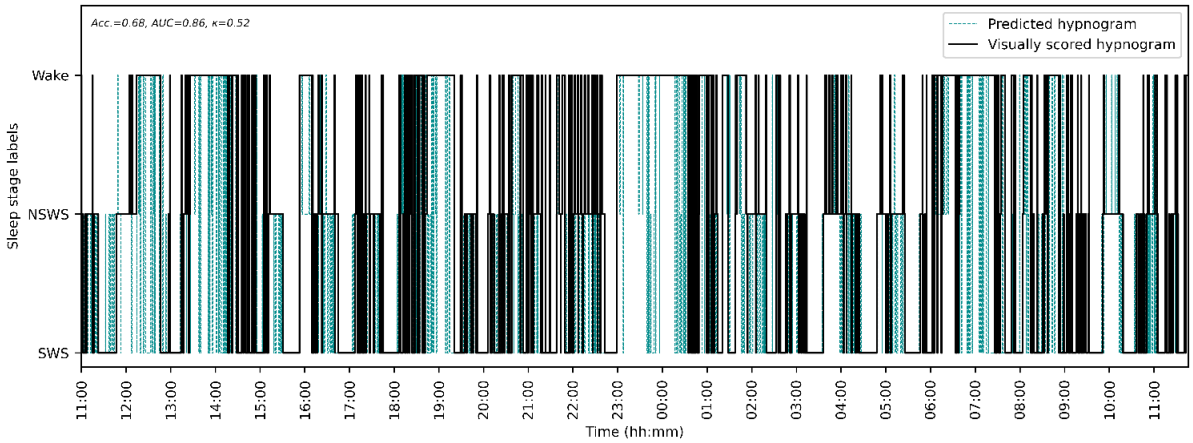


Figure S33. PICU patient F - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

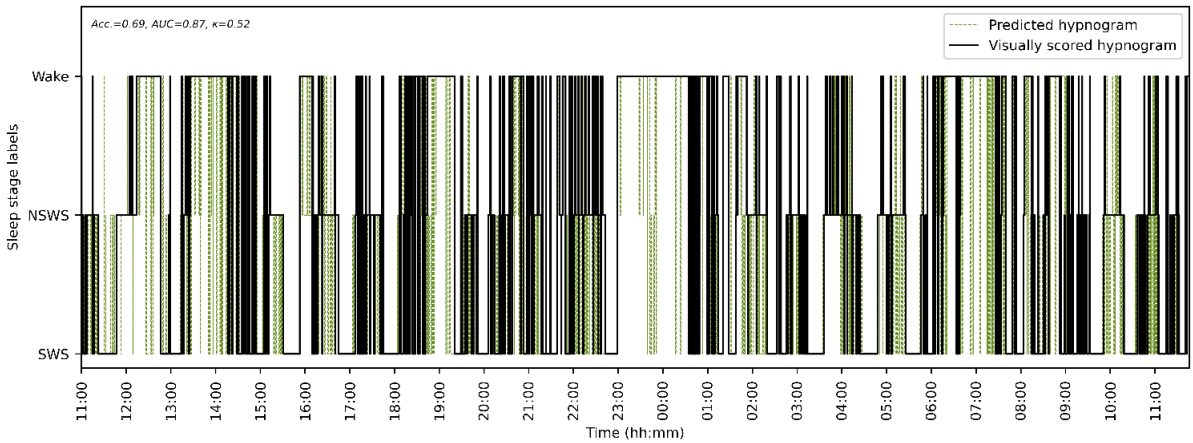


Figure S34. PICU patient F - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Patient G

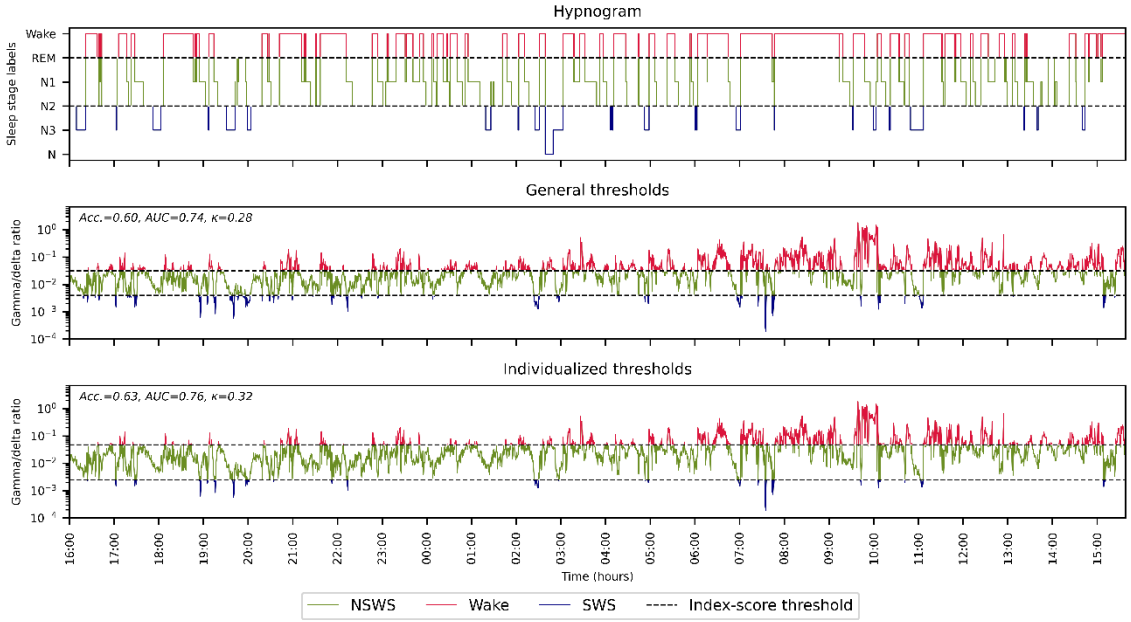


Figure S35. PICU patient G - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

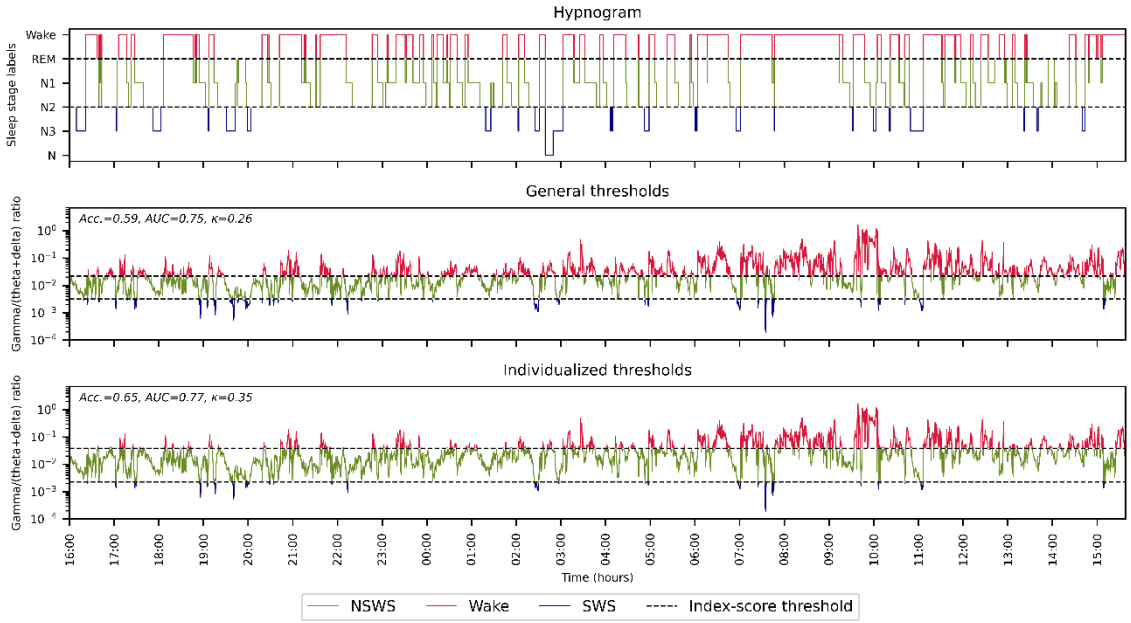


Figure S36. PICU patient G - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

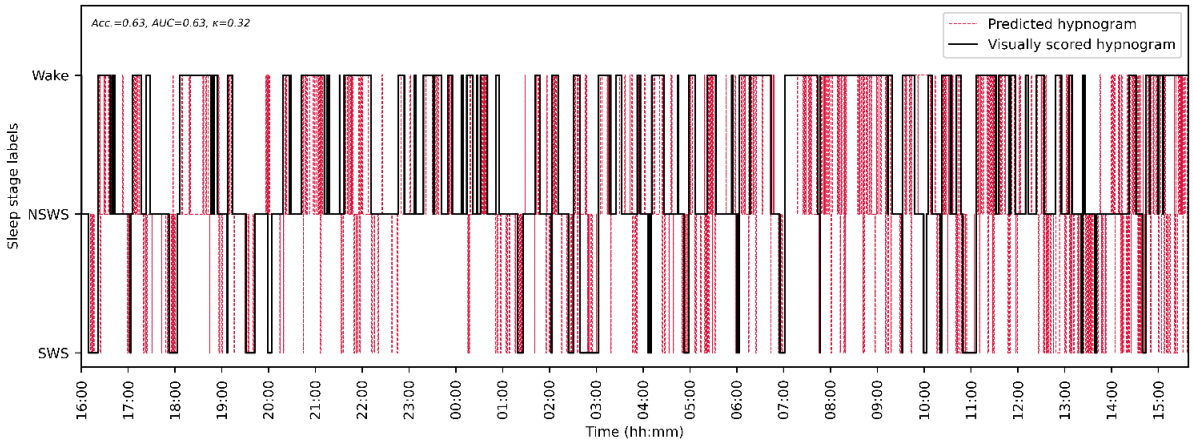


Figure S37. PICU patient G - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

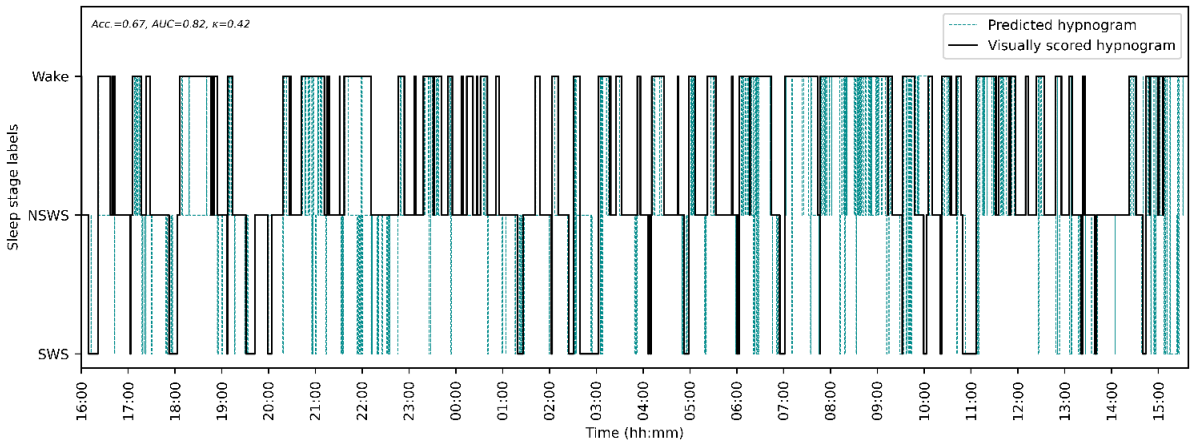


Figure S38. PICU patient G - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

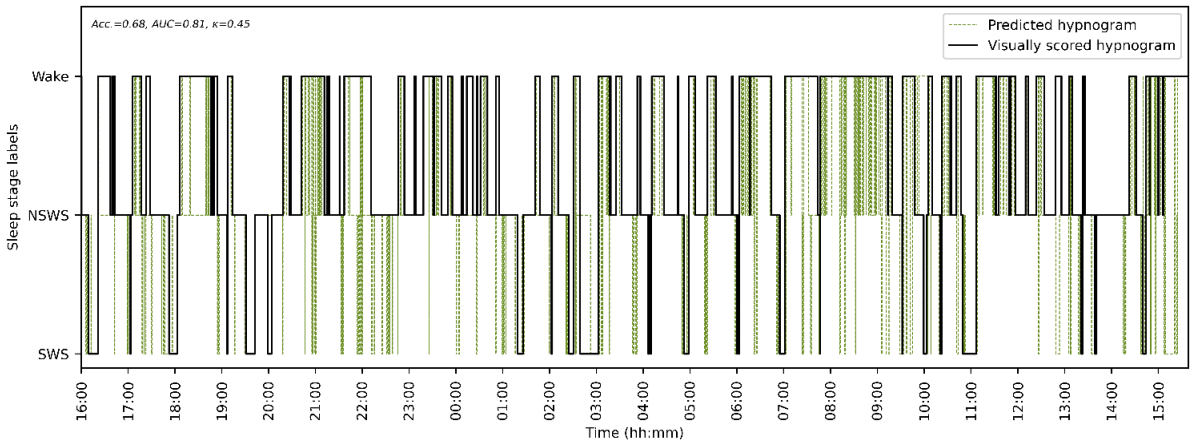


Figure S39. PICU patient G - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Patient H

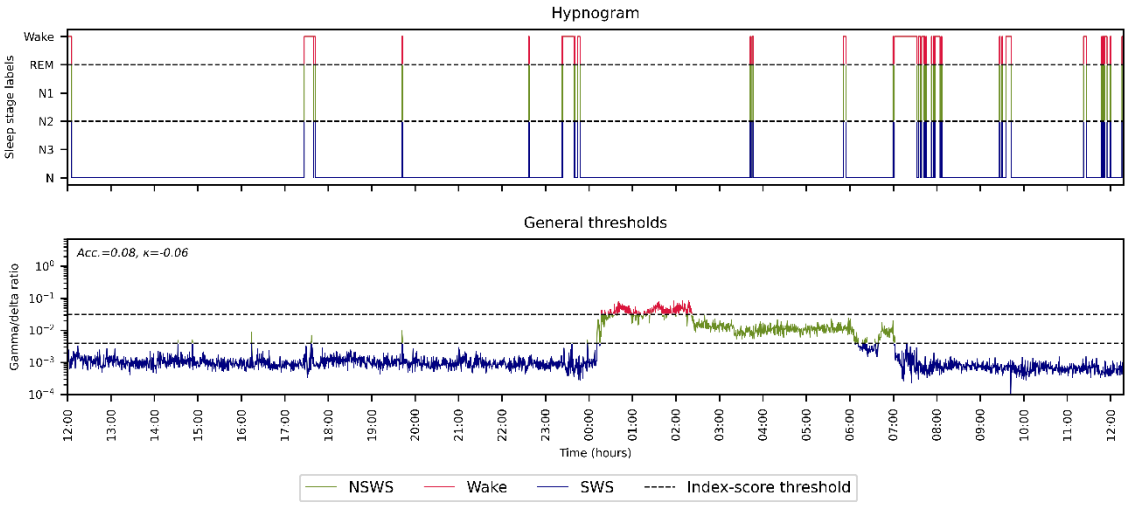


Figure S40. PICU patient H - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

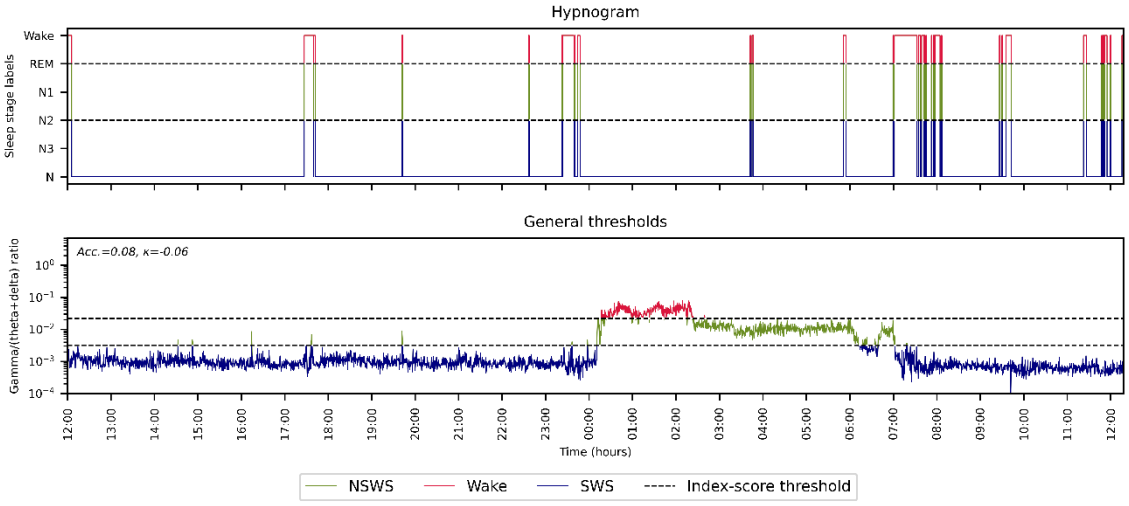


Figure S41. PICU patient H - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

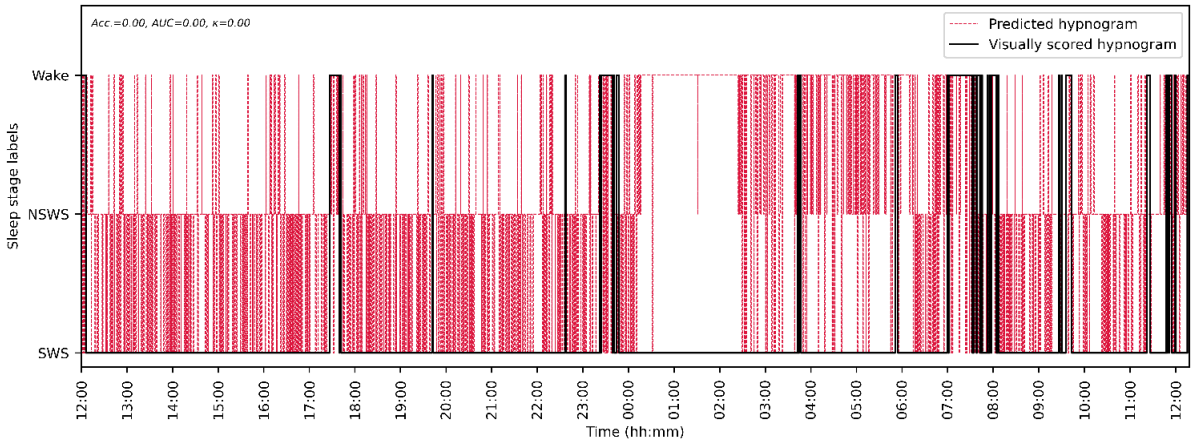


Figure S42. PICU patient H - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

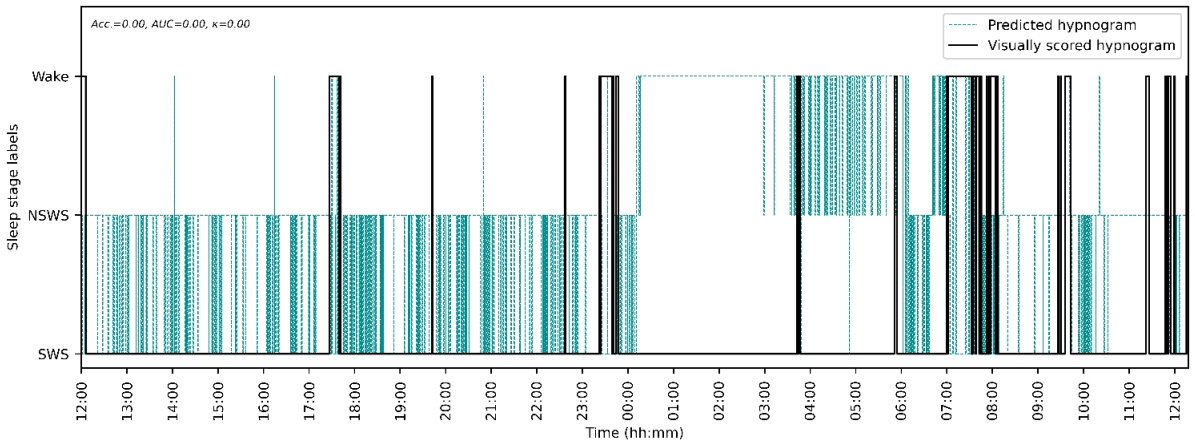


Figure S43. PICU patient H - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

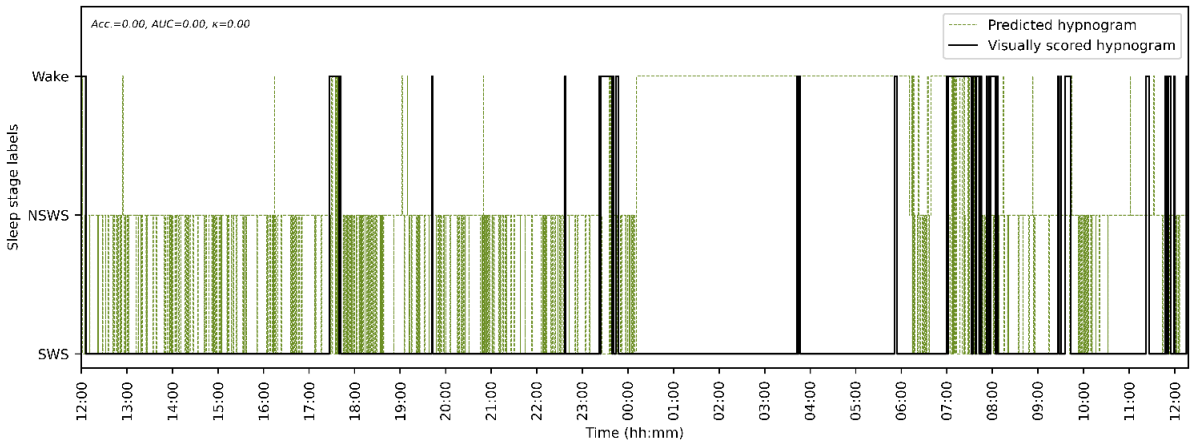


Figure S44. PICU patient H - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Patient I

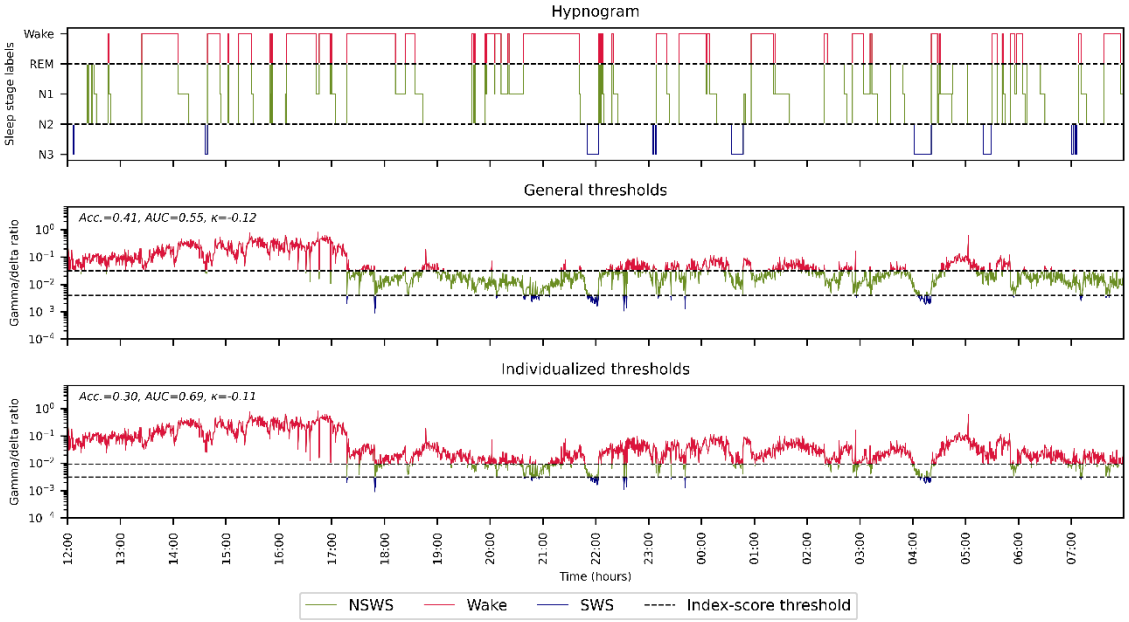


Figure S45. PICU patient I - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

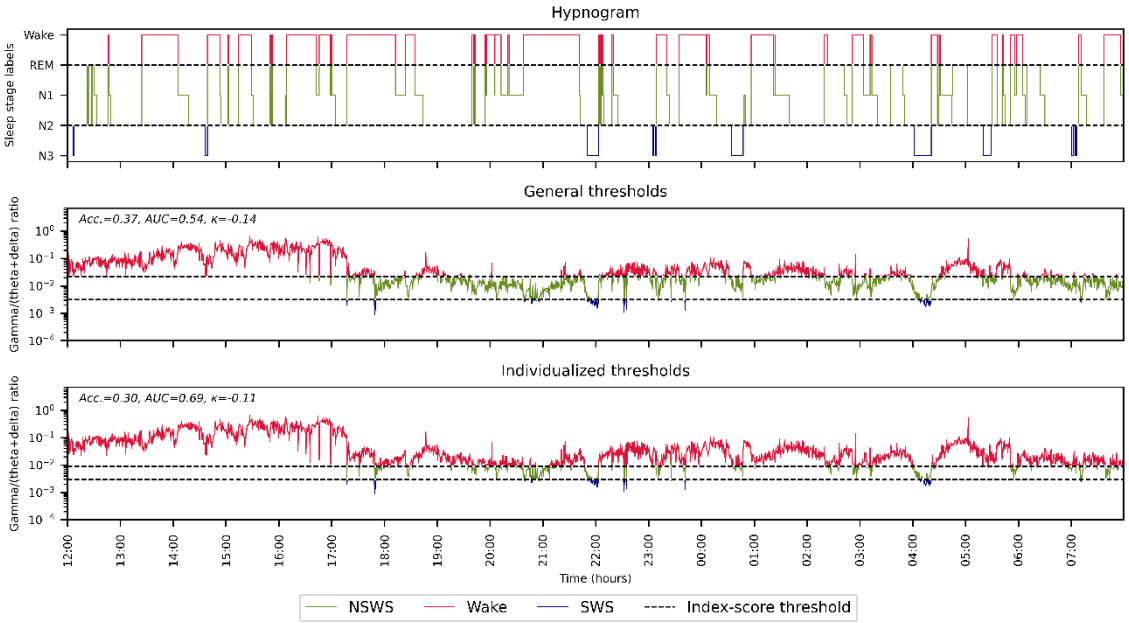


Figure S46. PICU patient I - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

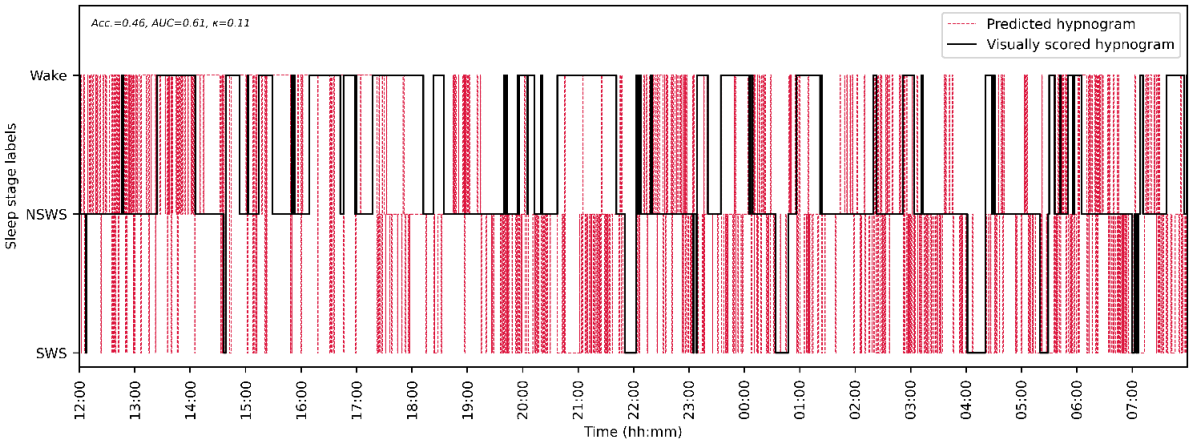


Figure S47. PICU patient I - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

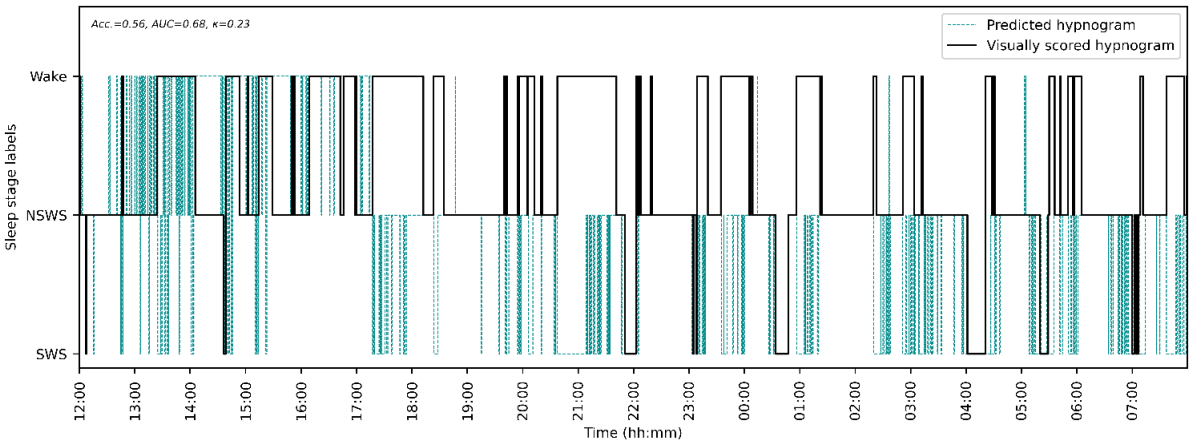


Figure S48. PICU patient I - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

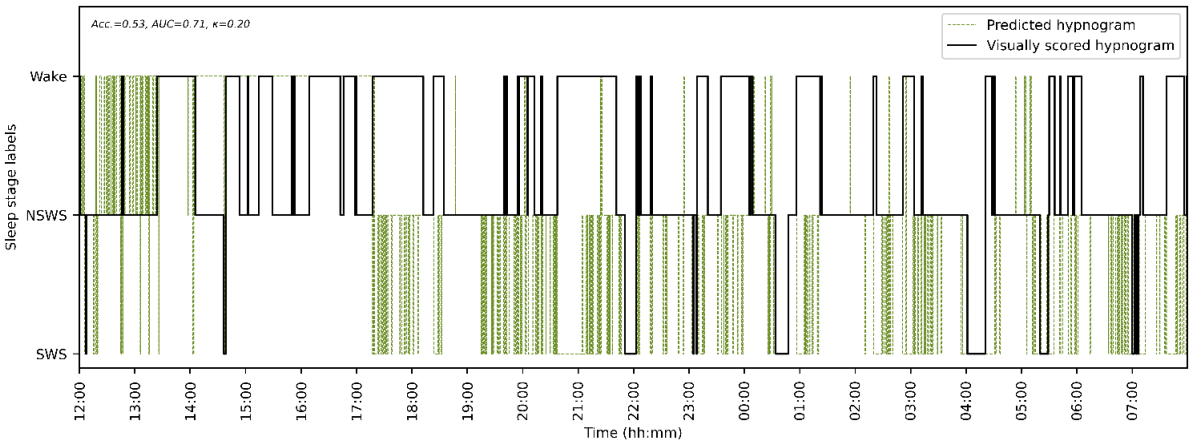


Figure S49. PICU patient I - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Patient J

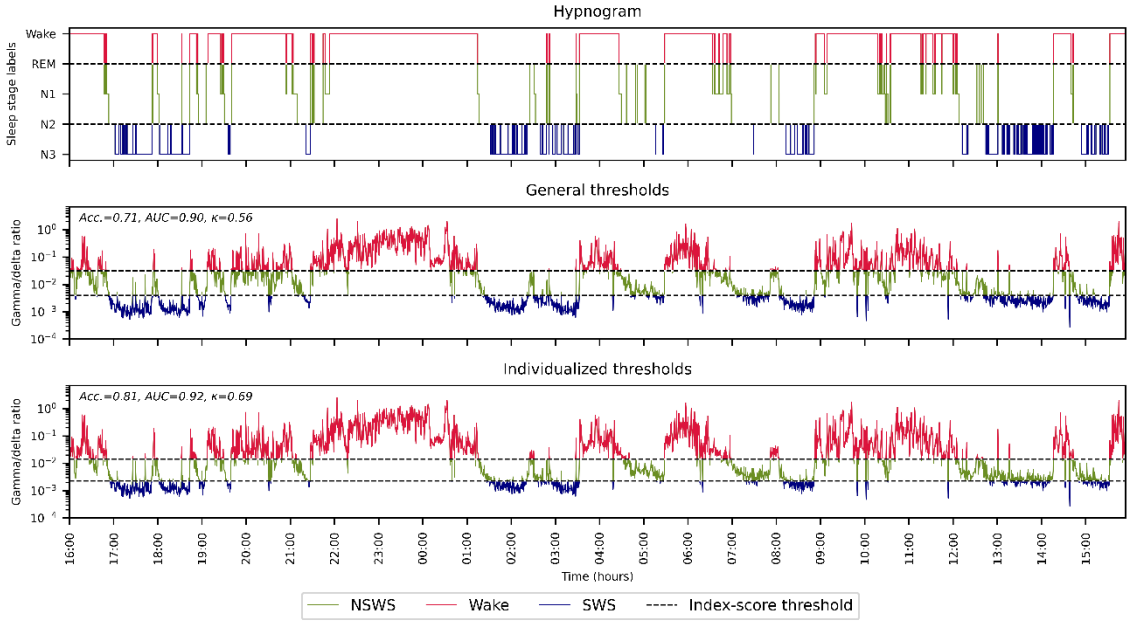


Figure S50. PICU patient J - Agreement of the visually scored hypnogram with the gamma/delta ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

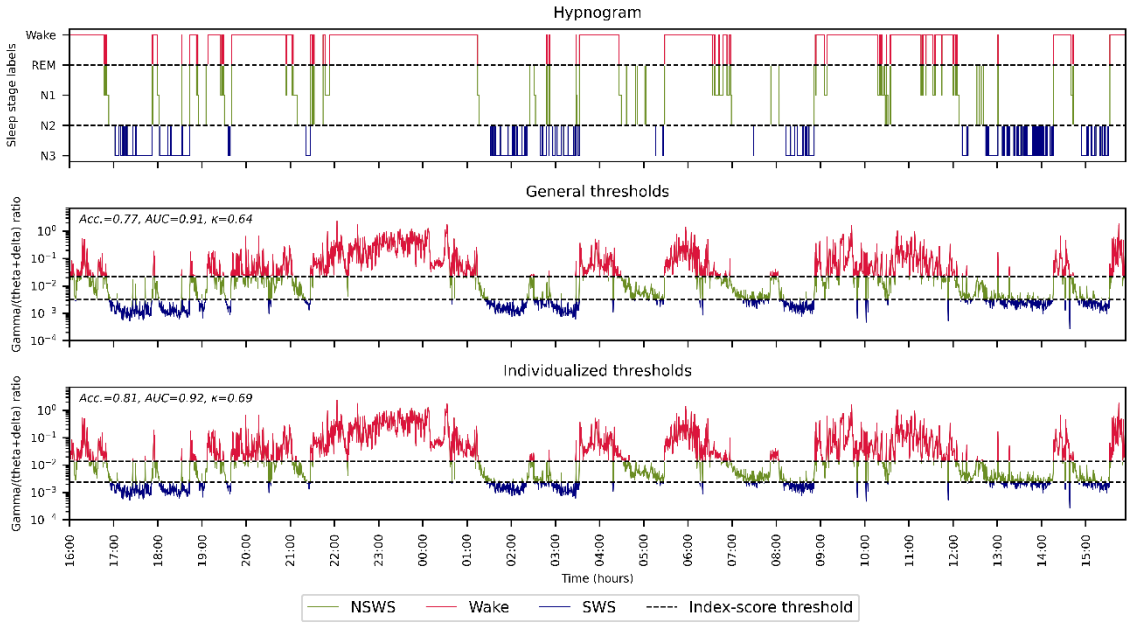


Figure S51. PICU patient J - Agreement of the visually scored hypnogram with the gamma/(theta + delta) ratio (with the final model, or general, and individualized thresholds) for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, κ = Cohen's kappa, NSWS = non slow wave sleep, REM = rapid eye movement, SWS = slow wave sleep

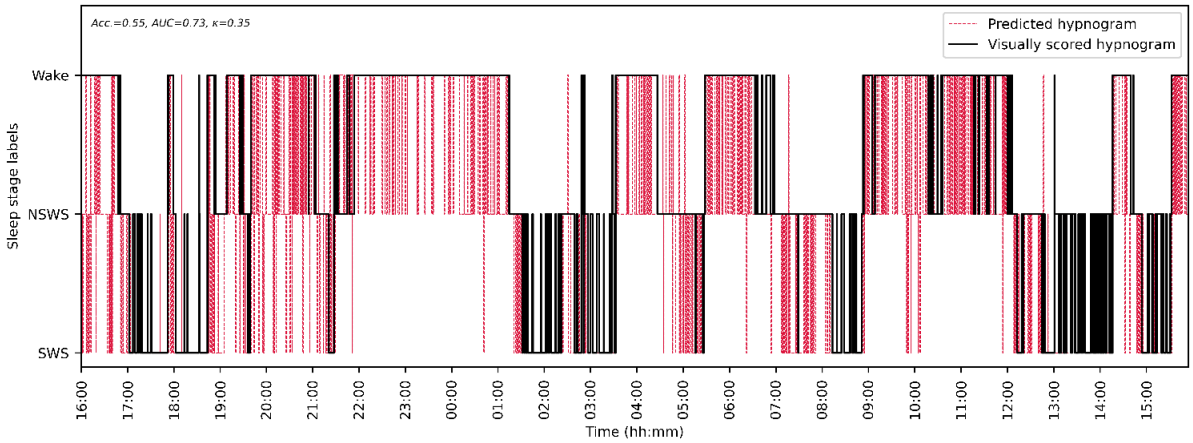


Figure S52. PICU patient J - Agreement of visually scored hypnogram and predicted hypnogram by DT model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, DT = decision tree, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep

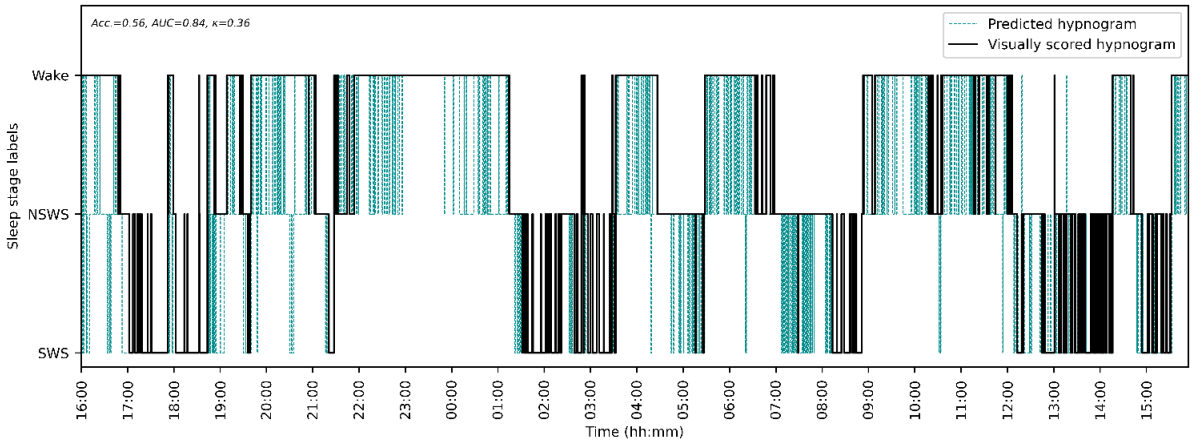


Figure S53. PICU patient J - Agreement of visually scored hypnogram and predicted hypnogram by SVM model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SVM = support vector machine, SWS = slow wave sleep

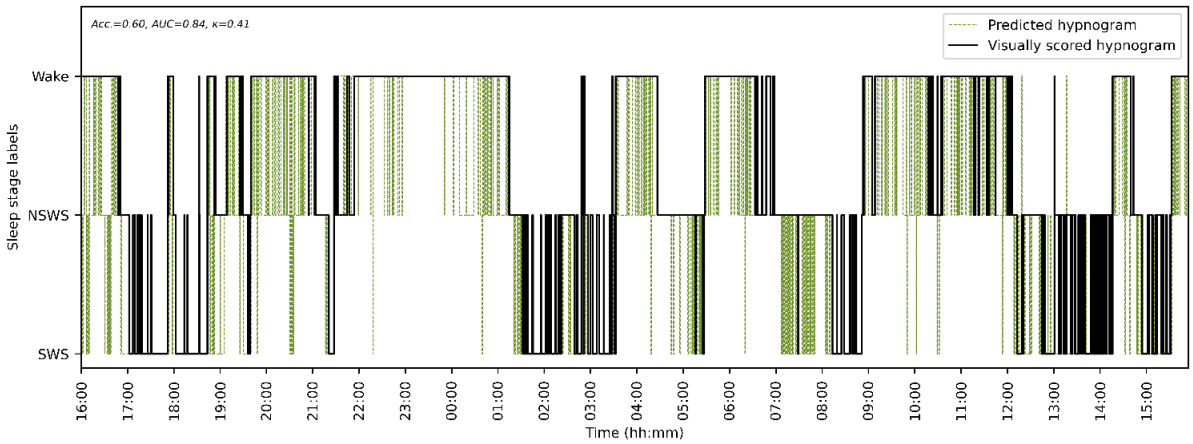


Figure S54. PICU patient J - Agreement of visually scored hypnogram and predicted hypnogram by XGBoost model for three-state classification. Acc. = accuracy, AUC = area under the receiver operating characteristic (ROC) curve, PICU = pediatric intensive care unit, NSWS = non slow wave sleep, SWS = slow wave sleep, XGBoost = extreme gradient boosting

Table S17. Classification performance (training subgroup scores, accuracy) of the final index-based models per age category. Performance results were obtained by selecting age categories from the training data set (i.e., all reference PSG data). CV = cross-validation, PSG = polysomnography

	Gamma/delta		Gamma/(theta+delta)	
Age category	**Three-state**	**Four-state**	**Three-state**	**Four-state**
0-2 months	0.58	n/a	0.59	n/a
2-6 months	0.66	n/a	0.66	n/a
6-12 months	0.72	0.51	0.72	0.51
1-3 years	0.74	0.60	0.73	0.59
3-5 years	0.72	0.57	0.72	0.57
5-9 years	0.74	0.57	0.75	0.58
9-13 years	0.68	0.57	0.69	0.57
13-18 years	0.71	0.62	0.72	0.63
0-18 years	0.69	0.57	0.70	0.57

Table S18. Classification performance (training subgroup scores, accuracy) of the final machine learning models per age category. Performance results were obtained by selecting age categories from the training data set (i.e., 50,000 epochs of the reference PSG data). DT = decision tree, PSG = polysomnography, SVM = support vector machine, XGBoost = extreme gradient boosting

	DT		SVM		XGBoost	
Age category	**Three-state**	**Four-state**	**Three-state**	**Four-state**	**Three-state**	**Four-state**
0-2 months	0.94	n/a	0.68	n/a	0.73	n/a
2-6 months	0.94	n/a	0.73	n/a	0.79	n/a
6-12 months	0.96	0.82	0.79	0.72	0.81	0.79
1-3 years	0.94	0.80	0.78	0.76	0.82	0.79
3-5 years	0.95	0.63	0.79	0.69	0.82	0.69
5-9 years	0.93	0.80	0.76	0.74	0.83	0.81
9-13 years	0.95	0.82	0.74	0.74	0.82	0.78
13-18 years	0.94	0.83	0.79	0.77	0.84	0.82
0-18 years	0.96	0.81	0.76	0.74	0.81	0.80

Table S19. Classification performance (accuracy) of the final index-based models versus the per age category trained models, three-state classification. Performance results were obtained by selecting age categories from the training data set (i.e., all reference PSG data). CV results were obtained using 5-fold CV. CV = cross-validation, PSG = polysomnography

	Gamma/delta		Gamma/(theta+delta)	
Age category	**Final model thresholds - test score**	**Thresholds per age group – CV score**	**Final model thresholds - test score**	**Threshold per age group – CV score**
0-2 months	0.58	0.57 (± 0.01)	0.59	0.57 (± 0.01)
2-6 months	0.66	0.66 (± 0.01)	0.66	0.66 (± 0.01)
6-12 months	0.72	0.72 (± 0.00)	0.72	0.72 (± 0.01)
1-3 years	0.74	0.77 (± 0.00)	0.73	0.77 (± 0.00)
3-5 years	0.72	0.74 (± 0.01)	0.72	0.74 (± 0.01)
5-9 years	0.74	0.79 (± 0.01)	0.75	0.79 (± 0.01)
9-13 years	0.68	0.71 (± 0.01)	0.69	0.71 (± 0.01)
13-18 years	0.71	0.71 (± 0.02)	0.72	0.71 (± 0.01)
0-18 years	0.69	0.69 (± 0.02)	0.70	0.70 (± 0.02)

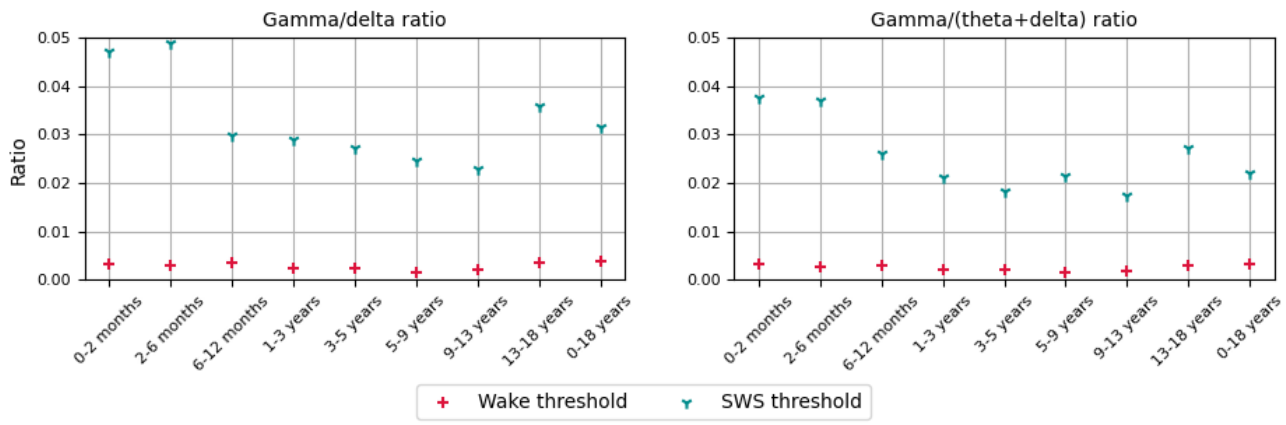


Figure S55. Variation in thresholds between age categories for the index-based models, three-state classification.

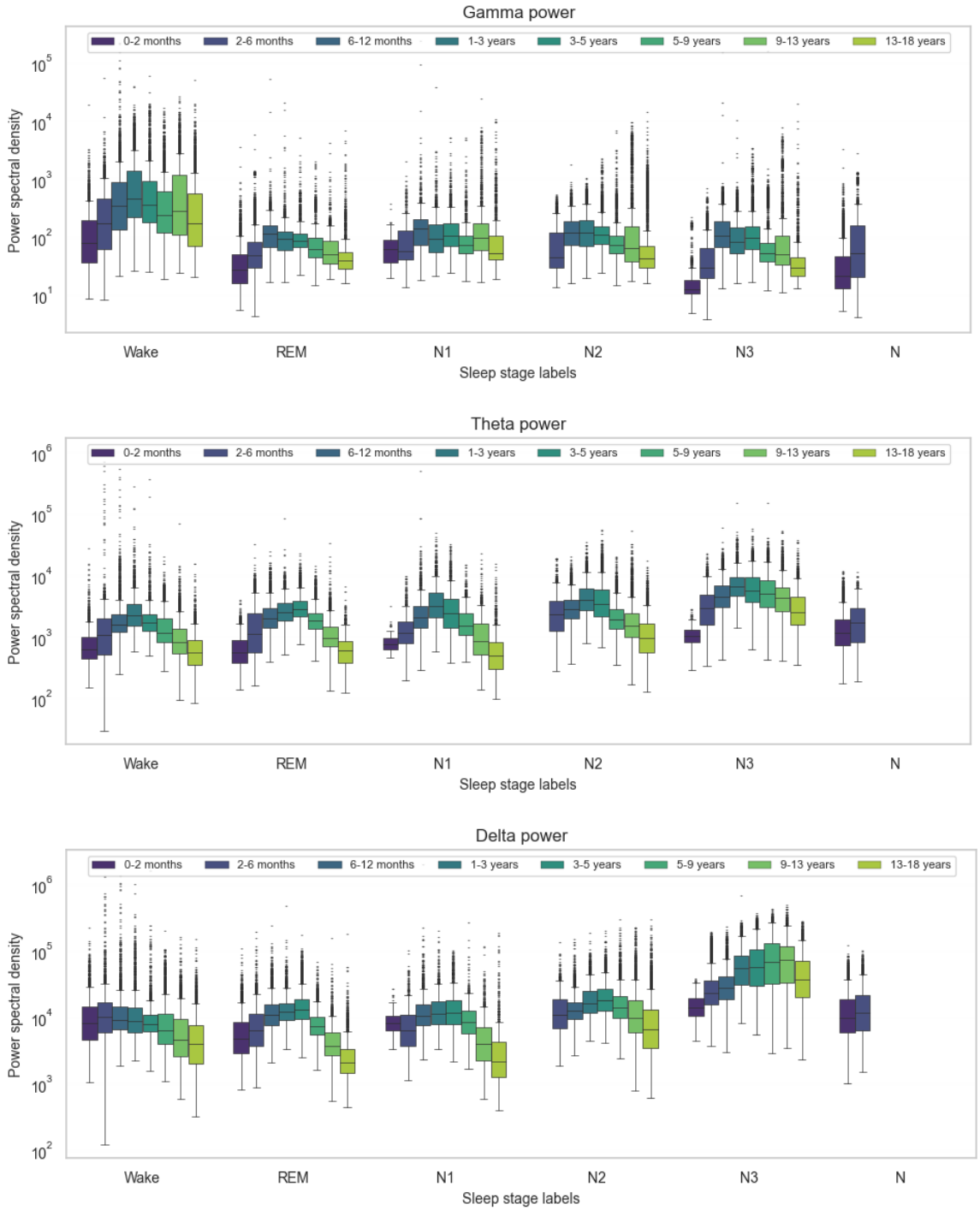


Figure S56. Absolute spectral powers across sleep stages per age category. The spectral powers were obtained from all reference PSG data using the F3-C4 channel. The boxes extend to the 25th and 75 interquartile range with the central line representing the median value. The whiskers extend to ± 2.7 SD. Points outside this range are defined as outliers, marked by the black plusses. The figures show a general increase in theta and delta power as sleep deepens, while gamma power decreases as sleep deepens. A large variation of spectral powers within age categories but also between age categories can be observed. Theta powers peaks across all sleep stages in the age categories 6-12 months and 1-3 years. Delta power during SWS increases with age. N1/N2/N3/N = non rapid eye movement sleep, PSG = polysomnography, REM = rapid eye movement sleep, SD = standard deviation, SWS = slow wave sleep

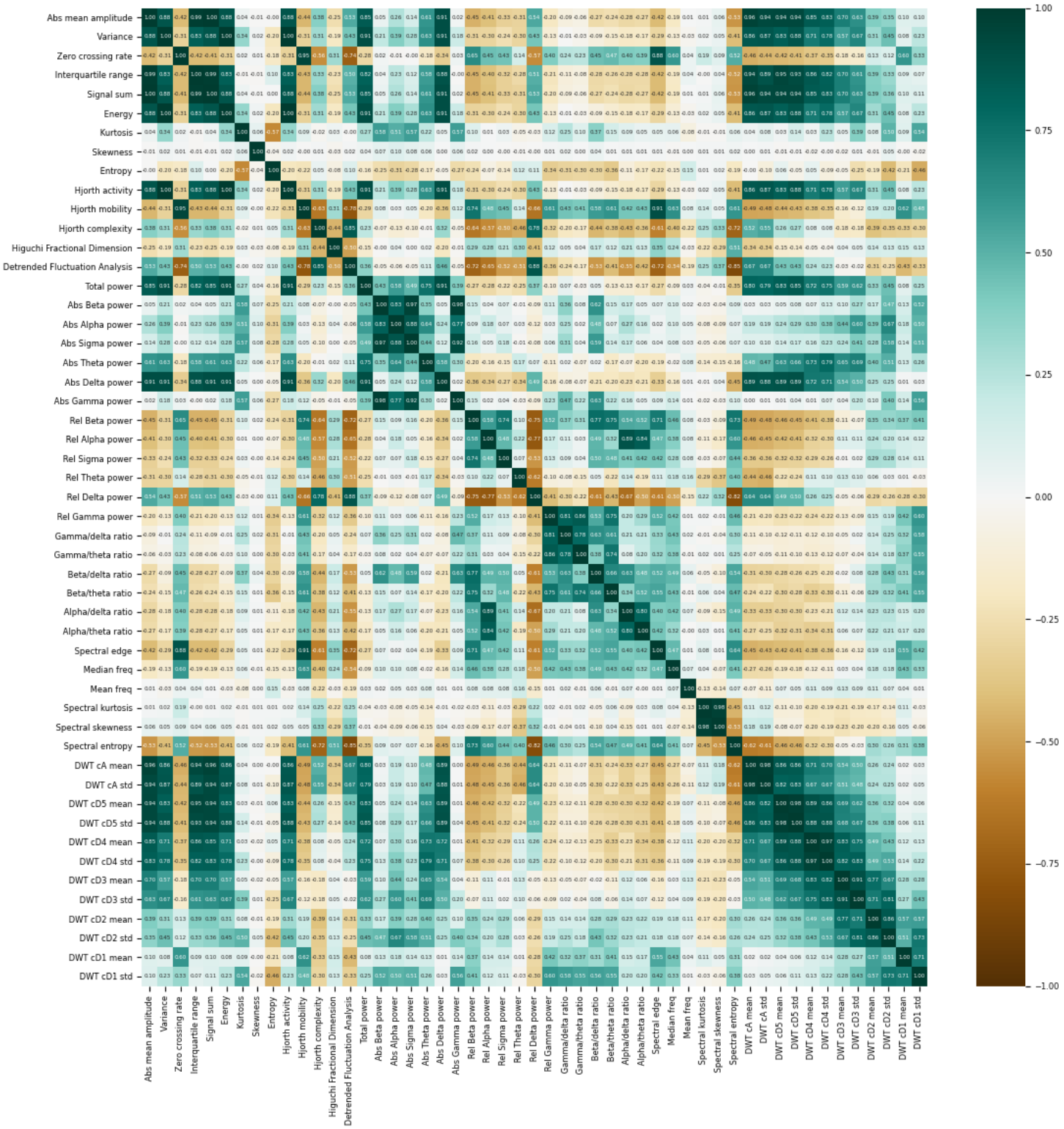


Figure S57. EEG feature correlation heatmap (F3-C3 channel). abs. = absolute, DWT = discrete wavelet transform, EEG = electroencephalogram, freq. = frequency, rel. = relative, SD = standard deviation