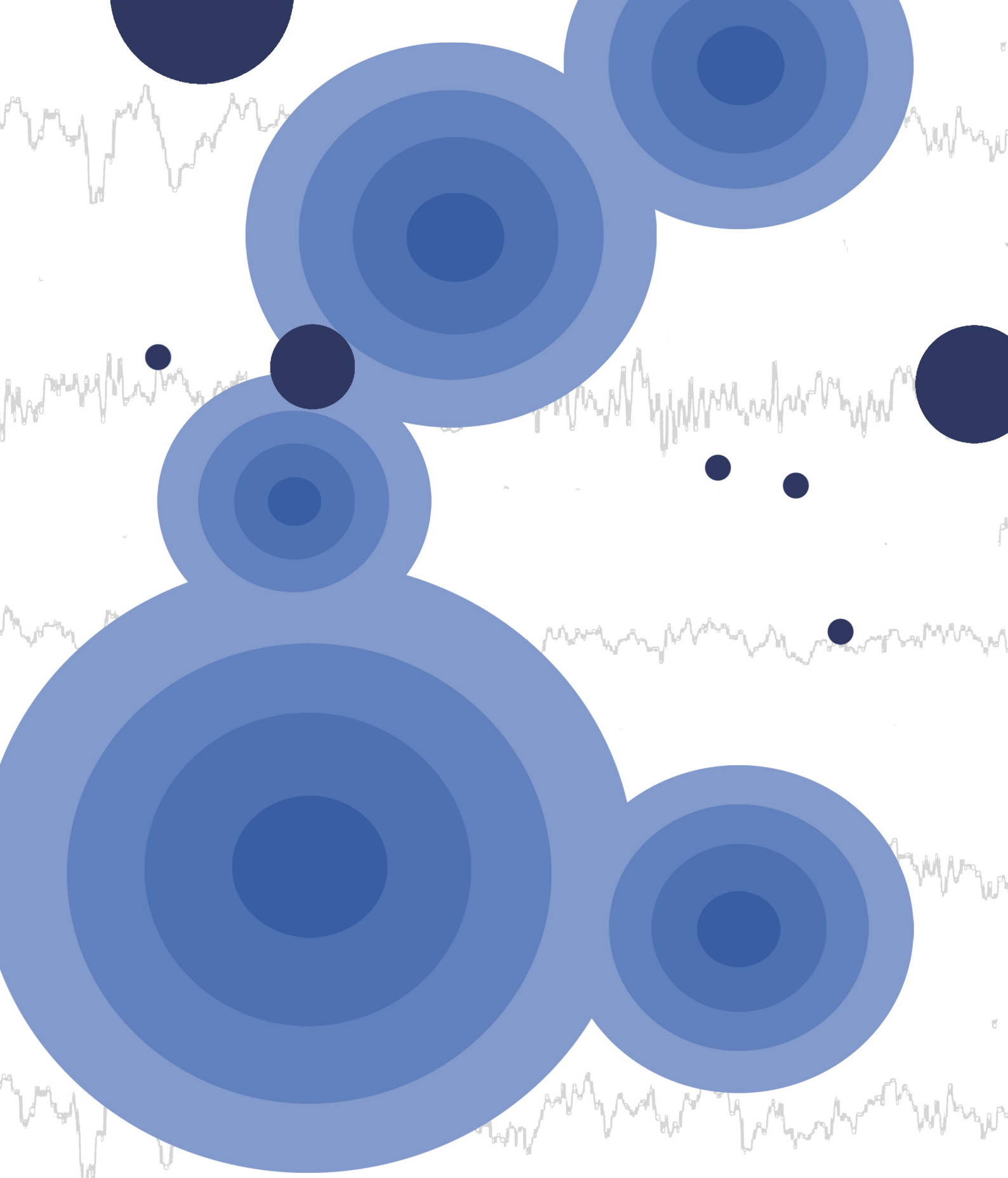

ANOMALY DETECTION IN SLEEP STAGING IN CRITICALLY ILL CHILDREN

MASTER THESIS

CARLIJN OERLEMANS

Anomaly Detection in Sleep Staging in Critically Ill Children.

- Master Thesis -

Carlijn Oerlemans
Student number: 4444906
January 2024

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

Technical Medicine

Leiden University; Delft University of Technology; Erasmus University Rotterdam

Master thesis project (TM30004; 35 ECTS)
Dept. of Pediatric Intensive Care, Erasmus MC
April 2023 – January 2024

Supervisor(s):
Dr. D.M.J. Tax, TU Delft
Dr. R.C.J. de Jonge, Erasmus MC
Dr. J.W. Kuiper, Erasmus MC
Drs. E. van Twist, Erasmus MC

Thesis committee members:
Prof. dr. K.F.M. Joosten (chair)
Dr. D.M.J. Tax, TU Delft
Dr. R.C.J. de Jonge, Erasmus MC
Dr. J.W. Kuiper, Erasmus MC

Universiteit Leiden

TU Delft

Erasmus Universiteit Rotterdam

Preface

With the completion of this graduation project, my studies in Technical Medicine at the Delft University of Technology, Erasmus University Rotterdam, and Leiden University, have come to an end. In 2015, I started in the second cohort of the Bachelor of Clinical Technology which felt like the right study choice for me. After my Bachelor's, I almost switched to Medicine since I experienced getting a lot of energy from contact with people and felt an intrinsic motivation to help people. However, I also was curious and felt intrigued by the technical innovations in healthcare. I wanted to learn more and gain a deeper understanding of the technical background. I decided to continue with my Master's and become a Clinical Technologist who can help people with technical innovations in healthcare. After another sidetrack to the track Imaging and Intervention, I started the Master track Sensing and Stimulation. Again, this felt like the perfect fit for me. The courses were inspiring and I was very enthusiastic about the internships. Every day I came home with so many stories and I appreciated the opportunities and experiences I gained during my studies. I developed a big interest in Acute Care and I did all my internships in this field. After an internship at the Pediatric Intensive Care Unit, I was delighted to start my graduation project here.

I want to thank my supervisors, David, Rogier, Jan Willem, and Eris for their valuable guidance throughout my research journey. Initially, delving into this graduation project was a little bit of a step outside my comfort zone. While I had experience with programming projects during my internships, developing a model from scratch was a new challenge. You could think: Why should you try this out during your graduation project? For me, it was driven by the motivation to explore if I could perform this task and if I would like to continue with similar work in my future job. Thank you David for dedicating an hour per week to guide me, for your patience and the drawings to help me understand the process. I enjoyed gaining more understanding of things and discussing the next steps together. You sometimes ended our meeting with the sentence: "You don't have to stress about it, enjoy the process and it will be all right". I always left our meetings motivated to learn more.

I want to thank Rogier, Jan Willem en Eris for your valuable insights, critical eye, and motivating attitude during the meetings and in your extensive feedback. You were also given a small glimpse of my stress from time to time. Eris, I could appreciate your calmness. Furthermore, I am grateful for the clinical experiences. I always get so much energy from days in the clinic. Also, I want to thank Koen for your willingness to participate in my thesis committee.

Most of all, I want to express my gratitude to my parents, my siblings, Jip, Heijke, Lous, my study friends, and my friends during this graduation project. I want to thank you not only for this period but for all the years of study. For all the hours on the phone listening and advising me about all the choices I had to make, every time all over again. My grandfather always motivated me to work hard and complete this study, he would be proud. Thank you all for the unconditional support and belief in me and for making my student time unforgettable.

Carlijn Romée Oerlemans
Rotterdam, December 2023

Contents

List of Abbreviations

A1	Left auricular electrode
A2	Right auricular electrode
AASM	American Association of Sleep Medicine
C3	Left central electrode
C4	Right central electrode
EEG	Electroencephalography
ED	Euclidean Distance
EMG	Electromyography
EOG	Electrooculography
F3	Left frontal electrode
F4	Right frontal electrode
Fpz	Pre-frontal electrode on midline sagittal plane
ICU	Intensive Care Unit
M1	Left mastoid electrode
M2	Right mastoid electrode
MC	Medical Center
MD	Mahalanobis Distance
MEC	Medical Ethics Committee
MREC	Medical Research Ethics Committee
Mvn	Multivariate normal distribution
N1	Non-rapid eye movement stage 1
N2	Non-rapid eye movement stage 2
N3	Non-rapid eye movement stage 3
N	Residual stage with characteristics of NREM which could not be classified as N1, N2 or N3
NICU	Neonatal Intensive Care Unit
NREM	Non-rapid eye movement
O3	Left occipital electrode
O4	Right occipital electrode
OSA	Obstructive sleep apnea
P3	Left parietal electrode
P4	Right parietal electrode
PDF	Probability density function
PDR	Posterior dominant rhythm
PELOD	Pediatric Logistic Organ Dysfunction Score
PICU	Pediatric Intensive Care Unit
POSTS	Positive sharp transients of sleep
PSG	Polysomnography
REM	Rapid eye movement
R	Rapid eye movement
SWS	Slow wave sleep
W	Wake

1 Abstract

Study objectives: Conventional sleep scoring is based on the scoring criteria of the American Association of Sleep Medicine (AASM) but may not be suited to describe sleep in critically ill children admitted to the Pediatric Intensive Care Unit (PICU). In this study, an anomaly detection model using Gaussian Models trained on sleep stages in data from non-critically ill children is developed to assess if polysomnography(PSG)-derived electroencephalography (EEG) data from critically ill children can be categorized into sleep stages based on these AASM scoring criteria.

Methods: A retrospective study at Erasmus MC Sophia Children's Hospital, using PSG recordings obtained in non-critically ill children between 2017 and 2021 and in critically ill children between 2020 and 2022.
Gaussian Models were individually trained for each sleep stage using data from non-critically ill children. Anomaly detection was carried out by computing the Mahalanobis Distances and assigning datapoints to specific sleep stages or categorizing them as anomalous. Errors were quantified by calculating the ratio of anomalous epochs to the total number of epochs. The trained Gaussian Models were applied to distinct sleep stages in the data from non-critically ill children. Subsequently, the models were applied to data from critically ill children to determine the categorization of their epochs. This was also analyzed over time and involved comparisons related to medication, mechanical ventilation, and the severity of illness assessed by the PELOD-2 score.

Results: In non-critically ill children the models obtained validation errors aligning with the margin error of the training set. The models could not fully differentiate the distinct sleep stages. In critically ill children, the majority of epochs were classified into multiple sleep stages. High error rates were evident for sleep stages N1, R, and N. Some patients exhibited elevated error rates specifically for sleep stage N1. REM sleep was reduced, consistent with findings from previous studies. In contrast, N3 sleep did not show a reduction. When compared to the sleep stage labels assigned by neurophysiologists, the model classified epochs into multiple sleep stages, while neurophysiologists frequently used the label N. A higher PELOD-2 score did not consistently correlate with an increased occurrence of anomalous classifications in the epochs of these patients to those with lower PELOD-2 scores.

Discussion: Overlap of sleep stages was observed in non-critically ill children. Epochs from critically ill children were classified into multiple sleep stages without clear associations in time or severity of illness. Building upon the established anomaly detection framework is recommended by employing more advanced anomaly detection methods using an informative feature selection. This study marks an initial step, indicating that applying the AASM scoring criteria may not be suitable for characterizing sleep in critically ill children.

2 Introduction

Critically ill children admitted to the Pediatric Intensive Care Unit (PICU) experience sleep disruptions, resulting from underlying illness and exposure to the PICU environment [1–7]. Additionally, the administration of centrally-acting medication can be disruptive to their normal sleep-wake rhythm [7]. It has been suggested that poor quality of sleep may contribute to delirium, compromised immune function, and prolonged mechanical ventilation in critically ill patients [6]. In addition to the crucial role of sleep in the recovery of critically ill individuals, the restorative benefits of sleep are fundamental for the neurocognitive development of children [7]. Bedside and real-time sleep monitoring at the PICU would allow for individually targeted interventions to improve the sleep of critically ill children. However, sleep in critically ill children may differ from healthy sleep presenting a challenge to accurately classify sleep in this population.

Conventionally, sleep is visually assessed using information derived from electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG) as part of polysomnography (PSG) [7–10]. Sleep in healthy people can be distinguished into rapid eye movement (REM) sleep and non-rapid eye movement (NREM) sleep. NREM is subdivided into stages N1 (NREM 1), N2 (NREM 2), and N3 (NREM 3), where N1 represents the lightest sleep and N3 the deepest sleep [11]. NREM serves a restorative and rest-facilitating role, while REM is considered crucial for brain development and the consolidation of memory [2, 11, 12]. In normal sleep, REM sleep and NREM sleep alternate cyclically.

A normal awake EEG demonstrates continuity, symmetry, and an anterior-posterior gradient. The anterior (frontal) side of the brain exhibits faster, lower amplitude frequencies and the posterior (occipital) side shows slower, higher amplitude frequencies. During a resting state with closed eyes, a posterior dominant rhythm (PDR) is commonly observed. Normal EEG is variable and shows reactivity to external stimuli. Based on the tracings, the state of wakefulness, drowsiness, or sleep stages can be determined. Examples of the EEG characteristics of the wakefulness and the sleep stages can be seen in Figure 2.1 and in the figures included in the Appendix. Drowsiness is seen as diffuse slowing and attenuation of the EEG with slow, roving eye movements. In the transition from drowsiness to N1 sleep, the PDR disappears, and positive sharp transients of sleep (POSTS) together with vertex waves can be seen. The N2 stage is characterized by sleep spindles and K-complexes, where sleep spindles are small symmetric bursts of higher-frequency activity and K-complexes are very high amplitude, symmetric waveforms. The K-complexes are the largest voltage events on a normal EEG that typically occur during auditory stimuli while asleep [11, 13]. The EEG during NREM typically transitions from alpha waves associated with wakefulness to shorter frequency theta waves associated with sleep [11]. In stage N3, an individual is least responsive to external stimuli indicating a high arousal threshold. The N3 sleep stage is referred to as slow wave sleep (SWS) because it is marked by diffuse and synchronized, high-amplitude, and low-frequency delta activity [11, 13]. The frequency of rapid eye movement (REM), the characteristic of REM sleep, increases overnight. This is observed as sharply contoured opposing waveforms arising in the frontal regions. REM sleep exhibits low-voltage and high-frequency EEG [11, 13].

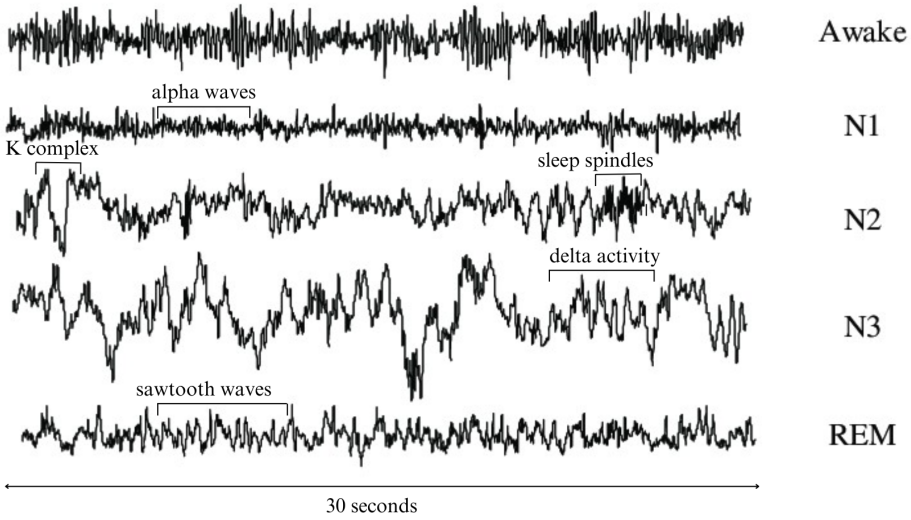

Figure 2.1: EEG signal for the different sleep stages [13–15]

Sleep stages are manually assigned to 30-second epochs of PSG recordings according to the scoring criteria of the American Association of Sleep Medicine (AASM) [8–10, 16]. Conventional sleep staging is based on the described typical EEG characteristics. Since atypical EEG findings are frequently present in critically ill patients, the use of conventional sleep stages in ICU patients has limited value [6, 7, 17–19]. Besides, the AASM scoring criteria are not validated for use in critical illness [6, 16]. Moreover, age-related development changes affect both PSG and EEG characteristics [13, 16].

Since the execution of PSG measurements and the manual assignment of sleep stages to the PSG recordings are done in retrospect and are burdensome, expensive, and time-consuming, many have sought to develop automatic sleep scoring methods [4–6, 8–10, 16, 20]. Little research to characterize EEG patterns in critically ill children has been performed and the development of accurate automatic sleep scoring tailored for this population remains a challenge. Classification models that have been developed and trained on data from non-critically ill children have shown poor performances when applied to data from critically ill children admitted to the PICU. This raises the question of whether the sleep of critically ill children is comparable to sleep in non-critically ill children [21]. Sleep-scoring algorithms developed for classifying the sleep of adults in the Intensive Care Unit (ICU) or neonates in the Neonatal Intensive Care Unit (NICU) are typically assessed by comparing their performance to conventional sleep staging. This evaluation often includes an assessment of the inter- and intra-observer variability associated with manual scoring [19, 22–27]. To the best of our knowledge, no studies have reassessed the appropriateness of employing conventional sleep staging based on the AASM scoring criteria for characterizing sleep in critically ill children within a PICU setting. There is a possibility that using alternative scoring criteria could provide a more suitable description of sleep in critically ill children admitted to the PICU.

To explore EEG patterns of sleep in critically ill children compared to non-critically ill children, we propose an alternative approach based on anomaly detection. Anomaly detection methods prove to be valuable in cases where only one class is well-represented and the other class is absent, poorly sampled, or not well-defined [28] [29]. These methods can effectively identify observations, events, or patterns that deviate from the rest of the data. Anomaly detection methods can operate with prior knowledge of a single class, allowing the determination of whether novel data can be assigned to a specific single class or not [29–33]. Importantly, these methods do not depend on uniform class sizes and abundant instances of anomalous data, making them highly flexible to each patient case [20, 30–34]. Considering that the sleep of non-critically ill children can be systematically categorized using the AASM scoring criteria, while the sleep of critically ill children is diverse and challenging to characterize, employing anomaly detection with non-critically ill children's data as the target class provides a tailored approach to explore anomalies in sleep patterns of critically ill children compared to the well-defined

non-critically ill class.

The primary objective of this study was to develop an anomaly detection model to assess whether PSG-derived EEG data from critically ill children could be categorized into sleep stages based on the AASM scoring criteria. The sleep stages of PSG-derived EEG data from non-critically ill children were treated as a single class and it was evaluated if the PSG-derived EEG data from critically ill children could be assigned to this specific class. A secondary aim was to explore the effectiveness of the anomaly detection model in identifying sleep stages as anomalous, considering each sleep stage as a single class. Additionally, the classification of PSG-derived EEG data from critically ill children into sleep stages over time was evaluated.

3
Data collection and preparation

3.1. Study population

For this master thesis, two independent datasets were used. The full methodology is described in Twist (2023) and Cramer (2023) [21, 35]. In short, the first dataset contained PSG recordings of non-critically ill children who underwent a PSG between 2017 and 2022 at the PICU of the Erasmus MC Sophia Children's Hospital, Rotterdam, the Netherlands, to diagnose or follow up on conditions such as obstructive sleep apnea (OSA) or neuromuscular disease. Considering the EEG changes during maturation, eight age categories were defined: 0-2 months, 2-6 months, 6-12 months, 1-3 years, 3-5 years, 5-9 years, 9-13 years, 13-18 years [13, 16, 21]. For patients born preterm (<37 weeks gestational age), age was corrected until the postnatal age of 2 years. Fifteen patients were included in each age category, resulting in a total of 120 PSG recordings [4, 5]. Due to the retrospective study design, informed consent was waived by the Medical Research Ethics Committee (MREC) of the Erasmus MC (MEC-2021-0121) [4, 5, 21]. The PSG data obtained from the recordings of non-critically ill children are further referred to as 'PSG data'.

The second dataset contained PSG recordings prospectively obtained from critically ill children admitted to the PICU of Erasmus MC Sophia Children's Hospital between 2020 and 2022. These patients were included in either a study investigating the circadian rhythm in children admitted to the PICU (Critical Clock: Netherlands Trial Register (NTR): NL8533) or a trial investigating the effect of continuous nutrition versus intermittent nutrition with an overnight fasting period in PICU patients (ContInNuPIC trial: NTR NL7877). Again, the full methodology of these studies is described by Cramer et al., Twist et al., and Veldscholte et al. [21, 35–37]. In summary, patients term born up to 18 years of age with an expected PICU stay of more than 48 hours were eligible for inclusion in these studies [5, 21]. Exclusion criteria for the ContInNuPIC trial mainly concerned contra-indications for prolonged fasting. For the Critical Clock study, the exclusion criteria were formulated to exclude pre-existing disruptions in circadian rhythm and to ensure the feasibility of blood draws for measuring melatonin and cortisol. The exclusion criteria were premature newborns, diagnosis of a syndrome associated with mental retardation, hydrocortisone use in the three days before admission, melatonin use in the 24 hours before admission, weight below two kilograms, and expected not to receive an arterial line during the study period. Both studies also excluded patients transferred from another ICU, patients previously included in the study, and patients who participated in trials with conflicting study procedures. Both studies were approved by review committees and MREC provided permission for the use of the patient data in this study (MEC-2021-0121) [5]. All participants and/or their parents or legal guardians gave informed consent to participate in the trials. Given the potential discomfort of the PSG measurement, separate consent was asked on the day of the measurement [5, 21]. The PSG data obtained from the recordings of critically ill children admitted to the PICU are further referred to as 'PICU data'. The patient characteristics including age, gender, PSG or PICU indication, medical history, and sedative or analgesic medication that was administered during the PSG measurement, were collected for all patients of both datasets [4, 5]. For the analysis of the PICU dataset, a Pediatric Logistic Organ Dysfunction Score 2 (PELOD-2) was calculated to describe the severity of the organ dysfunction of the patients [38, 39].

3.2. Data acquisition

In the target PSG dataset, all patients underwent a PSG using an eight-channel EEG, two EOG electrodes, and two chin EMG electrodes. The 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 [16]. The PSG measurements were recorded at a sampling frequency of 250 or 256 Hz, depending on the device used (Brain RT, OSG, Rumst, Belgium or Morpheus, Micromed Sp.A., Treviso, Italy) [4, 5, 21]. The PSG measurements started once patients went to bed at their habitual bedtime and

stopped in the morning. The last PSG recording was used whenever patients had undergone multiple PSG measurements [4, 21].

In various patients in the novel PICU dataset, the frontal, central, and occipital electrodes were only placed on one side of the brain to minimize the patient's discomfort. As described previously, the measurements were performed using the EEG electrodes: either F3, C3, and O1, or F4, C4, and O2, with the electrodes A1, A2, and Fz [21, 35]. Patients included in the PICU dataset were enrolled within the first 24 hours of PICU admission. 24-hour PSG measurements were performed on days 1, 3, 7, and 14, allowing for a one-day margin for logistical considerations. The majority of the patients in the PICU dataset underwent a single PSG measurement, either due to a short PICU stay or because there was no parental/legal guardian consent for a second PSG measurement. For the other patients, priority was given to PSG measurements performed on days when light or sound measurements were taken, or samples were collected for melatonin and cortisol assays [21, 36, 37].

The PSG recordings for patients in the target PSG dataset were scored by experienced sleep technologists under the supervision of an experienced clinical neurophysiologist. The sleep stages were assigned per 30-second epochs of the PSG recordings according to the AASM scoring criteria. For patients younger than 2 months, sleep was distinguished in REM and NREM and the recommended transitional sleep stage for these young infants was not used [21]. The N stage was used by sleep technicians for epochs that had the characteristics of NREM but could not be classified as either N1, N2, or N3 [4, 5, 21].

3.3. Preprocessing and feature extraction

3.3.1. Preprocessing

The preprocessing steps involved the detection and removal of various artifacts from the PSG recordings [4, 35]. Additionally, to enhance the data quality, a 16^{th}-order Butterworth band-pass filter was applied. This filter, with a frequency range of 0.5-48 Hz, removed irrelevant frequencies from the EEG signals. These preprocessing steps were uniformly applied to both datasets [4].

3.3.2. Feature selection

The initial feature selection and extraction steps are described in Hiemstra (2021). In essence, features were selected based on definitions and applications in sleep studies in adults and neonates. The calculated features were extracted from each epoch of the PSG recordings. An overview of the selected features can be found in the Supplementary Materials section A.2. [4].

In this study, a deliberate subset of selected features was employed for model development. The initial decision was to focus on the development of the model using EEG features since these reflect the processes regulating states of vigilance and controlling most physiologic parameters [40]. In the study by Twist et al. (2023), numerous features were computed across multiple unipolar and bipolar EEG derivations, resulting in a substantial number of features [4, 35]. Due to the associated high computational costs, a decision was made to develop the model by focusing on a single EEG derivation along with its corresponding features.

The recommended EEG derivations for sleep scoring are F4-A1, C4-A1, and O2-A1 with the back-up derivations F3-A2, C3-A2, and O1-A2 [16, 41]. The reference electrode was chosen at the opposite side of the head. The A1 and A2 electrodes here refer to the mastoid electrodes (also known as M1 and M2) [42]. Linking the electrodes to a mastoid reference avoids near-field and cancellation effects. Cancellation effects occur when the EEG activity of two electrodes has almost the same amplitude and electrical polarity leading to a loss or reduction in EEG amplitude [16]. Hiemstra (2021) extracted features from EEG derivations F3-A2, C3-A2, and O1-A2. In most critically ill patients, the electrodes were placed on a single side of the brain to minimize discomfort, and this predominantly occurred on the right side. Consequently, it was decided to use the features from the EEG derivations of the right side of the brain for model development. For some critically ill patients, the electrodes were placed on the left side of the brain. Assuming both sides of the brain are symmetric, features from EEG derivations of the left side were selected for model application in these patients.

The AASM recommends the use of at least the central EEG derivations (C4-M1, C4-Fpz) with the frontal and occipital derivations as a backup [16, 43]. This makes these central derivations likely to be present in a variety of PSG recordings and therefore a strategic choice for model development [43]. Arousals, defined as temporary disruptions of sleep by wakefulness or brief increases in alertness due to external stimuli or spontaneous changes in wakefulness levels, are more effectively detected by central derivations than frontal derivations [16]. Additionally, sleep spindles and vertex waves are most effectively detected over the central regions. Consequently, it was decided to choose EEG derivation C3-A2 for model development. Since K-complexes are best detected in frontal regions and PDR maximal observed in the occipital regions, it was also assessed whether incorporating EEG derivations F3-A2 and O1-A2 would enhance the overall performance of the model [16].

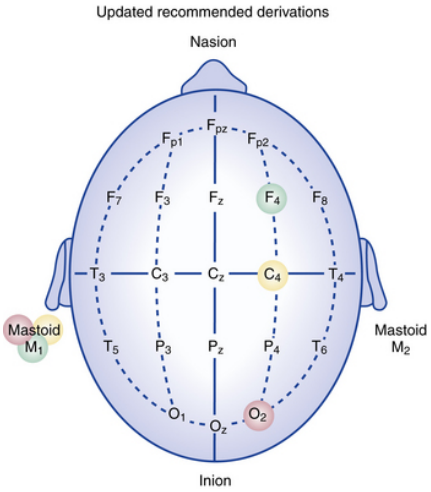

Figure 3.1: Recommended EEG derivations for sleep scoring [42].

3.4. Sleep stages

In this study, the term 'sleep stages' encompasses wake, N1, N2, N3, REM, and N. The 'N' label is assigned to epochs with characteristics of NREM sleep, but where no distinction can be made between N1, N2, or N3 due to atypical or absence of typical EEG characteristics.

4
Model development

4.1. Anomaly detection model and theory

In every sleep stage, most of the selected features from EEG derivation C3-A2 were normally distributed. A figure showing the distributions of these features in epochs from sleep stage N2 is included in the Supplementary Materials section A.3. . The description of the training set $\{\mathbf{x}_i\}, \quad i = 1, \ldots, N$ was obtained using an anomaly detection method, where **x** represents a vector. An univariate normal distribution can describe the probability density function (PDF) of a single feature, for instance, age. PDF describes the probability of a continuous variable. It provides information about how likely different values of the feature are to occur. When there is more than one feature (for example age, gender, and birthplace) a multivariate normal distribution (mnv) PDF can be determined. This describes the distribution of a random vector **x** including different features. These features may be correlated to one another. The mvn PDF can be described with the formula:

$$f(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{p/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

$\mathbf{x}$: The n-dimensional random vector.
$\boldsymbol{\mu}$: The mean vector.
Σ : The covariance matrix.
$\det(\Sigma)$: The determinant of the covariance matrix.
$(\mathbf{x}-\boldsymbol{\mu})^T$: Transpose of the vector $\mathbf{x}-\boldsymbol{\mu}$.
Σ^{-1} : Inverse of the covariance matrix.
p : The number of dimensions.

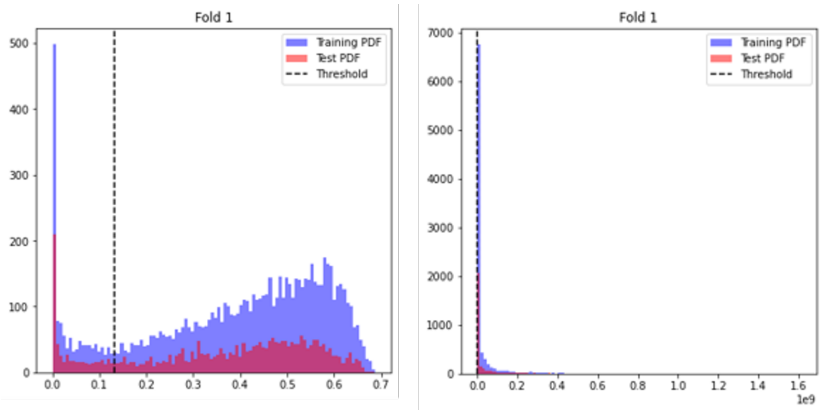

Figure 4.1: Histograms of probability density functions of N1 sleep (outliers not removed). Left the histogram of the probability density functions of four selected features and at the right is the histogram of the probability density functions of all 50 selected features for EEG channel C3-A2.

A low probability density means that the observation can be considered as an outlier according to the distribution. In figure Figure 4.1, examples of histograms of PDFs are illustrated. The PDF of all the 50 selected features from the EEG channel happened to show a high peak at low probability density. This can be explained by the quadratic term $d^2 = \mathbf{x}_1^2 + \mathbf{x}_2^2 + ... + \mathbf{x}_n^2$ in the exponent of mvn

PDF $\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$, where **x** is a *d*-dimensional vector. As the dimensionality increases, d^2 becomes larger, and exp$(-d^2$ becomes very small or zero. This results in the PDF containing numerous zeros or small values, leading to a concentration of probability at low-density levels. To illustrate the normality or anomaly of data set compared to the training set, it was chosen to compute the Mahalanobis Distance (MD).

When assessing the potential anomaly of a data point compared to a dataset, the metrics Euclidean Distance (ED) and MD can quantify the separation between this data point and the dataset. The ED is a straightforward measure that calculates the straight-line distance between a data point and the center of a dataset. The formula of the ED between is:

$$d_E = \sqrt{\sum_{i=1}^{N}(\mathbf{x}_i - \boldsymbol{\mu}_i)^2}$$

where $\mathbf{x}_i$ represents the values of the features of the datapoint along the horizontal axis, and $\boldsymbol{\mu}_i$ represents the corresponding values of the features of the dataset's center along the vertical axis in the coordinate plane. N is the number of dimensions. However, if features are correlated to one another, the ED may present inaccuracies and give misleading information about the proximity of the data point to the actual distribution of the dataset [44]. MD offers a more robust metric by normalizing the ED using the inverse covariance matrix. The covariance indicates how features vary together [45]. When dealing with datasets where features have different scales and are correlated, which is typically the case in real-world datasets, MD is a more reliable metric since it adjusts for the orientation and scaling of the features. Hence, MD ensures a robust measure of distance that reflects the underlying structure of the dataset [44]. The formula of MD is:

$$D = \sqrt{(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

where D is the MD, **x** the feature vector of the data point, $\boldsymbol{\mu}$ the mean vector containing the means of each feature, and Σ^{-1} the inverse covariance matrix of the features [44]. $(\mathbf{x}-\boldsymbol{\mu})$ ensures that the distribution is centered at the origin and the inverse covariance matrix Σ^{-1} normalizes the features by their variance. It considers the covariance and the correlation of the features, giving equal weight to all features. In this study, a Gaussian Model for anomaly detection using MD is developed. In this context, a datapoint corresponds to an epoch.

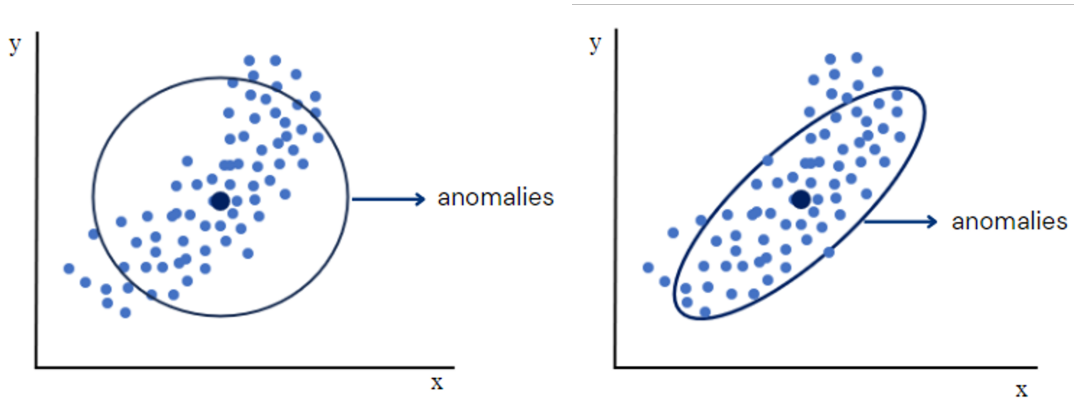

Figure 4.2: Difference between Euclidean Distance (left) and Mahalanobis Distance (right)

To assess anomalies using this Gaussian Model, a threshold needs to be set. This threshold accounts for potential lower-quality epochs, balancing anomaly detection and model performance. This threshold is based on a specific percentile of the MDs within the training set. Subsequently, the MDs of test data can be compared against this threshold to identify anomalies. An epoch is classified as an anomaly if the calculated MD for this epoch exceeds the threshold. Conversely, if the MD falls below the threshold, the epoch can be assigned to the training set. This is visually illustrated in Figure 4.3 and Figure 4.2, where the training set and the anomalies are separated by the threshold.

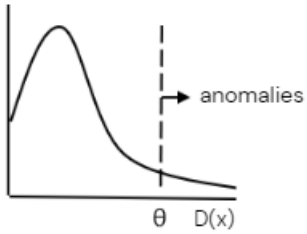

Figure 4.3: Anomaly detection using Mahalanobis Distance

4.2. Gaussian Model development

For each sleep stage, the features were extracted for each 30-second epoch and were used as input data for the Gaussian Model. The model was trained using five-fold cross-validation with grouping on patient level ensuring data integrity. To address redundancy due to 100% feature correlation, a small value was added to the diagonal of the covariance matrix. Consequently, five trained Gaussian Models of all sleep stages were produced as visualized in Figure 4.4.

Subsequently, each trained Gaussian Model was applied to PSG data from other sleep stages to evaluate the errors of the model. The errors represent the fraction of anomalous epochs relative to the total number of epochs in the dataset. For the validation set, this is similar to the false negatives in a confusion matrix as this set is derived from the same dataset as the training set. The errors of the distinct sleep stages correspond to the true negatives in a confusion matrix.

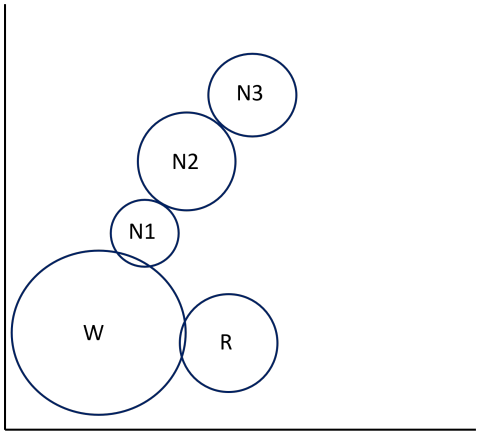

Figure 4.4: Representation of the Gaussian Models of all sleep stages in the feature space

It was assessed whether the Gaussian Model could be improved by performing subanalyses. As described in chapter 3, it was evaluated whether the model's performance could be enhanced by using different bipolar EEG derivations for feature selection. The performance of the model using EEG derivation F3-A2, O1-A2, and the average of F3-A2, O1-A2, and C3-A2 was evaluated. The selected features followed a Gaussian distribution. However, these were characterized by large outliers. To improve the representation of the underlying structure of the distribution of the training set, the outlier values were removed from the features. A percentage of outlier values for each feature was removed, followed by the exclusion of epochs containing missing values from the dataset. The impact of excluding different percentages of outlier values on the model's performance was evaluated. The outliers were only removed from the training set. Additionally, it was evaluated whether normalizing the feature values would enhance the model's performance despite the normalization process in the MD calculation. Normalization was performed with a mean equal to zero and a standard deviation equal to one.

Furthermore, cross-validation with grouping was conducted to ensure the separation of patients across folds. Consequently, the five folds did not precisely contain the same number of epochs, con-

tributing to variation in the validation error. To check whether the validation error would be more accurate without defining the groups, cross-validation without group specifications was also performed.

4.3. Application of the Gaussian Model on data from critically ill children

After establishing confidence in the model's performance, the model could be applied to data from critically ill patients, referred to as PICU data. The labels assigned by the neurophysiologists were removed from the data to evaluate the classification of sleep stages performed by the model. The MD of each epoch of the PICU dataset was calculated using the mean parameter values derived from each trained Gaussian Model. Due to high computational costs, the MD of each epoch was not computed within each fold. For each epoch, it was determined whether the MD fell below or exceeded the threshold derived from the model to identify anomalies.

Similar to the errors of the distinct sleep stages in the model development, the errors of the PICU data were calculated. The proportion of epochs classified as anomalies among all epochs for a particular PICU patient was determined. This corresponds to the true negatives in a confusion matrix.

The framework of the entire model is illustrated in figure Figure 4.5. The sleep stages were used as input data to train the Gaussian Models. Subsequently, each trained Gaussian Model was applied to distinct sleep stages, indicated in the middle. Once confidence in the model was established, the Gaussian models were applied to the PICU data, as illustrated in pink.

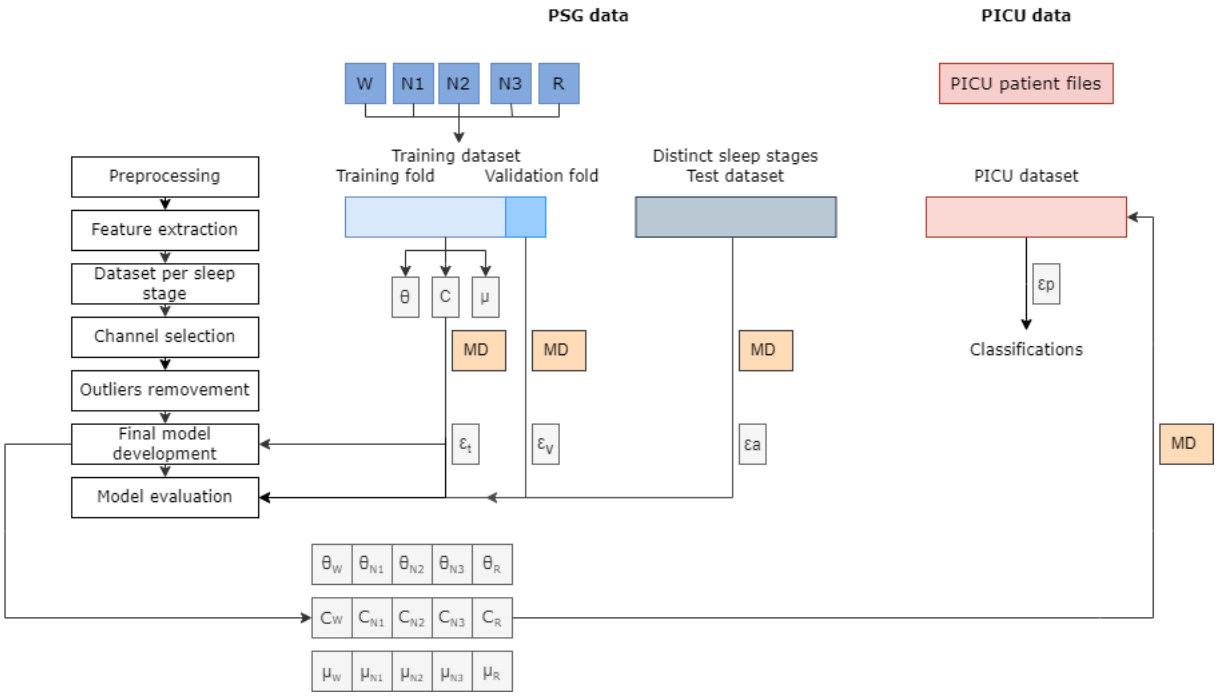

Figure 4.5: Framework of the model.
PSG = polysomnography; PICU = pediatric intensive care unit; W = wake; N1 = NREM1 = non-rapid eye movement stage 1; N2 = NREM2 = non-rapid eye movement stage 2; N3 = NREM3 = non-rapid eye movement stage 3; R = REM = rapid eye movement; MD = Mahalanobis Distance; θ = threshold; C = covariance matrix; μ = mean vector; εt = training error; εv = validation error; εo = distinct sleep stage error; εp = PICU error;

4.4. Analysis of the classification of data from critically ill children over time

Some critically ill patients underwent multiple PSG measurements. Initially, one measurement per patient was included in the analyses. Subsequently, the other available PSG measurement, not scored by the neurophysiologist, was incorporated into this analysis. Discrepancies between these two measurements, performed at different points in time, were assessed, taking into consideration the patient's medical condition, quantified by the PELOD score, and the administration of medication.

5 Polysomnography data and patient characteristics

A total of 120 non-critically ill children and 23 critically ill children admitted to the PICU were included in this study. From the non-critically ill children, 155199 epochs were obtained. The distribution of these epochs by sleep stage is illustrated in Table 5.1. The non-critically ill children had a median age of 36.5 $\pm$ 64 months and most of them underwent a PSG measurement for suspected or follow-up of airway obstruction (n=61) or neuromuscular disease (n=32). The patient characteristics are available in Table 5.2. Regarding the critically ill children, 67992 epochs were obtained. The critically ill children had a median age of 2 $\pm$ 8. The patient characteristics of the critically ill children are available in Table 5.2 and Table 5.3.

Table 5.1: Distribution epochs by sleep stage. *PSG = polysomnography; W = wakefulness; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

Sleep stages	PSG data	Percentages
W	37532	24%
N1	11308	7%
N2	29439	19%
N3	35591	23%
R	29478	19%
N	11851	8%
Total	155199	100%

Table 5.2: Patient characteristics of the non-critically ill and critically ill children. *PICU = Pediatric Intensive Care Unit; PSG = polysomnography; IQR = interquartile range.*

Patient characteristic	PSG dataset (n=120)	PICU dataset (n=23)
Median age $\pm$ IQR (months)	36.5 $\pm$ 64	2 $\pm$ 8
Males/females (number of patients)	62/58	10/13
PSG/PICU indication, % total of 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)	Cardiothoracic surgery: 26.1 (n=6) Neuromuscular disease: 8.7 (n=2) Heart disease/failure: 13.0 (n=3) Pulmonary disease: 26.1 (n=6) Sepsis: 8.7 (n=2) Others: 17.4 (n=4)

Table 5.3: Patient characteristics of the critically ill children. *PICU = Pediatric Intensive Care Unit; f = female; m = male; PELOD = performance of the pediatric logistic organ dysfunction.*

Patient	Gender	Age category	Day measurement after admission PICU	PICU indication	Medication during measurement	Intubated score	PELOD
PICU001	f	0-2 months	2	Abdominal surgery	-	No	3
PICU002	f	0-2 months	2	Pulmonary hypertension	Midazolam, opioids	Yes	9
PICU003	f	0-2 months	8	Pulmonary hypertension	Midazolam, opioids, esketamine	Yes	7
PICU004	f	0-2 months	2	Metabolic disease	Opioids, esketamine	No	6
PICU005	f	0-2 months	2	Cardiothoracic surgery	Opioids	Yes	8
PICU006	f	0-2 months	13	Cardiothoracic surgery	Midazolam	Yes	5
PICU007	f	0-2 months	8	Heart disease	-	No	2
PICU008	m	0-2 months	13	Heart disease	Midazolam	No	2
PICU009	m	0-2 months	13	Neuromuscular disease	Midazolam	Yes	7
PICU010	m	0-2 months	2	Respiratory infection	-	No	5
PICU012	f	2-6 months	3	Heart failure	Midazolam, esketamine	Yes	9
PICU013	m	2-6 months	6	Sepsis	Midazolam, opioids, esketamine	Yes	11
PICU014	m	2-6 months	2	Respiratory insufficiency	-	No	7
PICU015	f	2-6 months	7	Cardiothoracic surgery	Midazolam, opioids, esketamine	Yes	9
PICU016	f	2-6 months	7	Cardiothoracic surgery	Midazolam, opioids	Yes	14
PICU018	f	2-6 months	7	Cardiothoracic surgery	Midazolam	No	4
PICU020	m	6-12 months	1	Cardiothoracic surgery	Opioids	No	3
PICU021	m	1-3 years	1	Respiratory infection	Midazolam, opioids	Yes	7
PICU022	m	1-3 years	7	Respiratory insufficiency	Midazolam, opioids	Yes	7
PICU024	f	9-13 years	3	Sepsis	Midazolam, opioids	Yes	7
PICU025	m	13-18 years	3	Cerebral hematoma + post-resuscitation	Opioids	Yes	8
PICU026	m	13-18 years	12	Post-resuscitation	Opioids, esketamine	Yes	9
PICU027	f	13-18 years	2	Neuromuscular disease	Opioids	Yes	5

6
Results: model development

Following subanalyses, the best performing model was obtained using EEG derivation C3-A2 and excluding 0.2% of the outliers in the selected features. This led to the deletion of approximately 0.5% of the epochs in the training set. Normalization did not lead to significant improvement in the model's performance. Importantly, the normalization of feature values is integrated within the MD calculation.

The threshold of the Gaussian Model was set at 90%, allowing a 10% margin in the training set to classify sleep stages. The model demonstrated accurate performance using cross-validation without explicitly defining groups, resulting in validation errors hovering at 0.1. These validation errors align with the margin error in the training set. When implementing grouped cross-validation, the observed errors demonstrated a slightly extended range to the 0.1 benchmark. However, the decision was made to proceed with grouped cross-validation to maintain data integrity.

The results of the final model development using a threshold of 90% can be seen in Table 6.1. Each row represents a Gaussian Model trained on an indicated sleep stage. Each column signifies a sleep stage on which the Gaussian Model is applied. The values presented in this table illustrate the errors of the model, the number of epochs classified as anomalous compared to the total number of epochs for that sleep stage. The value for which the row and the column represent the same sleep stage signifies the validation error. The validation error is approximately 0.1 as expected since this data is derived from the same dataset as the training set. The remaining errors are expected to be large, approaching 1.0, since these are obtained from the application of a trained Gaussian Model on distinct sleep stages. As an illustration, consider the Gaussian Model trained on sleep stage N1 in row two, which exhibits an error of 0.757 when applied to sleep stage N3 in column four. This signifies that 0.76% of the epochs in sleep stage N3 can be considered anomalous, and 24% of the epochs can be assigned to sleep stage N1. The Gaussian Models consistently classify a substantial portion of epochs in sleep stage wake as anomalous, as indicated by the errors exceeding 0.5. Contrarily, the Gaussian Model specifically trained on sleep stage wake exhibits low errors across various sleep stages. This implies that only a limited number of epochs in these sleep stages are considered anomalous, while the majority can be assigned to sleep stage wake.

Table 6.1: Errors of the Gaussian Models applied on distinct sleep stages. *W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	W	N1	N2	N3	R	N
W	0.105	0.086	0.084	0.038	0.052	0.030
N1	0.885	0.082	0.783	0.757	0.468	0.130
N2	0.577	0.282	0.086	0.388	0.097	0.177
N3	0.683	0.417	0.238	0.110	0.191	0.106
R	0.708	0.413	0.495	0.717	0.116	0.184

The theoretical relationships among sleep stages can be approximated by analyzing these errors. A plausible scenario illustrating the orientation of sleep stages is visualized in Figure 6.1. The fractions are expressed as percentages, with each noted in the color corresponding to the trained Gaussian Model for the sleep stage presented in the row. As an illustration, sleep stage N2 is visually presented in pink. All percentages highlighted in pink correspond to the errors associated with the row representing sleep stage N2 when applied to the respective columns. To provide an example, 58% of sleep stage wake lies outside sleep stage N2, given that the error is 0.577. Noticeably, the sleep stage wake is illustrated as a large circle enclosing the majority of the data from the other sleep stages.

To optimize the visualization of selected features extracted from sleep stages in a lower-dimensional space, Principal Component Analysis (PCA) was conducted. A principal component delineates a direction in the feature space where the data exhibits the most variability [46]. The first two principal components are illustrated in Figure 6.2, as these components encapsulate the greatest variance within the dataset. As can be observed, the majority of the data is located in the bottom left corner of the scatter plot, and again overlap is seen for the sleep stages with wake varying the most.

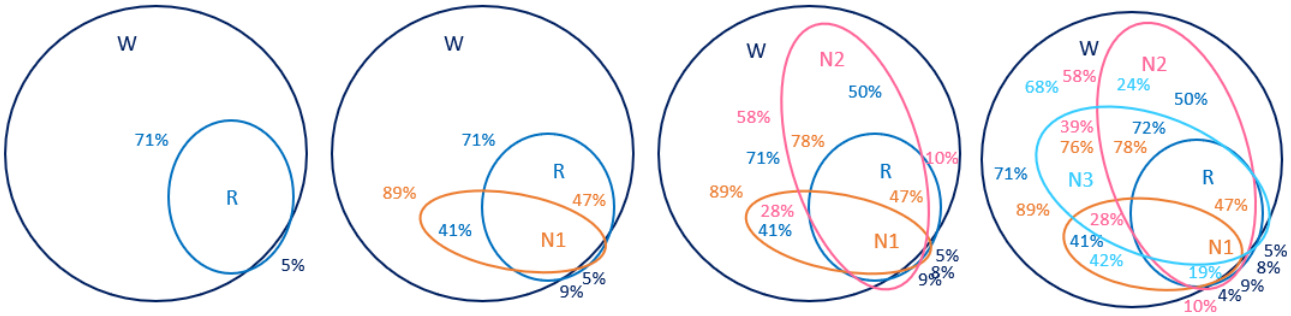

Figure 6.1: Visualizations of the anomaly errors. *W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement.*

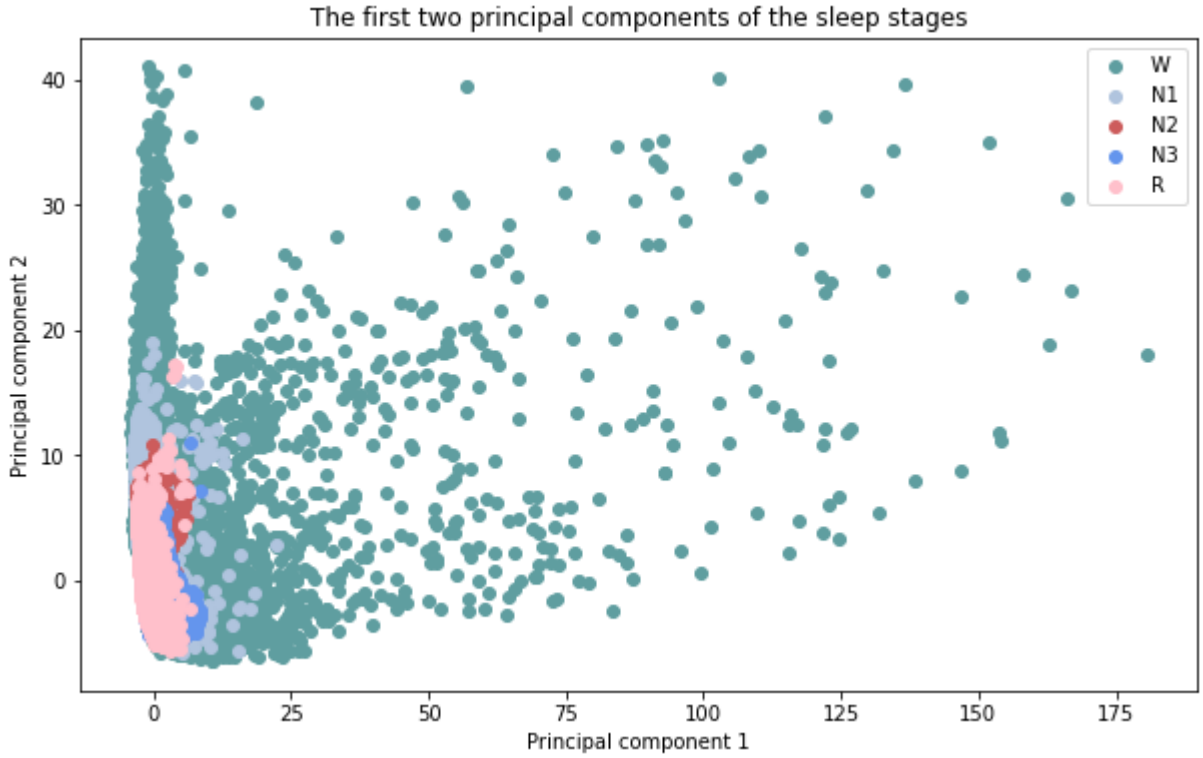

Figure 6.2: Visualization of the first two principal components of the selected features in the sleep stages. *W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement.*

7
Results: critically ill children

For all epochs from the PICU data, it was determined whether the calculated MD exceeded or fell below the threshold. Again, the error signifies the number of epochs categorized as anomalous to the total number of epochs. These errors can be read from Table 7.1. To assess the potential assignment of PICU data to sleep stage N, the Gaussian Model was also trained specifically on sleep stage N (final row). Visual representations in Figure 7.1 show histograms of the MDs of the two distinct PICU files, alongside the MDs for the training and validation sets of the N1 stage. Notably, the majority of the epochs in these patient files exceed the threshold and can be categorized as anomalous. The errors associated with these patient files, PICU002 and PICU0018, are 0.925 and 0.927, respectively. Conversely, patient files with errors approaching zero indicate that nearly all epochs align with the sleep stage specified in the corresponding row.

Patients with a PICU indication of heart disease exhibited notably high error rates in sleep stages N1, N2, R, and N. The patient with heart failure demonstrated elevated error rates in N2 and R, with relatively high errors in N1 and N. Those with a PICU indication of cardiothoracic surgery showed relatively high error rates in sleep stage N. Patient PICU009 displayed high errors across all classes, while a patient with a neuromuscular disease, similar to PICU009, exhibited relatively high errors in R and N and medium errors in N2 and N3. Patients with respiratory infection or respiratory insufficiency had the highest errors in sleep stages R and N. Patient PICU024 showed very high errors in sleep stages N1, N2, and R, with relatively high errors in sleep stage N. Another patient with sepsis only had a high error rate in sleep stage N. The patient with the highest PELOD-2 score did show high error rates for stages N1, N2, R, and N.

The distribution of epochs per patient, categorized into different sleep stages, is visually represented in Figure 7.2. The prevalence of epochs classified into multiple sleep stages leads to overlapping percentages of the total number of epochs in each sleep stage. The count of epochs assigned to a given number of sleep stages is detailed in Figure 7.3. Importantly, the y-axis ranges differ between figures due to the accumulation of epochs in the distribution figure resulting from classification into multiple sleep stages. The segments with a value of zero in this bar plot indicate unclassified epochs. For instance, 68% of the epochs in patient file PICU009 remained unclassified. In Figure 7.2, it is evident that certain patient files lack epochs classified into sleep stages N2 and REM. The proportion of epochs categorized as stages wake and N1 remained relatively consistent across all patients. The specific count of epochs corresponding to the data presented in these figures can be found in the Supplementary Materials (section A.5). The total count of epochs classified into each sleep stage can be read from Table 7.2. It's important to note that the percentages overlap as epochs are classified into multiple sleep stages. When comparing the distribution of epochs by sleep stage in PICU data with that of PSG data (Table 5.1 and Table 7.2, it becomes evident that, in both distributions, wake and N3 are the most prominent stages. Stages N1 and N have the least number of epochs in PSG data which are stages R and N in PICU data.

Table 7.1: Errors of the Gaussian Models of the different sleep stages applied on data from critically ill children. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	P001	P002	P003	P004	P005	P006	P007	P008	P009	P010	P012	P013	P014	P015	P016	P018	P020	P021	P022	P024	P025	P026	P027
W	0.000	0.000	0.000	0.000	0.000	0.000	0.044	0.001	0.995	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.044	0.000	0.000	0.000	0.000	0.000	0.000
N1	0.289	0.925	0.201	0.053	0.543	0.149	1.000	0.961	1.000	0.258	0.554	0.056	0.300	0.053	0.201	0.927	0.000	0.588	0.431	0.998	0.262	0.258	0.289
N2	0.558	0.940	0.365	0.125	0.655	0.308	1.000	0.988	1.000	0.497	0.964	0.188	0.359	0.125	0.365	1.000	0.000	0.638	0.612	0.997	0.309	0.497	0.557
N3	0.400	0.029	0.132	0.055	0.025	0.103	0.506	0.118	1.000	0.185	0.258	0.020	0.018	0.055	0.132	0.247	0.506	0.007	0.014	0.518	0.011	0.185	0.667
R	0.772	0.941	0.506	0.228	0.762	0.418	1.000	1.000	1.000	0.810	1.000	0.334	0.714	0.228	0.506	1.000	0.000	0.755	0.849	0.996	0.524	0.810	0.772
N	0.878	0.862	0.683	0.523	0.734	0.757	0.819	0.758	0.684	0.794	0.675	0.899	0.741	0.523	0.683	0.845	0.819	0.775	0.821	0.699	0.695	0.793	0.878

**P refers to PICU in the column names.*

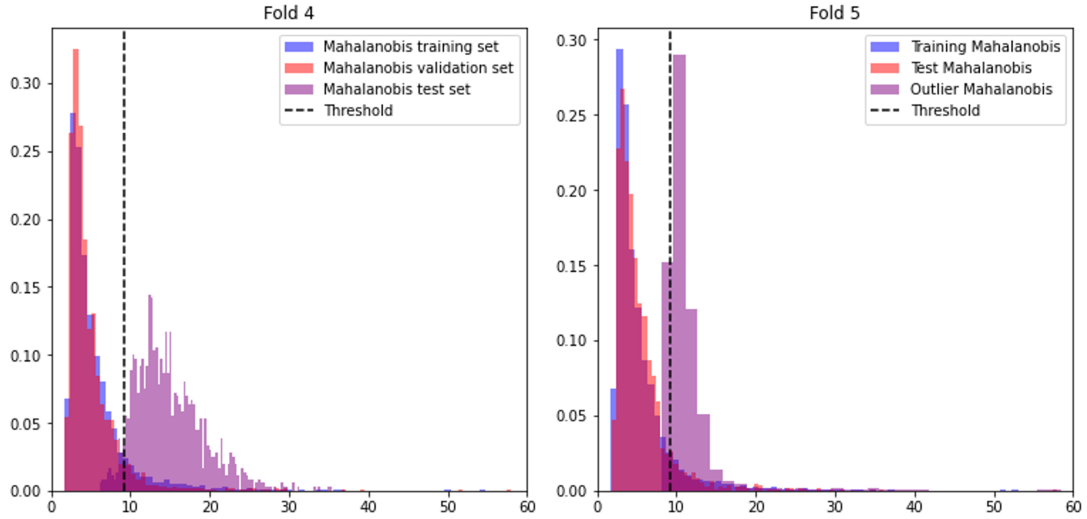

Figure 7.1: Examples of histograms of the Mahalanobis Distances for PICU002 (left) and PICU018 (right) together with the Mahalanobis Distances for the training and validation set of sleep stage N1

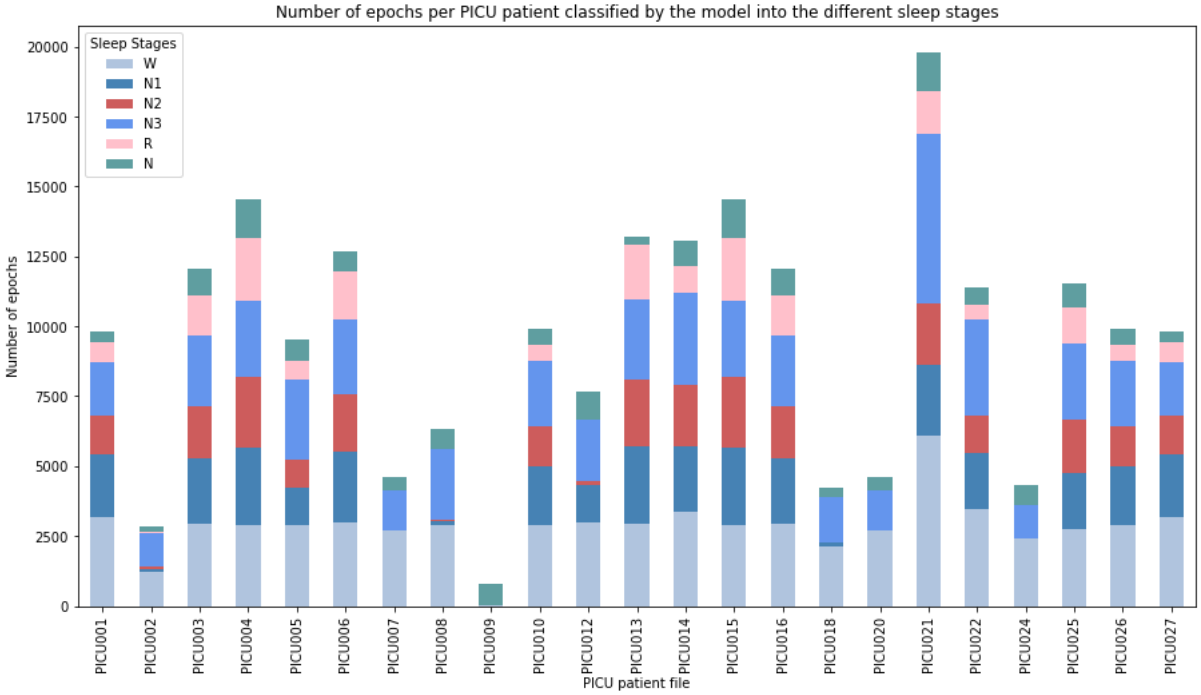

Figure 7.2: Distribution of epochs by sleep stage. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

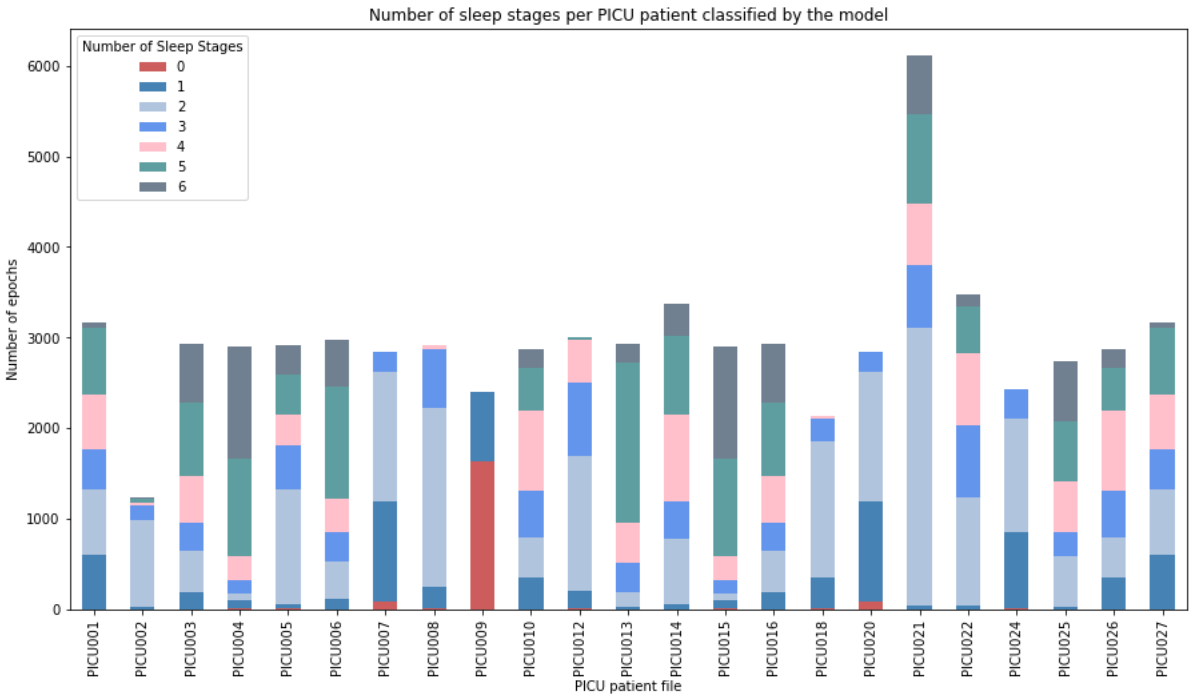

Figure 7.3: Number of sleep stages per PICU patient classified by the model. *PICU = Pediatric Intensive Care Unit.*

Table 7.2: Distribution of epochs by sleep stage: total counts. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

Sleep stages	PICU data	Percentages
W	65339	96%
N1	36131	53%
N2	27755	41%
N3	54490	80%
R	18643	27%
N	16784	25%
Total	67992	

For all PICU patients, it was analyzed whether discernible patterns in the temporal evolution of classifications could be identified. Figure 7.4 presents the MDs for PICU003, illustrating instances with moderate errors for the sleep stages. The striped red line represents the threshold for the Gaussian Model. The errors for sleep stage W, N1, N2, N3, R, and N were 0.000, 0.201, 0.365, 0.132, 0.506, and 0.683, respectively. Notably, clusters of patterns can be discerned in the MDs of this patient, characterized by periods of MD enlargement followed by subsequent reduction. Epochs with MDs exceeding the threshold were classified as anomalous, and the ratio of anomalous epochs to the total number of epochs represents the error rate for the patient file. For some patients, the clusters in the MDs were less prominent, and in other patients, it was observed that nearly all epochs were classified as anomalous. Figures of these examples are available in the Supplementary Materials section A.6 (Figure A.14, Figure A.15, Figure A.16).

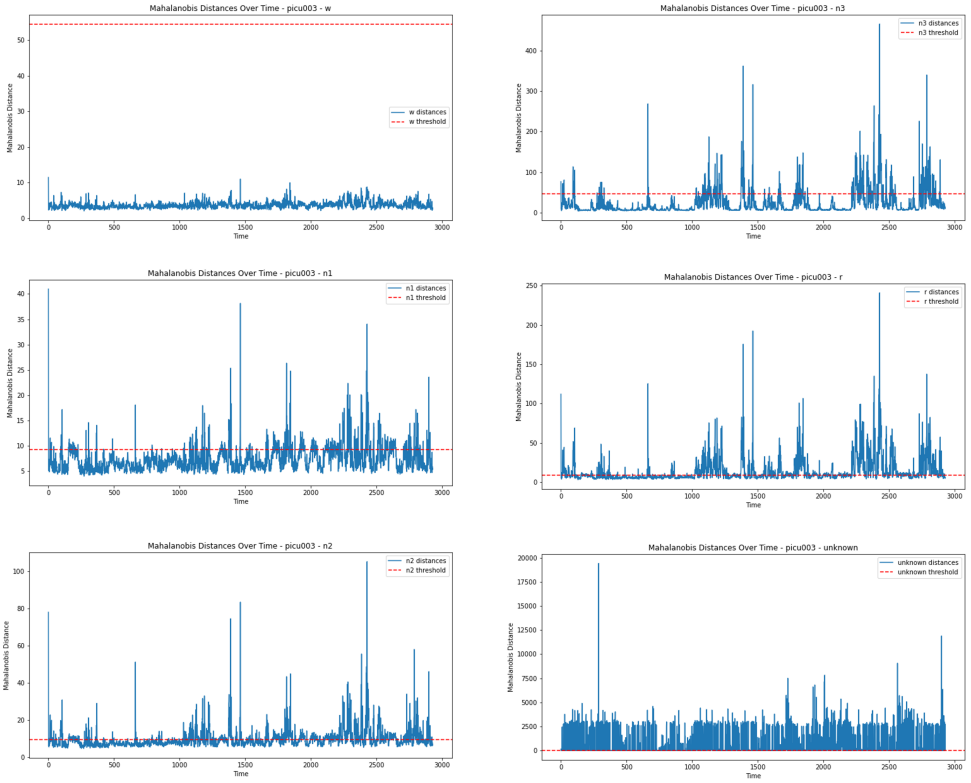

Figure 7.4: Mahalanobis Distances over time for PICU003. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

The results of manual sleep scoring by neurophysiologists are available in Figure 7.5. These results can be compared to the model's classification previously shown in Figure 7.2. Given the hypothesis that PICU data may not align with the sleep stages defined by AASM criteria, it's crucial not to treat the labels assigned by the neurophysiologists as absolute truth. Notably, discrepancies in the classification of epochs into sleep stage N are evident when comparing the bar plots of the model and of manual scoring. As can be observed, in 11 of the 23 patients, no labels of N1 and N3 were assigned to the epochs by the neurophysiologists. They more often assign sleep stage N to epochs. The total number of epochs assigned per sleep stage can be read from Table 7.3.

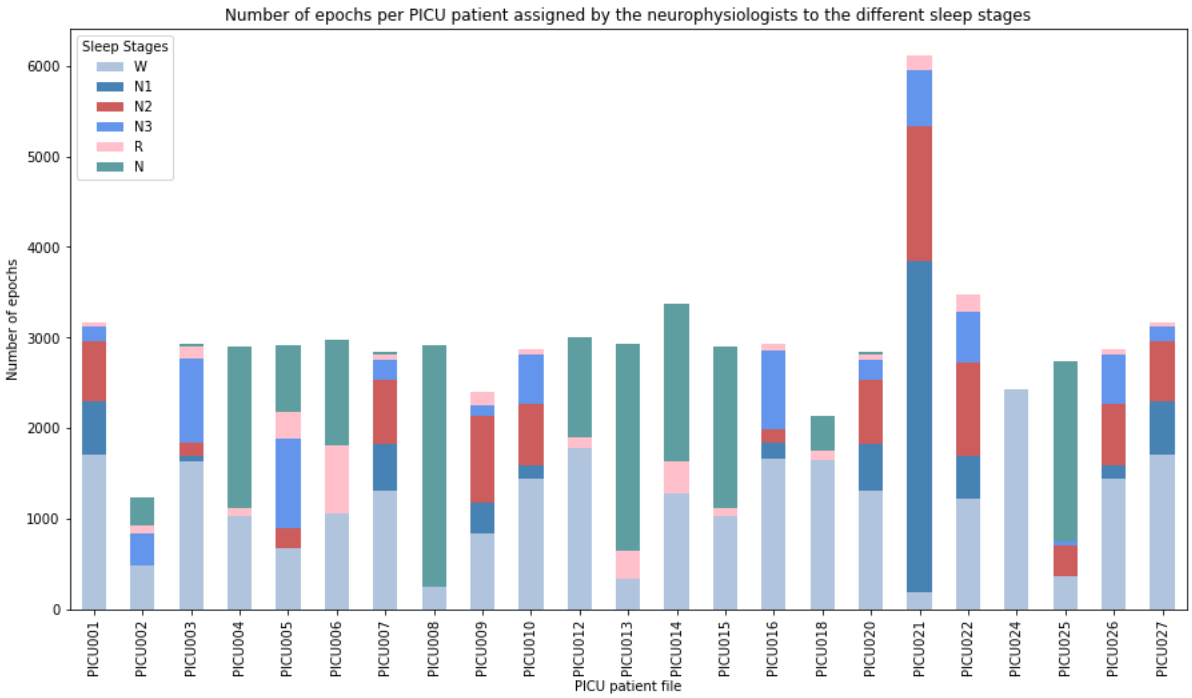

Figure 7.5: Number of epochs per PICU patient assigned by the neurophysiologists to the different sleep stages. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

Table 7.3: Labels assigned by the neurologists. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

Sleep stages	PICU data	Percentages
W	22704	40%
N1	3429	6%
N2	6235	11%
N3	5203	9%
R	3005	5%
N	16016	28%
Total	56592	

8

Results: classifications over time

When another PSG recording of PICU patients was available, it was included in this analysis to assess discrepancies between two measurements at different points in time. From six patients in total, a second measurement was available. The patient characteristics can be read from Table 8.1. Note that in three patients the PELOD-2 score is reduced during the second recording as compared to the first recording. The lowest PELOD-2 score is two, observed in the second recording of PICU000, a patient who was also not on mechanical ventilation during this recording. Three patients maintained the same PELOD-2 score in both recordings, although one of them received different medications during these measurements. The duration between the two recordings for all patients was 4 days or more.

Table 8.1: Characteristics PICU patients with longitudinal measurements. *PICU = Pediatric Intensive Care Unit; PELOD = pediatric logistic organ dysfunction.*

Patient file	Day of measurement	PICU indication	Medication	Intubation	PELOD-2 score
PICU000A*	3	Cardiac diseases	Midazolam	Yes	7
PICU000B*	10	Cardiac diseases	Midazolam	No	2
PICU005A	2	Cardiothoracic surgery	Opioids	Yes	8
PICU005B*	7	Cardiothoracic surgery	Midazolam	Yes	5
PICU006A*	6	Cardiothoracic surgery	Midazolam	Yes	5
PICU006B	13	Cardiothoracic surgery	Midazolam	Yes	5
PICU021A	1	Respiratory infection	Midazolam, opioids	Yes	7
PICU021B*	7	Respiratory infection	Midazolam, opioids	Yes	6
PICU022A*	3	Respiratory insufficiency	Midazolam, opioids	Yes	7
PICU022B	7	Respiratory insufficiency	Midazolam, opioids	Yes	7
PICU026A*	3	Post-resuscitation	Midazolam	Yes	9
PICU026B	12	Post-resuscitation	Opioids, esketamine	Yes	9

** These recordings were not included in the previous analyses of critically ill children.*

The errors corresponding to each recording are detailed in Table 8.2. Surprisingly, discrepancies between measurements PICU006A and PICU006B can be observed, despite their similar patient characteristics. Furthermore, a discrepancy in the number of epochs classified into stages N1, N2, and R is observed between measurements PICU000A and PICU000B, which also exhibit a difference in PELOD-2 scores. Intriguingly, a higher PELOD-2 score does not necessarily correlate with a higher error, as can be observed in the recordings PICU005A and PICU005B. During the second recording, this patient received midazolam instead of opioids. It is noticeable that errors were higher during measurements PICU005B and PICU026A, during which the patient received midazolam, compared to measurements in which they received opioids and/or esketamine. This was more prominent in patient PICU026.

The distributions of epochs by sleep stage and the counts of epochs assigned to a given number of sleep stages are illustrated in Figure 8.1 and Figure 8.2. Again it is important to notice that the ranges of the y-axis differ due to the accumulation of the counts of epochs in the figure illustrating the distribution of epochs by sleep stage. The recordings of PICU000, PICU006, and PICU026 exhibit significant differences in the sleep stages their epochs are assigned to. The three remaining patients exhibited a consistent pattern in the proportions of sleep stages across both recordings. Regarding the patients with recordings with and without midazolam administration, the epochs of recording PICU026A were categorized into fewer distinct sleep stages than recording PICU026B, aligning with the larger margin of error. However, the differences in errors between recording PICU005A and PICU005B were not substantial enough to visually impact the distribution of classifications.

Table 8.2: Errors of the Gaussian Models of the different sleep stages applied on data from critically ill children. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	PICU000A	PICU000B	PICU005A	PICU005B	PICU006A	PICU006B	PICU021A	PICU021B	PICU022A	PICU022B	PICU026A	PICU026B
W	0.055	0.011	0.000	0.000	0.230	0.000	0.000	0.119	0.000	0.000	0.048	0.000
N1	0.729	0.380	0.543	0.706	0.994	0.149	0.588	0.455	0.364	0.431	0.991	0.258
N2	0.921	0.448	0.655	0.745	0.994	0.308	0.638	0.560	0.548	0.612	0.997	0.497
N3	0.186	0.116	0.025	0.122	0.988	0.103	0.007	0.266	0.013	0.014	0.457	0.185
R	0.988	0.598	0.762	0.783	0.994	0.418	0.755	0.559	0.741	0.849	1.000	0.810
N	0.875	0.769	0.734	0.720	0.777	0.757	0.775	0.621	0.667	0.821	0.707	0.793

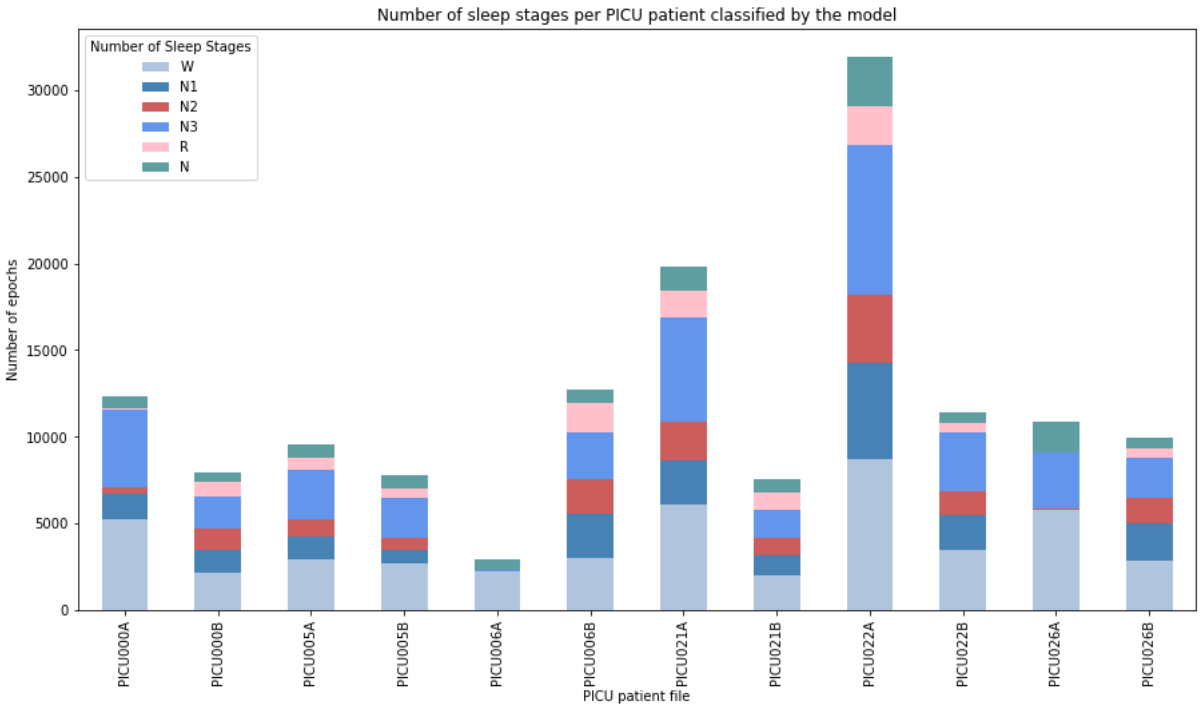

Figure 8.1: Epoch distribution across sleep stages per PICU patient. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

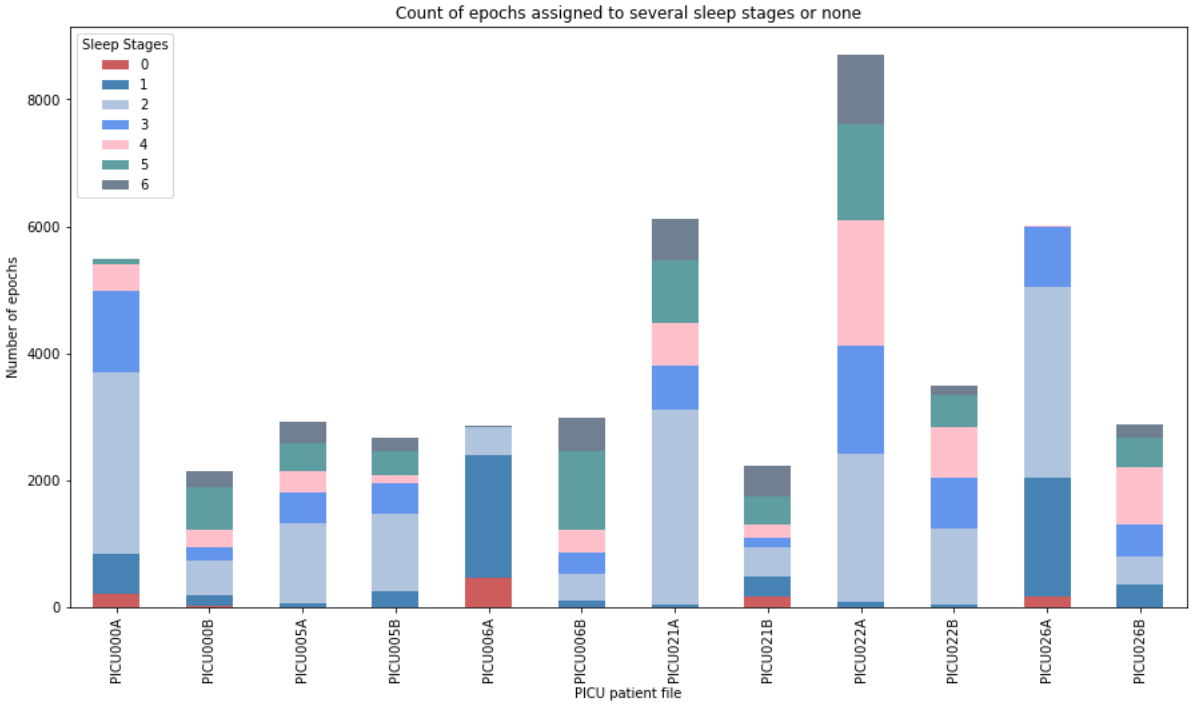

Figure 8.2: Number of sleep stages per PICU patient classified by the model. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

9 Discussion

The present study introduces an alternative approach to assess whether PSG-derived EEG data from critically ill children could be categorized into sleep stages based on the AASM scoring criteria by employing anomaly detection using Gaussian Models. The models trained on the sleep stages of PSG-derived EEG data from non-critically ill children using EEG derivation C3-A2 for feature selection resulted in validation errors aligning with the margin error of 10% of the training set. However, these models experienced challenges in effectively differentiating the sleep stages, resulting in a large overlap of epochs between sleep stages. A comparable trend emerged when applying these models to data from critically ill children, where a significant proportion of their epochs were categorized into multiple sleep stages. Interestingly, a higher PELOD-2 score did not consistently correlate with an increased number of anomalous classifications in the epochs of these patients to those with lower PELOD-2 scores.

To the best of our knowledge, this is the first study that explores sleep during critical illness and its alignment with conventional AASM sleep stages using anomaly detection methods. Previous studies that aimed to develop automated sleep scoring algorithms for non-critically ill and critically ill children already suggested that critically ill children may experience different sleeping patterns compared to non-critically ill children [21, 35]. Anomaly detection emerges as a valuable approach to identifying potential deviations in EEG patterns between critically ill and non-critically ill children. Unlike classification methods relying on balanced class labels in the dataset, anomaly detection models only use labels from the target class, in this case, PSG recordings of non-critically ill children. Evaluation of anomaly detection models varies from classification methods, and standard metrics like precision and recall may not directly apply. Previous studies in medical applications using anomaly detection methods typically used actual labels from 'abnormal data' to assess performance using metrics such as area under the curve, sensitivity, and specificity [30, 31, 33, 34, 47, 48]. The study's hypothesis challenges the assumption that critically ill children's data align with AASM sleep stages, and thus, labels assigned by neurophysiologists cannot be considered absolute truth. The anomaly detection model solely relies on prior knowledge of labels from PSG recordings of non-critically ill children. The model focuses on capturing true anomalies while minimizing false positives. This evaluation was conducted on the validation set, making direct comparisons to traditional classification methods challenging.

Previous studies evaluating interrater reliability in manual sleep scoring for healthy adults reported a Cohen's kappa of 0.76 for overall sleep staging. Per sleep stage, the Cohen's kappa values were 0.70, 0.24, 0.57, 0.57, and 0.69 for the W, N1, N2, N3, and R stages, respectively [49]. The moderate Cohen's kappa value of N2 can be a result of confusion with stage N1. Typically, after a K complex and a sleep spindle are detected once, the stage is scored as N2 even if the N2 features are no longer feasible. The poor agreement of stage N1 can be caused by the fact that the transition from wake to N1 can be difficult to recognize, especially in patients exhibiting spare alpha activity. Scoring discrepancies are most common in epochs of transition from one stage to another [49]. The overall Cohen's kappa values for patients in an intensive care unit was 0.51-0.56, consistent with challenges highlighted in previous studies regarding sleep quantification in critical care settings [21, 49, 50]. Considering the complexities observed in interrater reliability, achieving a distinct separation between sleep stages may prove challenging, particularly in critically ill children admitted to the PICU. While Twist et al. (2023) successfully classified two- and three-state sleep stages in non-critically ill children, the anomaly detection model in our study faced difficulties in effectively distinguishing between stages, reflecting the inherent complexities of sleep stage classification. Upon evaluating the errors of the Gaussian Models across various sleep stages, it became apparent that the models trained on a specific sleep stage did not consistently produce a comparable error rate when applied to the same sleep stage as if it were treated distinctly. Notably, a Gaussian Model trained on the wake stage primarily assigned epochs from other

sleep stages to wake. Conversely, models trained on other stages exhibited larger errors when applied to wake, indicating the majority of the wake class was considered anomalous by these models. This inconsistency suggests that the feature values of wake may span a wide range, resulting in an overlap with feature values from other stages. Factors such as muscle activity, influencing gamma activity in EEG, may impact the estimation of gamma activity representing wake [51]. Conversely, Gaussian Models trained on N1 and R demonstrated substantial errors when applied to wake, N2, and N3 but resulted in a low error when treated as distinct sleep stages for other Gaussian Models. This suggests that the feature values for N1 and R may fall within a more restricted range.

Once confidence was established in the model's accurate validation errors using EEG derivation C3-A2 for feature selection, these models were subsequently applied to PICU data to assess epoch classifications. In the majority of the PICU patients, the error rate was high for sleep stages N2, R, and N. Additionally, many patients had a high error rate for N1. A significant proportion of epochs were classified into multiple sleep stages, partly due to the overlapping nature found in the sleep stages of non-critically ill children, as indicated by the model. The findings of this study align with prior research on sleep patterns in the PICU. Earlier studies on critically ill children in the PICU reported a significant reduction or absence of REM sleep, as well as a reduced proportion of N3 sleep [21, 52–54]. In our investigation, a notable error rate was observed for sleep stage R in the majority of the patients, indicating that the model assigned a small number of epochs to this stage. Conversely, error rates for N3 were not as high for most patients. It's important to note that comparing the percentages of sleep stage distribution in critically ill children with those in non-critically ill children is challenging due to the overlap resulting from the classification of epochs into multiple sleep stages. Despite this challenge, it is notable that the proportion of N3 sleep is not reduced in various critically ill children. Both non-critically and critically ill children predominantly exhibit N3 as the most prevalent or one of the most prevalent sleep stages. The distribution of sleep stages in PICU patients reveals that most of the epochs were classified as wake and N3, followed by N1 and N2, with the fewest epochs assigned to stages R and N. This trend mirrors the distribution of the PSG dataset, where wake and N3 were predominant, followed by N2 and R, and the fewest epochs were assigned to N1 and N. Surprisingly, only eight patients including three with more than one epoch, had unclassified epochs. Some epochs were classified into a single sleep stage, while others were categorized into multiple sleep stages, leading to overlapping percentages. Neurophysiologists primarily label data as wake and N where the model might classify epochs into multiple sleep stages. In critically ill children, when an epoch exhibits characteristics of NREM but cannot be classified into N1, N2, or N3, neurophysiologists assign the label N to this epoch. Notably, a substantial portion of the PICU data was only classified into stages REM, N, or wake by neurophysiologists. Epochs exhibiting atypical EEG characteristics are typically assigned to sleep stage N by neurophysiologists, resulting in a stage with a diverse range of feature values. However, the Gaussian Model of sleep stage N is trained using PSG data, likely possessing distinct EEG characteristics compared to PICU patients labeled as N by neurophysiologists. The neurophysiologists assigned the label N to a notably larger proportion of epochs compared to the model's classifications. According to the AASM scoring criteria, epochs should be classified into the predominant sleep stage when characteristics of multiple sleep stages are detected within the epoch [16]. Contrary to the model classifying epochs into multiple sleep stages, neurophysiologists typically assign epochs to a single sleep stage.

Previous studies reported atypical sleep patterns in mechanically ventilated patients [6, 52]. A study focused on children with severe acute bronchopneumonia during mechanical ventilation, exploring various sedation levels, revealed that those under sedation exhibited slower EEG activity during sleep, diffuse delta wave activity, and a loss of N1 and REM stages. Notably, the deep sedated group demonstrated a significantly higher incidence of REM stage loss and a decrease in N2 sleep compared to the lightly sedated group [52]. This validates a prior report suggesting that benzodiazepines could decrease REM and increase N2 sleep in ventilated children [55]. Furthermore, the report indicated that the sleep cycle returned to normal without N1 or REM loss after weaning [52]. Unlike the referenced study, our analysis did not consider the level of sedation. Instead, a longitudinal analysis was conducted to assess variations in classifications over time in the context of underlying illness and medication. Among patients with two available PSG recordings, one was mechanically ventilated during the first measurement and not during the second. Interestingly, this patient exhibited a similar number of epochs classified as N1. Although REM sleep was absent during mechanical ventilation, it reappeared

after weaning, aligning with the findings of the aforementioned study. The PELOD-2 score was lower during the PSG recording after weaning from mechanical ventilation. However, when assessing all PICU patients, those on mechanical ventilation did not show higher error rates than non-ventilated patients. Among patients with two available PSG recordings, higher error rates were observed during the recording in which midazolam was administered. For one patient, significantly higher errors for stages N1, N2, and R were noted during the measurement with midazolam compared to the measurement without it. Surprisingly, in all critically ill patients, an increase in the PELOD-2 score did not necessarily correlate with higher error rates. It is crucial to emphasize that, even though these comparisons are drawn among sleep stages, these stages might not depict the sleep states in critically ill children. The disparities between the results of prior studies and the findings in the present study also suggest the challenges associated with defining the EEG patterns of critically ill children.

The present study has several strengths worth mentioning. It established a robust framework for anomaly detection, laying a foundation for future developments in this area. Given that prior research indicates potential differences in the sleep of critically ill children compared to non-critically ill children, a thorough exploration of sleep patterns in critically ill children becomes imperative. Anomaly detection proves to be a valuable tool in identifying potential deviations in EEG patterns among critically ill children, particularly considering that the data from critically ill children is poorly sampled and not well-defined. The assessment of results on a per-sleep-stage basis provides valuable insights into the potential divergence in sleep patterns among critically ill children as compared to non-critically ill children.

However, certain limitations of the present study warrant also need to be addressed. The methodology for feature selection, as outlined in van Twist et al. (2023), involved the calculation of features from various EEG channels, which was adopted for model development in this study. Unfortunately, the use of all these features resulted in excessively high computational costs, needing a reduction in the number of features. Traditional approaches for feature selection based on importance or principal component analysis were considered inadequate, as the anomaly detection model does not distinguish between two classes but rather differentiates one class from all other data. These techniques risk excluding features that may not provide significant information about a specific sleep stage but could play a crucial role in highlighting noteworthy deviations within the overall dataset. Therefore, the feature selection process in this study was based on clinical and PSG technical considerations. The AASM recommends the use of at least one of the following EEG electrode combinations for sleep scoring: C4-A1, C3-A2, C4-Fpz, or C3-Fpz [16]. Fpz-Cz is the most commonly used EEG channel in current tools for sleep staging based on single-channel design [56]. However, the Fpz electrode was not placed during the PSG measurements in this study. C3-A2 is a frequently chosen single EEG channel in previous studies focused on sleep staging [56]. Given the AASM recommendations and the available EEG electrode combinations, EEG derivation C3-A2 or C4-A1 was selected for feature selection in this study. Consequently, model development was conducted using a single EEG channel for feature selection to minimize computational demands. Other studies have also proposed automatic sleep scoring methods relying solely on EEG, employing either multi-channel or single-channel analysis for efficiency reasons [40]. While previous studies achieved accuracies above 75% with single-channel EEG information, it is acknowledged that the performance of single-channel EEG-based automatic scoring is slightly lower than that of multi-channel EEG or multi-signal approaches, especially in REM sleep [40].

A fundamental limitation is the constrained amount of data, precluding subgroup analysis. Obtaining more data from critically ill children is imperative to gain deeper insights into their sleep patterns. Additionally, this study lacked a comprehensive investigation of the data quality of the PICU files, as the preprocessing steps were adopted from a previous study and not reassessed. The artifact detection algorithm should be re-evaluated to optimize data quality. Since PICU data epochs are analyzed independently and completeness of the PICU data does not have to be maintained, clusters of epochs with low quality could be excluded. Furthermore, it is possible that some of the additional PSG recordings included for the longitudinal analysis were not part of the initial analyses due to low quality. Therefore, considering the selective deletion of certain epochs from the PICU data is worthwhile. Moreover, the mean values of the parameters from the cross-validation with the Gaussian Models were used to calculate the MDs of the epochs and categorize them due to efficiency reasons. Calculating the MDs of

these epochs within the cross-validation might be more robust. Obtaining more PSG recordings per patient is crucial for gaining deeper insights into differences in sleep staging over time and for varying degrees of illness. However, it is essential to consider that the initial patient inclusion at the PICU was challenging due to the severity of their illness, and obtaining consent for multiple measurements was met with hesitancy from parents.

After achieving a substantial increase in data volume, it is advisable to conduct subgroup analyses. To evaluate pediatric EEGs, it is crucial to assess whether observed patterns align with maturational age [16]. Therefore, integrating age categories into anomaly detection models is imperative, establishing classes for each sleep stage within each age category to facilitate comparisons among critically ill children within the same age range. Given the evolving nature of EEG characteristics during development, incorporating age-specific classes enhances model accuracy [16]. In the present study, a discernible pattern of sleep stage fluctuation occurring in clusters was noted among various patients. The temporal evolution of classifications demonstrated a non-random formation of clusters. Introducing a sliding window could be a valuable enhancement to evaluate how the model's performance might improve. Typically, manual scoring incorporates information from preceding and subsequent epochs when classifying a given epoch. Assessing the model's performance by using the preceding epoch as a priori information for the current epoch, and updating this context iteratively, could provide valuable insights. However, real-time sleep scoring may face challenges due to the contextual importance in sleep scoring. Additionally, further exploration of the feature selection process is recommended. Selecting features that effectively represent the model's objective is crucial. Reassessing the chosen features for EEG derivations and evaluating whether additional features should be incorporated is beneficial. Exploring outcomes using features derived from amplitude-integrated (aEEG) electrodes, given their non-invasive nature, is worthwhile. Although aEEG lacks a mastoid reference, potentially impacting cancellation effects, evaluating model performance using EEG derivation C3-C4 for feature selection can serve as an initial step. EMG helps to classify the wake stage from the rest of the sleep stage. However, its contribution to accuracy is negligible [4, 43]. EEG plays a vital role in discerning between different NREM stages, while EOG assists in distinguishing between REM and NREM stages. Integrating EOG signals into single EEG channels for feature selection may enhance the model's performance, particularly for distinguishing REM sleep [57].

In the context of anomaly detection using Gaussian Models, it is acknowledged that a Gaussian may not fully capture the entire training dataset. The simplicity and interpretability of the Gaussian Model make it suitable for smaller datasets, yet its limitations prompt the exploration of more advanced models. Building upon this framework with Gaussian Mixture Models (GMM) is recommended. GMM, representing a weighted sum of Gaussian component densities, can provide a more nuanced description of the training set by approximating arbitrarily shaped densities. Developing more advanced models aligns with the necessity for a larger dataset to establish robustness. [58]. An illustration of the GMM in red and the Gaussian in blue is given in Figure 9.1. Additionally, comparing the results of supervised learning with those of anomaly detection could give insights into model performance and distinctions between the two approaches. It is conceivable that a Gaussian Model encapsulates datapoints of two sleep stages within its description of the training set, whereas a supervised learning model might delineate a boundary between the two classes as illustrated in Figure 9.2.

In conclusion, the current study effectively established an anomaly detection framework using Gaussian Models to analyze the sleep patterns of critically ill children. Anomaly detection is a valuable approach in this context, given the well-defined sleep of non-critically ill children according to the AASM scoring criteria, whereas characterizing sleep in critically ill children represents challenges. Overlap in sleep stages was seen in data from non-critically ill children and epochs from critically ill children were classified into multiple sleep stages. The discrepancy between our model classifying epochs into multiple sleep stages and the neurophysiologists' typical assignment to sleep stage N suggests a potential limitation in the applicability of the AASM scoring criteria to critically ill children. While our implementation of the Gaussian Model has been successful, it would be imperative to explore more advanced anomaly detection models. Additionally, a critical reassessment of the feature selection process is imperative, focusing on features that are both informative and distinctively suited to the objectives of the anomaly detection model. After refining this anomaly detection framework and building confidence in

its application, conclusive insights into the sleep patterns of critically ill children admitted to the Pediatric Intensive Care Unit can be drawn. This understanding will contribute significantly to advancing knowledge in the sleep of critically ill children and, consequently, potentially in tailoring interventions to individually target and improve their sleep.

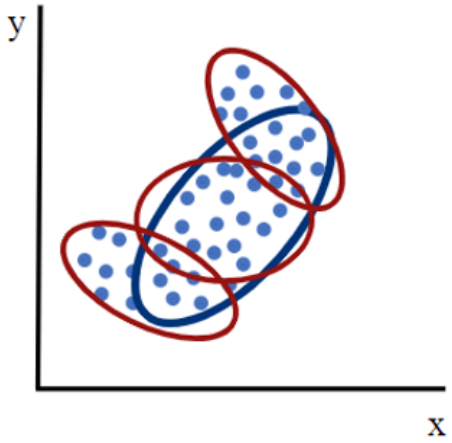

Figure 9.1: Example of a Gaussian (blue) and Gaussian Mixture Models (red).

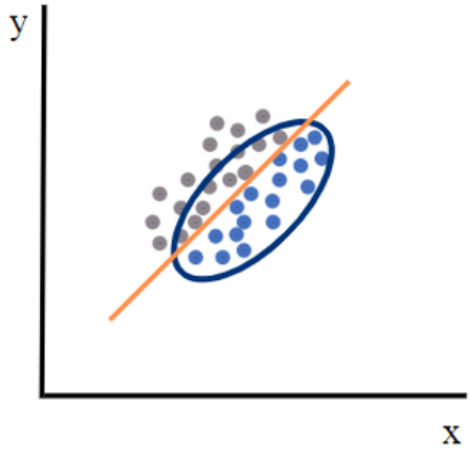

Figure 9.2: Example of Gaussian (blue) and a boundary obtained by supervised learning (orange).

References

[1] Y. Boyko et al. “Atypical sleep in critically ill patients on mechanical ventilation is associated with increased mortality”. In: *Sleep and Breathing* 23.1 (Sept. 2018), pp. 379–388. DOI: 10.1007/s11325-018-1718-3. URL: https://doi.org/10.1007/s11325-018-1718-3.

[2] K. Davis, K. Parker, and G. Montgomery. “Sleep in infants and young children”. In: *Journal of Pediatric Health Care* 18.3 (May 2004), pp. 130–137. DOI: 10.1016/s0891-5245(03)00150-0. URL: https://doi.org/10.1016/s0891-5245(03)00150-0.

[3] E. Kakar et al. “Sleep assessment in critically ill adults: A systematic review and meta-analysis”. In: *Journal of Critical Care* 71 (Oct. 2022), p. 154102. DOI: 10.1016/j.jcrc.2022.154102. URL: https://doi.org/10.1016/j.jcrc.2022.154102.

[4] F. Hiemstra. *Automated EEG-based sleep monitoring in critically ill children*. Unpublished. 2021.

[5] A. Meester. *Automated sleep monitoring based on vital signs in critically ill children*. Unpublished. 2022.

[6] P.M. Watson et al. “Atypical Sleep in Ventilated Patients”. In: *Critical Care Medicine* 41.8 (Aug. 2013), pp. 1958–1967. DOI: 10.1097/ccm.0b013e31828a3f75. URL: https://doi.org/10.1097/ccm.0b013e31828a3f75.

[7] S.R. Kudchadkar et al. “Temporal characteristics of the sleep EEG power spectrum in critically ill children”. In: *Journal of Clinical Sleep Medicine* 11.12 (Dec. 2015), pp. 1449–1454. DOI: 10.5664/jcsm.5286. URL: https://doi.org/10.5664/jcsm.5286.

[8] R. Stowe and O. Afolabi-Brown. “Pediatric polysomnography—A review of indications, technical aspects, and interpretation”. In: *Paediatric Respiratory Reviews* 34 (Apr. 2020), pp. 9–17. DOI: 10.1016/j.prrv.2019.09.009. URL: https://doi.org/10.1016/j.prrv.2019.09.009.

[9] J.V. Rundo and R. Downey. *Polysomnography*. Elsevier BV, Jan. 2019, pp. 381–392. DOI: 10.1016/b978-0-444-64032-1.00025-4. URL: https://doi.org/10.1016/b978-0-444-64032-1.00025-4.

[10] D.W. Carley and S.S. Farabi. “Physiology of Sleep”. In: *Diabetes Spectrum* 29.1 (Feb. 2016), pp. 5–9. DOI: 10.2337/diaspect.29.1.5. URL: https://doi.org/10.2337/diaspect.29.1.5.

[11] E. Bathory and S. Tomopoulos. “Sleep regulation, physiology and development, sleep duration and patterns, and sleep hygiene in infants, toddlers, and Preschool-Age children”. In: *Current Problems in Pediatric and Adolescent Health Care* 47.2 (Feb. 2017), pp. 29–42. DOI: 10.1016/j.cppeds.2016.12.001. URL: https://doi.org/10.1016/j.cppeds.2016.12.001.

[12] M. Carno et al. “Sleep monitoring in children during neuromuscular blockade in the pediatric Intensive care unit”. In: *Pediatric Critical Care Medicine* 5.3 (May 2004), pp. 224–229. DOI: 10.1097/01.pcc.0000124024.92280.f9. URL: https://doi.org/10.1097/01.pcc.0000124024.92280.f9.

[13] D. Valentine. *Learning EEG*. 2020. URL: https://www.learningeeg.com/.

[14] L. Fraiwan et al. “Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier”. In: *Computer Methods and Programs in Biomedicine* 108.1 (Oct. 2012), pp. 10–19. DOI: 10.1016/j.cmpb.2011.11.005. URL: https://doi.org/10.1016/j.cmpb.2011.11.005.

[15] S. Khalighi et al. “Automatic sleep staging: a computer assisted approach for optimal combination of features and polysomnographic channels”. In: *Expert Systems with Applications* 40.17 (Dec. 2013), pp. 7046–7059. DOI: 10.1016/j.eswa.2013.06.023. URL: https://doi.org/10.1016/j.eswa.2013.06.023.

[16] M. Grigg-Damberger et al. “The visual scoring of sleep and arousal in infants and children”. In: *Journal of Clinical Sleep Medicine* 03.02 (Mar. 2007), pp. 201–240. DOI: 10.5664/jcsm.26819. URL: https://doi.org/10.5664/jcsm.26819.

[17] A.B. Cooper et al. “Sleep in critically ill patients requiring mechanical ventilation”. In: *Chest* 117.3 (Mar. 2000), pp. 809–818. DOI: 10.1378/chest.117.3.809. URL: https://doi.org/10.1378/chest.117.3.809.

[18] N. Freedman et al. “Abnormal Sleep/Wake cycles and the effect of environmental noise on sleep disruption in the intensive care unit”. In: *American Journal of Respiratory and Critical Care Medicine* 163.2 (Feb. 2001), pp. 451–457. DOI: 10.1164/ajrccm.163.2.9912128. URL: https://doi.org/10.1164/ajrccm.163.2.9912128.

[19] C. Ambrogio et al. “Assessment of sleep in Ventilator-Supported Critically Ill Patients”. In: *Sleep* 31.11 (Nov. 2008), pp. 1559–1568. DOI: 10.1093/sleep/31.11.1559. URL: https://doi.org/10.1093/sleep/31.11.1559.

[20] H. Phan and K. Mikkelsen. “Automatic sleep staging of EEG signals: recent development, challenges, and future directions”. In: *Physiological Measurement* 43.4 (Apr. 2022), 04TR01. DOI: 10.1088/1361-6579/ac6049. URL: https://doi.org/10.1088/1361-6579/ac6049.

[21] A.B.G. Cramer et al. *Children in the pediatric intensive care unit experience limited REM sleep and frequent awakenings, and exhibit atypical electroencephalograms*. Unpublished work. 2023.

[22] S. Vacas et al. “The feasibility and utility of continuous sleep monitoring in critically ill patients using a portable electroencephalography monitor”. In: *Anesthesia and Analgesia* 123.1 (July 2016), pp. 206–212. DOI: 10.1213/ane.0000000000001330. URL: https://doi.org/10.1213/ane.0000000000001330.

[23] L. Reinke et al. “Intensive care unit Depth of sleep: proof of concept of a simple electroencephalography index in the non-sedated”. In: *Critical Care* 18.2 (Jan. 2014), R66. DOI: 10.1186/cc13823. URL: https://doi.org/10.1186/cc13823.

[24] A.H. Ansari 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* 17.1 (Jan. 2020), p. 016028. DOI: 10.1088/1741-2552/ab5469. URL: https://doi.org/10.1088/1741-2552/ab5469.

[25] A. Piryatinska et al. “Automated Detection of Neonate EEG sleep Stages”. In: *Computer Methods and Programs in Biomedicine* 95.1 (July 2009), pp. 31–46. DOI: 10.1016/j.cmpb.2009.01.006. URL: https://doi.org/10.1016/j.cmpb.2009.01.006.

[26] H. Ghimatgar et al. “Neonatal EEG Sleep stage Classification based on Deep learning and HMM”. In: *Journal of Neural Engineering* 17.3 (June 2020), p. 036031. DOI: 10.1088/1741-2552/ab965a. URL: https://doi.org/10.1088/1741-2552/ab965a.

[27] P. A. Estévez et al. “Polysomnographic pattern recognition for automated classification of sleep-waking states in infants”. In: *Medical and Biological Engineering and Computing* 40.1 (Jan. 2002), pp. 105–113. DOI: 10.1007/bf02347703. URL: https://doi.org/10.1007/bf02347703.

[28] Z. Cheng, C. Zou, and J. Dong. “Outlier detection using isolation forest and local outlier factor”. In: Sept. 2019. DOI: 10.1145/3338840.3355641. URL: https://doi.org/10.1145/3338840.3355641.

[29] S.S. Khan and M.G. Madden. “One-class classification: taxonomy of study and review of techniques”. In: *The Knowledge Engineering Review* 29.3 (Jan. 2014), pp. 345–374. DOI: 10.1017/s026988891300043x. URL: https://doi.org/10.1017/s026988891300043x.

[30] L. Shan et al. “Abnormal ECG detection based on an adversarial autoencoder”. In: *Frontiers in Physiology* 13 (Sept. 2022). DOI: 10.3389/fphys.2022.961724. URL: https://doi.org/10.3389/fphys.2022.961724.

[31] A. Karasmanoglou, M. Antonakakis, and M. Zervakis. “ECG-Based Semi-Supervised Anomaly Detection for Early Detection and Monitoring of Epileptic Seizures”. In: *International Journal of Environmental Research and Public Health* 20.6 (Mar. 2023), p. 5000. DOI: 10.3390/ijerph20065000. URL: https://doi.org/10.3390/ijerph20065000.

[32] C. Zhang et al. “Anomaly detection in ECG based on trend symbolic aggregate approximation”. In: *Mathematical Biosciences and Engineering* 16.4 (Jan. 2019), pp. 2154–2167. DOI: 10.3934/mbe.2019105. URL: https://doi.org/10.3934/mbe.2019105.

[33] N. Twomey et al. “Automated Detection of Perturbed Cardiac Physiology During Oral Food Allergen Challenge in Children”. In: *IEEE Journal of Biomedical and Health Informatics* 18.3 (May 2014), pp. 1051–1057. DOI: 10.1109/jbhi.2013.2290706. URL: https://doi.org/10.1109/jbhi.2013.2290706.

[34] J. Qin et al. “A novel temporal generative adversarial network for electrocardiography anomaly detection”. In: *Artificial Intelligence in Medicine* 136 (Feb. 2023), p. 102489. DOI: 10.1016/j.artmed.2023.102489. URL: https://doi.org/10.1016/j.artmed.2023.102489.

[35] E. Van Twist et al. “An EEG-based sleep index and supervised machine learning as a suitable tool for automated sleep classification in children”. In: *Journal of Clinical Sleep Medicine* (2023). DOI: 10.5664/jcsm.10880. URL: https://doi.org/10.5664/jcsm.10880.

[36] K. Veldscholte et al. “Continuous versus intermittent nutrition in pediatric intensive care patients: Protocol for a randomized controlled trial”. In: *JMIR Research Protocols* 11.6 (June 2022), e36229. DOI: 10.2196/36229. URL: https://doi.org/10.2196/36229.

[37] K. Veldscholte et al. “Intermittent feeding with an overnight fast versus 24-h feeding in critically ill neonates, infants, and children: an open-label, single-centre, randomised controlled trial”. In: *Clinical Nutrition* 42.9 (Sept. 2023), pp. 1569–1580. DOI: 10.1016/j.clnu.2023.07.010. URL: https://doi.org/10.1016/j.clnu.2023.07.010.

[38] S. Leteurtre et al. “99 PELOD-2: an update of the Pediatric Logistic Organ Dysfunction score”. In: *Archives of Disease in Childhood* 97.Suppl 2 (Oct. 2012), A28. DOI: 10.1136/archdischild-2012-302724.0099. URL: https://doi.org/10.1136/archdischild-2012-302724.0099.

[39] *SFAR - Société française d’Anesthésie et de réanimation*. URL: https://sfar.org/scores2/pelod2.php.

[40] I. Lambert and L. Peter□Derex. “Spotlight on sleep stage classification based on EEG”. In: *Nature and Science of Sleep* Volume 15 (June 2023), pp. 479–490. DOI: 10.2147/nss.s401270. URL: https://doi.org/10.2147/nss.s401270.

[41] W.R. Ruehland et al. “The 2007 AASM Recommendations for EEG Electrode Placement in Polysomnography: Impact on Sleep and Cortical Arousal Scoring”. In: *SLEEP* 34.1 (Jan. 2011), pp. 73–81. DOI: 10.1093/sleep/34.1.73. URL: https://doi.org/10.1093/sleep/34.1.73.

[42] U. Themes. *Sleep stages and scoring technique*. July 2016. URL: https://neupsykey.com/sleep-stages-and-scoring-technique/.

[43] R. Vallat and W. Matthew. “An open-source, high-performance tool for automated sleep staging.” In: *Elife 10* (2021). URL: https://doi.org/10.7554/eLife.70092.

[44] S. Prabhakaran. *Mahalanobis Distance; Understanding the math with examples (Python)*. Mar. 2022. URL: https://www.machinelearningplus.com/statistics/mahalanobis-distance/.

[45] S. Cansiz. “Multivariate Outlier detection in Python - towards data science”. In: (May 2023). URL: https://towardsdatascience.com/multivariate-outlier-detection-in-python-e946cfc843b3.

[46] I.T. Jolliffe and J. Cadima. “Principal component analysis: a review and recent developments”. In: *Philosophical Transactions of the Royal Society A* 374.2065 (Apr. 2016), p. 20150202. DOI: 10.1098/rsta.2015.0202. URL: https://doi.org/10.1098/rsta.2015.0202.

[47] S.M. You et al. “Unsupervised automatic seizure detection for focal-onset seizures recorded with behind-the-ear EEG using an anomaly-detecting generative adversarial network”. In: *Computer Methods and Programs in Biomedicine* 193 (Sept. 2020), p. 105472. DOI: 10.1016/j.cmpb.2020.105472. URL: https://doi.org/10.1016/j.cmpb.2020.105472.

[48] L. Cousyn, V. Navarro, and M. Chavez. “Outliers in clinical symptoms as preictal biomarkers”. In: *Epilepsy Research* 177 (Nov. 2021), p. 106774. DOI: 10.1016/j.eplepsyres.2021.106774. URL: https://doi.org/10.1016/j.eplepsyres.2021.106774.

[49] Y.J. Lee et al. “Interrater Reliability of Sleep Stage Scoring: A Meta-analysis”. In: *Journal of Clinical Sleep Medicine* 18.1 (Jan. 2022), pp. 193–202. DOI: 10.5664/jcsm.9538. URL: https://doi.org/10.5664/jcsm.9538.

[50] R. Elliott et al. “Characterisation of sleep in intensive care using 24-hour polysomnography: an observational study”. In: *Critical Care* 17.2 (Jan. 2013), R46. DOI: 10.1186/cc12565. URL: https://doi.org/10.1186/cc12565.

[51] J.F. Hipp and M. Siegel. “Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG”. In: *Frontiers in Human Neuroscience* 7 (Jan. 2013). DOI: 10.3389/fnhum.2013.00338. URL: https://doi.org/10.3389/fnhum.2013.00338.

[52] X. Zhao et al. “Sleep cycle in children with severe acute bronchopneumonia during mechanical ventilation at different depths of sedation”. In: *BMC Pediatrics* 22.1 (Oct. 2022). DOI: 10.1186/s12887-022-03658-8. URL: https://doi.org/10.1186/s12887-022-03658-8.

[53] R. H. Al-Samsam and P. M. Cullen. “Sleep and adverse environmental factors in sedated mechanically ventilated pediatric intensive care patients”. In: *Pediatric Critical Care Medicine* 6.5 (Sept. 2005), pp. 562–567. DOI: 10.1097/01.pcc.0000165561.40986.a6. URL: https://doi.org/10.1097/01.pcc.0000165561.40986.a6.

[54] L.A. Dervan, J.E. Wrede, and R.S. Watson. “Sleep architecture in mechanically ventilated pediatric ICU patients receiving Goal-Directed, dexmedetomidine- and opioid-based sedation”. In: *Journal of Pediatric Intensive Care* 11.01 (Nov. 2020), pp. 032–040. DOI: 10.1055/s-0040-1719170. URL: https://doi.org/10.1055/s-0040-1719170.

[55] F. Roche-Campo et al. “Comparison of sleep quality with mechanical versus spontaneous ventilation during weaning of critically ill tracheostomized patients*”. In: *Critical Care Medicine* 41.7 (July 2013), pp. 1637–1644. DOI: 10.1097/ccm.0b013e318287f569. URL: https://doi.org/10.1097/ccm.0b013e318287f569.

[56] S. Zhao et al. “Evaluation of a Single-Channel EEG Based Sleep Staging Algorithm”. In: *International Journal of Environmental Research and Public Health* 19.5 (Mar. 2022), p. 2845. DOI: 10.3390/ijerph19052845.

[57] M. Sharma, M. Darji J.and Thakrar, and U.R. Acharya. “Automated identification of sleep disorders using Wavelet-based features extracted from electrooculogram and electromyogram signals”. In: *Computers in Biology and Medicine* 143 (Apr. 2022), p. 105224. DOI: 10.1016/j.compbiomed.2022.105224. URL: https://doi.org/10.1016/j.compbiomed.2022.105224.

[58] C.R. Oerlemans. *Sleep patterns in critically ill children admitted to the pediatric intensive care unit: anomaly detection methods in hiqh-frequency data*. Unpublished. 2023.

[S]

A
Supplementary Materials

A.1. Electroencephalography characteristics of the different sleep stages

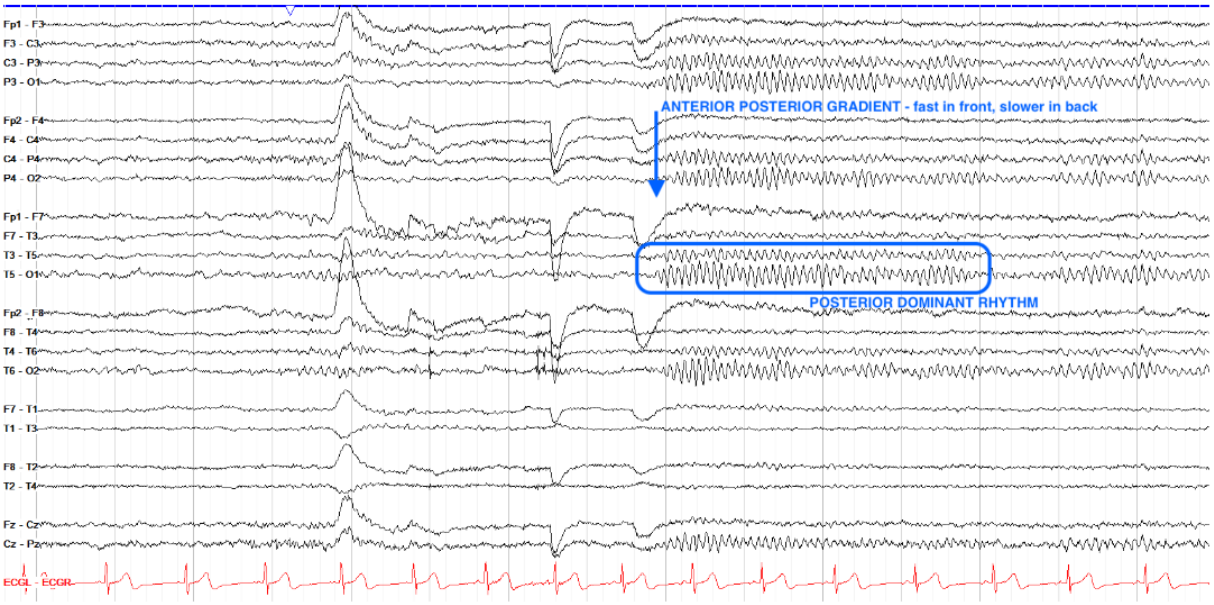

Figure A.1: Normal EEG during wakefulness [13]. *EEG = electroencephalography.*

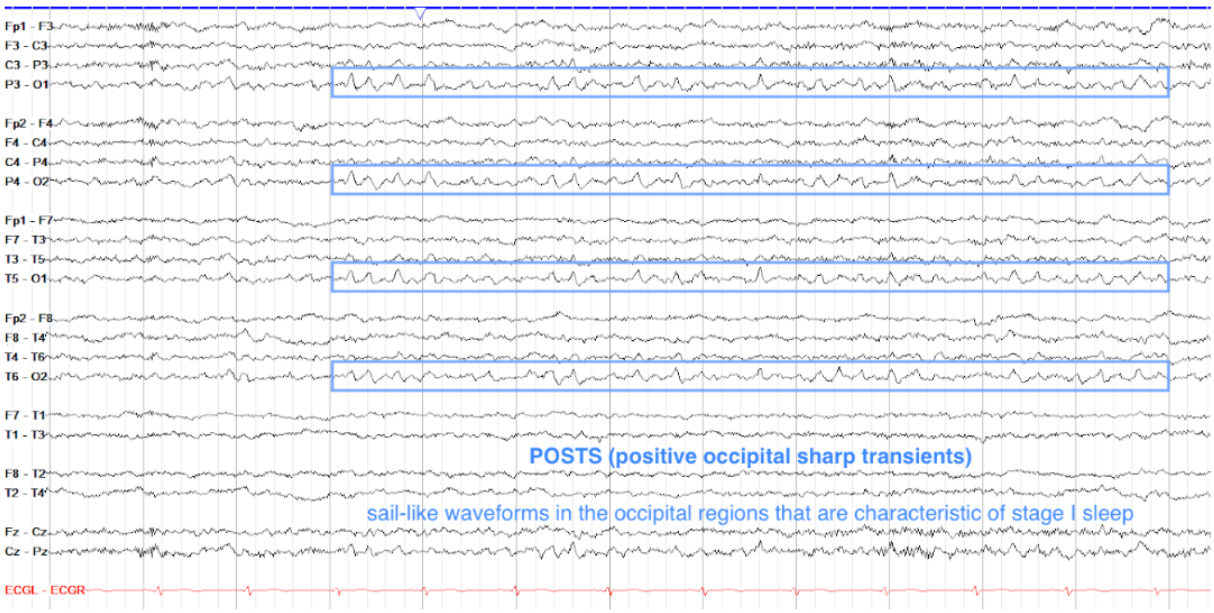

Figure A.2: Positive occipital sharp transients of sleep (POSTS) arise in N1 sleep and can persist in later sleep stages [13]. *N1 = NREM stage 1 = non-rapid eye movement stage 1.*

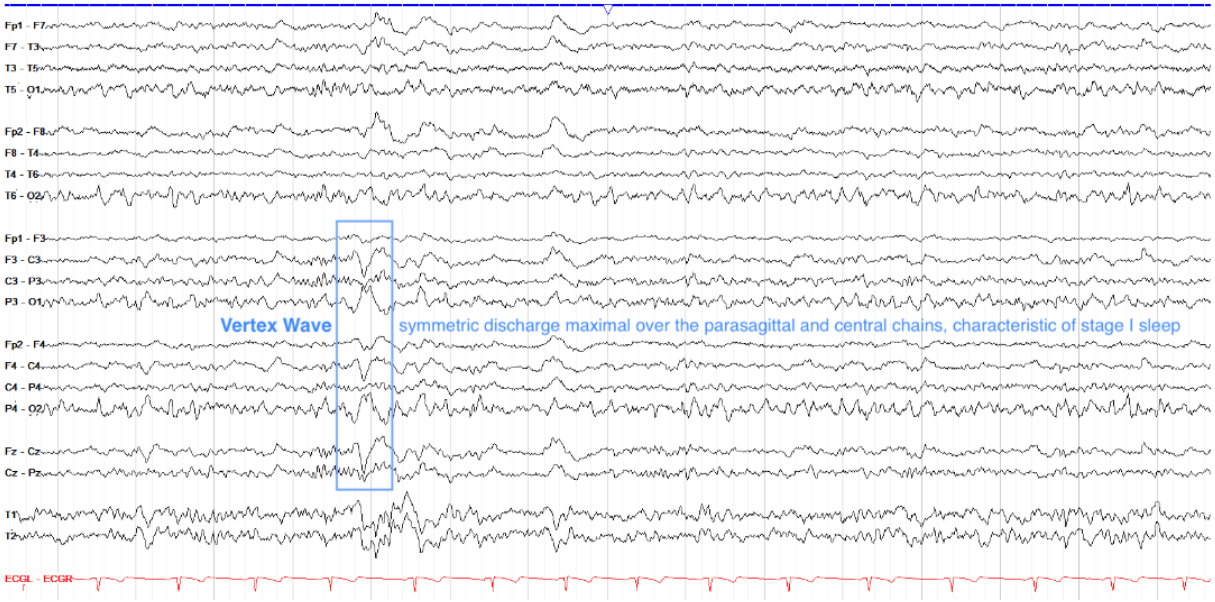

Figure A.3: Vertex waves seen in N1 sleep [13]. *N1 = NREM stage 1 = non-rapid eye movement stage 1.*

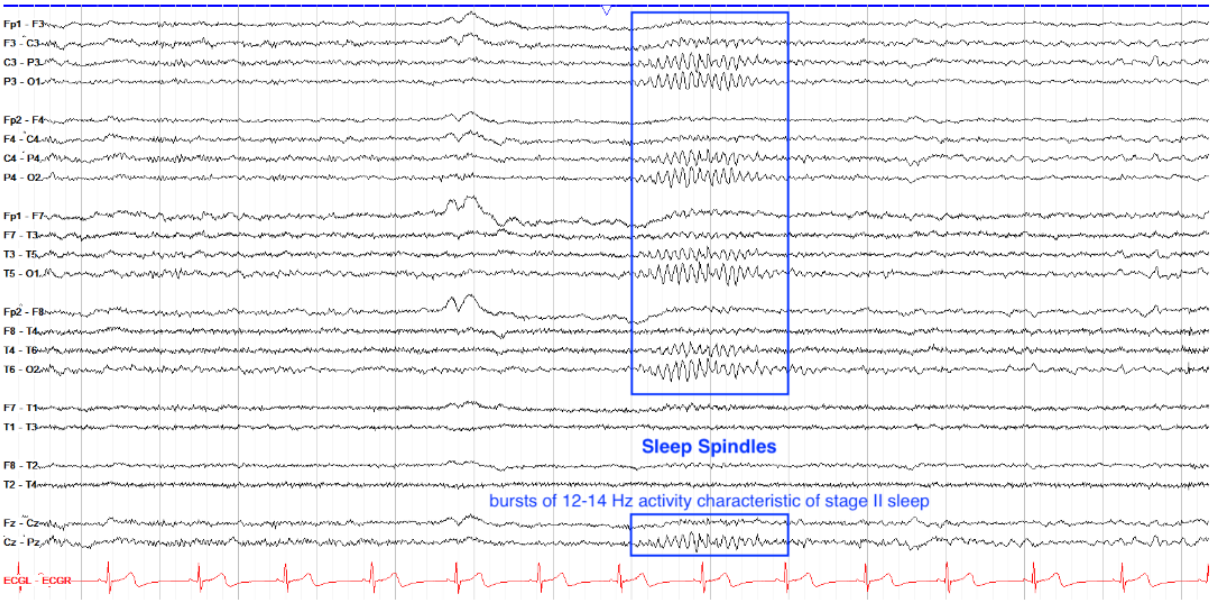

Figure A.4: Sleep spindles characterizing N2 sleep [13]. *N2 = NREM stage 2 = non-rapid eye movement stage 2.*

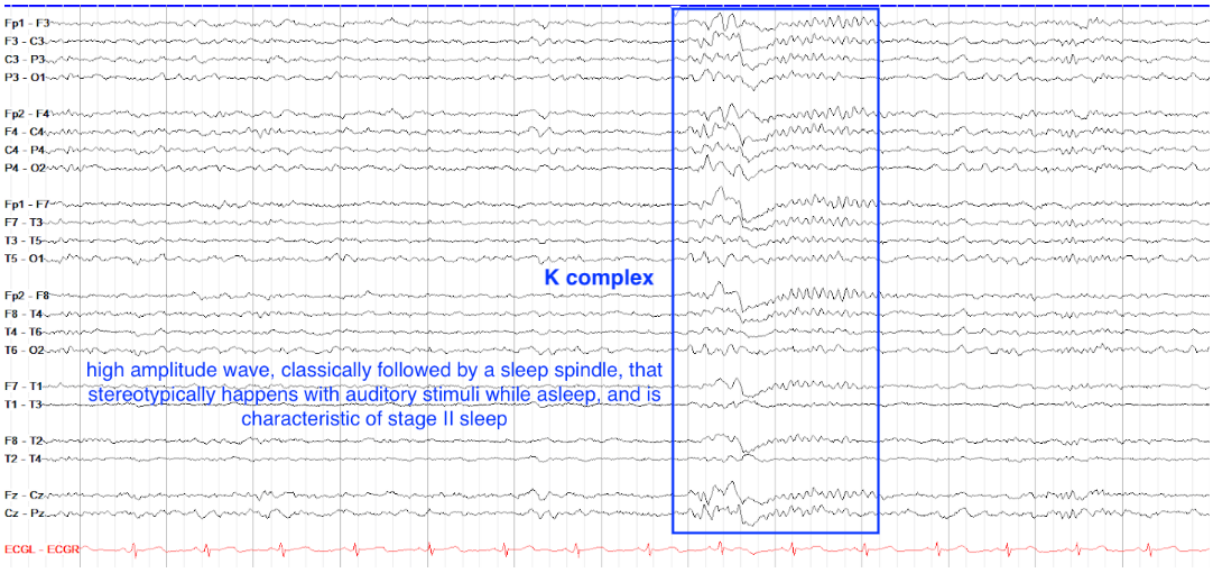

Figure A.5: K complexes characterizing N2 sleep [13]. *N2 = NREM stage 2 = non-rapid eye movement stage 2*.

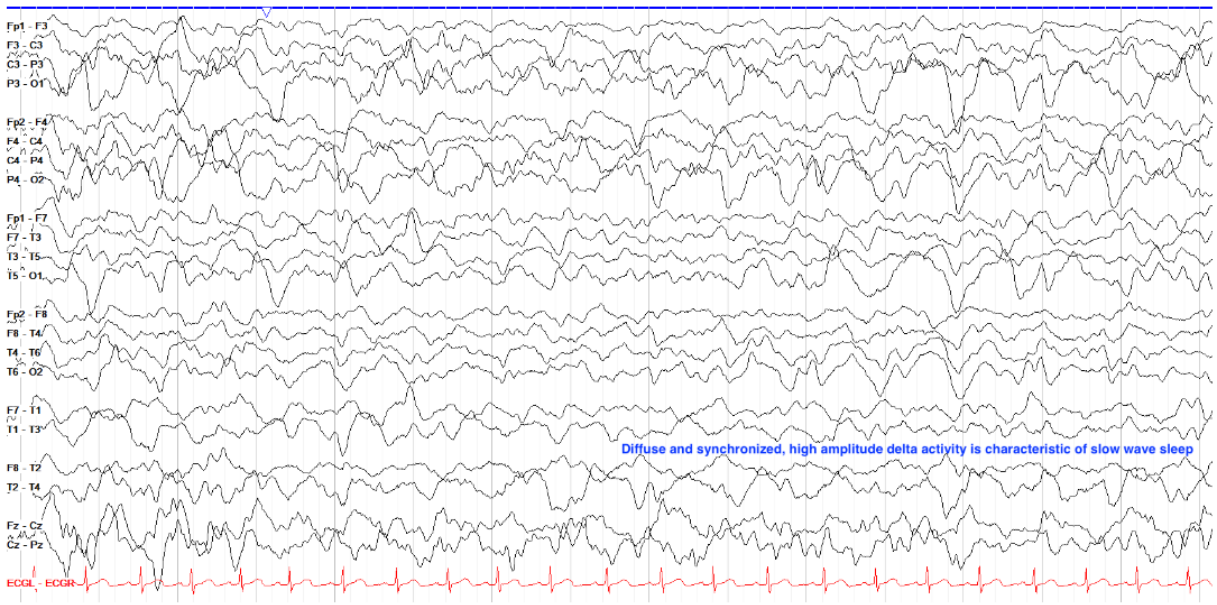

Figure A.6: Delta activity seen in N3 sleep [13]. *N3 = NREM stage 3 = non-rapid eye movement stage 3*.

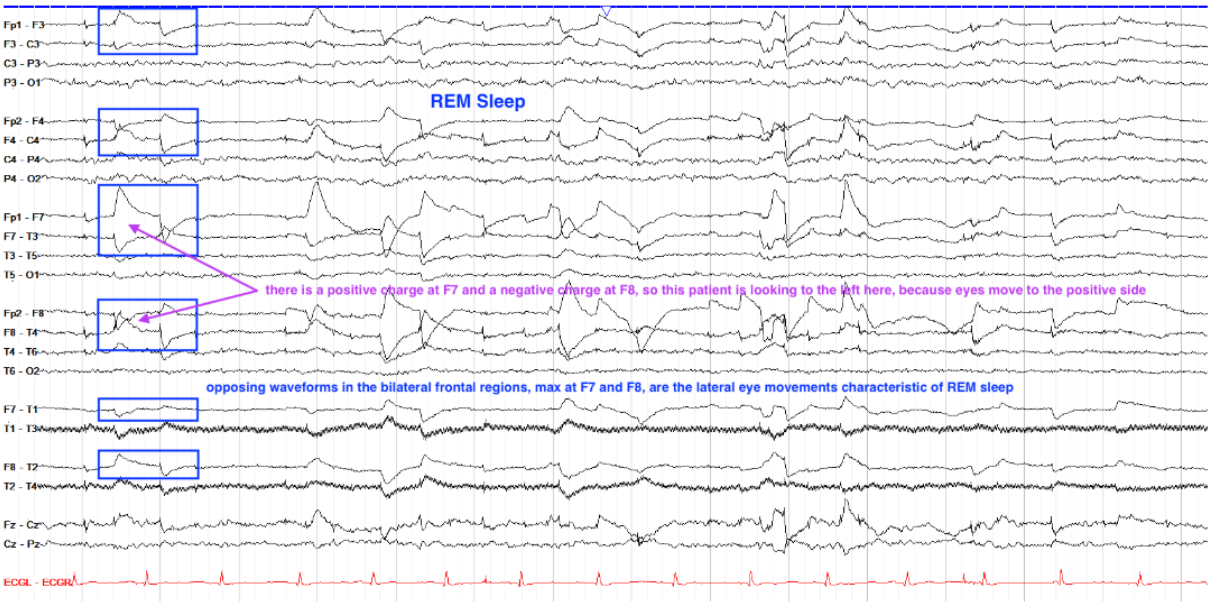

Figure A.7: Very sharply countered opposing waveforms in R sleep [13]. *R = REM = rapid eye movement.*

A.2. Feature selection

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)	

Figure A.8: Overview of all calculated features from each epoch of all PSG recordings. *PSG = polysomnography.*

Details and formulas of the feature calculation can be found in Hiemstra (2021) [4].

A.3. Distributions of selected features

Figure A.8, Figure A.9, Figure A.10, Figure A.11, Figure A.12 and Figure A.13 illustrate examples of the distributions of the selected features from EEG derivation C3-A2 for N2 sleep.

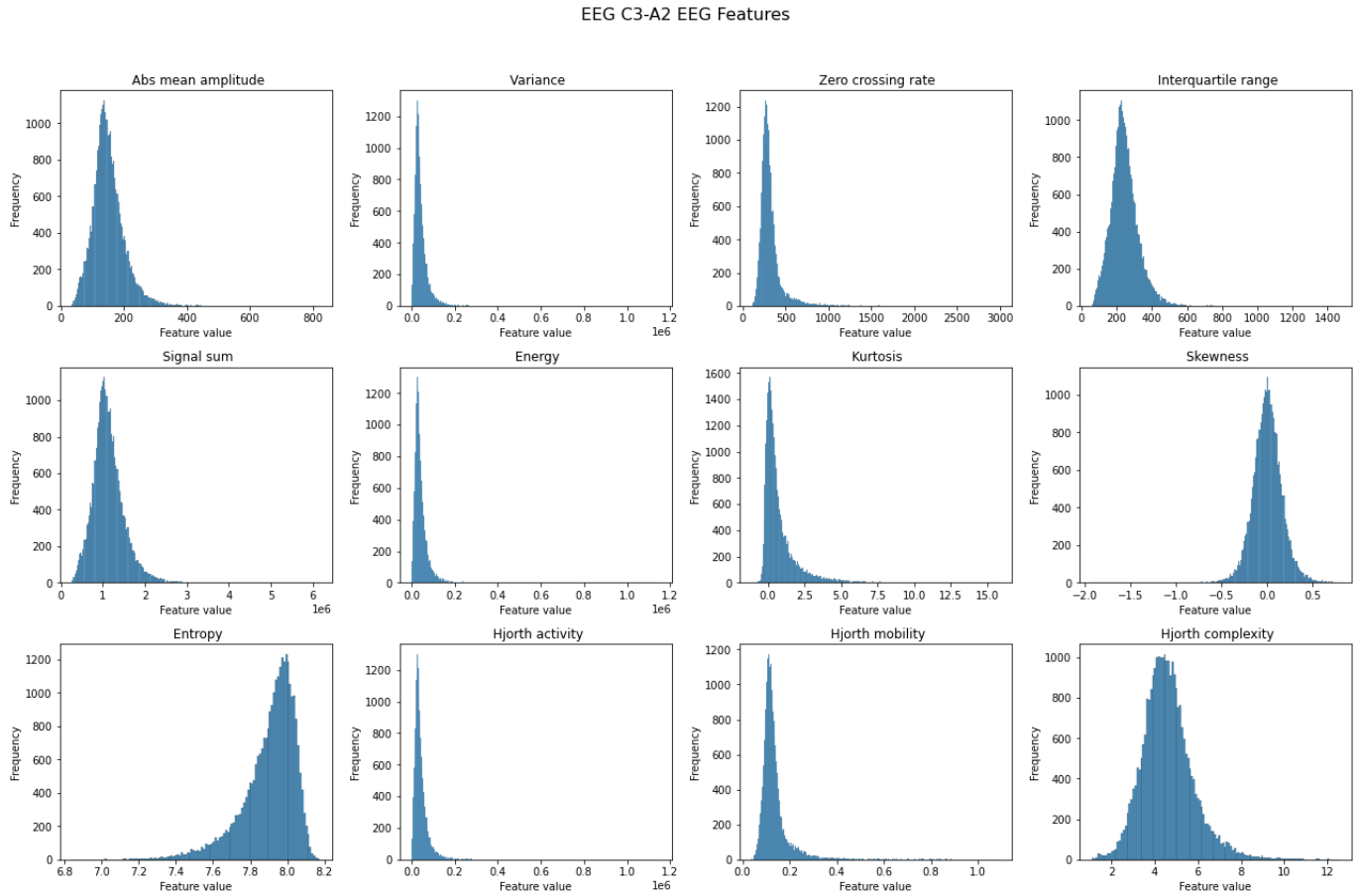

Figure A.9: Feature distributions

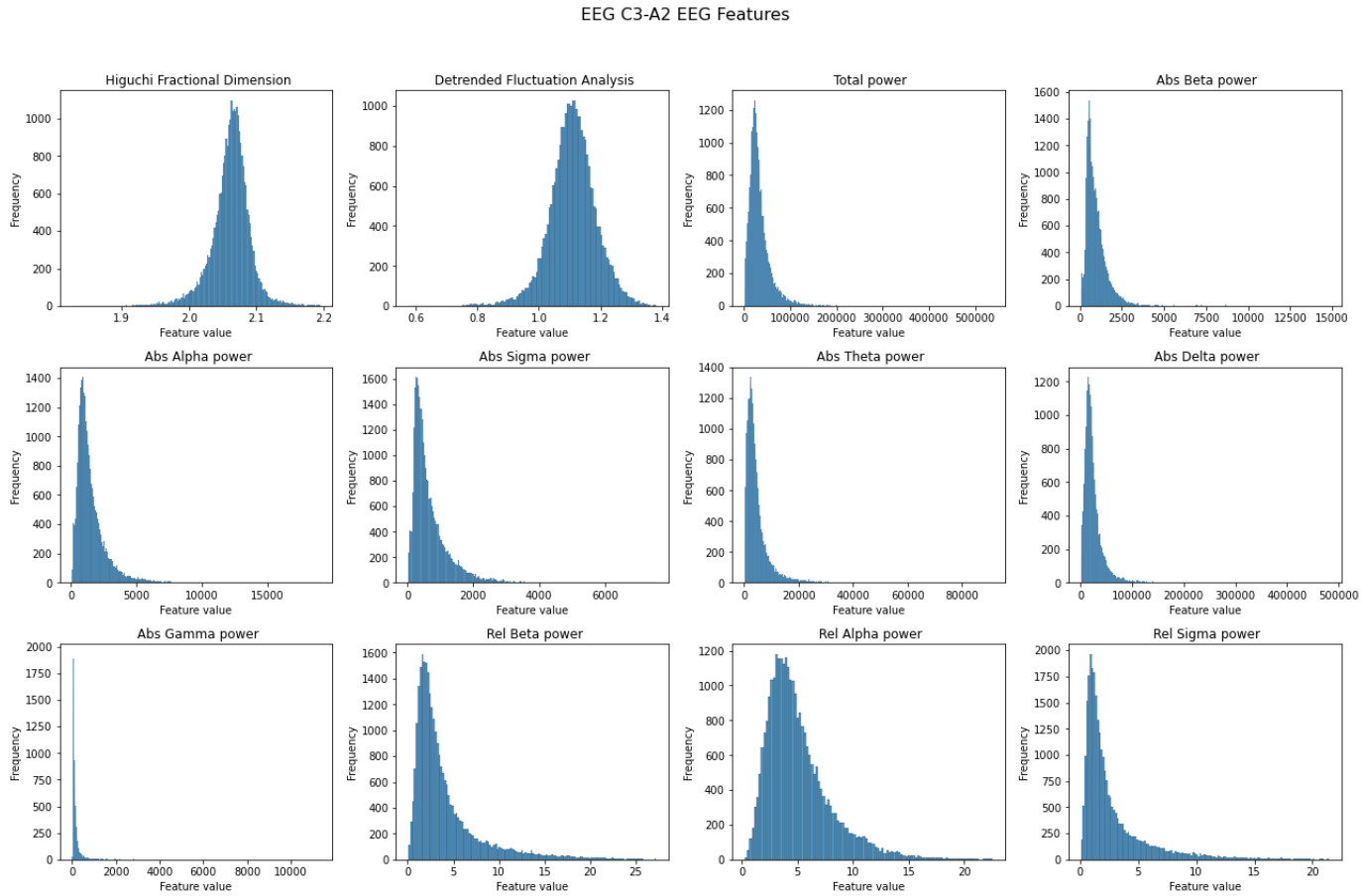

Figure A.10: Feature distributions

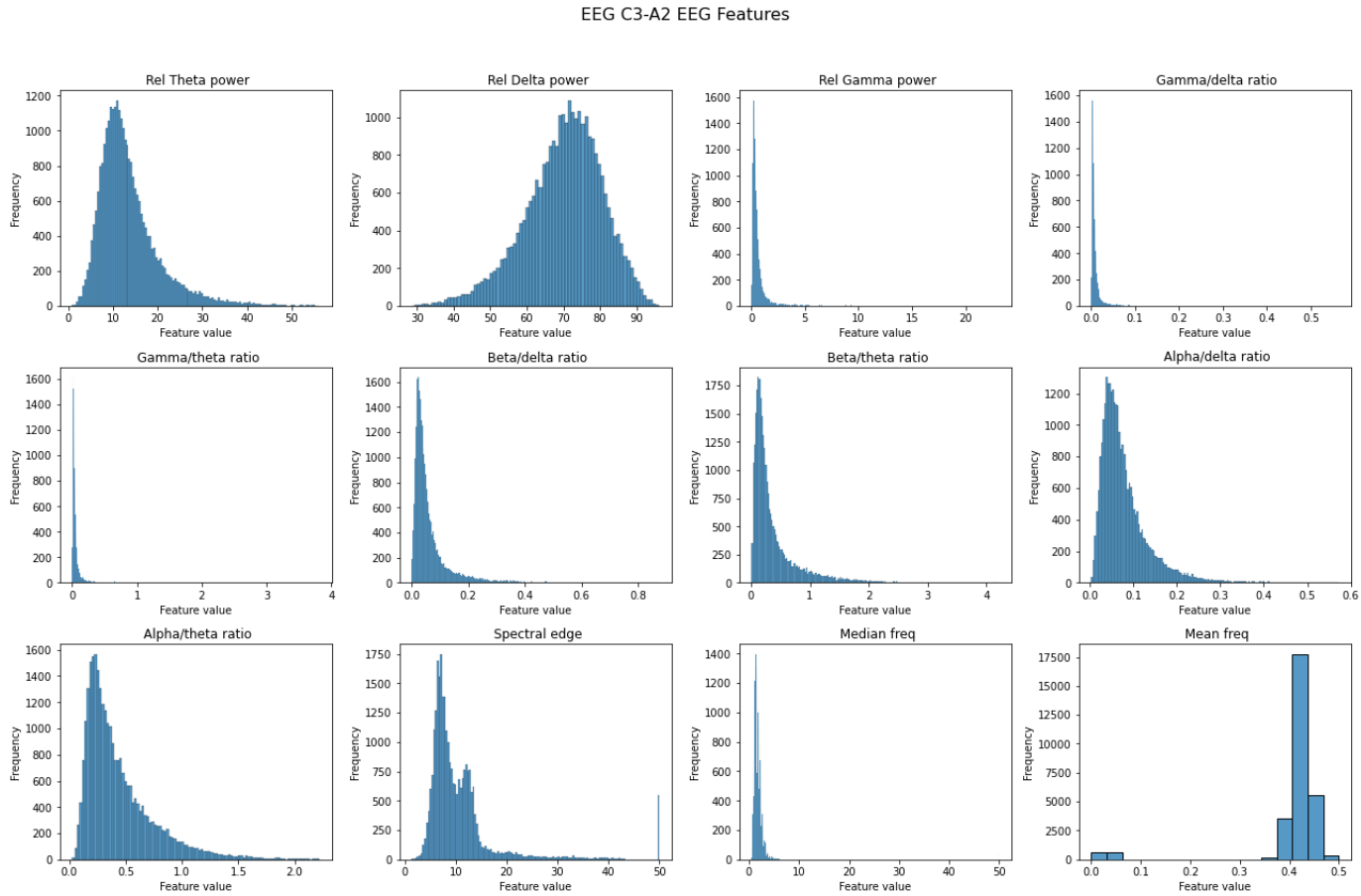

Figure A.11: Feature distributions

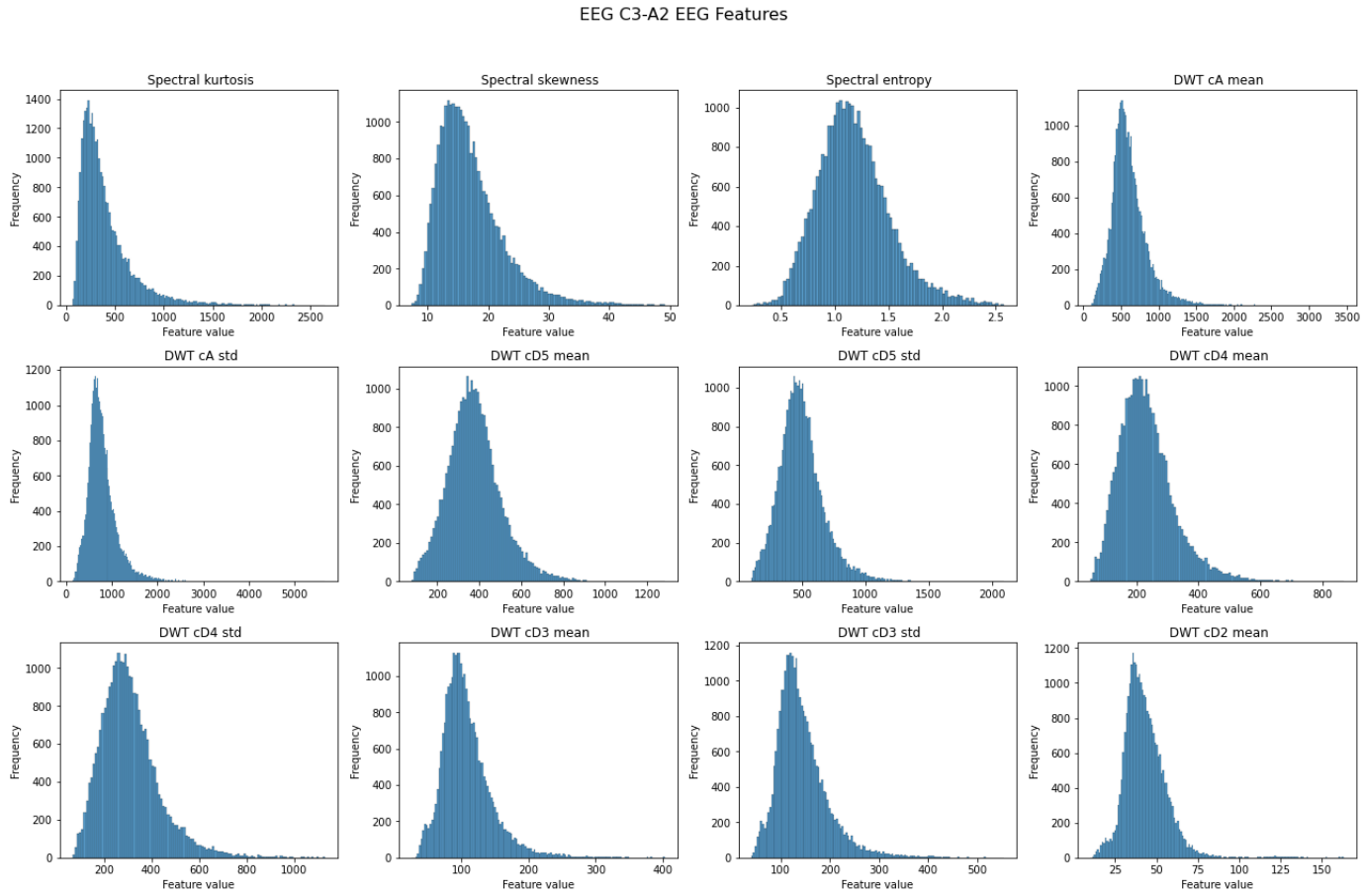

Figure A.12: Feature distributions

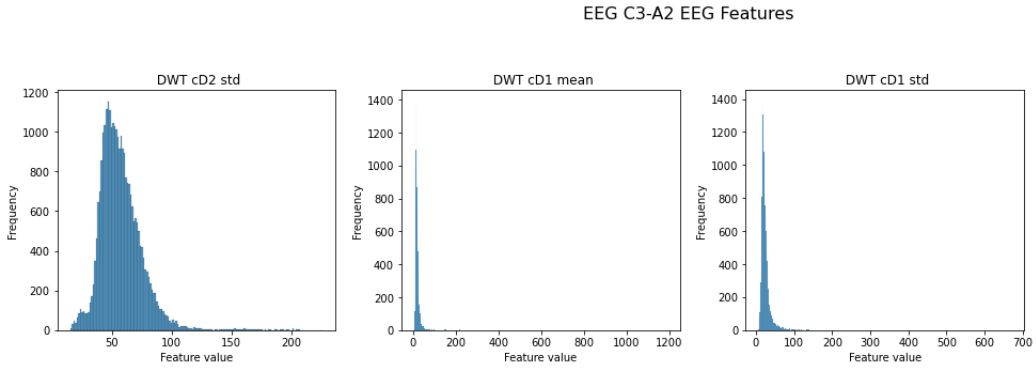

Figure A.13: Feature distributions

A.4. Results subanalyses

A.4.1. Performance evaluation using different electroencephalography derivations

The model has been trained using various EEG derivations along with their corresponding features to assess whether this approach enhances performance.

Table A.1: Errors of sleep stage predictions using a Gaussian fit trained specifically for a particular sleep stage - EEG derivation F3-A2. *EEG = electroencephalography; PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	W	N1	N2	N3	R	N
W	0.112	0.115	0.114	0.066	0.066	0.032
N1	0.885	0.202	0.781	0.820	0.457	0.157
N2	0.344	0.200	0.067	0.282	0.107	0.128
N3	0.801	0.502	0.300	0.099	0.213	0.108
R	0.703	0.396	0.449	0.634	0.116	0.143

Table A.2: Errors of sleep stage predictions using a Gaussian fit trained specifically for a particular sleep stage - EEG derivation O1-A2. *EEG = electroencephalography; PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	W	N1	N2	N3	R	N
W	0.124	0.071	0.054	0.032	0.039	0.012
N1	0.764	0.165	0.585	0.644	0.431	0.144
N2	0.373	0.194	0.067	0.226	0.114	0.153
N3	0.569	0.348	0.196	0.097	0.217	0.152
R	0.619	0.312	0.205	0.410	0.096	0.138

Table A.3: Errors of sleep stage predictions using a Gaussian fit trained specifically for a particular sleep stage - EEG derivation C3-A2, F3-A2, and O1-A2. *EEG = encephalography; PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	W	N1	N2	N3	R	N
W	0.114	0.091	0.084	0.036	0.052	0.025
N1	0.845	0.150	0.717	0.740	0.452	0.144
N2	0.431	0.225	0.073	0.299	0.106	0.152
N3	0.685	0.422	0.245	0.102	0.207	0.122
R	0.677	0.373	0.383	0.587	0.110	0.155

A.4.2. Performance evaluation after removing outliers in feature values

Deleting 0.2% of the outlying values from the selected features and subsequent deletion of epochs with missing data resulted in the errors reported in table Table A.4.

Table A.4: Number of epochs removed from the original dataset after deleting the 0.2% outlying values per selected feature. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	Original number of epochs	Number of epochs removed	Percentage removed epochs
W	37532	1365	3.64%
N1	11308	499	4.41%
N2	29439	817	2.78%
N3	35591	1042	2.93%
R	29478	811	2.75%
N	11851	499	4.21%

A.4.3. Performance evaluation using normalized feature values

The errors in sleep stage predictions, obtained from a Gaussian fit specifically trained for a particular sleep stage using normalized feature values from the EEG derivation C3-A2, are illustrated in table Table A.5.

Table A.5: Errors of sleep stage predictions using a Gaussian fit trained specifically for a particular sleep stage using normalized feature values. *EEG = electroencephalography; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	W	N1	N2	N3	R
W	0.111	0.062	0.029	0.040	0.021
N1	0.383	0.105	0.101	0.392	0.058
N2	0.632	0.260	0.108	0.469	0.139
N3	0.809	0.446	0.277	0.114	0.279
R	0.679	0.317	0.343	0.648	0.124

A.4.4. Evaluation of the cross-validation process within the model

The calculated errors of the model using cross-validation without defining groups can be read from table Table A.6. It can be seen that the errors of the validation set are approximately 0.1.

Table A.6: Cross-validation without defining groups. *W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	W	N1	N2	N3	R
W	0.100	0.082	0.078	0.045	0.033
N1	0.886	0.097	0.782	0.757	0.469
N2	0.584	0.285	0.100	0.400	0.099
N3	0.829	0.509	0.289	0.100	0.233
R	0.702	0.407	0.491	0.712	0.099

A.5. Results model application to data from critically ill children

Table A.7: Epoch distribution across sleep stages per PICU patient. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	W	N1	N2	N3	R	N	Total number of epochs
PICU001	3157	2246	1397	1894	719	385	3157
PICU002	1229	92	74	1194	73	170	1229
PICU003	2930	2341	1862	2543	1448	930	2930
PICU004	2895	2744	2534	2736	2236	1382	2896
PICU005	2909	1331	1003	2836	692	775	2910
PICU006	2972	2528	2058	2665	1729	721	2972
PICU007	2711	0	0	1402	0	515	2837
PICU008	2913	113	35	2571	0	707	2916
PICU009	12	0	0	0	0	758	2399
PICU010	2869	2129	1443	2339	546	592	2869
PICU012	2999	1341	107	2230	0	978	3005
PICU013	2931	2768	2380	2872	1951	295	2931
PICU014	3370	2359	2162	3311	965	874	3370
PICU015	2895	2744	2534	2736	2236	1382	2896
PICU016	2930	2341	1862	2543	1448	930	2930
PICU018	2129	156	0	1604	0	331	2130
PICU020	2711	0	0	1402	0	515	2837
PICU021	6108	2515	2211	6068	1496	1376	6108
PICU022	3479	1980	1351	3431	525	624	3480
PICU024	2423	4	7	1168	9	729	2423
PICU025	2741	2024	1894	2712	1305	837	2741
PICU026	2869	2129	1443	2339	546	593	2869
PICU027	3157	2246	1398	1894	719	385	3157
Total	65339 (96%)	36131 (53%)	27755 (41%)	54490 (80%)	18643 (27%)	16784 (25%)	67992

Table A.8: The total count of epochs per PICU patient classified into each sleep stage. *PICU = Pediatric Intensive Care Unit*

	0	1	2	3	4	5	6
PICU001	0	599	723	437	604	738	56
PICU002	0	27	955	162	25	51	9
PICU003	0	186	450	316	515	818	645
PICU004	1	91	84	148	265	1078	1229
PICU005	1	49	1276	474	348	441	321
PICU006	0	104	424	320	370	1243	511
PICU007	84	1097	1437	219	0	0	0
PICU008	1	242	1982	637	48	6	0
PICU009	1636	756	7	0	0	0	0
PICU010	0	352	433	517	892	469	206
PICU012	5	195	1483	822	472	28	0
PICU013	0	24	157	329	442	1770	209
PICU014	0	48	734	411	949	872	356
PICU015	1	91	84	148	265	1078	1229
PICU016	0	186	450	316	515	818	645
PICU018	1	340	1508	260	21	0	0
PICU020	84	1097	1437	219	0	0	0
PICU021	0	33	3067	694	688	983	643
PICU022	0	44	1194	797	796	511	138
PICU024	0	847	1250	322	2	2	0
PICU025	0	22	557	275	556	658	673
PICU026	0	351	433	519	890	471	205
PICU027	0	599	723	436	605	738	56

Table A.9: Number of epochs per PICU patient assigned to each sleep stage by the neurophysiologists. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

	W	N1	N2	N3	R	N	Total number of epochs
PICU001	1698	603	658	166	32	0	3157
PICU002	484	0	0	355	82	308	1229
PICU003	1625	63	147	925	138	32	2930
PICU004	1030	0	0	0	87	1779	2896
PICU005	675	0	224	980	303	728	2910
PICU006	1055	0	0	0	758	1159	2972
PICU007	1303	516	704	230	63	21	2837
PICU008	237	0	0	0	0	2679	2916
PICU009	830	338	958	128	145	0	2399
PICU010	1441	145	676	541	66	0	2869
PICU012	1780	0	0	0	117	1108	3005
PICU013	326	0	0	0	318	2287	2931
PICU014	1282	0	0	0	350	1738	3370
PICU015	1030	0	0	0	87	1779	2896
PICU016	1667	169	147	868	79	0	2930
PICU018	1652	0	0	0	93	385	2130
PICU020	1303	516	704	230	63	21	2837
PICU021	189	3660	1486	613	160	0	6108
PICU022	1220	476	1030	562	192	0	3480
PICU024	2423	0	0	0	0	0	2423
PICU025	368	0	329	52	0	1992	2741
PICU026	1441	145	676	541	66	0	2869
PICU027	1698	603	658	166	32	0	3157

A.6. Mahalanobis Distances over time

In Figure A.14, the MDs of PICU010 for stages N1 and N3, corresponding to moderate errors, also fluctuate in clusters. For comparison, Figure A.15 presents another patient file with instances having both low and high errors. In this figure, the clusters are less prominent. Patient PICU009 has remarkably high errors compared to the other PICU patients. The computed MDs are illustrated in Figure A.16. While a small portion of the epochs is assigned to sleep stage N, the majority of epochs consistently were categorized as anomalous by all other Gaussian Models, with MDs surpassing the thresholds.

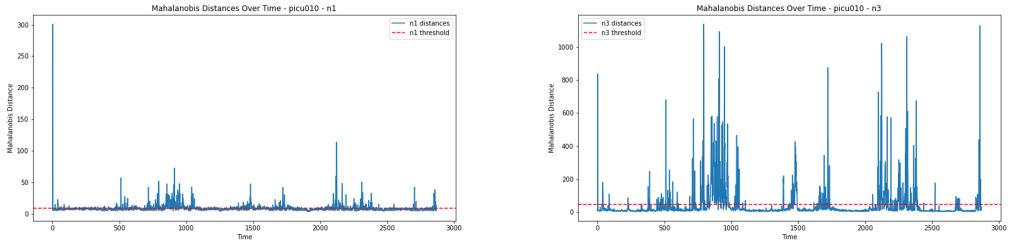

Figure A.14: Mahalanobis Distances over time for PICU010. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

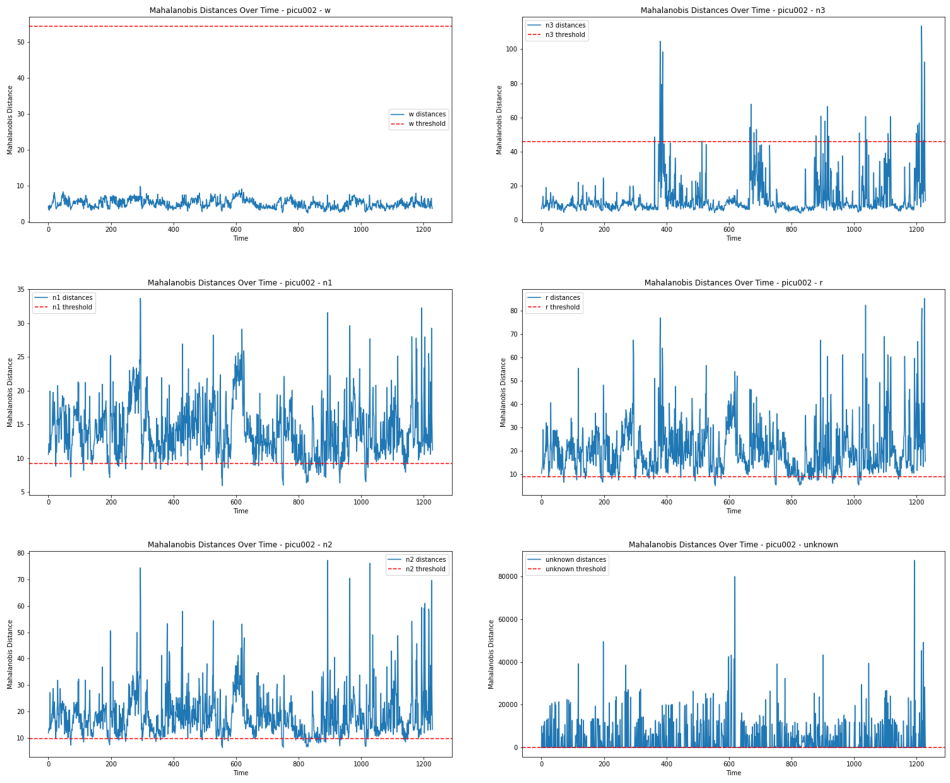

Figure A.15: Mahalanobis Distances over time for PICU002. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*

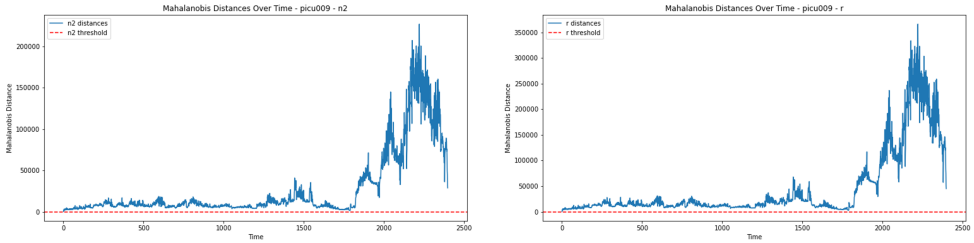

Figure A.16: Mahalanobis Distances over time for PICU009. *PICU = Pediatric Intensive Care Unit; W = wake; N1 = NREM stage 1 = non-rapid eye movement stage 1; N2 = NREM stage 2; N3 = NREM stage 3; R = REM = rapid eye movement; N = N-stage which has characteristics of NREM but could not be classified as N1, N2, or N3.*