

# Feasibility of wearable devices and machine learning for sleep classification in children with Rett syndrome: A pilot study

DIGITAL HEALTH  
Volume 9: 1–13  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076231191622  
journals.sagepub.com/home/dhj



Miroslava Migovich<sup>1</sup> , Akshith Ullal<sup>2</sup>, Cary Fu<sup>3</sup>, Sarika U Peters<sup>3,4</sup>  
and Nilanjan Sarkar<sup>1,2</sup>

## Abstract

Sleep is vital to many processes involved in the well-being and health of children; however, it is estimated that 80% of children with Rett syndrome suffer from sleep disorders. Caregiver reports and questionnaires, which are the current method of studying sleep, are prone to observer bias and missed information. Polysomnography is considered the gold standard for sleep analysis but is labor and cost-intensive and limits the frequency of data collection for sleep disorder studies. Wearable digital health technologies, such as actigraphy devices, have shown potential and feasibility as a method for sleep analysis in Rett syndrome, but have not been validated against polysomnography. Furthermore, the collected accelerometer data has limitations due to the rigidity, periodic limb movement, and involuntary muscle contractions prevalent in Rett syndrome. Heart rate and electrodermal activity, along with other physiological signals, have been linked to sleep stages and can be utilized with machine learning to provide better resistance to noise and false positives than actigraphy. This research aims to address the gap in Rett syndrome sleep analysis by comparing the performance of a machine learning model utilizing both accelerometer data and physiological data features to the gold-standard polysomnography for sleep analysis in Rett syndrome. Our analytical validation pilot study ( $n = 7$ ) found that using physiological and accelerometer features, our machine learning models can differentiate between awake, non-rapid eye movement sleep, and rapid eye movement sleep in Rett syndrome children with an accuracy of 85.1% when using an individual model. Additionally, this work demonstrates that it is feasible to use digital health technologies in Rett syndrome, even at a young age, without data loss or interference from repetitive movements that are characteristic of Rett syndrome.

## Keywords

Rett syndrome, sleep analysis, wearable physiological sensors, machine learning

Submission date: 15 April 2023; Acceptance date: 13 July 2023

## Introduction

At an estimated prevalence of 1 in 10,000 females, Rett syndrome (RTT) accounts for up to 10% of genetically linked severe intellectual disabilities in females.<sup>1–3</sup> RTT is associated with a spontaneous mutation in the methyl CPG binding protein 2 (MeCP2) gene,<sup>3</sup> located on the X-chromosome at Xq28. RTT is characterized by regression with loss of acquired spoken language and volitional hand use, disrupted or absent ambulation, and repetitive hand movements.<sup>4</sup> Associated clinical features include

<sup>1</sup>Department of Mechanical Engineering, Vanderbilt University, Nashville, TN, USA

<sup>2</sup>Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA

<sup>3</sup>Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>4</sup>Vanderbilt Kennedy Center, Nashville, TN, USA

### Corresponding author:

Miroslava Migovich, Vanderbilt University, Department of Mechanical Engineering, 2400 Highland Ave, Room 518, Nashville, TN 37240, USA.  
Email: miroslava.migovich@vanderbilt.edu



seizures, autonomic and breathing abnormalities, growth failure, scoliosis, gastrointestinal and nutritional symptoms, and impaired sleep.<sup>5,6</sup>

Sleep is regarded as essential for the well-being and health of children and is critical to many somatic, psychological, and cognitive processes.<sup>7,8</sup> However, sleep problems are highly prevalent in RTT, with studies showing that around 80% of children with RTT suffer from sleep disorders.<sup>9,10</sup> Disturbances, such as night-time laughing and screaming, sleep walking, and night terrors<sup>11</sup> adversely impact the quality of life for both the child with RTT and their families.<sup>12,13</sup> Current methods of sleep evaluation in the home are limited to sleep diaries and caregiver-completed questionnaires,<sup>14,10</sup> however, these methods are subjected to a number of biases, including recall and observer bias. Caregivers could miss night wakings<sup>15</sup> or misinterpret a child laying with their eyes closed as the child being asleep. Due to the severe communication impairments that accompany RTT, self-reports of sleep disturbances are not possible. Caregiver-independent, objective methods of sleep assessment are necessary to further explore sleep trends and the progression of sleep disturbances in this population. The gold standard for obtaining objective sleep measures is polysomnography (PSG), a procedure that uses a number of sensors to measure and record brain waves, oxygen levels, respiration rate, and heart rate, along with leg and eye movement to determine and classify sleep stages.<sup>16</sup> While highly effective in measuring sleep quality, PSG is resource intensive and impractical for longitudinal assessments. The unreliability of caregiver reports and the resource requirements of PSG highlights the need for new methods of effective sleep assessment that are suitable for longitudinal use in RTT research.

Recent research has explored the use of digital health technologies (DHTs), such as wearable accelerometers to measure body movements from the wrist (actigraphy), as a method of sleep analysis.<sup>17</sup> Actigraphy devices have shown promise when used with children with neurodevelopmental disorders (NDDs), such as Down syndrome<sup>18</sup> and autism spectrum disorders,<sup>19</sup> due to their ease of use and ability to capture data within the home environment for a prolonged period of time. Researchers have determined the feasibility of actigraphy for sleep analysis in RTT,<sup>20</sup> however, analytical validation of actigraphy against gold standard PSG has yet to be examined in RTT literature. In addition, individuals with RTT have a higher prevalence of involuntary muscle contractions, rigidity,<sup>21</sup> and increased prevalence of periodic limb movements during sleep,<sup>22</sup> which could potentially limit the interpretability of actigraphy alone. Currently, available models for sleep analysis and sleep scoring with wearable devices are not applicable for children with RTT due to the fact that the models are designed with data from typically developing adults<sup>23</sup> and greater sensitivity and specificity are required for this population. *These limitations can be*

*addressed with the addition of physiological data collection and a model specifically trained on data from children with RTT.*

Distinct hormonal patterns and the underlying sub-cortical network of brain structures that govern sleep significantly influence physiology,<sup>24,25</sup> meaning that changes in physiological signals can be correlated to sleep stages. As early as 1968, fluctuations in electrodermal activity (EDA) were found to increase during the late stages of non-rapid eye movement (NREM) sleep and decrease during rapid eye movement (REM) sleep.<sup>26</sup> In 1973 Aldredge et al. determined that heart rate averages trended higher in REM sleep and lower in NREM sleep with the variance in heart rate decreasing with the depth of sleep.<sup>27</sup> Other researchers found that heart rate variability (HRV) changes during sleep are highly individualized and vary based on the basal autonomic activity of each individual.<sup>28</sup> Sleep quality has been associated with both HR and HRV.<sup>29</sup> These works show that meaningful and distinct sleep characteristics can be determined by collecting both heart rate and inter-beat interval data. More recent studies have found that during NREM, many of the measurable physiological processes decrease when compared to being awake. These processes include brain activity, respiration, body temperature, and blood pressure. Alternatively, these signals show an uptick in measured values during REM sleep.<sup>25</sup> Wearable devices, such as the Empatica E4, make it possible to collect many of these physiological measures in a non-invasive way. As discussed previously, wearable devices that measure accelerometer data have been able to distinguish sleep vs wake periods,<sup>30</sup> however, these sensors struggle to differentiate between NREM and REM sleep. By combining accelerometer data with physiological data, it is possible to train a machine learning algorithm to predict sleep stages in children with RTT.

Work in the field of machine learning and sleep state analysis has focused on automating the labeling process of PSG<sup>31</sup> or reducing the need for a full PSG sleep study by combining in-home monitoring with machine learning.<sup>32,33</sup> In their work, Mikkelsen et al.<sup>34</sup> evaluated the performance of a machine-learning algorithm based on input from a mobile around the ear electroencephalography (EEG) when compared to an actigraphy, using PSG as the ground truth. It was found that the EEG alone outperformed the actigraphy and was acceptable compared to the PSG, however, 85% of the participants reported that the around the ear EEG negatively influenced their sleep, to some degree. More often, sleep quality is considered to be the target measure. Studies have explored machine learning for sleep quality both with commercial smartwatches and clinical actigraphy with positive results.<sup>35,36</sup> Commercial wearable devices, such as FitBit, claim to track sleep using their integrated sensors. However, studies have shown that the FitBit algorithms tend to overestimate total

sleep time and struggle to correctly estimate light and deep sleep.<sup>37,38</sup> It has been shown that adding HRV data and additional body movement measures did increase the accuracy of FitBit's algorithm.<sup>38</sup> While these studies show the promise of combining machine learning and physiological data for sleep analysis, current work is based on typically developed adults<sup>38,35</sup> and does not translate to the sleep patterns of children. The existing techniques and data for automated sleep analysis are even less applicable to children with special needs, such as RTT. While other works have shown promise using automated algorithms with physiology and accelerometer data to differentiate between autism<sup>39</sup> and to classify high severity and low severity RTT,<sup>40</sup> to our knowledge, there exist no studies that utilize automated algorithms with physiological and accelerometer data for sleep analysis of children with RTT.

This paper fills a gap in the research literature by examining the analytical validation of a wearable sensor-based sleep analysis against gold-standard PSG in RTT. We validate the combination of physiological and accelerometer data by training a machine learning algorithm on extracted features in order to output sleep metrics. This process follows the best practices for analytical validation as presented by Goldsack et al.<sup>41</sup> Our work also considers the impact of feature selection and parameters on the accuracy of machine learning algorithms for sleep analysis.

## Methods

### Participants

Seven participants (age range 4–16 years, mean age 7.22 years, standard deviation  $\pm 3.66$  years) with a diagnosis of Classic RTT (genetically confirmed with MECP2 mutations in all cases) were recruited, with parental consent given, according to the approval of the Institutional Review Board (IRB) of Vanderbilt University Medical Center. All participants received a structured evaluation conducted by a licensed child neurologist that assessed their clinical state using the Revised Motor-Behavioral Inventory (R-MBA).<sup>42</sup> The R-MBA is a clinician-reported outcome measure that has been used to assess children, adolescents, and adults with RTT. The revised version of this scale consists of 24 items, of which 21 items load onto five different factors (Motor Dysfunction, Functional Skills, Social Skills, Aberrant Behavior, and Respiratory Behaviors). The remaining three items are retained from the original MBA due to clinical relevance. Items are captured on a 5-point Likert scale. Higher total scores indicate increased disease severity. The R-MBA is psychometrically sound and shows a positive relationship with parent-reported items, age, and mutation subtype. Data were collected following the procedure detailed in the following subsection.

### Data collection

Physiological data were collected using the Empatica E4 device.<sup>43,44</sup> The device was shipped to participants, and they wore the device on their wrist continuously for two days. On the third night, overnight PSG was performed through the Vanderbilt Sleep Core while the participant concurrently wore the E4 device. A standard PSG protocol with monitoring of respiratory effort, blood oxygen saturation, nasal airflow, heart rate, electromyography, EEG, and electrooculography was conducted using Nihon Kohden Polysmith Sleep Systems.<sup>45</sup> The PSG studies were scored visually in 30-second epochs with analysis and interpretation performed by a board-certified sleep medicine neurologist with expertise in sleep measures for NDDs at the Vanderbilt Sleep Research Core.

The E4 collects data from four main sensors: a photoplethysmography (PPG) sensor, an EDA sensor, a 3-axis accelerometer, and an infrared thermopile. The PPG measures volumetric variations of blood circulation using red and green light.<sup>46,47</sup> The lights are oriented towards the wrist skin, which allows the light to be absorbed and reflected. A photodetector then measures the reflected light. The reflection measurements during green light exposure are generally a sequence of valleys caused by high light absorption during a heartbeat. The measured valleys are correlated to heartbeats and are used to estimate heart rate. The red light provides a reference light level for canceling out motion artifacts and allowing for maximization of pulse wave detection.<sup>47</sup> Empatica uses a proprietary algorithm in order to extract the blood volume pulse (BVP) from the PPG signal. The resultant BVP output is stored in a CSV file with a sampling rate of 64 Hz. Interbeat interval (IBI) and heart rate are computed from the BVP and output to CSV files. The IBI data are output intermittently with 1/64 second resolution while the heart rate file contains the average heart rate values over the span of 10 seconds, sampled at 1 Hz.

Innervating signals from the brain cause changes in the permeability of sweat glands on the skin, which can be measured as changes in electrical conductance on the skin surface.<sup>48</sup> The E4 uses a minuscule amount of current between two electrodes to measure these changes as the pores on the wrist fill with sweat.<sup>49</sup> The EDA data, measured in the conductance unit of microSiemens ( $\mu S$ ), are sampled at 4 Hz and stored in a CSV file. The EDA complex is composed of the baseline tonic skin conductance level (SCL) and the phasic skin conductance responses (SCR) that result from neuronal activity from the sympathetic nervous system.<sup>50</sup> SCL is extracted as the raw level of conductance of the resultant EDA CSV file given by the E4. Phasic changes are caused by associated stimuli and are represented by peaks in the measured data as there is an abrupt increase in the skin conductance.<sup>49</sup> During the preprocessing of the data, the SCL and SCR

data are separated and stored in independent data frames. The process of extracting these from the raw EDA data will be discussed in the subsequent section.

An onboard 3-axis micromachined microelectromechanical system accelerometer is used to measure linear motion without a fixed reference.<sup>51</sup> The E4 provides a measurement of acceleration in the unit of 1/64 g at 32 Hz by measuring the continuous gravitational force (g) that is applied in each of the three spatial dimensions ( $x$ ,  $y$ , and  $z$ ).

Temperature data are sampled at 4 Hz using an infrared thermopile on a scale of  $-40^{\circ}$  to  $115^{\circ}$  C.

All of the generated CSV files from each sensor are zipped and downloaded from the Empatica Data Manager before preprocessing.

### Data processing

The sleep data generated from the PSG are stored in a CSV file in 30 s epochs with six possible labels. The stages of sleep include lights on awake (L), lights off awake (W), sleep stage N1 (N1), sleep stage N2 (N2), sleep stage N3 (N3), and REM (R). The timestamps along with the labels are imported into a Jupyter Notebook and the labels are converted into numerical labels, starting with zero and ending with five, so that they are compatible with the scikit-learn.<sup>52</sup> After initial testing, it was determined that without eye movement data, differentiation between all six labels was beyond the capabilities of the current work. Therefore, the six classes are consolidated into three broader classes. Lights on awake and lights off awake are combined into an awake category, designated with the label 100. N1-N3 are combined into a non-REM sleep category, labeled 010. REM sleep remains as its own category, labeled with 001 following the rules of one-hot-encoding, used for categorical data.<sup>53</sup> The consolidation of sleep stages into 3 classes follows the procedure set in previous works by Korkalainen et al.<sup>54</sup> and Sun et al.<sup>55</sup> The distribution of labels can be seen in Figure 2.

Each 30 s epoch is resorted into the 3 resultant labels and stored in a data frame with the timestamp.

After unzipping the physiological data, each CSV file is loaded into the Jupyter Notebook.<sup>56,57</sup> The CSV files given begin with a Unix timestamp that is converted into Universal Time Coordinated time and the initial timestamp along with the sampling rates are used to generate timestamps for the length of the collected data. Using the generated timestamps, the labels and physiological data are concatenated into a data frame. Features based on prior work in physiological data research are then extracted from the physiological data.<sup>58,59</sup> Using filters, the SCL and SCR are extracted from the EDA data. Following the method presented in Bian et al.,<sup>58</sup> a low-pass filter with a 0.5 Hz cutoff frequency removes noisy data. A high pass filter with a 0.05 Hz cutoff frequency is then used to isolate the SCL baseline which is stored in a data frame.

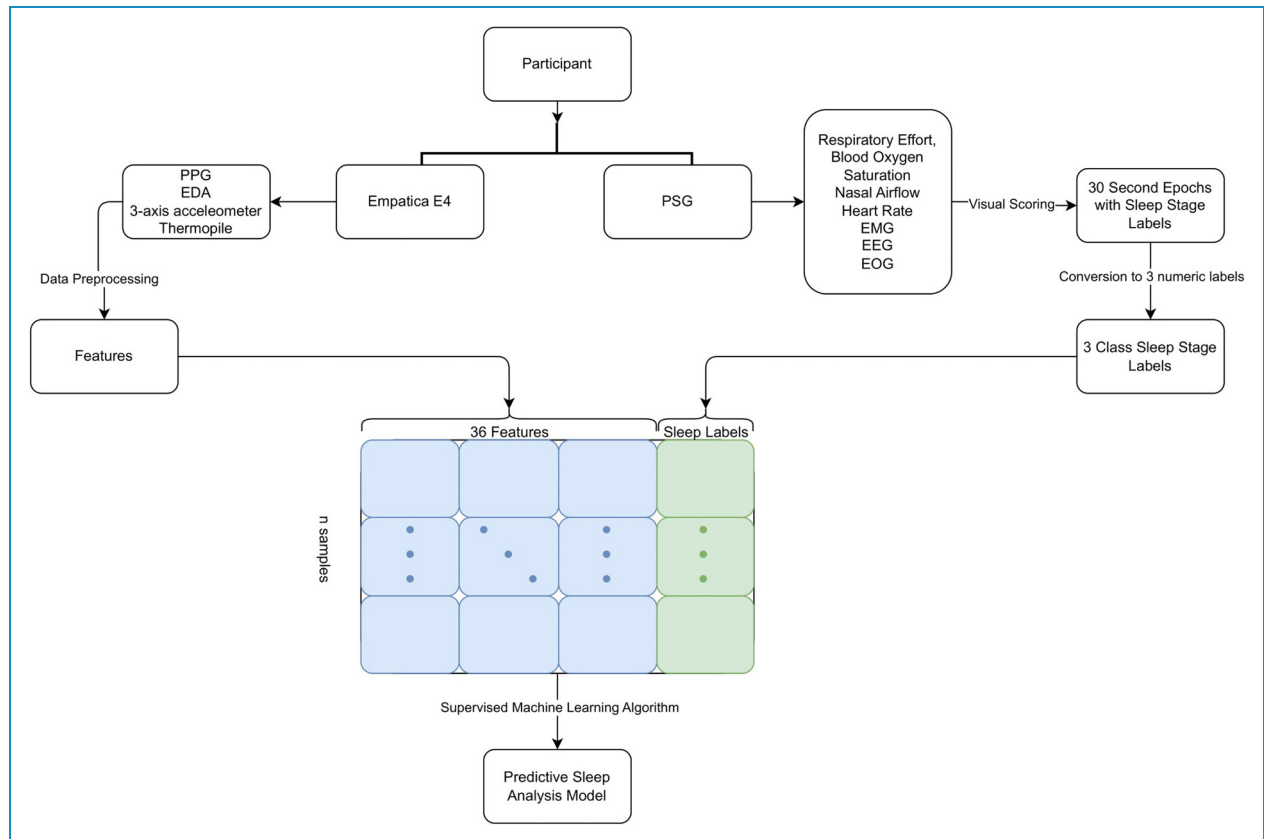
The isolated SCL level is subtracted from the filtered data to find the SCRs, which are also stored in a data frame. To match the standard PSG, 30-second epochs without overlapping windows are applied to the physiological data. Interpolation is used to account for any missing data caused by the different sampling rates of the sensors on the E4, as detailed in the “Data collection section. Once all the data are synced and interpolated, the standard deviation and mean of each window is calculated. Multiple features were derived from each sensor in order to explore the full extent of changes to physiology during sleep, which sets the stage for future work that evaluates sleep quality along with sleep stages in children with RTT. The initial features extracted are presented in Table 2 and a graphical overview of the data collection and data processing procedure is depicted in Figure 1.

### Feature selection

Feature selection in machine learning is used to eliminate redundant features or features that may be unnecessary. By reducing the number of features, the resultant models are less likely to over-fit and the training time is optimized.<sup>60</sup> Model accuracies are also increased as the model no longer has to parse through noisy data. When training models that lack a large amount of training data, feature selection reduces the search space, making the resulting model more accurate.<sup>61</sup>

To begin with the feature selection, a dataset of all the available data from the participants was created, and the permutation importance of each feature was generated. Knowing that physiological features are often colinear, we utilized hierarchical clustering on the Spearman rank-order correlations of the features, as detailed in the Permutation Importance page<sup>62</sup> on scikit-learn.<sup>52</sup> This allowed us to determine if feature reduction was possible without loss of information with the available features. The correlation between features was found and plotted and a distance matrix was created. The distance matrix was used to create a dendrogram using Ward’s linkage<sup>63</sup> for hierarchical clustering. The resulting dendrogram (seen in Figure 3) allowed us to choose a feature from each cluster and create a new training set using only the selected features. A comparison of the accuracy between the model generated from all the features and the model generated from the selected features showed a 2% drop in accuracy, indicating that the reduction of features would not negatively affect the performance of the model.

The features chosen from the dendrogram included heart rate standard deviation and mean, BVP standard deviation and mean, the mean of acceleration in the  $X$ ,  $Y$ , and  $Z$  directions, the temperature mean, and the standard deviation of the skin conductance response. These nine features were able to produce a model that performed well on the group data when no separation based on individual participants



**Figure 1.** Process of collecting data and creating a predictive sleep analysis model.

was made. However, due to the heterogeneity of individuals' physiological data and sleep patterns, when creating models for each individual, the nine features did not capture enough information. Since the initial dendrogram analysis showed that some features could be reduced, we then used  $f$  regression, which employs univariate linear regression, along with  $K$  best feature selection to find the lowest number of features that could be used without jeopardizing the performance of the model for each individual. From these, we were able to identify that reducing to below 23 features caused the performance of the models to drop. In order to ensure that there was no cross-over between the training data and testing data,  $K$  best feature selection was done only on the training data, after the test and training split was complete. Each model was trained on the training data consisting of 23 features. Table 2 Column 3 shows the frequency of each feature being used in an individual model.

### Machine learning

Machine learning uses computational algorithms in order to build models that can represent a given dataset.<sup>64</sup> The most commonly used method of machine learning for practical applications is supervised machine learning. Supervised machine learning algorithms produce hypotheses and

**Table 1.** R-MBA clinical severity scores of each participant.

Participant ID	R-MBA Clinical severity
1398352	8
1411836	10
1458535	24
1471409	45
1502801	25
1503635	13
1623448	7

ID: Identification; R-MBA: Revised Motor-Behavioral Inventory

predictions by learning the general pattern found in data and correlating the patterns to the provided labels.<sup>65</sup> For supervised machine learning, a labeled dataset has to be provided. A subset of the data, the training set, is used to train the model using both the features and the labels. The

**Table 2.** Physiological features extracted and how many times each feature was used after feature selection.

Physiological signal	Initial feature extracted	Number of times used in individual model
PPG	Heart rate mean	Two
	Heart rate standard deviation	Six
	Heart rate max	Two
	Heart rate min	Two
	BVP mean	
	BVP standard deviation	Two
	BVP max	Five
	BVP min	Five
	IBI mean	Five
	IBI standard deviation	Two
	IBI max	Four
	IBI min	Six
EDA	SCL mean	Five
	SCL standard deviation	Four
	SCL max	Four
	SCL min	Five
	SCR mean	One
	SCR standard deviation	Four
	SCR max	Three
	SCR min	Two
3-Axis accelerometer	X-direction acceleration mean	One
	X-direction acceleration standard deviation	Seven
	X-direction acceleration max	Seven
	X-direction acceleration min	Six
	Y- acceleration mean	Six
	Y-direction acceleration standard deviation	Six
	Y-direction acceleration max	Six

(continued)



Table 2. Continued.

Physiological signal	Initial feature extracted	Number of times used in individual model
	Y-direction acceleration min	Six
	Z-direction acceleration mean	Three
	Z-direction acceleration standard deviation	Six
	Z-direction acceleration max	Five
	Z-Direction acceleration min	Seven
Temperature	Temperature mean	Seven
	Temperature standard deviation	Six
	Temperature max	Six
	Temperature min	Seven

PPG: photoplethysmography; EDA: electrodermal activity; BVP: blood volume pulse; IBI: interbeat interval; SCL: skin conductance level; SCR: skin conductance response.

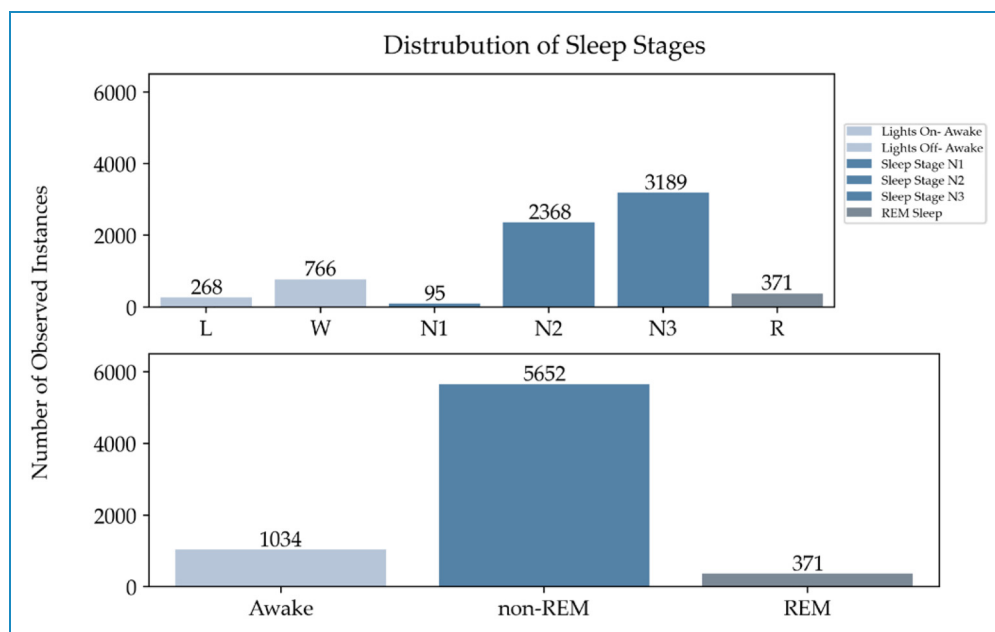
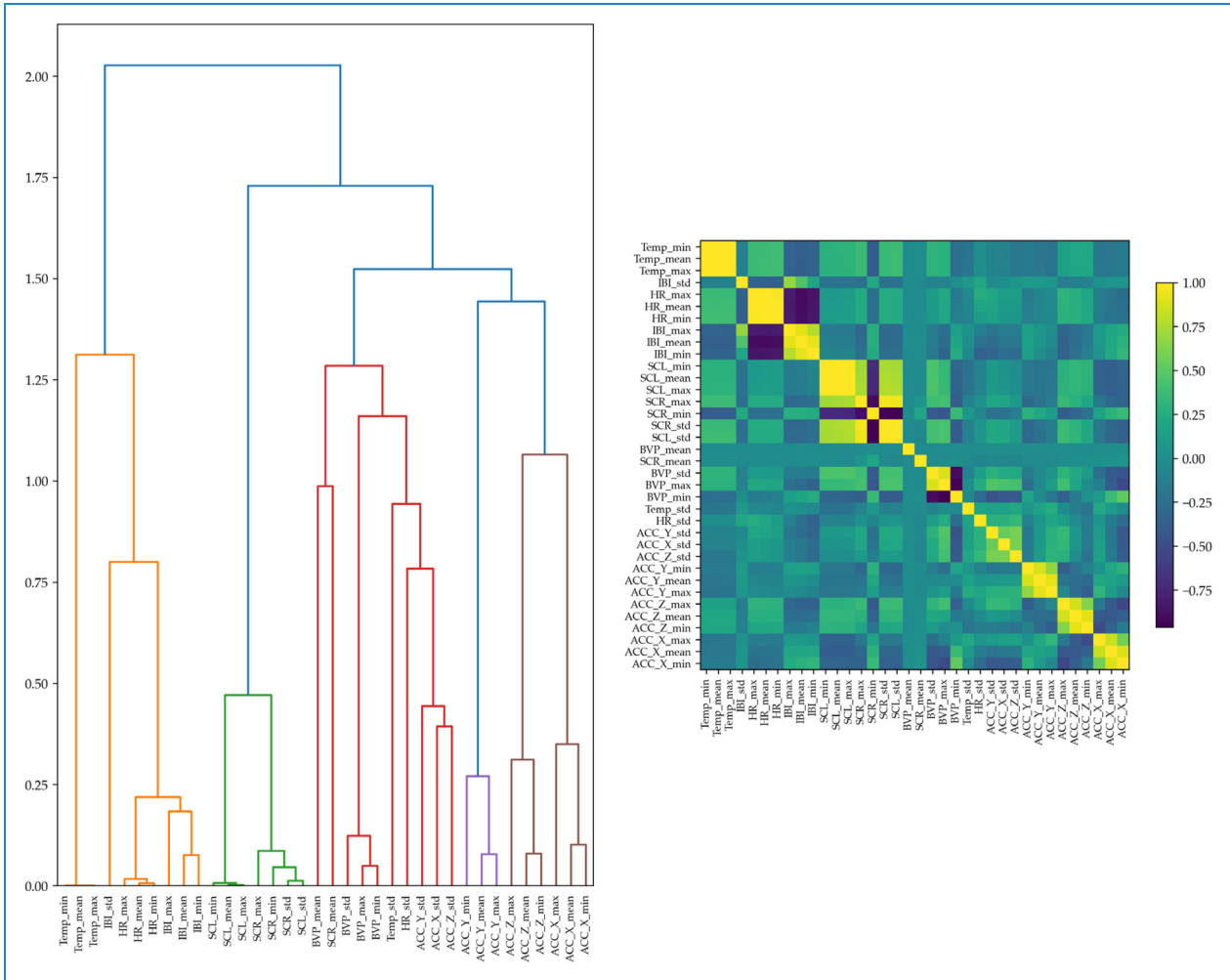


Figure 2. Distribution of sleep stages shown with all six labels and how they are recategorized into 3 classes.

remaining data, the test set, is used to evaluate the model. The test set labels are removed and the unlabeled features are predicted on by the model. The predicted labels are then compared to the ground truth labels.<sup>66</sup> Variation in physiological signals is expected and can be attributed to a variety of factors, including age, activity level, and, in the case of this work, neurological disorders. When features

are varied but possess fundamental qualities that can distinguish the different classes, i.e., physiological signals, supervised machine learning is especially applicable.<sup>67</sup> For this reason, supervised machine learning was used to create our predictive sleep analysis model.

The term 'individual model' refers to a model that is trained only on data from the participant that the model



**Figure 3.** Dendrogram of hierarchical clustering and heatmap of feature correlation used to determine starting point for feature selection.

will be predicting on. One well-known method of evaluating individual models is K-fold cross-validation, which prevents over-fitting and increases the robustness of the evaluation of the model.<sup>68</sup> K-fold cross-validation works by splitting the data into K-equal folds. One fold is held out as the test set each time and the remaining folds are used to train the model. When dealing with unbalanced classes within the data, literature suggests the use of stratified K-fold cross-validation. Stratified K-fold validation uses the same basic method as K-fold cross-validation but maintains the class ratio throughout the K folds of the original dataset.<sup>30</sup> For our work, the value of K was determined by the following equation:

$$K = \frac{L}{L * 0.2} \quad (1)$$

where L is the length of the dataset. This creates an 80/20 training-testing split within each of the folds, which is the most common split used for individual model evaluation. After the cross-validation was completed, the fold with

the best F1 score for each individual was extracted to generate a model for that participant. Individual models are the preferred method when working with highly individualized and varied data, such as physiological data. For this work, a support vector machine (SVM) classifier was used. SVM, an optimization approach, finds the best-fit line with the optimal separating hyperplane. The data are classified by maximizing the margin found between the classes in the feature space and the hyperplane boundary is used to classify the unseen data.<sup>69,70</sup> SVM has good generalization capabilities, making it ideal for classifying across different individuals and varying physiological signals. Other works have also shown the feasibility of SVM for sleep-wake classification from PPG data<sup>70</sup> and sleep stage classification from a single channel EEG.<sup>71</sup> The scikit-learn<sup>52</sup> SVM classifier was used with a radial basis function kernel and a regularization parameter of 1.0. The kernel coefficient, gamma, was set to “scale,” which uses  $1/(n\text{-features} * X.\text{var}())$  as the value of gamma. Lastly, a one-vs-rest decision function shape was used. For the



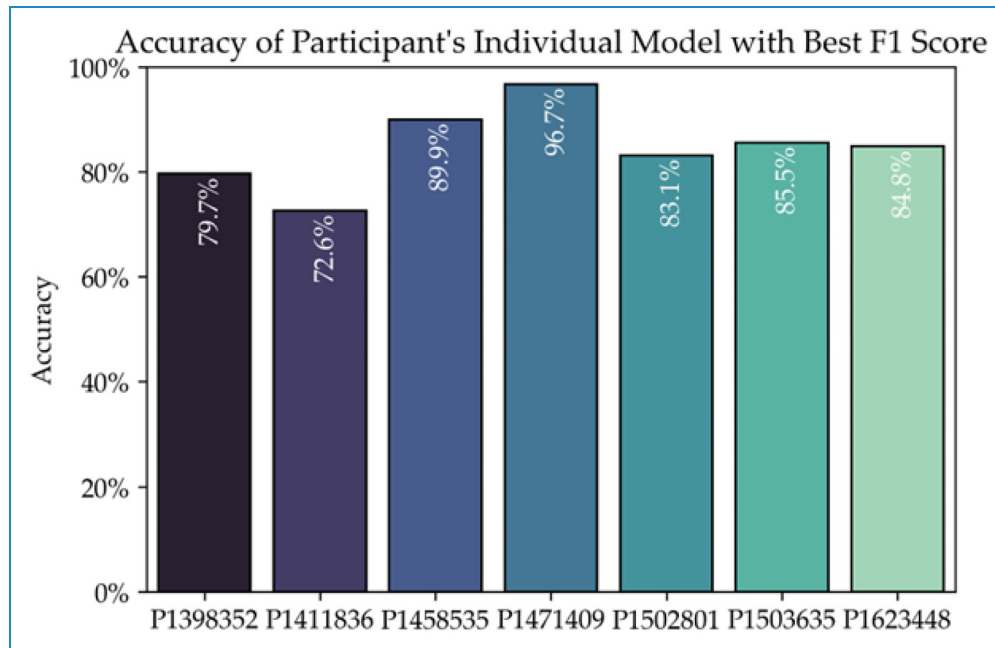


Figure 4. Accuracy of best F1 scored model for each participant.

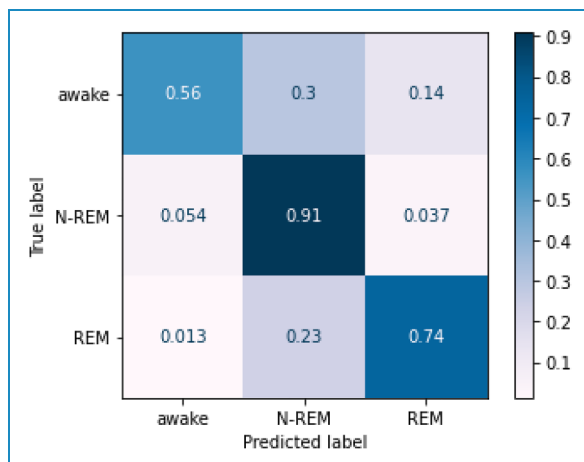


Figure 5. Confusion matrix for all predictions given by individual models.

individual models, the SVM was used within a pipeline along with an Edited Nearest Neighbors undersampling technique which was used to resample all classes, except the minority class. A Borderline SMOTE oversampling technique was used on the minority class in order to balance the classes.

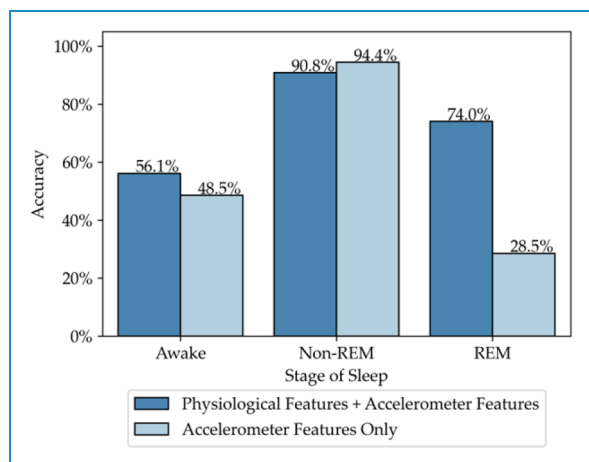
In order to evaluate the significance of the addition of physiological features, as opposed to using only accelerometer features as seen in previous studies with actigraphy, individual models using an SVM and stratified K-fold were developed using only features derived from the accelerometer data. This allowed for a direct comparison of individual models with and without physiological features.

## Results and discussion

In order to ensure that the resultant models were correctly predicting each class, the fold with the best F1 score from the stratified K-fold was extracted and stored for each subject. This resulted in the predictions given by each of the seven models having an accuracy of 85.1% and an F1 score of 84.4 when compared to the ground truth labels from the PSG. Each participant's model predicted with an accuracy between 72.6% and 96.7% as can be seen in Figure 4. By extracting the highest F1 models, we were able to address the overfitting issue we saw when looking at the models with the best accuracies. The confusion matrix in Figure 5 shows the confusion matrix for all the predictions given by the seven individual models when compared to the PSG labels. This shows that the individual models were able to predict all three sleep stages at above 50% accuracy.

Because we do not collect eye movement data, the transition between awake and Non-REM as well as the transition from Non-REM to REM sleep is particularly difficult to differentiate,<sup>72</sup> as can be seen by the lower accuracy of the awake sleep state.

While our seven participants showed a wide range of R-BMA severity scores, there were no noticeable trends or correlations between the disease severity and accuracy of the models. We did not explore clinical severity as part of this pilot study, since this initial work aims not to create a model for severity, but to provide an analytical validation of a proxy for sleep analysis against gold-standard PSG. The use of wearable sensors and machine learning for sleep analysis as a predictor for disease severity in RTT is beyond the scope of this current work, especially



**Figure 6.** Comparison of accuracy for each stage of sleep predictions using physiological features and accelerometer features vs only using accelerometer features.

given the small sample size and preliminary nature of this work. Using information gleaned from the dendrogram of hierarchical clustering and K best feature selection, the final model was trained using 23 input features. The final features for each model relied heavily on temperature and accelerometer data measures, with seven out of seven models using temperature and accelerometer features. Features derived from PPG were consistently used for five of the seven models. These features being used is consistent with other studies in RTT that explore HRV, temperature, and accelerometer measures.<sup>20,39,40,73</sup>

While the overall accuracy of the predictions given by the model utilizing only accelerometer data is comparable to the accuracy of the model with both physiological features and accelerometer features, when analyzing the distribution of predictions, it is clear that accelerometer data alone is unable to provide sufficient information for the model to predict the REM stage of sleep accurately. This can be seen in Figure 6, where the accelerometer-only model predicted REM sleep with 28.5% accuracy in contrast to 74% accuracy with the addition of physiological features.

Overall, the individual models with physiological features and accelerometer features perform well when compared to commercial products, which tend to have accuracies varying between 60% and 90% when doing epoch-to-epoch comparisons to PSG, depending on the class (awake, Non-REM, REM) being predicted.<sup>72,37</sup> Our work shows that the addition of physiological data improves the model fit and should be considered for future research in NDDs such as RTT. While actigraphy-based sleep analysis has been done,<sup>20</sup> in light of clinical features, related to temperature dysregulation<sup>73</sup> and HRV<sup>74</sup> in RTT, wearable devices that incorporate physiological data should be considered in future studies. The present work

provides evidence for the feasibility of using DHTs, such as wearable sensors, in RTT, even at young ages, as there was no data loss due to non-compliance with wearing the device. Although some previous studies have questioned the feasibility of DHTs in RTT due to repetitive hand movements and interference with accelerometer readings, the incorporation of physiological data in our work circumvented these concerns. It is also important to note that models developed for commercially developed devices rely on data from a large group of people. In contrast, our current sample size only consists of 7 individuals. The sample size of the current work is a limitation, however, more participants are being recruited and further analysis and validations are planned. The variations in accuracies for the individual participants may be due to the placement of the E4 during data collection, as it is on the wrist and the device not being flush against the skin can result in noisy data. The placement of the E4, as well as other wearable devices, may be adjusted to reduce the probability of data loss and reduce noise in the data for future studies. Future work will focus on the creation of group models and the possible addition of non-invasive methods for obtaining muscle activation near the eyes for REM detection. The expansion of the model to accommodate more classes, such as being able to differentiate between N1, N2, and N3 sleep stages, as well as being able to detect sleep apnea, is pending the collection of additional data.

## Conclusion

Our current work paves the way for sleep analysis from wearable sensors for children with Rett by providing evidence for the feasibility of DHTs in RTT for continuous monitoring and demonstrating the benefits of incorporating physiological features. This will expand the ability of clinicians to monitor and analyze how sleep patterns for children with Rett differ from other children, which may allow for new interventions to be explored which can better inform sleep-based interventions and support families. The individual models developed and validated are clinically significant as they allow for progression of sleep disturbances to be tracked longitudinally with more frequent data points. Compared to the current standard of care, where sleep is assessed via PSG far less frequently, our method also allows for sleep monitoring to occur in the child's natural environment. Our work establishes the viability, through analytical validation, of using wearable devices and machine learning for sleep analysis, which paves the way for the establishment of group models in future work to expand the reach of sleep analysis through wearable sensors.

This initial work analytically validates the use of wearable physiological sensors and accelerometer data as a method of sleep analysis in children with RTT as compared to gold-standard PSG. This addresses the gap in the

literature for affordable and non-invasive methods of sleep analysis for children with RTT. Our models are able to predict 3 sleep stages, awake, non-REM, and REM sleep, with around 85% accuracy with individual models when compared to the gold-standard PSG. During our model creation, we explored feature selection to reduce the search space and training time without reducing accuracy and F1 and reduced our features from 36 to 23, which we hope will inform decisions regarding the use of different types of DHT for NDDs such as RTT. However, we acknowledge that this research is in its preliminary stage, featuring a limited sample size. Therefore, it is important to exercise caution when interpreting the results and we emphasize the necessity for additional data before generalization of the results. Future work will expand the validation to other classes of sleep and additional sleep metrics. In addition, future studies will examine the clinical validation of these models to the most used clinical outcome measures in RTT with the goal of also demonstrating both analytical and clinical validation for use in future clinical trials in RTT.

**Acknowledgments:** We graciously acknowledge the children and families who participated in this study.

**Author Contributions:** Conceptualization was done by M.M., A.U., C.F., S.U.P., and N.S.; methodology was done by M.M., A.U., and N.S.; software was handled by M.M. and A.U.; validation was done by M.M., S.U.P., and N.S.; formal analysis was done by M.M.; investigation was carried out by C.F. and S.U.P.; resources were handled by S.U.P. and N.S.; data curation was done by M.M., A.U., C.F., and S.U.P.; writing—original draft preparation was done by M.M.; writing—review and editing was done by A.U., C.F., S.U.P., and N.S.; visualization was done by M.M.; supervision was done by S.U.P. and N.S.; project administration was done by C.F. and S.U.P.; funding acquisition was done by S.U.P. All authors have agreed to the submission of this manuscript.

**Declaration of conflicting interests:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by 1R21TR003942-01 to S.U.P.

**Guarantor:** M.M.

**Ethical approval:** This study was approved by the IRB (IRB Number 210217).

**Informed consent:** Informed consent was obtained from all participants.

**ORCID ID:** Miroslava Migovich  <https://orcid.org/0000-0002-4819-7901>

## References

- Banerjee A, Miller MT, Li K et al. Towards a better diagnosis and treatment of Rett syndrome: A model synaptic disorder. *Brain* 2019; 142: 239–248.
- Ehinger Y, Matagne V, Villard L et al. Rett syndrome from bench to bedside: Recent advances. *F1000Research* 2018; 7: 398.
- Petruti U, Dudman DC, Scosyrev E et al. Global prevalence of Rett syndrome: Systematic review and meta-analysis. *Syst Rev* 2023; 12: 5.
- Neul JL, Kaufmann WE, Glaze DG et al. Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol* 2010; 68: 944–950.
- Fu C, Armstrong D, Marsh E et al. Consensus guidelines on managing Rett syndrome across the lifespan. *BMJ Paediat Open* 2020; 4: e000717.
- Fu C, Armstrong D, Marsh E et al. Multisystem comorbidities in classic Rett syndrome: A scoping review. *BMJ Paediat Open* 2020; 4: e000731.
- Brand S and Kirov R. Sleep and its importance in adolescence and in common adolescent somatic and psychiatric conditions. *Int J Gen Med* 2011; 4: 425–442.
- Matricciani L, Paquet C, Galland B et al. Children’s sleep and health: A meta-review. *Sleep Med Rev* 2019; 46: 136–150.
- Leven Y, Wiegand F and Wilken B. Sleep quality in children and adults with Rett syndrome. *Neuropediatrics* 2020; 51: 198–205.
- Spruyt K. Sleep problems in individuals with Rett syndrome: a systematic review and meta-analysis. *Sleep Epidemiol* 2022; 2: 100027.
- Young D, Nagarajan L, de Klerk N et al. Sleep problems in Rett syndrome. *Brain Dev* 2007; 29: 609–616.
- Mori Y, Downs J, Wong K et al. Longitudinal effects of caregiving on parental well-being: The example of Rett syndrome, a severe neurological disorder. *Eur Child Adolesc Psychiatry* 2019; 28: 505–520.
- Palacios-Ceña D, Famoso-Pérez P, Salom-Moreno J et al. “Living an obstacle course”: a qualitative study examining the experiences of caregivers of children with Rett syndrome. *Int J Environ Res Public Health* 2019; 16: 41.
- Patel AA and Glaze DG. Chapter 34 - sleep and sleep disorders in Rett Syndrome. In: Watson RR and Preedy VR (eds) *Neurological modulation of sleep*. Academic Press, 2020, pp.339–345.
- Boban S, Leonard H, Wong K et al. Sleep disturbances in Rett syndrome: impact and management including use of sleep hygiene practices. *Am J Med Genet Part A* 2018; 176: 1569–1577.
- Rundo JV and Downey R. Polysomnography. *Handb Clin Neurol* 2019; 160: 381–392.
- Smith MT, McCrae CS, Cheung J et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American academy of sleep medicine systematic review, meta-analysis, and GRADE assessment. *J Clin Sleep Med: JCSM: Off Publ Am Acad Sleep Med* 2018; 14: 1209–1230.

18. Esbensen AJ, Hoffman EK, Stansberry E et al. Convergent validity of actigraphy with polysomnography and parent reports when measuring sleep in children with down syndrome. *J Intell Disabil Res: JIDR* 2018; 62: 281–291.
19. Moore M, Evans V, Hanvey G et al. Assessment of sleep in children with autism spectrum disorder. *Children (Basel, Switzerland)* 2017; 4: 72.
20. Merbler AM, Byiers BJ, Garcia JJ et al. The feasibility of using actigraphy to characterize sleep in Rett syndrome. *J Neurodev Disord* 2018; 10: 8.
21. Temudo T, Oliveira P, Santos M et al. Stereotypies in Rett syndrome: Analysis of 83 patients with and without detected MECP2 mutations. *Neurology* 2007; 68: 1183–1187.
22. Carotenuto M, Esposito M, D’Aniello A et al. Polysomnographic findings in Rett syndrome: A case-control study. *Sleep Breath* 2013; 17: 93–98.
23. Fiorillo L, Puiatti A, Papandrea M et al. Automated sleep scoring: a review of the latest approaches. *Sleep Med Rev* 2019; 48: 101204.
24. Carley DW and Farabi SS. Physiology of sleep. *Diabetes Spectrum: Publ Am Diabetes Assoc* 2016; 29: 5–9.
25. Sadeghi R, Banerjee T, Hughes JC et al. Sleep quality prediction in caregivers using physiological signals. *Comput Biol Med* 2019; 110: 276–288.
26. Koumans AJR, Tursky B and Solomon P. Electrodermal levels and fluctuations during normal sleep. *Psychophysiology* 1968; 5: 300–306.
27. Aldredge JL and Welch AJ. Variations of heart rate during sleep as a function of the sleep cycle. *Electroencephalogr Clin Neurophysiol* 1973; 35: 193–198.
28. Zemaityte D, Varoneckas G and Sokolov E. Heart rhythm control during sleep. *Psychophysiology* 1984; 21: 279–289.
29. Sajjadih A, Shahsavari A, Safaei A et al. The association of sleep duration and quality with heart rate variability and blood pressure. *Tanaffos* 2020; 19: 135–143.
30. Beattie Z, Oyang Y, Statan A et al. Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. *Physiol Meas* 2017; 38: 1968.
31. Sekkal RN, Bereksi-Reguig F, Ruiz-Fernandez D et al. Automatic sleep stage classification: from classical machine learning methods to deep learning. *Biomed Signal Process Control* 2022; 77: 103751.
32. Santaji S and Desai V. Analysis of EEG signal to classify sleep stages using machine learning. *Sleep Vigilance* 2020; 4: 145–152.
33. Piñero P, Garcia P, Arco L et al. Sleep stage classification using fuzzy sets and machine learning techniques. *Neurocomputing* 2004; 58–60: 1137–1143.
34. Mikkelsen KB, Ebajemito JK, Bonmati-Carrion MA et al. Machine-learning-derived sleep–wake staging from around-the-ear electroencephalogram outperforms manual scoring and actigraphy. *J Sleep Res* 2019; 28: e12786.
35. Arora A, Chakraborty P and Bhatia MPS. Analysis of data from wearable sensors for sleep quality estimation and prediction using deep learning. *Arab J Sci Eng* 2020; 45: 10793–10812.
36. Sathyanarayana A, Joty S, Fernandez-Luque L et al. Sleep quality prediction from wearable data using deep learning. *JMIR Mhealth Uhealth* 2016; 4: e125.
37. de Zambotti M, Goldstone A, Claudatos S et al. A validation study of fitbit charge 2™ compared with polysomnography in adults. *Chronobiol Int* 2018; 35: 465–476.
38. Haghayegh S, Khoshnevis S, Smolensky MH et al. Accuracy of wristband fitbit models in assessing sleep: Systematic review and meta-analysis. *J Med Internet Res* 2019; 21: e16273.
39. Iakovidou N, Lanzarini E, Singh J et al. Differentiating females with Rett syndrome and those with multi-comorbid autism spectrum disorder using physiological biomarkers: a novel approach. *J Clin Med* 2020; 9: 2842.
40. Suresha PB, O’Leary H, Tarquinio DC et al. Rett syndrome severity estimation with the BioStamp nPoint using interactions between heart rate variability and body movement. *PLoS ONE* 2023; 18: e0266351.
41. Goldsack JC, Coravos A, Bakker JP et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). *npj Digit Med* 2020; 3: 1–15.
42. Raspa M, Bann CM, Gwaltney A et al. A psychometric evaluation of the motor-behavioral assessment scale for use as an outcome measure in Rett syndrome clinical trials. *Am J Intellect Dev Disabil* 2020; 125: 493–509.
43. Garbarino M, Lai M, Bender D et al. Empatica E3 — A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In: *2014 4th international conference on wireless mobile communication and healthcare - transforming healthcare through innovations in mobile and wireless technologies (MOBIHEALTH)*. pp. 39–42. DOI:10.1109/MOBIHEALTH.2014.7015904.
44. Real-time physiological signals—E4 EDA/GSR sensor, 2022. <https://www.empatica.com/e4-wristband>.
45. Polysmith Sleep Systems, 2017. <https://eu.nihonkohden.com/en/products/neurology/polysmithsleepsystems.html>.
46. Castaneda D, Esparza A, Ghamari M et al. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int J Biosens Bioelectron* 2018; 4: 195–202.
47. Utilizing the PPG/BVP signal, 2021. <https://support.empatica.com/hc/en-us/articles/204954639-Utilizing-the-PPG-BVP-signal>.
48. Critchley H and Nagai Y. Electrodermal activity (EDA). In: Gellman MD and Turner JR (eds) *Encyclopedia of behavioral medicine*. New York, NY: Springer, 2013, pp.666–669.
49. Electrodermal Activity (EDA) | SpringerLink. [https://link.springer.com/referenceworkentry/10.1007/978-1-4419-1005-9\\_13](https://link.springer.com/referenceworkentry/10.1007/978-1-4419-1005-9_13).
50. Braithwaite DJJ. A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments.
51. Thinkology. Silicon Sensing | MEMS Accelerometers. <https://www.siliconsensing.com/technology/mems-accelerometers/>.
52. Pedregosa F, Varoquaux G, Gramfort A et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011; 12: 2825–2830.
53. Zheng A and Casari A. *Feature engineering for machine learning: principles and techniques for data scientists*. Sebastopol, CA: O’Reilly Media, Inc., 2018. ISBN 978-1-4919-5319-8.
54. Korkalainen H, Aakko J, Duce B et al. Deep learning enables sleep staging from photoplethysmogram for patients with suspected sleep apnea. *Sleep* 2020; 43: zsa098.
55. Sun H, Jia J, Goparaju B et al. Large-scale automated sleep staging. *Sleep* 2017; 40: zsx139.

56. Kluyver T, Ragan-Kelley B, Pérez F et al. Jupyter Notebooks— a publishing format for reproducible computational workflows. IOS Press, pp. 87–90. DOI:10.3233/978-1-61499-649-1-87.
  57. Project Jupyter. <https://jupyter.org>.
  58. Bian D, Wade J, Swanson A et al. Design of a physiology-based adaptive virtual reality driving platform for individuals with ASD. *ACM Trans Access Comput* 2019; 12: 2:1–2:24.
  59. Migovich M, Korman A, Wade J et al. Design and validation of a stress detection model for use with a VR based interview simulator for autistic young adults. In: Antona M and Stephanidis C (eds) *Universal access in human-computer interaction. Design methods and user experience*. Lecture Notes in Computer Science, Cham: Springer International Publishing, pp.580–588. ISBN 978-3-030-78092-0. DOI:10.1007/978-3-030-78092-0\_40.
  60. Guyon I and Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003; 3: 1157–1182.
  61. Kumar V and Minz S. Feature selection: A literature review. *SmartCR* 2014; 4: 211–229.
  62. Permutation Importance with Multicollinear or Correlated Features. [https://scikit-learn/stable/auto\\_examples/inspection/plot\\_permutation\\_importance\\_multicollinear.html](https://scikit-learn/stable/auto_examples/inspection/plot_permutation_importance_multicollinear.html).
  63. Ward's Linkage. <https://www.statistics.com/glossary/wards-linkage/>.
  64. Rebala G, Ravi A and Churiwala S. Machine learning definition and basics. In: Rebala G, Ravi A and Churiwala S (eds) *An introduction to machine learning*. Cham: Springer International Publishing, 2019, pp.1–17.
  65. Singh A, Thakur N and Sharma A. A review of supervised machine learning algorithms. In: *2016 3rd international conference on computing for sustainable global development (INDIACom)*. pp. 1310–1315.
  66. Jiang T, Gradus JL and Rosellini AJ. Supervised machine learning: a brief primer. *Behav Ther* 2020; 51: 675–687.
  67. El Naqa I and Murphy MJ. What is machine learning? In: El Naqa I, Li R and Murphy MJ (eds) *Machine learning in radiation oncology: theory and applications*. Cham: Springer International Publishing, 2015, pp.3–11.
  68. Berrar D. Cross-Validation. ISBN 978-0-12-809633-8, 2018. DOI:10.1016/B978-0-12-809633-8.20349-X.
  69. Cortes C and Vapnik V. Support-vector networks. *Mach Learn* 1995; 20: 273–297.
  70. Motin MA, Karmakar C, Palaniswami M et al. Photoplethysmographic-based automated sleep–wake classification using a support vector machine. *Physiol Meas* 2020; 41: 075013.
  71. Alickovic E and Subasi A. Ensemble SVM method for automatic sleep stage classification. *IEEE Trans Instrum Meas* 2018; 67: 1258–1265.
  72. Chinoy ED, Cuellar JA, Huwa KE et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep* 2021; 44: zsa291.
  73. Symons FJ, Byiers B, Hoch J et al. Infrared thermal analysis and individual differences in skin temperature asymmetry in Rett syndrome. *Pediatr Neurol* 2015; 53: 169–172.
  74. zSingh J, Ameenpur S, Ahmed R et al. An observational study of heart rate variability using wearable sensors provides a target for therapeutic monitoring of autonomic dysregulation in patients with Rett syndrome. *Biomedicines* 2022; 10: 1684.
-