




Article

A Deep Learning Strategy for Automatic Sleep Staging Based on Two-Channel EEG Headband Data

Amelia A. Casciola ^{1,†} , Sebastiano K. Carlucci ^{1,†}, Brianne A. Kent ^{2,3,†}, Amanda M. Punch ¹, Michael A. Muszynski ¹, Daniel Zhou ¹, Alireza Kazemi ⁴ , Maryam S. Mirian ², Jason Valerio ², Martin J. McKeown ^{2,*}  and Haakon B. Nygaard ^{2,*}

¹ Department of Electrical and Computer Engineering Capstone, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; amelia.casciola@gmail.com (A.A.C.); sebastiano.carlucci@gmail.com (S.K.C.); amandampunch@gmail.com (A.M.P.); michael.ampm@gmail.com (M.A.M.); danielzhou4970@gmail.com (D.Z.)

² Djavad Mowafaghian Centre for Brain Health, Division of Neurology, University of British Columbia, Vancouver, BC V6T 1Z3, Canada; bkent@sfu.ca (B.A.K.); maryam.mirian@ubc.ca (M.S.M.); jvale093@gmail.com (J.V.)

³ Department of Psychology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

⁴ Center for Mind and Brain, Department of Psychology, University of California, Davis, CA 95618, USA; kazemi@ucdavis.edu

* Correspondence: martin.mckeown@ubc.ca (M.J.M.); haakon.nygaard@ubc.ca (H.B.N.)

† Equal contribution.



Citation: Casciola, A.A.; Carlucci, S.K.; Kent, B.A.; Punch, A.M.; Muszynski, M.A.; Zhou, D.; Kazemi, A.; Mirian, M.S.; Valerio, J.; McKeown, M.J.; et al. A Deep Learning Strategy for Automatic Sleep Staging Based on Two-Channel EEG Headband Data. *Sensors* **2021**, *21*, 3316. <https://doi.org/10.3390/s21103316>

Academic Editor: Jinseok Lee

Received: 15 April 2021

Accepted: 6 May 2021

Published: 11 May 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Sleep disturbances are common in Alzheimer's disease and other neurodegenerative disorders, and together represent a potential therapeutic target for disease modification. A major barrier for studying sleep in patients with dementia is the requirement for overnight polysomnography (PSG) to achieve formal sleep staging. This is not only costly, but also spending a night in a hospital setting is not always advisable in this patient group. As an alternative to PSG, portable electroencephalography (EEG) headbands (HB) have been developed, which reduce cost, increase patient comfort, and allow sleep recordings in a person's home environment. However, naïve applications of current automated sleep staging systems tend to perform inadequately with HB data, due to their relatively lower quality. Here we present a deep learning (DL) model for automated sleep staging of HB EEG data to overcome these critical limitations. The solution includes a simple band-pass filtering, a data augmentation step, and a model using convolutional (CNN) and long short-term memory (LSTM) layers. With this model, we have achieved 74% ($\pm 10\%$) validation accuracy on low-quality two-channel EEG headband data and 77% ($\pm 10\%$) on gold-standard PSG. Our results suggest that DL approaches achieve robust sleep staging of both portable and in-hospital EEG recordings, and may allow for more widespread use of ambulatory sleep assessments across clinical conditions, including neurodegenerative disorders.

Keywords: deep learning; EEG headband; sleep staging; machine learning; neurodegenerative disease; sleep

1. Introduction

It is increasingly recognized that sleep abnormalities often accompany neurodegenerative disorders and in some cases are considered a core manifestation of the disease [1]. Research is now focused on determining whether sleep disturbances antedate disease onset and/or represent a biomarker of disease progression [1].

The gold standard in sleep assessment is polysomnography (PSG). However, there are several factors that limit the usefulness of PSG for studying sleep in patients with neurodegenerative diseases. First, it is relatively expensive, and thus many PSG studies are statistically under-powered. Second, the unnatural environment and discomfort associated with the numerous electrodes and wires may disturb the subject, and results using PSG

may therefore not accurately reflect sleep in the home environment [2]. Third, patients with dementia are prone to delirium and it is not always ethically feasible for all clinical populations to undergo inpatient PSG [3]. Finally, the proportion of subjects in need of formal sleep assessments outweighs the capacity of accredited PSG sleep laboratories, limiting access to diagnostic services.

Because of these limitations there is a desire to move toward inexpensive, comfortable, in-home sleep measurements. However, with portable EEG measurement devices, the data quality is often a key concern. PSG provides high-quality data collected under tightly controlled conditions from multiple simultaneous recordings, monitored in real time by trained technicians. In contrast, portable headsets are used in uncontrolled environments and are limited in the type and amount of data collected (e.g., number of simultaneous EEG electrodes and physiological measurements). A common form of in-home EEG measurement device is a headband (HB), where the electrodes are placed on a fabric band wrapping the circumference of the subject's head. In headbands, electrodes are often placed on a subject's face and forehead, where hair is not a concern. This concentration over frontal regions causes the data to be susceptible to corruption by ocular artifacts created by eye blinks, flutters, and eye movements as well as muscle artifacts from the frontalis muscle [4]. However, this electrode placement functions remarkably well in detecting slow wave power [5], which is critically important in the assessment of Alzheimer's disease [6].

Computer-aided diagnosis has been used effectively for some time in various health-care applications [7–11], and several deep learning (DL) methods and neural network architectures have been proposed to automate sleep stage scoring using single, double, or multichannel PSG EEG data [12]. During the sleep staging process, the EEG signal is divided into periods of Wake, rapid eye movement (REM) and non-REM sleep (NREM). The NREM stage is further split into stages N1, N2 and N3. Convolutional neural networks (CNN), recurrent neural networks (RNN) or a mixture of both have been applied with a high degree of success in EEG processing for sleep stage classification [13–15]. Convolutional neural networks are particularly efficient at extracting time-invariant and localized information [16]. In EEG processing, convolutional layers replace the task of manual feature engineering by relying on the network's training ability to learn relevant features. Whereas recurrent neural networks impose a structure to learn relationships between consecutive data points in a sequence.

However, current machine learning (ML) models built for PSG data typically fail to generalize well on EEG data collected from a HB. The significant difference in the distribution between HB and PSG data is reinforced by using a covariate shift similarity test [17]. The test identifies a covariate shift between two datasets based on whether the origin of any sample from either set can be correctly identified. A simple classifier was trained to predict whether a sample was collected with PSG or HB and was able to correctly predict the origin 99% of the time. This highlights the differences in characteristics and distribution between both types of EEG. These differences explain why models built for PSG cannot necessarily be applied to HB data with an expectation of the same degree of success. Here we present a model that is more robust than current automated sleep staging solutions and is able to perform sleep staging accurately on both PSG and HB data.

Limited studies applying ML approaches to HB recordings have been promising. We found only one publicly available HB dataset, collected using the Dreem Headband (Dreem, San Francisco, CA, USA) [18], and only a handful of studies on sleep staging with HB EEG [2,5,19–22]. For example, Levendowski and colleagues [22] compared simultaneous recordings from 47 participants, using a commercially available three-channel EEG recording device (Sleep Profiler, Advanced Brain Monitoring, Carlsbad, CA, USA) and PSG, and found that the overall agreement between the technologist-scored PSG and the automated sleep staging of the EEG device was 71.3% ($\kappa = 0.67$). Given that the five technologists in the study had an overall interscorer agreement of 75.9% ($\kappa = 0.70$), it was concluded that the autoscored frontal EEG was as accurate as human-staged PSG, except for sleep stage N1. Similarly, Lucey and colleagues [5] compared the same three-channel EEG de-

vice (Sleep Profiler, Advanced Brain Monitoring, Carlsbad, CA, USA) with simultaneous PSG in 29 participants and found similar agreement ($\kappa = 0.67$). However, in this study, the two technologists had extremely high interscorer agreement for the PSG data ($\kappa = 0.97$, 80.8–98.8% depending on stage), so there was a substantial drop compared to the autoscored EEG. Of the HB sleep staging solutions studied, only one attempted deep learning. Arnal and colleagues achieved an accuracy of 83.5%, compared to the interscorer agreement of 86.4% [2], using a long short-term memory (LSTM)-based model. This model used 25 nights of HB EEG collected with the Dreem Headband, and incorporated heart rate, respiration rate and respiration rate variability into their algorithm.

Here, we present an automatic sleep staging method for two-channel HB EEG data based on a CNN + LSTM architecture. The model was trained and tested with data from 12 overnight EEG recordings from 12 different participants. We achieve a mean accuracy of 74% when sleep staging HB data, compared to 68% when using a traditional ML pipeline. Additionally, when balanced accuracy is taken into consideration, the deep learning method outperforms traditional automated techniques by 20%. The increase in balanced accuracy implies consistent performance across all sleep stages.

Our model's predictive ability on low-quality data indicates its potential in clinical and research applications. By advancing the usability of EEG headband data, this model can dramatically expand formal sleep staging in an ambulatory setting, paving the way for larger studies in key populations such as Alzheimer's disease. Additionally, the DL model's predictive power may also benefit future work on computer-aided diagnosis of neurodegenerative diseases.

2. Materials and Methods

2.1. Data Collection

The data was collected from a group of 12 subjects overnight at the University of British Columbia Hospital Blackmore Centre for Sleep Disorders. The subjects (6 male, 6 female) were between the ages of 21 and 61. Participants were simultaneously monitored with a PSG setup, as well as a Cognionics 2-channel EEG headband (Cognionics, San Diego, CA, USA). Data collection was approved by Vancouver Coastal Health Authority-UBC and University of British Columbia Office of Research Ethics (ethics protocol H16-00925).

An overnight attended level 1 PSG was performed using standard electrode montages. This included EEG (channels F3-M2, F4-M1, C3-M2, C4-M1, O1-M2 and O2-M1), electrooculography (EOG), electromyography (EMG), electrocardiography (ECG), oronasal thermal airflow sensor, pulse oximetry, respiratory inductance plethysmography acoustic sensor and audio-equipped video PSG as recommended by American Academy of Sleep Medicine (AASM) criteria (version 2.4, 2017) [23]. The headband setup was performed according to the Cognionics manufacturer instructions [24]. Both PSG and HB EEG recordings used adhesive foam electrodes (Kendall Medi-Trace Mini, Davis Medical Electronics, Vista, USA). The overnight PSG was scored by a Senior Polysomnographic Technologist and interpreted by a board-certified sleep physician using standard AASM criteria based on 6 EEG channels, 2 EOG channels and 1 EMG channel [23]. The manually scored PSG labels were used for both the PSG and HB recordings. To control for any possible temporal misalignment between the two systems, we performed both automatic correlation checking of lagged signals and manual inspection. In Figure 1, we report the distribution of the assigned sleep stage labels. The dataset is highly unbalanced toward the N2 and Wake classes, the N2 class being twice more frequent than the Wake class and roughly seven times more frequent than the N1 class. This label imbalance is to be expected, since the majority of sleep is usually spent in the N2 stage [25].

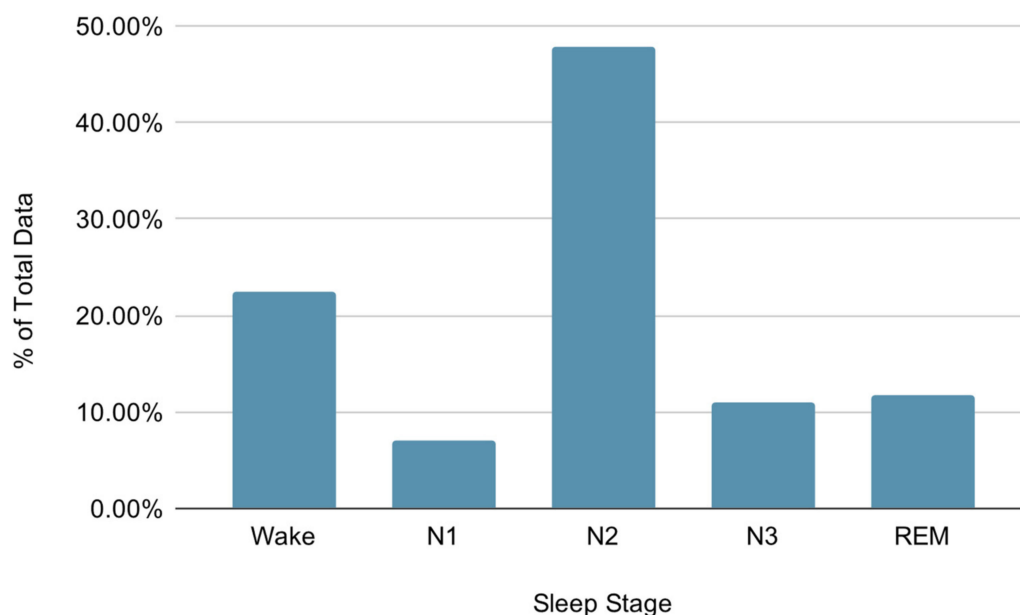


Figure 1. Distribution of sleep stages assigned by the Senior Polysomnographic Technologist (Wake: 22.45%, N1: 6.95%, N2: 47.86%, N3: 11.07%, rapid eye movement (REM): 11.66%).

The HB EEG recordings were taken between nodes F3-A1 (channel 1) and F4-A2 (channel 2). The characteristics of the raw recorded HB data are reported in Table 1.

Table 1. Headband data characteristics.

Bandwidth	Sample Rate	Amplifier Gain	Resolution	Noise
0–131 Hz	500 samples/sec	6	24 bits/sample	0.7 μ V

2.2. Data Preprocessing

Slightly different preprocessing pipelines were applied to PSG and HB data, as presented in Sections 2.2.1 and 2.2.2.

2.2.1. Preprocessing PSG Data

Following the sleep scoring, only two PSG channels (F3-M2 and F4-M1) were used for the remainder of the study, to most closely match the electrode placement of the HB. A Finite Impulse Response (FIR) bandpass filter with cutoff frequencies of 0.5 and 12 Hz was applied to these collected PSG channels. The data were then downsampled to 25 Hz and separated into nonoverlapping 30-s epochs.

2.2.2. Preprocessing HB Data

The HB EEG data was visually inspected to identify segments which were corrupted beyond interpretation due to hardware error and a section of 50 min was removed from one subject. The erroneous epochs were likely due to electrodes losing contact with the skin surface.

The same bandpass filtering and epoching as described in Section 2.2.1 was applied to the HB data.

2.3. HB Data Cleaning

In addition to preprocessing, we implemented a simple data screening algorithm to automatically remove corrupted epochs from the HB data only. An example of a corrupted epoch is shown in Figure 2. There are a few key indicators that the signal on the right (channel 2) is not a valid EEG signal. The first is that the amplitude of the channel 2 signal

largely exceeds the amplitude of an average EEG signal (generally below 100 μV) [26]. The second is that there is no visual similarity between the channel 1 and channel 2 signals.

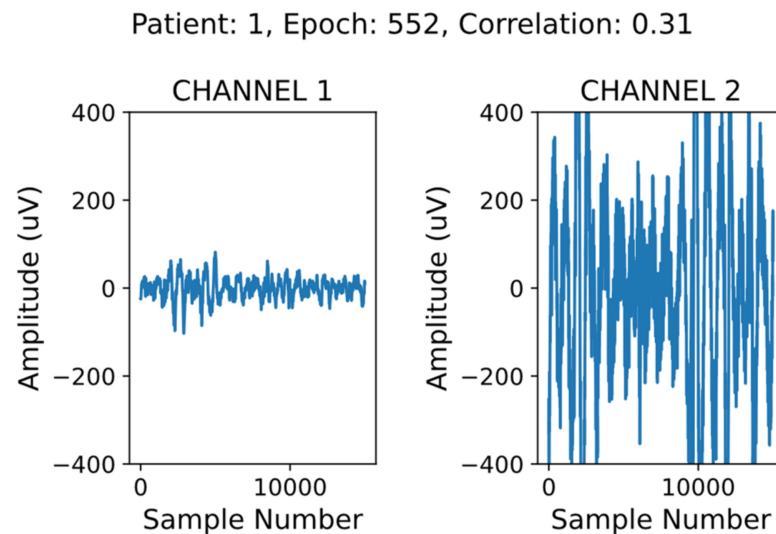


Figure 2. Valid electroencephalogram (EEG) signal shown on the left, corrupted data on the right.

To remove the corrupted data, we filtered the epochs where the correlation between the signals from channel 1 and channel 2 was low (below 0.9), and the mean amplitude of at least one channel was high (above 40 μV). We removed only the data in the corrupted channel, duplicating the valid channel in its place. This method aims to minimize the amount of valid data being discarded, while making use of the expected correlation between mirroring channels [27].

A grid search was performed in order to determine the best correlation and amplitude thresholds. First, ranges of valid amplitude and correlation thresholds were selected by manual parameter adjustment and inspection of the filtered epochs. Next, a grid search was performed on each combination of minimum correlation from 0.6 to 0.9 in steps of 0.1 and maximum channel amplitude from 30 μV to 60 μV in steps of 10 μV . For each of these 16 threshold combinations, we performed leave-one-out cross validation 12 times, once for each subject.

2.4. Data Augmentation

It has been shown that data augmentation can substantially increase accuracy when using deep learning for EEG analysis [28]. We implemented overlapping windows in order to artificially increase the size of our dataset, an approach which has yielded promising results in other literature [29–31]. In this method, we concatenated all contiguous 30-s epochs of the same sleep stage in the time domain. The concatenated blocks were then redivided into new 30-s epochs, all overlapping by 75%. This process is illustrated in Figure 3. We applied this method only on training data to avoid corrupting the prediction results.

To determine the optimal overlap percentage, we generated artificially augmented datasets using overlap percentages of 75%, 50% and 20%. The optimal overlap percentage was determined after averaging the results of leave-one-out cross validation over the 12 participants.

2.5. Deep Learning Model Architecture

Based on a review of existing deep learning models for sleep staging using EEG signals, the most promising architectures for low-quality EEG data were CNN and CNN + LSTM models [32–36]. We implemented a CNN + LSTM model based on Bresch’s model for single-channel EEG [37], with the architecture shown below in Figure 4. The model was

created using Python 3 and the Keras framework with TensorFlow backend version 2.2, and summarized in Table A1.

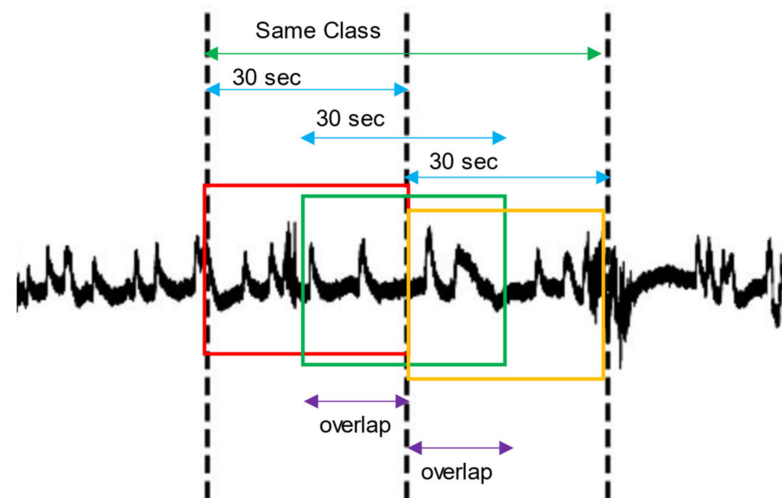


Figure 3. Overlapping windows applied over the EEG signal (black line). The dotted vertical lines delimit two 30-s epochs of the same class. The red, green and yellow rectangles correspond to the newly generated epochs after applying a specified overlap (purple).

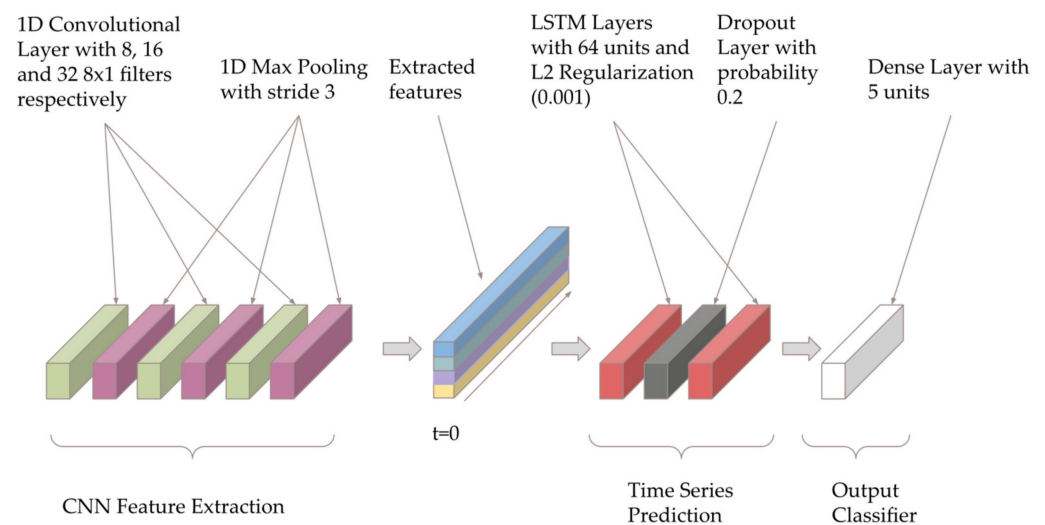


Figure 4. Convolutional and long short-term memory (CNN + LSTM) model architecture inspired by [37].

The model contains three convolutional layers (to learn the useful spatial information), followed by two LSTM layers (to extract influential temporal information) and a dense layer. The first convolutional layer has eight filters, and this number doubles with each convolutional layer. The two LSTM layers have 64 units, with the final dense layer only containing five, as well as a softmax output activation. The convolutional layers are particularly effective for extracting time-invariant and localized information. Their function in the model is similar to the task of manual feature extraction in traditional machine learning methods. The LSTM layers then impose a structure to learn relationships between the consecutive data points extracted by the CNN layers.

Implementing regularization is a common method of reducing overfitting in deep learning models [38,39]. In our model, the LSTM layers are regularized to by adding a dropout of 0.2. By randomly assigning different nodes to a weight of zero at each iteration, dropout can simulate a range of architectures, as opposed to one static model layout. This

minimizes the risk of certain nodes developing large weights based on the training data, reducing generalizability [40]. To further reduce the model's reliance on individual features we implemented L2 weight decay with parameter 0.001.

2.6. Model Training

To guide the training of the model, we use the categorical cross entropy loss function. When the dataset is unbalanced (the distribution is skewed toward some labels), the model's predictions tend to become biased toward the most common labels. Our dataset is highly unbalanced with the most frequent class, sleep stage 2, comprising 47% of the data. To address this, we apply a per-class weight to the categorical cross entropy loss function:

$$L(y, \hat{y}) = - \sum_{i=1}^m w(y_i) y_i \log(\hat{y}_i) \quad (1)$$

where y and \hat{y} are the true labels and the predictions, respectively, and $L(y, \hat{y})$ is the loss. The relative weight of label y_i in the dataset (the number of y_i labels divided by the number of examples in the dataset) is defined as $w(y_i)$. The parameters of the model are updated at each training iteration using an Adam optimizer [41] with an initial learning rate of 5×10^{-5} , $\beta_1 = 0.99$ and $\beta_2 = 0.999$.

Training was performed using leave-one-out cross validation, for 200 epochs per fold, using a batch size of 64. This process was completed using a NVIDIA Titan RTX GPU (Nvidia, Santa Clara, CA, USA), and a 4.0 GHz Intel processor (Intel, Mountain View, USA). Each training epoch took about 3 s to complete, with the entire process taking approximately 2 h.

2.7. Traditional Sleep Staging Techniques

As well as using our deep learning model for sleep staging, we also attempted multiple traditional processing pipelines. The highest accuracy of these supervised modeling techniques was achieved using extracted features and an ensemble bagged trees model.

After a similar preprocessing pipeline to that described in Section 2.2, we extracted features in the following categories:

- Frequency domain;
- Time domain;
- Higher-order statistical analysis (HOSA)-based;
- Wavelet-based.

In total, 62 features per EEG channel were extracted and are listed in Table 2. The computed values for the features were normalized using Min-Max feature scaling for a final value between 0 and 1.

Table 2. Type and number of features extracted from each EEG channel.

Feature Category	Feature Group	Feature Size
Frequency Domain	RSP	11
	HP	15
	SWI	3
Time Domain	Hjorth	3
	Skewness	1
	Kurtosis	1
HOSA	Bi-Spectrum	20
Wavelet	Relative Power	8
Total Features		62

As a classic baseline approach, the ensemble-bagged trees model was trained on these features using leave-one-out cross validation.

3. Results

3.1. Deep Learning Model Performance

Combining the model architecture, data cleaning and data augmentation methods described previously, we performed leave-one-out cross validation for each participant. We achieved a mean prediction accuracy of 74.01% (standard deviation 10.32%) with HB data and 76.98% (standard deviation 10.05%) with PSG data. These results are summarized in Figure 5.

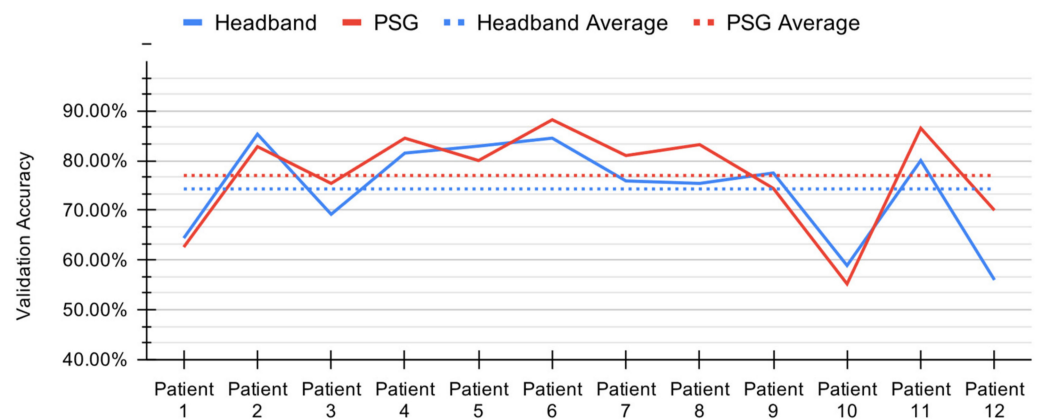


Figure 5. Per-patient validation accuracy during leave-one-out cross validation for the convolutional and long short-term memory (CNN+LSTM) model with headband (HB) and polysomnography (PSG).

The similarity in sleep staging accuracy achieved with HB and PSG demonstrates this model’s robustness to variance in data source and thus quality. In Table 2, we include the confusion matrices for each subject for both HB and PSG.

Furthermore, we report the average stage-wise performance of the deep learning CNN+LSTM model over 12 folds, also known as the percent correct recall. The results shown in Table 3 highlight the model’s similar performance across sleep stages N2, N3, REM and Wake despite the unbalanced label frequencies. Especially notable is the high accuracy achieved for sleep stage N3, which comprises only 11% of the dataset, and may be particularly important for risks of dementia [6]. The accuracies achieved for stage N1 are considerably lower, as is commonly the case with both manual sleep staging and automated methods [42,43].

Table 3. Average stage-wise performance over all folds for the deep learning model.

Data	N1	N2	N3	REM	Wake	Accuracy	Balanced Accuracy
HB	29.80%	74.87%	84.02%	73.96%	80.60%	74.01%	68.65%
PSG	31.08%	77.82%	85.27%	75.38%	84.64%	77.00%	70.84%

To provide an interpretation of the model’s predictive behavior, we followed a similar approach to Chambon and colleagues [13]. We analyzed whether the model learned to discriminate between sleep stages based on frequency bands, which is a common approach among human sleep scorers. For example, the relative spectral density of the δ band (0.5–4 Hz) is often used by human scorers to classify epochs as “deep sleep” (sleep stage N3). We studied how the model’s predictive behavior changes with respect to the following frequency bands: δ (0.5–4 Hz), θ (6–8 Hz) and $\alpha + \beta$ (>8 Hz). We filtered the test data into the aforementioned frequency bands and passed these individually into the pre-trained model.

The model associates higher frequencies with the Wake stage, and its predictive power for sleep stage 3 does not decrease when given the delta band of a signal (Figure 6). These results point to the use of frequency bands for the sleep stage prediction of Wake and N3.

However, the results indicate that the model does not learn to classify sleep stages 1 and 2 based on the presence of theta frequencies.

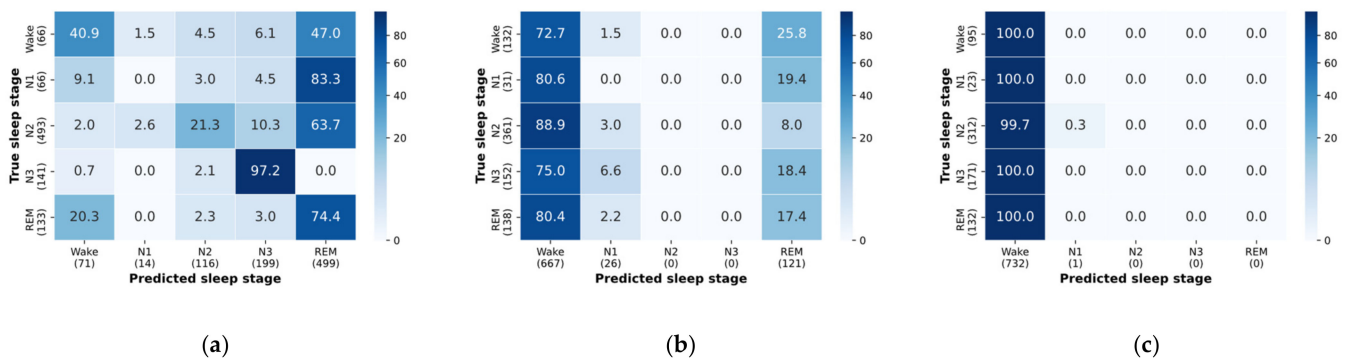


Figure 6. Confusion matrices for inputs passed through the following bandpass filters: (a) delta; (b) theta; (c) alpha and beta.

3.2. Baseline Model Performance

The ensemble bagged trees method presented in Section 2.7 achieved an accuracy of 67.53% (standard deviation 11.94%) with HB data, and accuracy of 73.07% (standard deviation 6.46%) with PSG data, as shown in Figure 7.

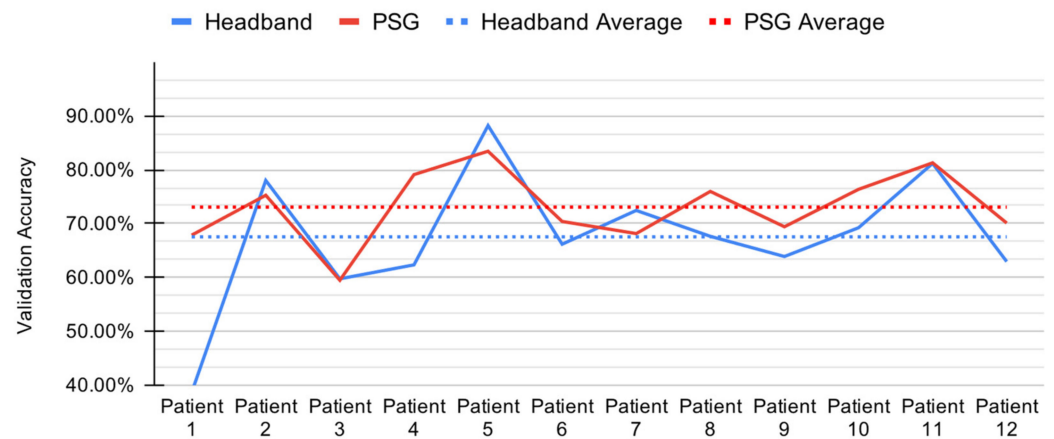


Figure 7. Per-patient validation accuracy during leave-one-out cross validation for the ensemble bagged trees model with headband (HB) and polysomnography (PSG).

We report the average stage-wise performance of the ensemble bagged trees method over all 12 folds, shown in Table 4. This model performs best on N2 and Wake stages for both HB and PSG data, while performing comparatively poorly on N1, N3 and REM sleep stages.

Table 4. Average stage-wise performance over all folds for the ensemble bagged trees method.

Data	N1	N2	N3	REM	Wake	Accuracy	Balanced Accuracy
HB	4.11%	82.32%	49.65%	28.26%	80.03%	67.53%	48.88%
PSG	13.13%	88.36%	35.90%	55.17%	83.49%	73.06%	55.21%

The balanced accuracy, which is computed as the mean of all percent recalls of the labels, is also reported with 48.88% and 55.21% on HB and PSG data, respectively.

The low accuracies for N1, N3 and REM and high accuracies for N2 and Wake are justified by the dataset bias toward the N2 and Wake labels. This implies that the ML model is overfitting the dataset, and does not provide any meaningful predictions on the data.

4. Discussion

We have shown that a DL sleep staging model achieves 74% accuracy on low-quality headband EEG data, compared to 77% with gold-standard PSG. The model performs well across all sleep stages, leading to a balanced accuracy of almost 20% more than any machine learning sleep staging method attempted. We also show that this model achieves an especially high accuracy for sleep stage N3, with acceptable performance for REM sleep classification, both of which may be highly relevant to the pathophysiology of neurodegenerative disorders [1,6]. These results were attained without any extensive preprocessing or special artifact removal procedures. Our findings address the major barrier preventing more widespread use of HB for ambulatory sleep assessment, largely related to poor performance in automated classification tasks.

4.1. Deep Learning Model Comparison

As previously noted, the architecture of our deep learning solution was adapted from a model created by Bresch and colleagues [37]. However, our implementation has a few key differences. Firstly, we reduce overfitting by regularizing the LSTM layers with dropout and weight decay. Secondly, the CNN output in Bresch's architecture is flattened, which discards the temporal structure of the signal. Instead of flattening the CNN output, we directly pass the time sequence of the extracted features to the downstream LSTM layers.

Another key difference between our model and Bresch's is the input data. For our use-case, we have two-channel EEG data, downsampled to 25 Hz as opposed to one 100 Hz channel. Finally, due to our substantially smaller dataset, we include a comprehensive data cleaning and data augmentation preprocessing pipeline, before passing the signal to the deep learning model, something which is not done by Bresch and colleagues.

Using Bresch's architecture before our adjustments resulted in a mean accuracy over 12 folds of 70.83% (standard deviation 10.55%), while after our adjustments, the results improved by almost 4%.

4.2. Baseline Model Comparison

We present a comparison between the results of the CNN+LSTM model and the ML model described in the previous sections. The accuracies and balanced accuracies shown in Tables 3 and 4 demonstrate the deep learning model's superior adaptability to changes in data quality. The DL model shows a 2% difference between balanced accuracies when trained on PSG versus trained on HB. The ML model, on the other hand, shows a difference of 6%. The smaller gap between PSG and HB results points to a more robust model, indicating that data quality is not a significant variable with respect to sleep staging performance.

When comparing the two models' results directly, the DL model has a higher accuracy and balanced accuracy both with HB and PSG data. Specifically, we see an additional increase of 6.50% in accuracy and 19.77% in balanced accuracy using HB data, when compared to the ML model. Similarly, we see an additional increase of 3.94% in accuracy and 15.63% in balanced accuracy using PSG data. The higher accuracy indicates an improvement in overall predictive capability, while the higher balanced accuracy indicates an improvement in the model's generalizability across all sleep stages.

Stage-wise comparison of the model performances show that the DL model's N2 accuracy decreased slightly from 82.32% on HB (and 88.36% on PSG) to 74.87% on HB (and 77.82% on PSG), in exchange for a significantly greater increase in the under-represented labels (N1, N3 and REM). The largest of these reported increases is in stage N3, where an additional 34.37% on HB data (and 49.37% on PSG data) is gained.

4.3. Electrode Comparison

The EEG recordings in this study used gel electrodes that balance ease of use, patient comfort, and signal quality (Kendall Medi-Trace Mini, Davis Medical Electronics, Vista, USA). New developments of dry and semi-dry electrodes have been promising [44–47],

and while we anticipate similar sleep staging results using our DL model with dry and semi-dry electrode variants, additional studies are needed to test this hypothesis.

5. Conclusions

We demonstrate that a simplified two channel HB EEG device can be used to accurately stage sleep when combined with a DL data analysis model. The DL model outperforms traditional machine learning approaches, and may allow for wider application of ambulatory sleep assessments across clinical populations, including neurodegenerative disorders.

Author Contributions: Conceptualization, B.A.K., M.J.M., M.S.M. and H.B.N.; methodology, B.A.K., A.A.C., S.K.C., M.A.M., D.Z., J.V., M.J.M. and H.B.N.; software, A.A.C., A.K., M.A.M., S.K.C., D.Z. and A.M.P.; validation, A.A.C. and S.K.C.; formal analysis, A.K., S.K.C. and A.A.C.; investigation B.A.K., M.J.M., J.V., M.S.M. and H.B.N.; resources M.J.M. and H.B.N.; data curation, B.A.K., A.K., M.A.M.; writing—original draft preparation, A.A.C., S.K.C., A.M.P., M.A.M. and D.Z.; writing—review and editing, all authors; visualization, S.K.C., A.A.C., A.M.P. and M.A.M.; supervision M.J.M., M.S.M. and H.B.N.; project administration, M.J.M., M.S.M. and H.B.N.; funding M.J.M. and H.B.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by the John Nichol Chair in Parkinson’s Research. B.A.K. was funded by Michael Smith Foundation for Health Research, Killam Trusts, and CIHR Banting.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Vancouver Coastal Health Authority-UBC and University of British Columbia Office of Research Ethics (protocol code H16-00925 approved on 7 March 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding authors. The source code created in this study can be found at: <https://bitbucket.org/cap131/headbandsleepscorer/src/master/> (accessed on 2 May 2021).

Acknowledgments: The authors of this study acknowledge the project guidance given by Terry Lee and Mohammad Kazemi at the University of British Columbia, through the Electrical and Computer Engineering Capstone Program.

Conflicts of Interest: H.B.N. is a paid, independent consultant to Biogen and Roche Canada. M.J.M. has received honoraria from Sunovion Pharmaceuticals and Paladin Labs, Inc.

Appendix A

Table A1. Convolutional and long short-term memory (CNN + LSTM) model summary.

Layer Type	Output Shape	Param #
Conv1D	(None, 2993, 8)	136
Activation (ReLU)	(None, 2993, 8)	0
MaxPooling1D	(None, 997, 8)	0
Conv1D	(None, 990, 16)	1040
Activation (ReLU)	(None, 990, 16)	0
MaxPooling1D	(None, 330, 16)	0
Conv1D	(None, 323, 32)	4128
Activation (ReLU)	(None, 323, 32)	0
MaxPooling1D	(None, 107, 32)	0
LSTM	(None, 107, 64)	24,832
LSTM	(None, 64)	33,024
Dense	(None, 5)	325
Total Params	63,485	
Trainable Params	63,485	
Non-Trainable Params	0	

Table 2. Confusion matrices for deep learning model predictions on each subject.

		Headband (HB)					Polysomnography (PSG)						
Subject 1	True sleep stage	Wake (112)	54.5	4.5	8.0	0.0	33.0	Wake (112)	70.5	0.9	0.9	0.0	27.7
	N1 (67)	1.5	9.0	35.8	0.0	53.7	N1 (67)	10.4	7.5	4.5	0.0	77.6	
	N2 (236)	0.8	2.1	75.4	9.3	12.3	N2 (236)	0.4	2.1	65.7	7.2	24.6	
	N3 (9)	0.0	0.0	11.1	88.9	0.0	N3 (9)	0.0	0.0	44.4	55.6	0.0	
	REM (76)	2.6	3.9	0.0	2.6	90.8	REM (76)	2.6	6.6	0.0	0.0	90.8	
		Predicted sleep stage					Predicted sleep stage						
		Wake (66)	N1 (19)	N2 (212)	N3 (32)	REM (171)	Wake (89)	N1 (16)	N2 (163)	N3 (22)	REM (210)		
Subject 2	True sleep stage	Wake (118)	85.6	5.9	5.1	0.8	2.5	Wake (118)	91.5	3.4	1.7	0.0	3.4
	N1 (43)	2.3	18.6	32.6	0.0	46.5	N1 (43)	0.0	23.3	30.2	0.0	46.5	
	N2 (379)	0.0	0.3	90.8	4.5	4.5	N2 (379)	0.0	7.1	83.6	2.6	6.6	
	N3 (122)	0.0	0.0	5.7	94.3	0.0	N3 (122)	0.0	0.8	9.0	90.2	0.0	
	REM (102)	2.9	5.9	8.8	0.0	82.4	REM (102)	0.0	5.9	7.8	0.0	86.3	
		Predicted sleep stage					Predicted sleep stage						
		Wake (105)	N1 (22)	N2 (380)	N3 (133)	REM (124)	Wake (108)	N1 (48)	N2 (351)	N3 (120)	REM (137)		
Subject 3	True sleep stage	Wake (128)	88.3	4.7	3.9	0.0	3.1	Wake (128)	84.4	0.0	14.1	0.0	1.6
	N1 (88)	5.7	30.7	46.6	1.1	15.9	N1 (88)	2.3	25.0	70.5	0.0	2.3	
	N2 (385)	0.8	3.4	66.8	12.2	16.9	N2 (385)	0.0	2.3	92.5	3.1	2.1	
	N3 (35)	0.0	0.0	11.4	85.7	2.9	N3 (35)	0.0	0.0	40.0	60.0	0.0	
	REM (124)	15.3	2.4	2.4	0.0	79.8	REM (124)	8.9	16.1	21.8	0.0	53.2	
		Predicted sleep stage					Predicted sleep stage						
		Wake (140)	N1 (49)	N2 (310)	N3 (78)	REM (183)	Wake (121)	N1 (51)	N2 (477)	N3 (33)	REM (78)		

Table 2. Cont.

		Headband (HB)					Polysomnography (PSG)						
Subject 4	True sleep stage	Wake (205)	65.9	20.5	4.9	0.5	8.3	Wake (205)	88.3	5.9	2.0	0.0	3.9
	N1 (51)	11.8	39.2	31.4	0.0	17.6	N1 (51)	11.8	54.9	15.7	0.0	17.6	
	N2 (539)	1.9	4.8	88.9	2.2	2.2	N2 (539)	0.2	6.5	89.4	1.5	2.4	
	N3 (96)	1.0	0.0	9.4	89.6	0.0	N3 (96)	2.1	0.0	18.8	79.2	0.0	
	REM (123)	2.4	3.3	7.3	0.0	87.0	REM (123)	4.1	14.6	8.1	0.0	73.2	
		Wake (155)	N1 (92)	N2 (523)	N3 (99)	REM (145)	Wake (195)	N1 (93)	N2 (522)	N3 (84)	REM (120)		
		Predicted sleep stage					Predicted sleep stage						
Subject 5	True sleep stage	Wake (312)	85.3	5.8	1.0	0.3	7.7	Wake (312)	90.4	7.4	0.6	0.0	1.6
	N1 (26)	3.8	19.2	3.8	0.0	73.1	N1 (26)	0.0	42.3	11.5	0.0	46.2	
	N2 (322)	0.0	1.6	80.1	3.7	14.6	N2 (322)	0.0	8.1	65.5	5.0	21.4	
	N3 (96)	0.0	0.0	3.1	96.9	0.0	N3 (96)	0.0	0.0	0.0	100.0	0.0	
	REM (41)	0.0	4.9	0.0	0.0	95.1	REM (41)	0.0	7.3	0.0	0.0	92.7	
		Wake (267)	N1 (30)	N2 (265)	N3 (106)	REM (129)	Wake (282)	N1 (63)	N2 (216)	N3 (112)	REM (124)		
		Predicted sleep stage					Predicted sleep stage						
Subject 6	True sleep stage	Wake (95)	81.1	6.3	3.2	0.0	9.5	Wake (95)	71.6	8.4	6.3	1.1	12.6
	N1 (23)	17.4	26.1	21.7	0.0	34.8	N1 (23)	0.0	34.8	34.8	0.0	30.4	
	N2 (312)	0.0	5.4	86.2	6.4	1.9	N2 (312)	0.0	1.0	94.9	3.5	0.6	
	N3 (171)	0.6	0.0	1.2	95.9	2.3	N3 (171)	0.0	0.0	1.2	98.8	0.0	
	REM (132)	1.5	10.6	9.1	0.0	78.8	REM (132)	0.8	2.3	16.7	0.0	80.3	
		Wake (84)	N1 (43)	N2 (291)	N3 (184)	REM (131)	Wake (69)	N1 (22)	N2 (334)	N3 (181)	REM (127)		
		Predicted sleep stage					Predicted sleep stage						

Table 2. Cont.

		Headband (HB)					Polysomnography (PSG)				
Subject 7	Wake (240)	96.7	0.4	2.1	0.4	0.4	96.7	0.0	1.2	0.8	1.2
	N1 (80)	30.0	33.8	25.0	0.0	11.2	22.5	11.2	35.0	0.0	31.2
	N2 (357)	4.8	11.2	71.7	9.5	2.8	4.2	2.0	85.7	7.0	1.1
	N3 (110)	0.9	0.9	12.7	85.5	0.0	0.9	0.0	10.9	88.2	0.0
	REM (73)	21.9	16.4	1.4	0.0	60.3	4.1	4.1	19.2	0.0	72.6
		Wake (290)	N1 (81)	N2 (296)	N3 (129)	REM (64)	Wake (269)	N1 (19)	N2 (363)	N3 (124)	REM (85)
Subject 8	Wake (66)	72.7	6.1	15.2	1.5	4.5	90.9	1.5	4.5	1.5	1.5
	N1 (66)	10.6	33.3	33.3	0.0	22.7	30.3	24.2	33.3	0.0	12.1
	N2 (493)	4.1	7.1	77.3	7.1	4.5	2.8	5.5	87.0	3.7	1.0
	N3 (141)	0.0	0.0	0.7	99.3	0.0	0.0	0.0	3.5	96.5	0.0
	REM (133)	13.5	11.3	9.8	0.0	65.4	6.8	9.8	3.0	0.0	80.5
		Wake (93)	N1 (76)	N2 (427)	N3 (176)	REM (127)	Wake (103)	N1 (57)	N2 (463)	N3 (155)	REM (121)
Subject 9	Wake (132)	75.0	10.6	8.3	0.8	5.3	53.0	8.3	12.1	1.5	25.0
	N1 (31)	6.5	38.7	45.2	0.0	9.7	0.0	54.8	29.0	0.0	16.1
	N2 (361)	0.3	5.0	82.8	7.2	4.7	1.1	7.2	77.6	10.0	4.2
	N3 (152)	0.0	0.0	5.3	94.7	0.0	0.7	0.0	0.0	99.3	0.0
	REM (138)	2.2	23.2	18.8	0.0	55.8	0.0	23.2	13.0	0.0	63.8
		Wake (105)	N1 (76)	N2 (358)	N3 (171)	REM (104)	Wake (75)	N1 (86)	N2 (323)	N3 (189)	REM (141)

Table 2. Cont.

		Headband (HB)					Polysomnography (PSG)						
Subject 10	True sleep stage	Wake (191)	88.5	8.4	1.0	0.5	1.6	Wake (191)	93.2	4.7	0.5	0.0	1.6
	N1 (48)	20.8	70.8	0.0	0.0	8.3	20.8	43.8	18.8	0.0	16.7		
	N2 (491)	18.7	28.1	41.8	3.9	7.5	28.7	28.9	35.8	1.2	5.3		
	N3 (73)	4.1	0.0	4.1	90.4	1.4	1.4	4.1	5.5	89.0	0.0		
	REM (136)	26.5	14.7	0.0	0.0	58.8	36.0	5.9	0.0	0.0	58.1		
		Predicted sleep stage					Predicted sleep stage						
		Wake (310)	N1 (208)	N2 (210)	N3 (86)	REM (125)	Wake (379)	N1 (183)	N2 (190)	N3 (71)	REM (116)		
Subject 11	True sleep stage	Wake (504)	80.6	7.1	1.8	0.0	10.5	Wake (504)	95.0	0.6	0.2	0.0	4.2
	N1 (35)	48.6	25.7	0.0	0.0	25.7	51.4	37.1	2.9	0.0	8.6		
	N2 (192)	0.0	7.3	76.0	0.0	16.7	1.0	3.1	74.0	10.9	10.9		
	N3 (91)	0.0	0.0	85.7	14.3	0.0	0.0	0.0	15.4	84.6	0.0		
	REM (49)	0.0	10.2	0.0	0.0	89.8	0.0	10.2	2.0	0.0	87.8		
		Predicted sleep stage					Predicted sleep stage						
		Wake (423)	N1 (64)	N2 (233)	N3 (13)	REM (138)	Wake (499)	N1 (27)	N2 (159)	N3 (98)	REM (88)		
Subject 12	True sleep stage	Wake (86)	93.0	1.2	1.2	2.3	2.3	Wake (86)	80.2	0.0	4.7	1.2	14.0
	N1 (128)	24.2	12.5	34.4	0.8	28.1	4.7	14.1	47.7	1.6	32.0		
	N2 (548)	11.1	2.6	60.6	19.3	6.4	1.5	2.2	82.1	10.2	4.0		
	N3 (11)	9.1	0.0	18.2	72.7	0.0	0.0	0.0	18.2	81.8	0.0		
	REM (23)	13.0	17.4	26.1	0.0	43.5	4.3	0.0	30.4	0.0	65.2		
		Predicted sleep stage					Predicted sleep stage						
		Wake (176)	N1 (35)	N2 (385)	N3 (117)	REM (83)	Wake (84)	N1 (30)	N2 (524)	N3 (68)	REM (90)		

References

1. Kent, B.A.; Feldman, H.H.; Nygaard, H.B. Sleep and Its Regulation: An Emerging Pathogenic and Treatment Frontier in Alzheimer's Disease. *Prog. Neurobiol.* **2021**, *197*, 101902. [CrossRef] [PubMed]
2. Arnal, P.J.; Thorey, V.; Ballard, M.E.; Hernandez, A.B.; Guillot, A.; Jourde, H.; Harris, M.; Guillard, M.; Beers, P.V.; Chennaoui, M.; et al. The Dreem Headband as an Alternative to Polysomnography for EEG Signal Acquisition and Sleep Staging. *bioRxiv* **2019**, 662734. [CrossRef]
3. Malkani, R.; Attarian, H. Sleep in Neurodegenerative Disorders. *Curr. Sleep Med. Rep.* **2015**, *1*, 81–90. [CrossRef]
4. Dora, C.; Biswal, P.K. An Improved Algorithm for Efficient Ocular Artifact Suppression from Frontal EEG Electrodes Using VMD. *Biocybern. Biomed. Eng.* **2020**, *40*, 148–161. [CrossRef]
5. Lucey, B.P.; Mclelland, J.S.; Toedebusch, C.D.; Boyd, J.; Morris, J.C.; Landsness, E.C.; Yamada, K.; Holtzman, D.M. Comparison of a Single-Channel EEG Sleep Study to Polysomnography. *J. Sleep Res.* **2016**, *25*, 625–635. [CrossRef]
6. Jackson, M.L.; Cavuoto, M.; Schembri, R.; Doré, V.; Villemagne, V.L.; Barnes, M.; O'Donoghue, F.J.; Rowe, C.C.; Robinson, S.R. Severe Obstructive Sleep Apnea Is Associated with Higher Brain Amyloid Burden: A Preliminary PET Imaging Study. *J. Alzheimer's Dis.* **2020**, *78*, 611–617. [CrossRef]
7. Santos-Mayo, L.; San-Jose-Revuelta, L.M.; Arribas, J.I. A Computer-Aided Diagnosis System With EEG Based on the P3b Wave During an Auditory Odd-Ball Task in Schizophrenia. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 395–407. [CrossRef]
8. Ashfaq Khan, M.; Kim, Y. Cardiac Arrhythmia Disease Classification Using LSTM Deep Learning Approach. *Comput. Mater. Contin.* **2021**, *67*, 427–443. [CrossRef]
9. Giger, M.; Suzuki, K. Computer-Aided Diagnosis (CAD). In *Biomedical Information Technology*; Elsevier: Amsterdam, The Netherlands, 2007; pp. 359–374. ISBN 978-0-12-373583-6.
10. Pineda, A.M.; Ramos, F.M.; Betting, L.E.; Campanharo, A.S.L.O. Quantile Graphs for EEG-Based Diagnosis of Alzheimer's Disease. *PLoS ONE* **2020**, *15*, e0231169. [CrossRef]
11. Hulbert, S.; Adeli, H. EEG/MEG- and Imaging-Based Diagnosis of Alzheimer's Disease. *Rev. Neurosci.* **2013**, *24*, 563–576. [CrossRef]
12. Roy, Y.; Banville, H.; Albuquerque, I.; Gramfort, A.; Falk, T.H.; Faubert, J. Deep Learning-Based Electroencephalography Analysis: A Systematic Review. *J. Neural. Eng.* **2019**, *16*, 051001. [CrossRef]
13. Chambon, S.; Galtier, M.; Arnal, P.; Wainrib, G.; Gramfort, A. A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series. *arXiv* **2017**, arXiv:1707.03321. [CrossRef]
14. Neng, W.; Lu, J.; Xu, L. CRRSsleepNet: A Hybrid Relational Inductive Biases Network for Automatic Sleep Stage Classification on Raw Single-Channel EEG. *Brain Sci.* **2021**, *11*, 456. [CrossRef]
15. Patanaik, A.; Ong, J.L.; Gooley, J.J.; Ancoli-Israel, S.; Chee, M.W.L. An End-to-End Framework for Real-Time Automatic Sleep Stage Classification. *Sleep* **2018**, *41*. [CrossRef]
16. Ismail Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.-A. Deep Learning for Time Series Classification: A Review. *Data Min. Knowl. Disc.* **2019**, *33*, 917–963. [CrossRef]
17. Bickel, S.; Brückner, M.; Scheffer, T. Discriminative Learning under Covariate Shift. *J. Mach. Learn. Res.* **2019**, *10*, 2137–2155.
18. AWS S3 Explorer. Available online: <https://dreem-octave-irba.s3.eu-west-3.amazonaws.com/index.html> (accessed on 24 March 2021).
19. Griessenberger, H.; Heib, D.P.J.; Kunz, A.B.; Hoedlmoser, K.; Schabus, M. Assessment of a Wireless Headband for Automatic Sleep Scoring. *Sleep Breath* **2013**, *17*, 747–752. [CrossRef]
20. Zhang, Z.; Guan, C. An Accurate Sleep Staging System with Novel Feature Generation and Auto-Mapping. In Proceedings of the 2017 International Conference on Orange Technologies (ICOT), Singapore, 8–10 December 2017; pp. 214–217.
21. Dong, H.; Supratak, A.; Pan, W.; Wu, C.; Matthews, P.M.; Guo, Y. Mixed Neural Network Approach for Temporal Sleep Stage Classification. *IEEE Trans. Neural. Syst. Rehabil. Eng.* **2018**, *26*, 324–333. [CrossRef]
22. Levendowski, D.J.; Ferini-Strambi, L.; Gamaldo, C.; Cetel, M.; Rosenberg, R.; Westbrook, P.R. The Accuracy, Night-to-Night Variability, and Stability of Frontopolar Sleep Electroencephalography Biomarkers. *J. Clin. Sleep Med.* **2017**, *13*, 791–803. [CrossRef]
23. Berry, R.B.; Brooks, R.; Gamaldo, C.; Harding, S.M.; Lloyd, R.M.; Quan, S.F.; Troester, M.T.; Vaughn, B.V. AASM Scoring Manual Updates for 2017 (Version 2.4). *J. Clin. Sleep Med.* **2017**, *13*, 665–666. [CrossRef]
24. Dry EEG Headset | CGX | United States. Available online: <https://www.cgxsystems.com/> (accessed on 27 April 2021).
25. Patel, A.K.; Reddy, V.; Araujo, J.F. Physiology, Sleep Stages. In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2021.
26. Aurlien, H.; Gjerde, I.O.; Aarseth, J.H.; Eldøen, G.; Karlsen, B.; Skeidsvoll, H.; Gilhus, N.E. EEG Background Activity Described by a Large Computerized Database. *Clin. Neurophysiol.* **2004**, *115*, 665–673. [CrossRef]
27. Shukla, A.; Majumdar, A. Exploiting Inter-Channel Correlation in EEG Signal Reconstruction. *Biomed. Signal Process. Control* **2015**, *18*, 49–55. [CrossRef]
28. Lashgari, E.; Liang, D.; Maoz, U. Data Augmentation for Deep-Learning-Based Electroencephalography. *J. Neurosci. Methods* **2020**, *346*, 108885. [CrossRef]
29. O'Shea, A.; Lightbody, G.; Boylan, G.; Temko, A. Neonatal Seizure Detection Using Convolutional Neural Networks. In Proceedings of the 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), Tokyo, Japan, 25–28 September 2017; pp. 1–6.

30. Majidov, I.; Whangbo, T. Efficient Classification of Motor Imagery Electroencephalography Signals Using Deep Learning Methods. *Sensors* **2019**, *19*, 1736. [[CrossRef](#)]
31. Ullah, I.; Hussain, M.; Qazi, E.-H.; Aboalsamh, H. An Automated System for Epilepsy Detection Using EEG Brain Signals Based on Deep Learning Approach. *Expert Syst. Appl.* **2018**, *107*, 61–71. [[CrossRef](#)]
32. Sors, A.; Bonnet, S.; Mirek, S.; Vercueil, L.; Payen, J.-F. A Convolutional Neural Network for Sleep Stage Scoring from Raw Single-Channel EEG. *Biomed. Signal Process. Control* **2018**, *42*, 107–114. [[CrossRef](#)]
33. Tsinalis, O.; Matthews, P.M.; Guo, Y.; Zafeiriou, S. Automatic Sleep Stage Scoring with Single-Channel EEG Using Convolutional Neural Networks. *arXiv* **2016**, arXiv:1610.01683.
34. Vilamala, A.; Madsen, K.H.; Hansen, L.K. Deep Convolutional Neural Networks for Interpretable Analysis of EEG Sleep Stage Scoring. *arXiv* **2017**, arXiv:1710.00633.
35. Xu, Z.; Yang, X.; Sun, J.; Liu, P.; Qin, W. Sleep Stage Classification Using Time-Frequency Spectra From Consecutive Multi-Time Points. *Front. Neurosci.* **2020**, *14*. [[CrossRef](#)] [[PubMed](#)]
36. Supratak, A.; Dong, H.; Wu, C.; Guo, Y. DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG. *IEEE Trans. Neural. Syst. Rehabil. Eng.* **2017**, *25*, 1998–2008. [[CrossRef](#)] [[PubMed](#)]
37. Bresch, E.; Großekathöfer, U.; Garcia-Molina, G. Recurrent Deep Neural Networks for Real-Time Sleep Stage Classification From Single Channel EEG. *Front. Comput. Neurosci.* **2018**, *12*. [[CrossRef](#)]
38. Xie, L.; Wang, J.; Wei, Z.; Wang, M.; Tian, Q. DisturbLabel: Regularizing CNN on the Loss Layer. In Proceedings of the 2016 Institute of Electrical and Electronics Engineers (IEEE) Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4753–4762.
39. Zheng, Q.; Yang, M.; Tian, X.; Jiang, N.; Wang, D. A Full Stage Data Augmentation Method in Deep Convolutional Neural Network for Natural Image Classification. *Discret. Dyn. Nat. Soc.* **2020**, *2020*, e4706576. [[CrossRef](#)]
40. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
41. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
42. Rosenberg, R.S.; Van Hout, S. The American Academy of Sleep Medicine Inter-Scorer Reliability Program: Sleep Stage Scoring. *J. Clin. Sleep Med.* **2013**, *09*, 81–87. [[CrossRef](#)]
43. Fiorillo, L.; Puiatti, A.; Papandrea, M.; Ratti, P.-L.; Favaro, P.; Roth, C.; Bargiotas, P.; Bassetti, C.L.; Faraci, F.D. Automated Sleep Scoring: A Review of the Latest Approaches. *Sleep Med. Rev.* **2019**, *48*, 101204. [[CrossRef](#)]
44. Mathewson, K.E.; Harrison, T.J.L.; Kizuk, S.A.D. High and Dry? Comparing Active Dry EEG Electrodes to Active and Passive Wet Electrodes. *Psychophysiology* **2017**, *54*, 74–82. [[CrossRef](#)]
45. Li, G.-L.; Wu, J.-T.; Xia, Y.-H.; He, Q.-G.; Jin, H.-G. Review of Semi-Dry Electrodes for EEG Recording. *J. Neural Eng.* **2020**, *17*, 051004. [[CrossRef](#)]
46. Li, G.; Wu, J.; Xia, Y.; Wu, Y.; Tian, Y.; Liu, J.; Chen, D.; He, Q. Towards Emerging EEG Applications: A Novel Printable Flexible Ag/AgCl Dry Electrode Array for Robust Recording of EEG Signals at Forehead Sites. *J. Neural Eng.* **2020**, *17*, 026001. [[CrossRef](#)]
47. Li, G.; Wang, S.; Li, M.; Duan, Y.Y. Towards Real-Life EEG Applications: Novel Superporous Hydrogel-Based Semi-Dry EEG Electrodes Enabling Automatically ‘charge–Discharge’ Electrolyte. *J. Neural Eng.* **2021**, *18*, 046016. [[CrossRef](#)]