2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-0126-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIBM52615.2021.9669890

Sleep Stage Classification Using Non-Invasive Bed Sensing and Deep Learning

Nikhil Vyas
Dept. of Electrical Eng. and Computer Science
University of Missouri
Columbia, MO, USA
nv4vb@umsystem.edu

Kelly Ryoo
Dept. of Computer Science
Cornell University
Ithaca, NY, USA
kyr5@cornell.edu

Hosanna Tesfaye
Dept. of Information Technology
University of Missouri
Columbia, MO, USA
hhthk7@umsystem.edu

Ruhan Yi
Dept. of Electrical Eng. and Computer Science
University of Missouri
Columbia, MO, USA
ry793@umsystem.edu

Marjorie Skubic
Dept. of Electrical Eng. and Computer Science
University of Missouri
Columbia, MO, USA
SkubicM@missouri.edu

***Abstract*—Sleep stage classification can be used to monitor sleep quality and diagnose sleep disorders. Sleep disorders can be correlated to health conditions such as Alzheimer's and Parkinson's disease. This project uses a hydraulic bed sensor positioned under the mattress, as well as a deep learning approach, for sleep stage classification. Our motivation is to provide an automatic, non-invasive and more accessible method of classifying sleep stages by using deep learning to analyze data gathered from the hydraulic bed sensor. The test subjects for this project were elderly patients with sleep disorders. Polysomnography (PSG) data, the current gold standard, was also collected in a Sleep Lab to serve as the ground truth for the bed sensor data. In this study, sleep stages are categorized into 3 categories: Wake, Rapid Eye Movement (REM), and Non-Rapid Eye Movement (NREM). This paper uses a Convolutional Neural Network (CNN)- Long-Short Term Memory (LSTM) hybrid model with 2 CNNs of different filter sizes for feature extraction. These features are then fed into the LSTM for classification. Our results show an average accuracy of about 76% using the leave-one-subject-out (LOSO) validation. These results are promising and show that the hydraulic bed sensor combined with a deep learning approach is capable of providing an automatic and non-invasive method of classifying sleep stages.**

Keywords: classification, sleep stages, in-home monitoring, deep learning

1. Introduction

Sleep occupies a significant p ortion o f p eople's lives and is one of the key factors that contributes to a person's health. Good sleep can enhance one's neurological and physiological systems and bring about better mood and increased productivity. Conversely, sleep loss and sleep disorders have been associated with many health consequences in the long term, such as increased risk of hypertension, diabetes, obesity, depression, heart attack, and stroke. Sleep disorders can also be associated with many neurological and physiological diseases, such as Alzheimer's Disease [1] and Parkinson's Disease [2] or other ailments such as obstructive lung disease, sleep apnea, and restless leg syndrome [3]. As such, it is crucial to monitor sleep and conduct sleep studies to measure sleep quality and diagnose sleep disorders.

Polysomnography (PSG) is a comprehensive study of biological data consisting of numerous signals that capture various changes in the body. PSG is the most accurate and comprehensive method used in sleep studies to monitor sleep and diagnose sleep disorders. Traditionally, the extraction of PSG is carried out in sleep labs and consists of placing many electrodes, sensors, and masks on a patient's body surface. Through all these sensors, the system is able to simultaneously extract multiple biological signals from the patients. The extraction process also includes a technician who monitors the patient throughout the night and manually annotates the sleep stages based on 30-second epochs. The sleep scoring follows the American Academy of Sleep Medicine (AASM) Manual to differentiate between 5 main sleep stages: Wake, Rapid Eye Movement (REM), Non-Rapid Eye Movement (NREM)1, NREM2, and NREM3. In this study, we combined NREM1-3 into one sleep stage, NREM. Although the PSG system has many great advantages such as being very accurate and comprehensive, there are some disadvantages in its system. The extraction of PSG is usually held in sleep labs, which makes it less accessible to the general public. The extraction also requires that many devices are placed onto the patient's body. This is quite an invasive process that can affect the normal sleep patterns of the patients, resulting in the altering of sleep study results. Lastly, the classification of the data requires a technician

who monitors the devices and classifies the data manually, which is a very time consuming process.

In contrast, the study described here uses a non-invasive hydraulic bed sensor as a way to extract Ballistocardiography (BCG), which is a measurement of the body motion generated by the ejection of the blood at each cardiac cycle [4]. A deep learning approach is then used to classify 3 sleep stages: Wake, REM, and NREM. This provides an automatic, non-invasive, and longitudinal method of classifying sleep stages and overcomes the challenges of the traditional methods of sleep classification and PSG extraction. Another novelty of this study is that the dataset consists mainly of older adults, which tend to have noisier BCG signals compared to young adults due to problems in their cardiovascular system and stiffness in their arteries.

2. Related Works

2.1. Sleep Stage Classification using Pressure Sensors

As mentioned previously, using the traditional method of PSG to extract sleep data can be a very invasive and uncomfortable process. Thus, there have been efforts to develop a method to extract biological features from subjects in a non-invasive manner. A particular study conducted in 2010 at the VTT Technical Research Centre of Finland extracted heart-beat intervals and movement activity from healthy subjects using a bed sensor. Emfit sensor foils were placed into the bed mattress. A hidden Markov model was used for classification and achieved an accuracy of about 79%. Unlike our study, however, this project primarily uses a time-variant autoregressive model and hidden Markov model rather than a deep learning model [5]. Another study in 2016 extracted three signals (respiratory effort, leg movement, and body movement) from the pressure recordings to classify sleep stages, ultimately achieving a 70.3% precision and 71.1% recall on average [6]. Finally, one study conducted in 2016 extracted the heart rate collected sleep data through an under-the-mattress piezoelectric sensor and a smartphone application. Signals collected included respiratory rate, body movement, and calculated sleep-parameter, which were compared simultaneously to data generated by PSG. This system found a linear correlation between total sleep time determined by the sensor and PSG. The overall accuracy for the study was 90.5% [7].

Similar to our project, these studies aimed to monitor sleep through a non-invasive method, specifically pressure sensing devices. However, these studies used healthy subjects while our subjects are older adults with sleep disorders.

2.2. Sleep Stage Classification using Hydraulic Bed Sensors

Previous studies have also attempted to develop a deep or machine learning algorithm to classify sleep stages automatically. We particularly focus on two studies conducted using hydraulic bed sensors in Center for Eldercare and Rehabilitation Technology (CERT).

The first study used both bed sensor data and synchronized PSG data that was obtained from five lab patients with a low Apnea-hypopnea index (AHI). Corresponding PSG data was used as ground-truth. This study also used a transfer learning approach in which sleep posture data was collected from 56 young healthy subjects. A neural network consisting of convolutional, max-pooling, batch normalization, Long-Short Term Memory (LSTM), dropout, rectified linear unit (ReLU), and activation layers was developed, which achieved an accuracy of 95.3% for awake, 84% for REM, 93.1% for NREM [8]. This study is quite similar to our study as it uses the same 5 patients with low AHI and has a deep learning approach for sleep stage classification. However, our study focuses purely on the 5 patients and does not use the transfer learning approach with posture data that this previous study uses.

The second study extracted 74 features (such as heart rate variability) from bed sensor data. Then, Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) were applied to the features, resulting in an accuracy of 85.3% (single layer, SVM) and 83.7 % (single layer, KNN) [9]. This study is different from our study as it uses a Machine Learning approach to sleep stage classification which requires manual selection of features as compared to a deep learning approach which is fully automatic.

2.3. Sleep Stage Classification using PSG Data

As PSG is the traditional method of measuring sleep stages, many studies have attempted to automate sleep stage classification using PSG data. A particular study [10] conducted in 2016 developed a convolutional neural network (CNN) to automatically score sleep stages using Electroencephalography (EEG) data from a publicly available PSG dataset (PhysioNet repository). There was no preprocessing layer, and the model consisted of convolutional, max-pooling, LSTM, and fully-connected layers. It also used regularization methods such as dropout. This model achieved an overall accuracy of 86.2% with a Cohen's Kappa value of 0.80.

There are several similar studies [11] - [15] that have shown the effectiveness of using EEG signals for automatic sleep stages. Studies [11], [12] and [13] used a deep learning approach as their classification method and achieved high accuracies of 86.6%, 88.4%-87.6%, and 92.67% respectively. One common trend in these papers was the use of a CNN and study [11] also used an additional LSTM layer in their model. Another study [14] used features based on REM microstructures for their classification and achieved high accuracy results ranging from 89% to 92.7%. Study [15] uses a deep learning approach analyzing an EEG signal obtained from a smartphone which acts as a wearable EEG sensor and achieved an accuracy of 83.5%.

There are also some studies [16], [17] that have used multiple PSG signals for their deep learning approach. One common point in these studies is that they used CNNs and

LSTMs in their deep learning models. Study [16] used EEG and Electrooculography (EOG) of healthy subjects as well as PSG from patients with clinical suspicion of Obstructive Sleep Apnea (OSA). They achieved an accuracy of 83.7% on the EEG dataset, 83.9% on the EEG+EOG dataset, and accuracy ranging from 75.5% to 84.5% for the OSA dataset based on the severity of OSA. Study [17] used two EOG channels, four EEG channels, and one EMG channel from the PSG of healthy subjects and achieved an F1-Score of 78.9%.

3. Methods

3.1. Hydraulic Bed Sensor

The Hydraulic Bed Sensor [18] can detect BCG movements throughout the subject's body without having to physically attach devices to them. The hydraulic bed sensor shown in Fig. 1, which was developed by CERT, utilizes four hydraulic bed transducers that are placed parallel to the body under the mattress. The transducers were evenly spaced in order to get accurate measurements of the body as the patient rests on the mattress. At the end of each transducer is the pressure sensor that converts the change of pressure applied onto the mattress into data that was sampled at a rate of 100 Hz. The BCG movements that the pressure sensors capture represent the motions of the body generated by the flow of blood in each cardiac cycle.

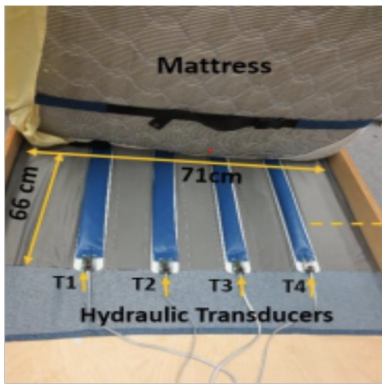

Figure 1. The hydraulic bed sensor utilizing 4 transducers T1-T4 to capture BCG movements

3.2. Data Collection

The data for this study consists of 5 subjects with low apnea-hypopnea index (AHI). Each subject's data includes all 3 sleep stages of Wake, REM, and NREM during the night. Data from the bed sensor and de-identified and synchronized PSG data were collected at the Boone Hospital Center (BHC) sleep center in Columbia, MO. The de-identified PSG data was annotated by a sleep technician at BHC according to the AASM standards and serves as the ground truth for the bed sensor data. The dataset consists of the elderly (66.2 ± 2.4 years) and people with sleep disorders and has a skew towards the NREM stage. This data was collected under the University of Missouri's Institutional Review Board (IRB) approval. Table 1 shows the distribution of the 5 subjects.

TABLE 1. DATA DISTRIBUTION OF 5 SUBJECTS

Subject	Gender / Age	WAKE(%)	REM(%)	NREM(%)
1	F / 69	21.57%	14.13%	64.30%
2	M / 66	14.95%	16.61%	68.44%
3	M / 68	33.76%	12.30%	53.94%
4	F / 66	46.98%	7.32%	45.70%
5	F / 62	29.85%	19.53%	50.63%

3.3. Data Preprocessing

The Bed sensor data was first filtered to remove noise and the respiration component. The Bed sensor data has 8 different signals: 4 raw signals and 4 hardware filtered signals. The hardware filtered signals are the filtered versions of the raw signals. In order to consolidate the data into a single channel signal, the first step was to preprocess the data using transducer selection of the raw signals. We use a 30 second window (1 epoch) and calculate the average amplitude of the raw signals in the window for each transducer. We then select the transducer with the highest average amplitude. Fig. 2 shows how the transducer selection identifies the transducer with the highest average for each epoch of the signal.

Then for each epoch, we select the 3000 hardware filtered data points of the transducer selected. After doing so, a sixth order Butterworth Bandpass filter with a cutoff frequency of 0.7 to 10Hz was implemented on the selected data. Fig. 3 displays the plot of the unfiltered signal versus the filtered signal. Initially, there were gaps in the data that affected the filtering process. These gaps could have arisen if the subject got up from the bed (to use the bathroom, for example), so the bed sensors did not record any data. To account for this issue, we removed the entire epoch if there were any gaps present in it.

4. Architecture

4.1. Design

Our architecture is a Convolutional Neural Network (CNN)- Long-Short Term Memory (LSTM) hybrid network which consists of 2 CNN Models attached to a LSTM Model. This model is inspired from [10], which used a similar model to classify sleep stages from single-channel EEG data.

The CNNs are used to extract features which are then input into the LSTM for classification. The 2 CNNs have different filter sizes to make sure that features on both the

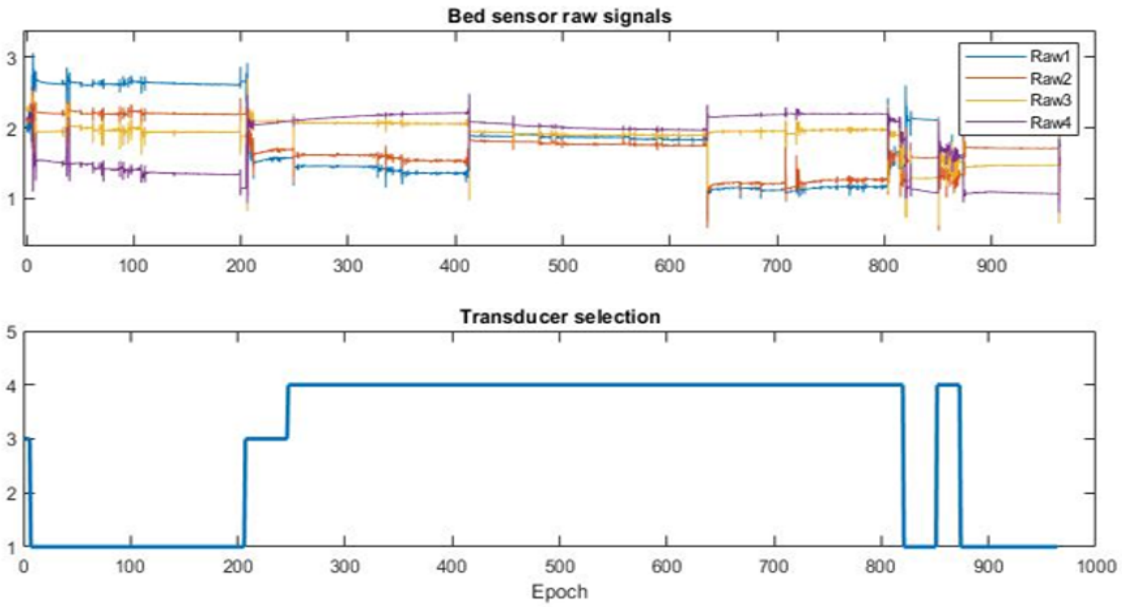

Figure 2. The signals from the 4 transducers Raw1-Raw4 (corresponding to the pressure on each transducer) are compared to each other based on 30 second time windows and the transducer with the highest average value is selected. The changes in pressure can be attributed to position changes by the subject.

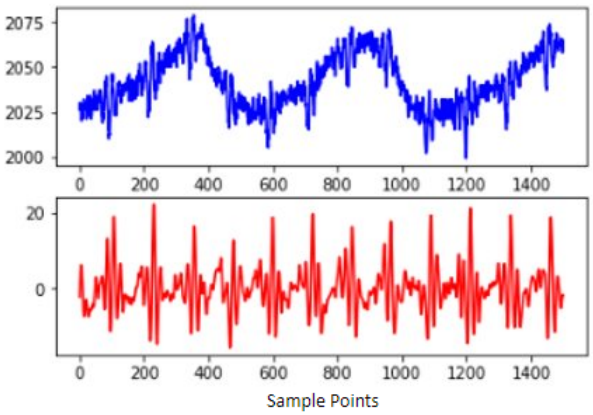

Figure 3. Plot of the unfiltered signal (Blue) versus the filtered signal (Red)

large and small scale are captured. The CNN with a starting filter size of 100 is responsible of extracting the heartbeat features that are present in the BCG, while the CNN with a starting filter size of 400 captures frequency features and general trends. Each CNN model consists of 4 convolution layers, 2 max-pooling layers, and 2 dropout layers. The max pooling layers downsample the inputs. The dropout layers provide a regularization method to help reduce overfitting. The outputs of each CNN are concatenated before sending it to the LSTM.

The LSTM can be trained to learn long term dependencies such as the sleep stage transition rules. The LSTM model contains a single bidirectional LSTM layer along with one linear layer. This layer facilitates learning from both past and new data.

A diagram of the architecture can be seen in Fig. 4. Each convolution layer gives information about the filter size, number of filters, and the stride size. Each max pooling layer shows a pooling size and the stride size. The dropout layers also show the dropout rate. The LSTM layer shows the hidden sizes of forward and backward LSTM.

4.2. Training

We decide our training and testing dataset through 2 methods. In the first method, the training and testing dataset for this experiment was decided by subject. The first 70% of data of each subject was used as training while the rest was used as testing. For our second method, we use the leave one subject out (LOSO) method. In this method, we have 5 combinations in which we use 4 subjects as our training dataset and 1 subject for our testing dataset. The LOSO method allows for a more comprehensive evaluation of the performance of the model in a more real world setting.

The CNNs are pretrained, meaning that they are trained before being connected to the LSTM, on balanced data. The balanced dataset was obtained by undersampling the training data. Both the small CNN and the large CNN are pretrained at the same time as they are put into one model. This is done by having the data pass through the 2 CNNs seperately, concatenating the outputs of the 2 CNNs, and then passing the concatenated output through a linear layer for classification.

The CNNs are then frozen to make sure they continue to extract features based on the class balanced data. This is done by not placing the CNN and LSTM in one model and letting the data pass through the separate CNNs first,

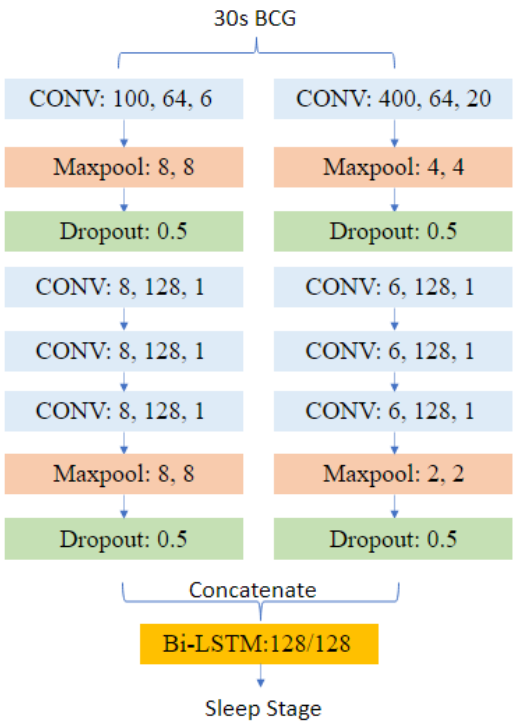

Figure 4. An overview of the network architecture. The layers of each model as well as their parameters. For convolution layer: filter size, number of filters, and the stride size. For maxpooling layer: pooling size and stride size. For LSTM layer: hidden sizes of forward/backward

concatenating their output, and then sending the output into the LSTM for classification. As such, when the CNN is connected to the LSTM only the LSTM is being trained.

After connecting, the LSTM is trained on sequential data to learn the transition rules that sleep experts use to identify the next possible sleep stages from a sequence of epochs. The final output for which class is predicted is done by sending the output of the LSTM layer into a softmax layer.

5. Results

We evaluated the performance of our model using per-class recall (RE), per-class precision (PR), per-class F1-Score (F1), overall accuracy (ACC), overall PR, and overall F1.

Shown in Fig. 5 is a confusion matrix of the testing results for one run of our model where the training and testing dataset were decided using the subject-wise 70-30 split method. The results show an overall accuracy of 75%, an overall precision of 76%, and an overall F1-score of 74%. The per-class metrics for this run are shown in Table 2.

TABLE 2. THE RESULTS BASED ON THE 70-30 SPLIT. SHOWS THE PER-CLASS RECALL, PRECISION, AND F1-SCORE

Class	RE(%)	PR(%)	F1(%)
Wake	53	81	64
REM	72	79	75
NREM	88	71	79

In Table 3, we have the evaluation metrics of the run using 5 runs of our model. The training and testing dataset in these 5 runs were decided using the LOSO method with each run having a different subject being left out. The overall average accuracy ranges from 70.3-78.3% with an average of 75.6%. The recall values for each class average out to be 67.4% for Wake, 62.6% for REM, and 82.2% for NREM. The results are skewed towards the NREM due to the LSTM being trained on the imbalanced dataset, which has NREM as its majority class. Also, the REM class is the minority class in the imbalanced dataset which accounts for its lower accuracy result. The precision scores for each class are fairly consistent with the average scores for the Wake, REM, and NREM classes being 69.2%, 76.6%, and 77.6%.

6. Conclusion and Future Work

In this paper, we proposed an automatic, non-invasive, and more accessible method of classifying sleep stages. We used a deep learning approach, specifically a CNN-LSTM hybrid model, to extract features automatically. We also used a hydraulic bed sensor to provide a non-contact extraction of heartbeat and respiration data, overcoming some of the limitations of PSG extraction. This allows the bed sensor to be non-invasive and be more accessible to people as they can be used in people's houses for longitudinal monitoring.

TABLE 3. THE RESULTS BASED ON THE LOSO METHOD. SHOWS OVERALL ACCURACY, PRECISION, AND F1-SCORE AND PER-CLASS RECALL, PRECISION, AND F1-SCORE

Subject	Overall Metrics(%)			Per-class RE(%)			Per-class PR(%)			Per-class F1(%)		
	ACC	PR	F1	WAKE	REM	NREM	WAKE	REM	NREM	WAKE	REM	NREM
1	**78.3**	**81**	**79**	**84**	65	79	58	75	**91**	69	70	**85**
2	77.5	79	78	58	66	85	47	**88**	84	52	**75**	**85**
3	74.9	77	74	60	42	**92**	**85**	76	71	70	55	80
4	76.8	77	77	74	58	82	82	79	72	**78**	67	77
5	70.3	71	70	58	**82**	73	74	65	70	65	73	72
Average	75.6	77.0	75.6	67.4	62.6	82.2	69.2	76.6	77.6	66.8	68.0	79.8

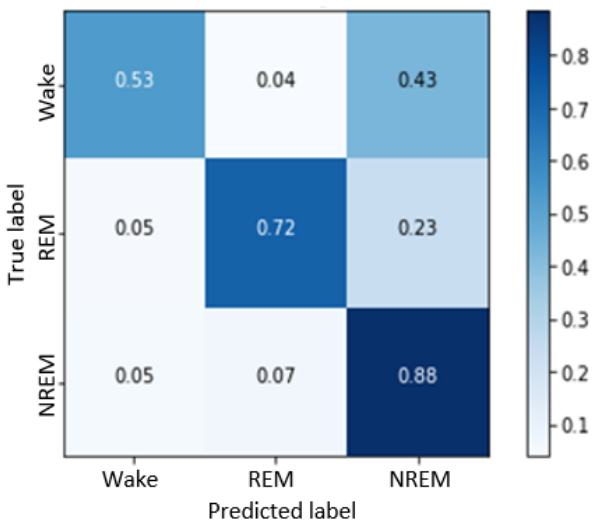

Figure 5. Confusion matrix of the results.

Our results show that our model is able to perform reasonably well in the classification of sleep stages. As such, we believe that our automatic and non-contact approach is an improvement upon the current standard which requires dedicated staff and wearable devices.

Our study achieved a final average accuracy of 75.6% including many novelties that add onto the current state of the art. Most studies mainly use young, healthy adults for their test subjects as well as the gold standard PSG data for sleep stage classification. Also, most of the results of these studies are using K-fold cross validation. In comparison, our study uses a more challenging database, as it contains elderly patients and people with sleep disorder, as well as a non-invasive method of extracting our data which brings additional challenges. We also use the LOSO method which is a more rigorous test.

In the future, we would like to expand the scope of the study. This study was conducted using only 5 subjects and we would like to train our model on a greater number of subjects, expanding the scope of our model. Another topic that we would like to look at is whether or not the respiration component of the bed sensor signal brings useful information for the classification of sleep stages. We also want to look at a transfer learning approach in which we pretrain the CNNs on EEG data.

7. Acknowledgment

This material is based upon work supported by the National Science Foundation under Award Number: CNS-1950873. This work is also supported by the National Institute of Nursing Research of the National Institutes of Health under award number R01NR016423.

References

[1] E. S. Musiek, D. D. Xiong, and D. M. Holtzman, "Sleep, circadian rhythms, and the pathogenesis of Alzheimer disease," Experimental & molecular medicine, vol. 47, no. 3, p. e148, 2015.

[2] A. Videnovic and D. Golombek, "Circadian and sleep disorders in Parkinson's disease," Experimental neurology, vol. 243, pp. 45-56, 2013.

[3] J. M. Parish, "Sleep-related problems in common medical conditions," Chest, vol. 135, no. 2, pp. 563-572, 2009.

[4] L. Giovangrandi, O. T. Inan, R. M. Wiard, M. Etemadi and G. T. A. Kovacs, "Ballistocardiography — A method worth revisiting," 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 4279-4282.

[5] J. M. Kortelainen, M. O. Mendez, A. M. Bianchi, M. Matteucci and S. Cerutti, "Sleep Staging Based on Signals Acquired Through Bed Sensor," in IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 3, pp. 776-785, May 2010.

[6] L. Samy, M. Huang, J. J. Liu, W. Xu and M. Sarrafzadeh, "Unobtrusive Sleep Stage Identification Using a Pressure-Sensitive Bed Sheet," in IEEE Sensors Journal, vol. 14, no. 7, pp. 2092-2101, July 2014.

[7] A. Tal, Z. Shinar, D. Shaki, S. Codish, and A. Goldbart, "Validation of contact-free sleep monitoring device with comparison to polysomnography," Journal of Clinical Sleep Medicine (JCSM), vol. 13, no. 3, pp. 517–522, 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27998378/.

[8] R. Gargees, J. M. Keller, M. Popescu, and M. Skubic, "Non-invasive Classification of Sleep Stages with a Hydraulic Bed Sensor Using Deep Learning," How AI Impacts Urban Living and Public Health Lecture Notes in Computer Science, pp. 73–82, Oct. 2019.

[9] R. Yi, M. Enayati, J. M. Keller, M. Popescu and M. Skubic, "Non-Invasive In-Home Sleep Stage Classification Using a Ballistocardiography Bed Sensor," 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 1-4, 2019.

[10] A. Supratak, H. Dong, C. Wu and Y. Guo, "DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 11, pp. 1998-2008, Nov. 2017.

[11] A. Supratak and Y. Guo, "TinySleepNet: An Efficient Deep Learning Model for Sleep Stage Scoring based on Raw Single-Channel EEG," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 641-644, 2020.

[12] J. Zhang, R. Yao, W. Ge, and J. Gao, "Orthogonal Convolutional Neural Networks for Automatic Sleep Stage Classification based on Single-Channel EEG", Computer Methods and Programs in Biomedicine. 183. 105089, 2019.

[13] E. Fernandez-Blanco, D. Rivero and A. Pazos, "Convolutional neural networks for sleep stage scoring on a two-channel EEG signal", Soft Comput. 24, 4067–4079, 2020.

[14] I. Rechichi, M. Zibetti, L. Borzí, G. Olmo, and L. Lopiano, "Single-channel EEG classification of sleep stages based on REM microstructure". Healthcare Technology Letters. 8, 2021.

[15] A. Koushik, J. Amores and P. Maes, "Real-time Smartphone-based Sleep Staging using 1-Channel EEG," 2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN), pp. 1-4, 2019.

[16] H. Korkalainen, J. Aakko, S. Nikkonen, S. Kainulainen, A. Leino, B. Duce, IO. Afara, S. Myllymaa, J. Toyras, T. Leppanen. "Accurate Deep Learning-Based Sleep Staging in a Clinical Population With Suspected Obstructive Sleep Apnea". IEEE J Biomed Health Inform. 24(7):2073-2081, 2020.

[17] A. Vetek, K. Müller and H. Lindholm, "A Compact Deep Learning Network for Temporal Sleep Stage Classification," 2018 IEEE Life Sciences Conference (LSC), pp. 114-117

[18] D. Heise and M. Skubic, ”Monitoring pulse and respiration with a non-invasive hydraulic bed sensor,” 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 2119-2123, 2010.