

# CSE6250: Sleep Study Project - Sleep Stage Classification With CNN Using Single Channel EEG Signal

Jon Cusick, Liyue Hu, Mikey Ling, Yi Lu  
Georgia Institute of Technology, Atlanta, Georgia

## Abstract

*Tsinalis, et al used one-dimensional data to train convolutional neural networks (CNNs) to analyze and classify electroencephalography (EEG) readings<sup>9</sup>. In hopes of shedding light on alternative methods of implementing the training process employed by Tsinalis, et al, we explored alternative CNN architectures using the same one-dimensional data to classify sleep stages using single-channel EEG signals, then we compared and analyzed our results to those achieved by Tsinalis, et al. Our goal with this study is to promote further investigation on this application of machine learning in the sleep-related healthcare space by producing similar results with a different approach to training the CNN.*

Our project presentation slides are available [here](#).

Our video presentation is available [here](#).

The code is available on Github [here](#).

## Introduction

Sleep is an essential part of healthy living and research shows poor sleep patterns can be directly related to a plethora of known health issues<sup>11</sup>. Medical technologies including EEG and electrocardiography (ECG) are effective measurements for capturing and recording activity within the brain and heart. These data are often used by human experts to manually analyze patients' sleep stages. Unfortunately, sleep stage analysis by human experts requires special training and expertise and thus is very expensive, inefficient, and inaccessible for the everyday person. Recent studies suggest the detection of abnormal sleep and circadian cycles could be an invaluable method of early detection for neurodegenerative diseases such as Alzheimer's and Parkinson's<sup>11</sup>. Luckily, wearable EEGs already exist in the market space, so developing a reliable software suite for analysis is the next necessary step towards successfully bringing reliable and affordable sleep cycle analysis to the masses.

## Related Work

Automatic EEG sleep stage classification has been a much studied task in recent years. The three common steps for classifying sleep stages are pre-processing, feature extraction, and classification. The most relevant work we are considering for this study is Tsinalis, et al, in which raw 1D signals were used without pre-processing as input<sup>9</sup>. This method achieved an overall accuracy of 74%, mean accuracy across individual stages of 82%, and mean  $F_1$ -score of 81%, using 20-fold cross-validation and class-balanced random sampling. Other related works on single-channel EEG classification have been done with a variety of methods including combining statistical features and similarities of complex networks (92% accuracy)<sup>1</sup>, relative wavelet energy with artificial neural networks to analyze epileptic EEG signals (95% accuracy)<sup>3</sup>, a genetic algorithm and a fuzzy classifier to classify spectral features (84.6% accuracy)<sup>4</sup>, visibility graph algorithms and an SVM (87.5% accuracy)<sup>6</sup>, an ensemble SVM-based system (86% accuracy)<sup>5</sup>, and a CNN with time-frequency based images (83-86% accuracy)<sup>10</sup>.

## Method

In this study, our best model was achieved with a CNN over one-dimensional EEG readings. We will describe the data set, CNN architectures and performance metrics and evaluation. We used the same classification standards described by Tsinalis, et al to prove the viability of our approaches. As such, we followed the sleep cycle classification guidelines provided in the American Academy of Sleep Medicine (AASM) manual<sup>2</sup>. Sleep has four stages: Rapid Eye Movement (stage R) and three non-R stages (N1, N2, and N3). An additional Wake stage (stage W) will be added for clarity.

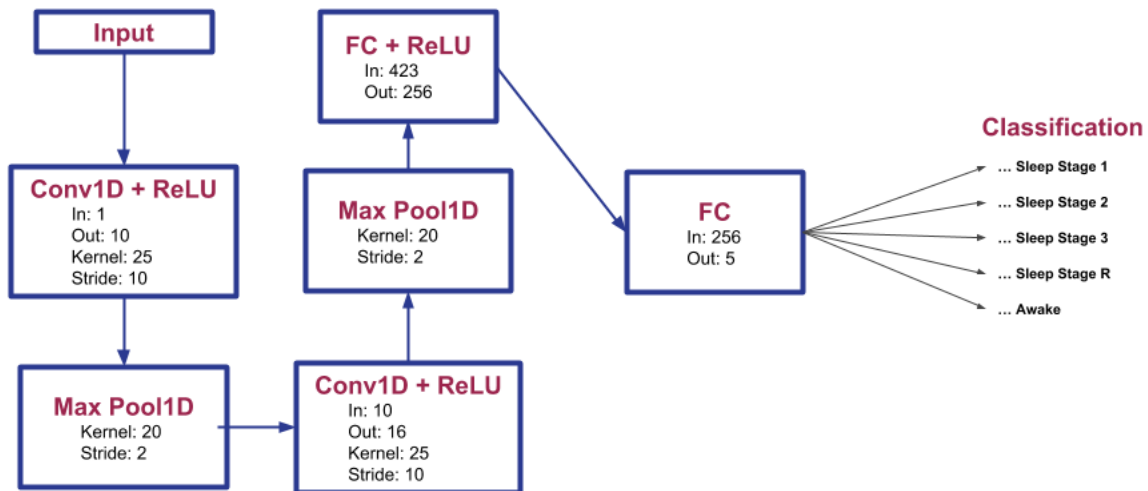
We compared results from our analysis with Tsinalis, et al's study. The success of our model will be measured using the same criteria of Tsinalis, et al: the Rechtschaffen and Kales sleep staging criteria<sup>8</sup>. We will use these criteria to calculate per-stage precision, per-stage sensitivity, per-stage  $F_1$ -score, and overall accuracy. These metrics will

be computed as part of a 20-fold cross validation. For each fold, a single patient will be used for testing, while the remaining 19 patients will have 15 patients randomly selected for training, with the remaining 4 being used for validation.

### CNN Architecture

For our best model, we used raw one dimensional EEG signal without pre-processing as input. Our CNN architecture, as shown in Figure 1, uses two sets of convolutional and pooling layers (C1, P1, C2, P2), one fully-connected layer (F1) and an output layer (F2). Compared to the Tsinalis et al implementation, we chose a model that was much simpler and therefore much faster to train, and we tried to achieve the best results from it. For a summary of the final hyper-parameters for the CNN architecture, please see Table 1.

As shown in Table 3, layer C1 contains 25 filters, so that the output of layer C1 is 25 filtered versions of the original input signal. These filtered signals are then sub-sampled in layer P1. The next layer C2 has 16 units and 25 filters, then it is fed into layer P2. Then, the results are fed into a fully-connected F1 layer, then finally into the fully-connected output layer which makes a prediction for one of the five sleep stages.



**Figure 1: Model Architecture**  
**Table 1: CNN over 1D EEG Architecture**

Layer	Layer Type	Units	Unit Type	Size	Stride	Output Size
Input						(1, 15000)
C1	convolutional	10	ReLU	25	10	(10, 1498)
P1	max-pooling			20	2	(10, 740)
C2	convolutional	16	ReLU	25	10	(16, 72)
P2	max-pooling			20	2	(16, 27)
F1	fully-connected	432	ReLU	256		(256)
Output	fully-connected	256	linear	5		(5)

## Other Attempts

Our original thought behind this paper was to feed the CNN with 2D image (2D array) made from the raw 1D signals. We calculated the min (-208) and max (209) EEG value across all \*PSG.edf files. As part of the pre-processing step, we increased each EEG value by 208 and converted the result to the nearest integer as the new value. For example, an EEG value of -1.1 would be pre-processed to 207. Therefore, each input to a CNN now consists of 15,000 non-negative integers. Next, we initiate a 417 by 15,000 2D array of zeros. 417 is the absolute difference between the min (-208) and the max (209) EEG values. For each of the 15,000 non-negative integers, we fill in the  $i$ -th row and  $j$ -th column of the 417 by 15,000 2D array with the value of 1, where  $i$  is the value of the non-negative integer and  $j$  is the index of the non-negative integer in the 15,000 non-negative integer array. The resulting 417 by 15,000 2D array (2D image) would then be the final input used to train a CNN with the corresponding sleeping stage label. After removing *Sleep Stage Movement* and *Sleep stage ?* labels, we have a total of 106,414 input-label training data.

We attempted to represent input data for a single patient as a two-dimensional array of 417 by 15,000. We then combined multiple sets of patients' nightly data into a three-dimensional matrix for input into the model. If a set of input data has 30,000 417 by 15,000 input data, then the three-dimensional matrix is 30,000 by 417 by 15,000. Despite our effort in achieving performant training process by using sparse matrix (Pytorch sparse API) to represent 2D input data, we ran into significant memory issue in training a neural network that's representative of Tsinalis, et al's architecture on a computer-optimized Linux machine with 144 GiB memory.

## Data

The EEG data used in this study was obtained from [physionet.org](http://physionet.org)<sup>7</sup>. The entire dataset consists of 20 healthy subjects, 10 male and 10 female, aged 25-34 years. There are two approximately 20-hour recordings per subject, apart from a single subject for whom there is only a single recording. Therefore, we retrieved 39 ( $2 \cdot 20 - 1$ ) \*PSG.edf files that record EEG value at 100 Hz and their corresponding 39 \*Hypnogram.edf files that provide sleeping stages (labels) of *Sleep stage 1*, *Sleep stage 2*, *Sleep stage 3*, *Sleep stage 4*, *Sleep stage R*, *Sleep stage W*, or *Sleep stage ?*. Labels are manually assigned by experts. Following guidelines given by the American Academy of Sleep Medicine, we combined *Sleep stage 4* into *Sleep stage 3*. We also removed the small amount of data for *Sleep stage Movement* and *Sleep stage ?*.

We analyzed the distribution of different sleep stages in each patient, as shown below in Figure 2.

We noticed that the data is heavily unbalanced, with the majority of epochs in the "W" stage. A heavily skewed data have the possibility of impacting model performance by favoring the most frequently occurred stage. To improve the prediction capability of the CNN, we decided to implement class-balanced random selection in our model. After playing around with the weights of the different stages, we were able to achieve best results by decreasing the weight of the "W" stage to that of the second frequently occurring stage, as shown in Figure 3. We first tried a truly balanced weighting and achieved worse results than we did from the final weighting.

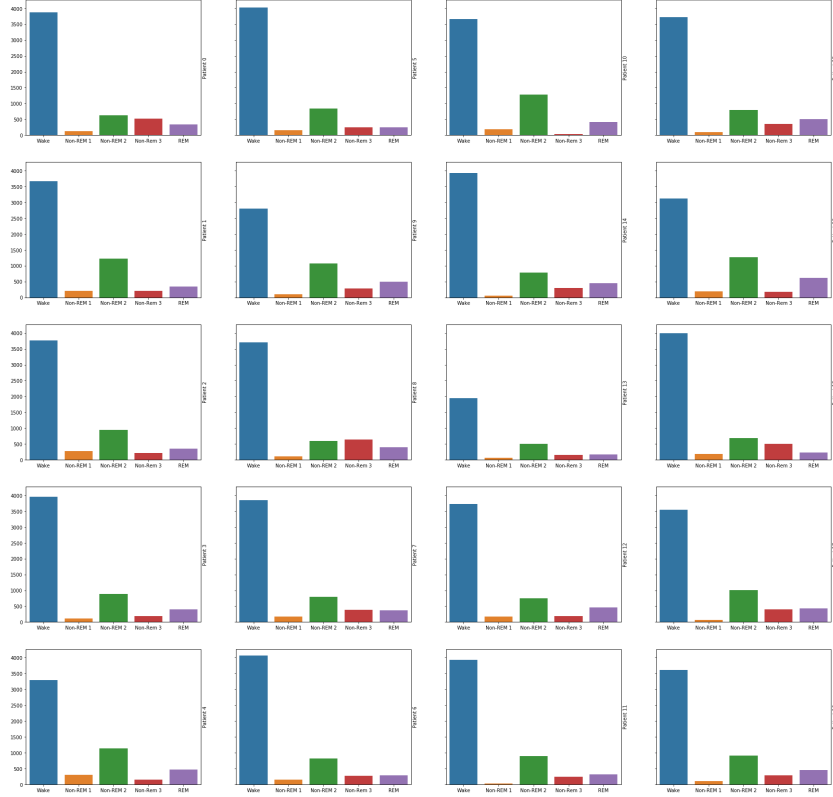
The original paper trains a CNN that classifies a 30-second EEG recording (epoch) into sleeping stages of *Sleep stage 1*, *Sleep stage 2*, *Sleep stage 3*, or *Sleep stage R*. In the paper, an epoch is represented by a 1D array of 3,000 ( $100\text{hz} \cdot 30$ ) EEG values. To account for the relationships and sequence patterns among various sleeping stages, each training input consists of 5 epochs: 2 proceeding epochs, the current epoch, and 2 succeeding epochs. Therefore, in the paper, the input to a CNN consists of 5 epochs (1D array of 15,000 EEG values). In our first model, we used the same input as described in the original paper.

Patient records were chosen at random for the test, train and validate data; specifically, we used 15 patient records for the training set, 1 patient records for the testing set and 4 patient records for the validation set.

## Result

We computed the confusion matrix for per-stage accuracy, as shown below in Figure 4.

We achieved best performance in classifying the Wake stage (93%), and had the worst accuracy for classifying Non-Rem 1 Stage (4%). Overall, the total accuracy across all five stages is 85%. This is higher than the (74%) overall



**Figure 2: Sleep Stages by Patient**

accuracy achieved by Tsinalis' study. However, our model performance are less even across the different sleep stages.

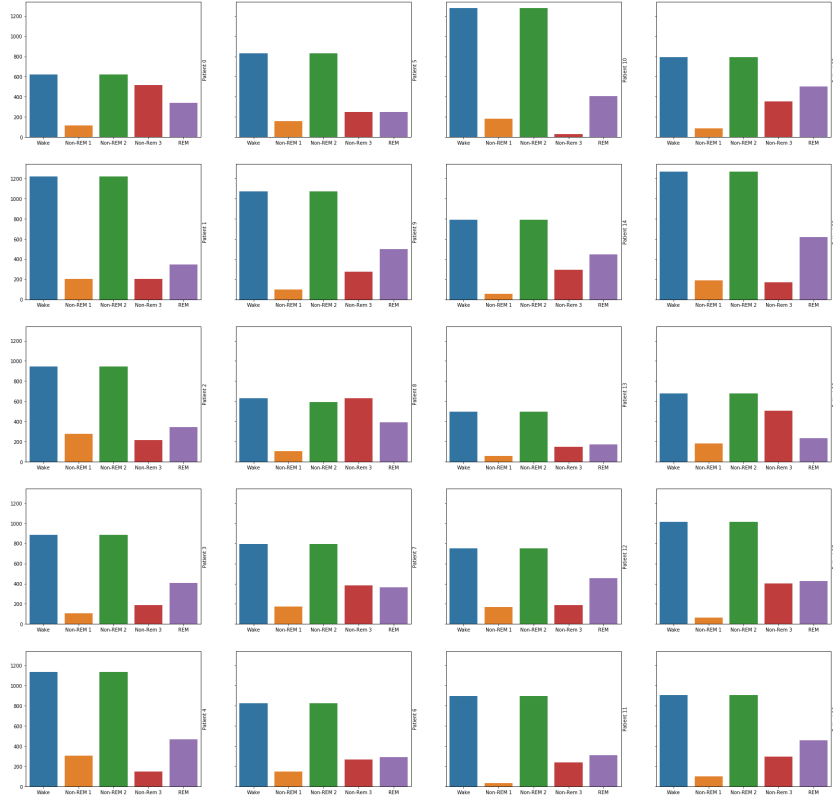
To evaluate our models, we computed per stage performance metrics, as shown in Table 2.

**Table 2: Model Evaluation**

Stage	Precision	Sensitivity	$F_1$ -score
Non-REM 1	17	4	7
Non-REM 2	82	85	77
Non-REM 3	55	85	62
REM	44	67	53
Wake	98	93	95

## Analysis

Non-REM 1 stage had the worst performance, likely due to the fact that it had so few occurrences. This caused our data set to be quite small and not have enough time during training to learn the full set of sleep stage patterns. Wake stage had the best performance, since brain activity is likely very different when awake as compared to any category of sleep. When we first implemented the CNN architecture without the class-balancing, we achieved (54%) accuracy. We were surprised by the effect of class-balance random sampling. Additionally, even though we achieved higher overall accuracy than Tsinalis' architecture, our model is by no means superior. In the real life, it is important to understand the underlying issue that we trying to solve. We get the sense that Tsinalis et al believed that it is more important to achieve higher mean accuracy measures over the different sleep stages as well as improving the worst stage performance, as seen in their focus on calculating the mean and worst performance metrics. If that is indeed the case, then our model would not be a good predictor for sleep classification. However, our goal was to find a simple

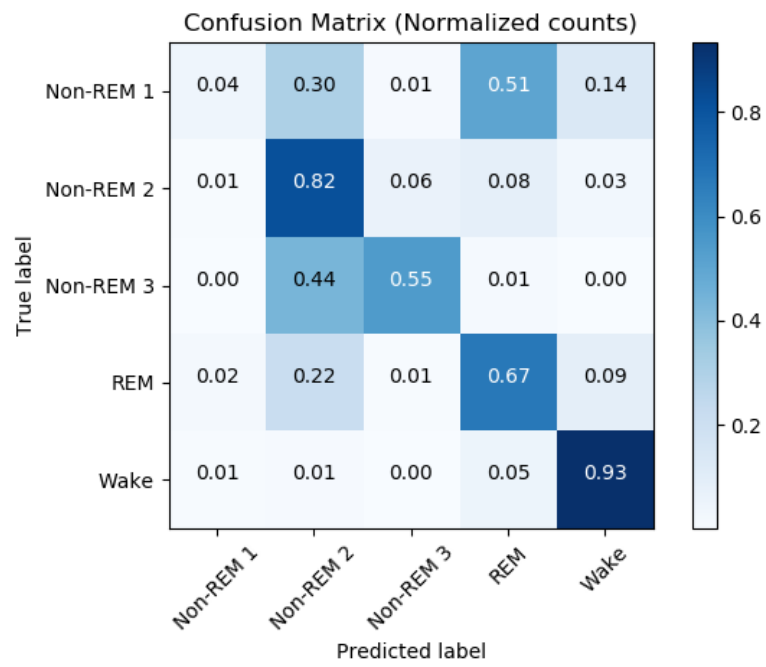


**Figure 3: Sleep Stages by Patient - Class-Balanced**

model with high overall accuracy, and we were able to achieve our goal. The other benefit of our model is that it is a lot simpler and therefore require less computing power than Tsinalis' model. We believe that our results provide a good starting point for future projects on this topic.

## Conclusion

The ultimate goal of this project is to further explore the possibility of making accurate and reliable sleep stage analysis more accessible for everyone. Today, wearable technology has already begun to make state of the art healthcare more accessible. For example, the iWatch can take EKG measurements and send them to a patient's cardiologist for further inspection. Creating a software package that can provide in-house, expert-grade sleep stage analysis would be a monumental step towards our ultimate goal of providing invaluable healthcare services to those who wouldn't be able to get it otherwise. Technology is in the midst of completely changing the way the world looks at healthcare and we would love for our work to contribute to this cause.



**Figure 4:** Per-stage Confusion Matrix

## References

1. Diykh, Mohammed and Yan Li. 2016. Complex networks approach for EEG signal sleep stages classification. *Expert Syst. Appl.* 63, C (November 2016), 241-248. DOI: <https://doi.org/10.1016/j.eswa.2016.07.004>
2. Iber, C., S. Ancoli-Israel, A. Chesson, and S. F. Quan. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Westchester, IL: American Academy of Sleep Medicine, 2007.
3. Guo, L., Rivero, D., Seoane, J.A., and Pazos, A. (2009). Classification of EEG signals using relative wavelet energy and artificial neural networks. *Proceeding of the first ACM/SIGEVO summit on genetic and evolutionary computation* (pp. 177-184). ACM.
4. Jo, H. G., Park, J. Y., Lee, C. K., An, S. K., and Yoo, S. K. (2010). Genetic fuzzy classifier for sleep stage identification. *Computers in Biology and Medicine*, 40, 629-634.
5. Koley, B., and Dey, D. (2012). An ensemble system for automatic sleep stage classification using single channel EEG signal. *Computers in biology and medicine*, 42 12, 1186-95.
6. Liu, C., Zhou, W. X., and Yuan, W. K. (2010). Statistical properties of visibility graph of energy dissipation rates in three-dimensional full developed turbulence. *Physica A: Statistical Mechanics and its Applications*, 389, 2675-2681.
7. PhysioNet: The Sleep-EDF database [Expanded]: <http://www.physionet.org/physiobank/database/sleep-edfx/> (Accessed October 2018).
8. Rechtschaffen, A., and A. Kales. (eds.). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. Washington, DC: Public Health Service, U.S. Government Printing Office, 1968.
9. O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou. Automatic sleep stage scoring with single-channel eeg using convolutional neural networks. *arXiv preprint arXiv:1610.01683*, 2016.
10. Vilamala, A., Madsen, K. H., and Hansen, L. K. (2017). Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring. *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, 1-6.
11. Wulff, K., S. Gatti, J. G. Wettstein, and R. G. Foster. Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nat. Rev. Neurosci.* 11(8): 589-599, 2010.
12. Zhang, H., McLoughlin, I. and Song, Y., 2015, April. Robust sound event recognition using convolutional neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (pp. 559-563). IEEE.