Simple and Comparable Sleep Stage Scoring using CNN.

Yifu Yuan, Ali Yousefi, and Abigail Albuquerque

Abstract

Recent advancements in neural networks have allowed us to model complex data and achieve accuracy levels higher than we have ever seen before. However, now a common occurrence noticed in the field is the increase in complexity of models, where there is minimal or no improvement in accuracy as compared to a simpler model. The complex models take more computing power and time to give us a prediction, often being impractical in daily use. This paper proposes a sleep stage scoring model, that uses a simple architecture that can reach comparable accuracy with several complex models.

Introduction

Recent advancements in machine learning, specifically in Convolutional Neural Networks, have allowed for reaching higher levels of accuracy in the modeling of data. However, the question that arises is whether a high level of complexity in our models is necessary, or if they are solely extra, unnecessary steps in a model whose similar accuracy can be achieved with a model which is much simpler. In this paper, we are comparing the efficiency and accuracy of Sleep Stage Scoring models in different papers, including "TinySleepNet", "DeepSleepNet" and "AttnSleep" as opposed to our model. We are proposing that a simpler model, using preprocessing methods can allow us to reach comparable levels of accuracy.

Sleep is an important necessity, vital for the health and well-being of many living beings, including humans. Healthy sleep is important for a balanced mood, mental health, cognitive functioning, cardiovascular, cerebrovascular, and metabolic health.[1] Sleep is a complex and dynamic process. Many biological processes occur during sleep. It allows the brain to store new information and get rid of toxic waste. It also helps nerve cells communicate and reorganize and allows the body to repair cells, restore energy, and release molecules like hormones and proteins.[2] The brain thus goes through different stages during sleep to help the brain recuperate and develop.[3]

In this paper, we are comparing models that are trying to predict sleep stages. Classifying sleep stages are important in the medical world to allow for the diagnosis of several physical and

[1] Watson NF, Badr MS, Belenky G, et al. Joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: methodology and discussion. *J Clin Sleep Med*. 2015;11(8):931–952.
[2] *Why Do We Sleep?* (2020, July 20). Healthline. https://www.healthline.com/health/why-do-we-sleep#why-do-we-sleep
[3] Suni, E. (2021, December 2). *Stages of sleep: What happens in a sleep cycle* (N. Vyas, Ed.). Sleep Foundation. https://www.sleepfoundation.org/stages-of-sleep

mental ailments. Abnormal sleep patterns and brain waves are observed in patients with ailments and sleep quality and time spent in each sleep stage may become altered by factors like aging, medications, brain injuries, depression, and circadian rhythm disorders.[4] For example, patients with sleep apnea experience reduced times in stage 3 and REM sleep because they experience airway collapse in deeper sleep states. Since they do not get any deep sleep, they experience an array of problems that come with the lack of sleep like lack of concentration, fatigue, obesity, and depression. Similarly, studies have demonstrated that patients with narcolepsy tend to skip the initial phases of sleep and go directly into REM sleep.

Sleep stage scoring currently relies heavily on the recognition of patterns by an expert in the field, which requires tedious amounts of work as well as money, and is prone to human error.[5] The motivation for building a sleep stage scoring model is that having a model that is efficient and simple, with a decent level of accuracy that allows for the removal of the skilled expert in the field, as well as the time it takes to classify each cycle allowing for overall efficiency in time and cost.

Sleep Stages

In a regular sleep cycle, there are 2 phases of sleep, namely REM (Rapid Eye Movement) sleep and NREM (Non-Rapid Eye Movement) sleep. The human body moves through these different phases of sleep in no specific order, often jumping from one stage to another abruptly. The body cycles through all these stages 4-6 times each night, for around 90 minutes each cycle.[6]

NREM consists of 3 stages - N1, N2, and N3, with each stage being a progressively deeper sleep. 75% of sleep is spent in the NREM stage.[4] The REM stage and the wake stage, together with the 3 NREM stages make the 5 stages of sleep that are being studied and classified in this paper. In-depth descriptions of each of these stages and their corresponding physical and brain activities are provided below.

[4] Patel, A. K., & Araujo, J. F. (2018b, October 27). *Physiology, Sleep Stages*. Nih.gov; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK526132/

[5] Sharma, M.; Tiwari, J.; Acharya, U.R. Automatic Sleep-Stage Scoring in Healthy and Sleep Disorder Patients Using Optimal Wavelet Filter Bank Technique with EEG Signals. *Int. J. Environ. Res. Public Health* 2021, *18*, 3087. https://doi.org/10.3390/ijerph18063087

[6] Patel, A. K., & Araujo, J. F. (2018, October 27). *Physiology, Sleep Stages*. Nih.gov; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK526132/

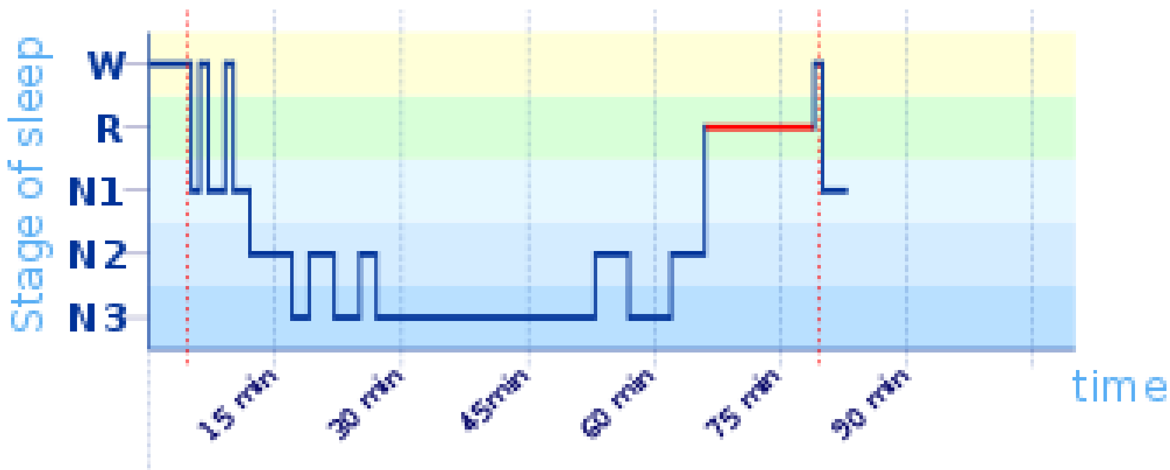

Figure 1: This hypnogram shows the typical pattern of a sleep cycle in a healthy adult.[7] The time spent in REM is indicated by the red band. Sleep moves through different stages, sometimes abruptly moving from one stage to another. Each cycle lasts about 90 minutes, and the body cycles through all 5 stages 4-6 times during a regular night's sleep.

Wake Stage or Stage W

The wake stage is also called the alert stage. During this stage, there is normal physiological activity in a healthy human, including a regular heart rate, brain activity, and the body is fully alert. The brain activity during this stage depends further on whether the eyes are open or closed. During eye-open wakefulness, beta waves predominate. Beta waves are high-frequency, low-amplitude brain waves. During quiet or relaxed wakefulness, or when individuals become drowsy and close their eyes, alpha waves become predominant.[8]

N1 or Stage 1

N1 is the lightest sleep stage that occurs when a person first falls asleep. During this stage, the body has not fully relaxed, and the brain activity is just starting to slow down. This stage usually lasts one to seven minutes and makes up about 5% of a regular night's sleep.[9] Stage 1 sleep is associated with both alpha and theta waves. Early N1 sleep produces alpha waves, which are low-frequency and high-amplitude electrical activities in the brain, which become synchronized in terms of their frequency and amplitude.[10]

---

[7] Wikipedia Contributors. (2019, November 13). Sleep cycle. Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Sleep_cycle

[8] Varga B, Gergely A, Galambos Á, Kis A. Heart Rate and Heart Rate Variability during Sleep in Family Dogs (*Canis familiaris*). Moderate Effect of Pre-Sleep Emotions. Animals (Basel). 2018 Jul 02;8(7)

[9] Suni, E. (2021, December 2). *Stages of sleep: What happens in a sleep cycle* (N. Vyas, Ed.). Sleep Foundation. https://www.sleepfoundation.org/stages-of-sleep

[10] *Stages of Sleep – Introductory Psychology*. (n.d.). Opentext.wsu.edu. https://opentext.wsu.edu/psych105/chapter/stages-of-sleep/

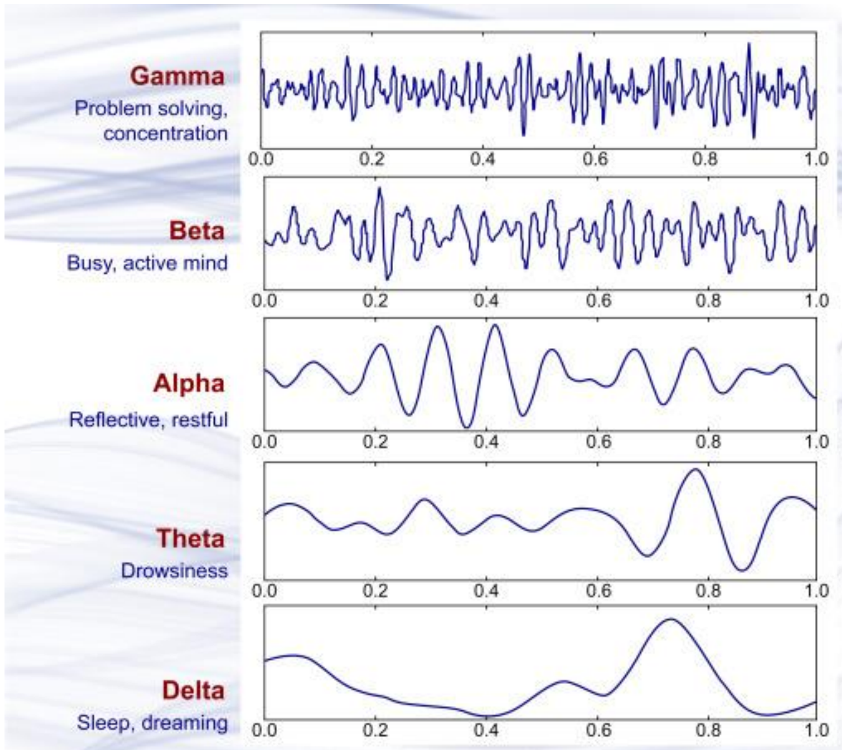

Figure 2: The image shows the brain wave samples with dominant frequencies belonging to gamma waves and beta, alpha, theta, and delta bands. [11]

N2 or Stage 2

During the N2 stage, the body is in a more subdued state. Physical effects noticed during this stage include the slowing down of breathing and heart rate, a drop in body temperature, and muscle relaxation.[6] The N2 stage is observed for the longest period of time during sleep, noticed during about 45% of sleep each night. Memories are formed, and learning takes place during the N2 stage, which is essential for cognition and other brain activity.[12] Brain activity slows down during this stage of sleep, and the body prepares to enter deep slow-wave sleep. The N2 stage involves theta waves, which are slower brain waves that have a frequency of about 5 to 8 Hz.[13]

Brain activity during this stage is characterized by sleep spindles, K-complexes, and often both. Sleep spindles are brief, powerful bursts of activity in the brain, that show a higher frequency of brain waves and are considered to be important for learning and memory. K-Complexes are

[11] Abhang, P. A., Gawali, B. W., & Mehrotra, S. C. (2016, January 1). *Chapter 2 - Technological Basics of EEG Recording and Operation of Apparatus* (P. A. Abhang, B. W. Gawali, & S. C. Mehrotra, Eds.). ScienceDirect; Academic Press. https://www.sciencedirect.com/science/article/abs/pii/B9780128044902000026

[12] Poe, G. R., Walsh, C. M., & Bjorness, T. E. (2010). Cognitive neuroscience of sleep. *Progress in Brain Research*, *185*, 1–19. https://doi.org/10.1016/B978-0-444-53702-7.00001-4

[13] *N2 Sleep: K Complexes and Sleep Spindles | Sleepopolis. (n.d.). Retrieved April 27, 2023, from https://sleepopolis.com/education/n2-sleep/#:~:text=Brain%20waves%20slow%20down%20during*

short periods of high amplitude (delta waves) brain activity that may occur in response to environmental stimuli.[8] They work in suppressing arousal in response to stimuli that the sleeping brain evaluates not to signal danger. In other words, if the brain determines that a stimulus is not important or dangerous, the K-complexes work towards keeping the body in a sleeping state. They also help in sleep-based memory consolidation.[14]

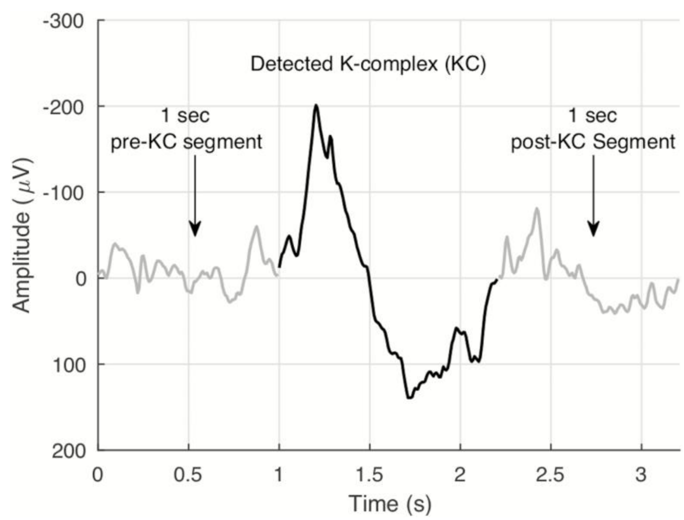

Figure 3: An example K-complex and the corresponding pre- and post-K-complex segments. [15]

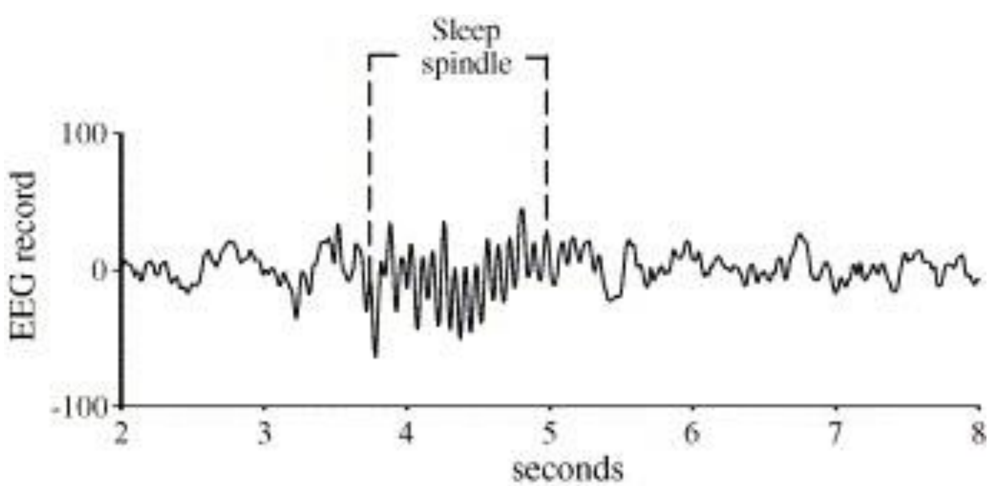

Figure 4: The figure shows a sleep spindle in an EEG recording across the corresponding time domain. [16]

---

[14] Gandhi, M. H., & Emmady, P. D. (2021). *Physiology, K Complex*. PubMed; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK557469/

[15] Parekh A, Mullins AE, Kam K, Varga AW, Rapoport DM, Ayappa I. Slow-wave activity surrounding stage N2 K-complexes and daytime function measured by psychomotor vigilance test in obstructive sleep apnea. Sleep. 2019 Mar 1;42(3):zsy256. doi: 10.1093/sleep/zsy256. PMID: 30561750; PMCID: PMC6424089.

[16] Ventouras, E. M., Monoyiou, E. A., Ktonas, P. Y., Paparrigopoulos, T., Dikeos, D. G., Uzunoglu, N. K., & Soldatos, C. R. (2005). Sleep spindle detection using artificial neural networks trained with filtered time-domain EEG: A feasibility study. *Computer Methods and Programs in Biomedicine*, *78*(3), 191–207. https://doi.org/10.1016/j.cmpb.2005.02.006

N3 or Stage 3

N3 is also known as slow-wave sleep because of the distinct pattern of synchronized waves observed in this stage. The brain activity during this period has an identifiable pattern of delta waves.[4] Delta waves have a much lower frequency and high amplitudes.[9] The first part of the delta wave before the peak is considered a 'down' state, during which the neurons in the neocortex of the brain rest and don't display much activity. The second half represents an 'up' state during which the neurons fire quickly and at a high rate. The high voltage delta waves are interspersed surges of what are called 'sharp' waves, which are said to be closely connected to memory processing. [4]

The N3 Stage is the deepest non-REM sleep stage, and it is the most difficult stage to be awoken from. The body relaxes even further during this stage, with muscle tone, pulse, and breathing rate decreasing further. It is believed that this stage of sleep is important for bodily recovery and growth.[3] This is also the stage where night terrors, bed wetting, and sleepwalking occur.

Rapid Eye Movement Stage (REM)

As the name suggests, the REM stage is marked by Rapid Movements of the Eyes. Dreaming is considered to occur during the REM stage, and therefore REM is not considered a restful sleep stage. It makes up the last 25% of a night's sleep. REM is associated with the paralysis of muscles, increased oxygen use in the brain, and increased irregular pulse and blood pressure.[9] Researchers hypothesize the paralysis of muscles during REM to be a safety measure to prevent a human from acting out their dreams and injuring themselves.[17] The brain waves observed during this stage are similar to the ones observed when a person is awake. There are mixed frequencies during this stage, including frequencies in the theta, beta, and gamma ranges.[18]

Methodology and Results

Dataset and Preprocessing

The dataset used to train our model is called Sleep-EDF-20 whose data comes from polysomnography (PSG) recordings. PSG monitors body functions like brain activity (EEG; electroencephalogram), eye movements (EOG; electrooculogram), muscle activity (EMG; electromyogram), and heart rhythm (ECG; electrocardiogram).

---

[17] Summer, J. (2021, December 16). *REM Sleep: What It Is and Why It Matters*. Sleep Foundation. https://www.sleepfoundation.org/stages-of-sleep/rem-sleep

[18] Cowdin, N., Kobayashi, I., & Mellman, T. A. (2014). Theta frequency activity during rapid eye movement (REM) sleep is greater in people with resilience versus PTSD. Experimental Brain Research, 232(5), 1479–1485. https://doi.org/10.1007/s00221-014-3857-5

The data includes 39 PSG recordings from 20 healthy subjects. All the EEG recordings in the PSG had a sampling rate of 100 Hz. These recordings were manually classified by sleep experts into one of 8 classes, which were (W, N1, N2, N3, N4, REM, MOVEMENT, UNKNOWN). The 'MOVEMENT' and 'UNKNOWN' periods were excluded as they did not belong to the five sleep stages. [21] Data was classified into an epoch, which is a 20-30 second period of each data point in the collection. For our model, we only use the EEG data from the PSG. There are a total of about 40,000 epochs in the dataset, with the highest number of epochs being classified as 'N2', following the trend seen in human sleep cycles. DC noise was removed from the data by taking the average of the signal values and then removing the mean from the raw data, then dividing the resulting values by the square root of itself.

Approach 1

Our first approach to sleep stage classification was to convert the raw EEGs into spectrograms for image classification. The feature extraction was performed using MATLAB and its toolbox called 'chronux'. With continuous brain monitoring becoming a more routine part of clinical investigation, many medical centers have adopted a quantitative EEG, in the form of a spectrogram to improve efficiency in identification and treatment as it is easier to translate and understand a spectrogram. Spectrograms are also known as color spectral arrays or color density spectral ways. The spectrogram is displayed as a three-way plot of time on the x-axis, frequency on the y-axis, and the intensity of the spectral power at that particular time and frequency as color, with brighter color indicating higher power levels.[19] Different durations for the sliding window and sampling rates were experimented with which affect the time and frequency respectively. The parameters for the spectrogram chosen for our model included a sampling rate of 100, with a frequency range for displaying the spectrogram between 0 and 30.

Spectrograms are able to compress EEG data, displaying long periods of data on a single screen. This enables us to reduce the time to train our model as well as decrease the complexity of our model significantly.

---

[19] Ng MC, Jing J, Westover MB. A Primer on EEG Spectrograms. J Clin Neurophysiol. 2022 Mar 1;39(3):177-183. doi: 10.1097/WNP.0000000000000736. PMID: 34510095; PMCID: PMC8901534.

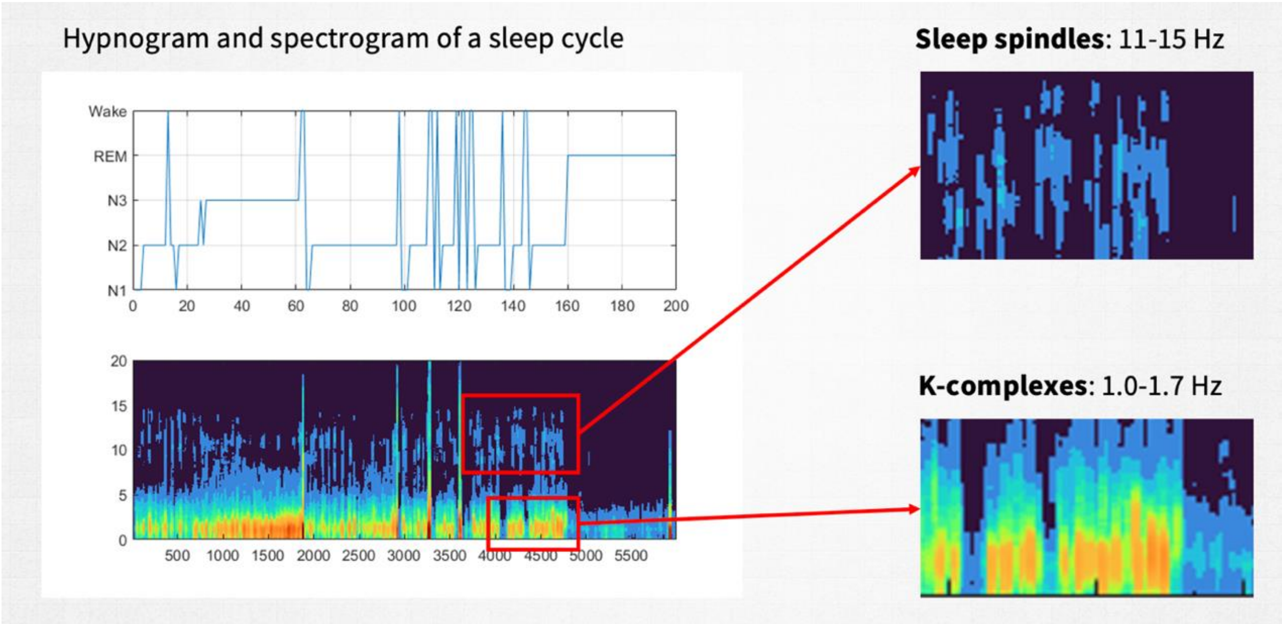

Figure 6: The figure above shows how different stages and phenomena in signals during stages of sleep can be captured in a spectrogram, and potentially be identified by a model.

Two different algorithms were used in this approach – Random Forests, as well as Convolutional Neural Networks (CNN). Table 1 shows the accuracy and outputs for different training and methods. As we can observe from Table 1 which shows the accuracy for different models in the first approach, the CNN-slicer model on a balanced training dataset through oversampling produces the best model in terms of both overall accuracy and the F1 score. The CNN slicer model takes 5 slices of each epoch and each slice lasts 8 seconds. There is a 3 second overlap between each slice. It then merges the different slices and gives the majority stage.

Training Dataset	Method	Overall Accuracy	F1(weighted)
Unbalanced training dataset (18 subjects, 38320 epochs)	CNN	0.737	0.715
	Random Forest	0.707	0.675
Balanced training dataset-under sampling (18 subjects, 11500 epochs, 2300 for each class)	CNN	0.611	0.633
Balanced training dataset-oversampling	CNN	0.73	0.742

(18 subjects, 75000 epochs, 15000 for each class)	CNN-slicer	**0.742 (0.717)**	**0.746 (0.721)**
	Random Forest	0.718	0.695

Table 1: The table shows the overall accuracy for the different methods used in approach 1, where we performed image classification. The testing dataset consisted of two novel subjects, totaling 4693 epochs.

Approach 2

Our second approach was to convert the raw EEG data into numerical spectrogram data and build models using CNN. The raw EEG signals were converted into a 2D matrix, with the dimension 138 by 28. The y-axis of 138 variables contained data about the frequency. The x-axis with 28 variables was the time interval, with each of the 28 instances being one second of data. The moving window size used for this task was of about 30 seconds, however since the ends were tapered off by a second each, the x-axis goes down to 28. In this model, the frequency range was limited to 3-30 Hz. Then numerical data on the size of the spectrogram was extracted from the raw data. The input to the CNN is the 2D matrix described above.

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 138, 28, 16)       160

 conv2d_1 (Conv2D)           (None, 138, 28, 32)       4640

 conv2d_2 (Conv2D)           (None, 138, 28, 32)       9248

 max_pooling2d (MaxPooling2D  (None, 69, 14, 32)       0
 )

 conv2d_3 (Conv2D)           (None, 67, 12, 32)        9248

 conv2d_4 (Conv2D)           (None, 65, 10, 64)        18496

 conv2d_5 (Conv2D)           (None, 63, 8, 64)         36928

 max_pooling2d_1 (MaxPooling  (None, 31, 4, 64)        0
 2D)

 conv2d_6 (Conv2D)           (None, 29, 2, 128)        73856

 max_pooling2d_2 (MaxPooling  (None, 14, 1, 128)       0
 2D)

 flatten (Flatten)           (None, 1792)              0

 dropout (Dropout)           (None, 1792)              0

 dense (Dense)               (None, 5)                 8965

=================================================================
```

Figure 7: The figure shows a summary from TensorFlow of the architecture of the CNN model used in the numerical classification. The model has 7 convolutional layers and 3 max pooling layers.

The architecture of the CNN used is fairly simple, and its summary can be seen in Figure 7. The CNN has a total of 7 convolutional layers, along with 3 max-pooling layers, and 1 drop-out layer. As we can see from Table 2, the CNN model with an unbalanced training dataset seems to give the best accuracy in terms of both the overall accuracy and the F1 score across both approaches.

Training Dataset	Method	Overall Accuracy	F1(weighted)
Unbalanced training dataset_Sc (18 subjects, 39527 epochs)	CNN	**0.79**	**0.772**
	CNN-slicer	0.76 (0.728)	0.736 (0.707)
Balanced training dataset_Sc (18 subjects, 79270 epochs)	CNN	0.72	0.729
	CNN-slicer	N/A	N/A

Table 2: The table shows the overall accuracy for the different methods used in approach 2, where we performed numerical classification using CNNs. The testing dataset consisted of two novel subjects, totaling 4693 epochs.

Comparing our approaches, we see that approach 2 using numerical data worked better than image classification and gave us a higher accuracy for its best model by 5%. Further, we investigated how these different methods were predicting different classes, and if the models were doing better in some classes rather than others.

As we can see from Tables 3 and 4, different models are able to classify certain classes better than others. In most models, the N2 stage is classified as the best and N1 as the worst. This is because, in a normal sleep cycle, the N2 stage is observed for about 50% of a night's sleep, however, N1 is observed the least making up about 5% of a regular night's sleep. This imbalance in nature is translated into our models and makes the precision during validation worse for some classes as compared to others.

Training Dataset	Method	Precision per-class				
		W	N1	N2	N3	REM
Unbalanced training dataset (18 subjects, 39527 epochs)	CNN	**0.707**	0.207	0.83	**0.941**	0.549
	Random Forest	0.594	**0.5**	0.784	0.936	0.594
Balanced training dataset (18 subjects, 79270 epochs)	CNN	0.603	0.191	**0.937**	0.67	0.554
Balanced training dataset-oversampling (18 subjects, 75000 epochs, 15000 for each class)	CNN	0.68	0.278	0.896	0.86	0.65
	CNN-slicer	0.676	0.328	0.912	0.87	0.582
	Random Forest	0.611	0.4	0.841	0.883	0.559

Table 3: The table shows the precision for the 5 different classes or stages using the image classification model. The testing dataset consisted of two novel subjects, totaling 4693 epochs. As we can see, different models classify different classes better than others.

Training Dataset	Method	Precision per-class				
		W	N1	N2	N3	REM
Unbalanced training dataset_Sc	CNN	**0.702**	0.291	0.885	**0.883**	**0.691**

(18 subjects, 39527 epochs)	CNN-slicer	0.666	**0.5**	0.909	0.831	0.586
Balanced training dataset_Sc (18 subjects, 79270 epochs)	CNN	0.597	0.239	**0.949**	0.773	0.668

Table 4: The table shows the precision for the 5 different classes or stages using the numerical classification model. The testing dataset consisted of two novel subjects, totaling 4693 epochs. As we can see, different models classify different classes better than others.

Other Deep-Learning-Based Sleep Stage Classification Models

<u>"TinySleepNet"</u>

The "TinySleepNet" model consists of a CNN with four convolutional layers, interleaved with two max-pooling and two dropout layers. It is then followed by an RNN consisting of a single LSTM layer followed by a dropout layer.[20] The model takes as an input a sequence of single-channel EEG epochs and produces a sequence of sleep stages of the same length in the many-to-many scheme. The first part of the network in the "TinySleepNet" model extracts the time-invariant features and builds a convolutional model with 128 filters. The second part of the model with the RNN is used to learn about the temporal information of the input signals, such as sleep transition rules that experts often use to determine future possible sleep stages. [20]

An issue with having a CNN with 128 filters is the increase in computational requirements of the network. Secondly, having temporal information about signals can mislead classification, especially when trying to classify sleep stages data that are abnormal. For example, if the network is trying to use temporal information to classify a person with narcolepsy, it is likely to incorrectly classify since the sleep stages do not follow the pattern observed in most sleep cycles.

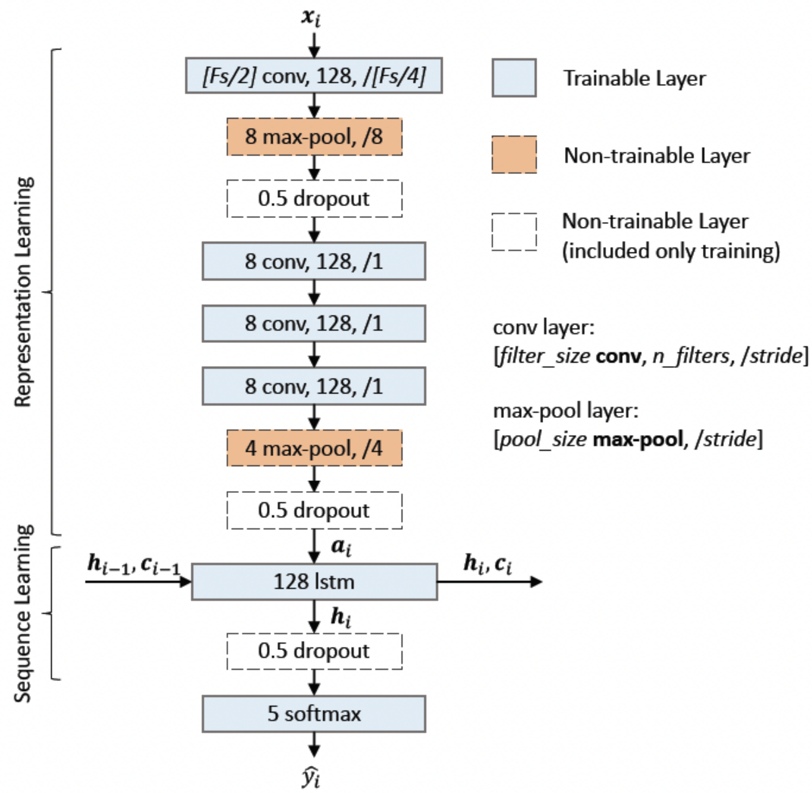

Figure 8: Overview of the architecture of TinySleepNet. Each rectangular box represents one layer in the model, and the arrows indicate the flow of data from the input, which is the raw single-channel EEG epochs to the output, that is the sleep stages.[20]

<u>"DeepSleepNet"</u>

The 'DeepSleepNet' model utilizes a complex CNN. The input for this model is raw-single-channel EEG epochs. It consists of two main parts, as shown in Figure 9. The first part is representation learning, which can be trained to learn filters to extract time-invariant features from the epochs. The second part is sequence residual learning, which can be trained to encode the temporal information such as stage transition rules from a sequence of EEG epochs in the extracted features.[21]

---

[20] 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)*, Montreal, QC, Canada, 2020, pp. 641-644, doi: 10.1109/EMBC44109.2020.9176741.

[21] Supratak, Akara & Dong, Hao & Wu, Chao & Guo, Yike. (2017). DeepSleepNet: a Model for Automatic Sleep Stage Scoring based on Raw Single-Channel EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering. PP. 10.1109/TNSRE.2017.2721116.

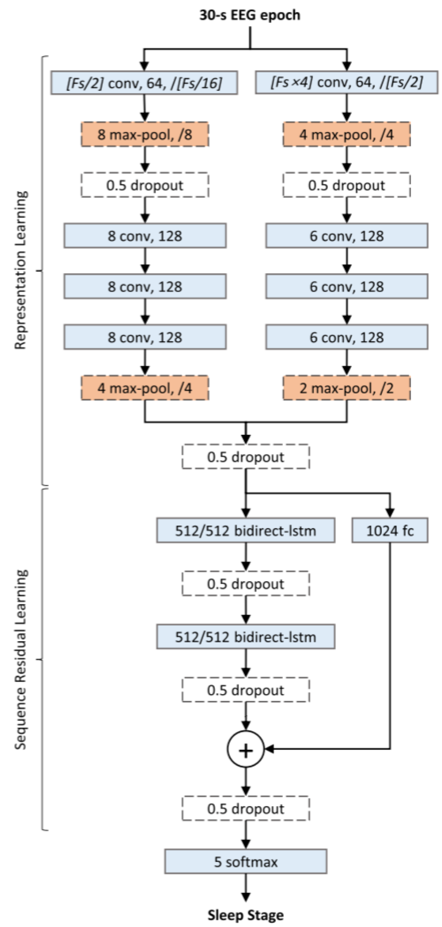

Figure 9: Overview of the architecture of DeepSleepNet consisting of two main parts: representation learning and sequence residual learning. The blue boxes represent trainable layers, the orange boxes non-trainable layers, and the white dotted outline boxes represent non-trainable layers only included in training.[22]

<u>“AttnSleep”</u>

‘AttnSleep‘ employs a very complex model including a multi-resolution convolutional neural network (MRCNN) and adaptive feature recalibration (AFR) along with a temporal context encoder (TCE). The authors of the paper that created ‘AttnSleep’ claim that the MRCNN can extract low and high-frequency features, the AFR can improve the quality of the extracted features and the TCE can capture the long-term dependencies in the input features.[23]

---

[22] 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)*, Montreal, QC, Canada, 2020, pp. 641-644, doi: 10.1109/EMBC44109.2020.9176741.

[23] E. Eldele et al., "An Attention-Based Deep Learning Approach for Sleep Stage Classification With Single-Channel EEG," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 29, pp. 809-818, 2021, doi: 10.1109/TNSRE.2021.3076234.

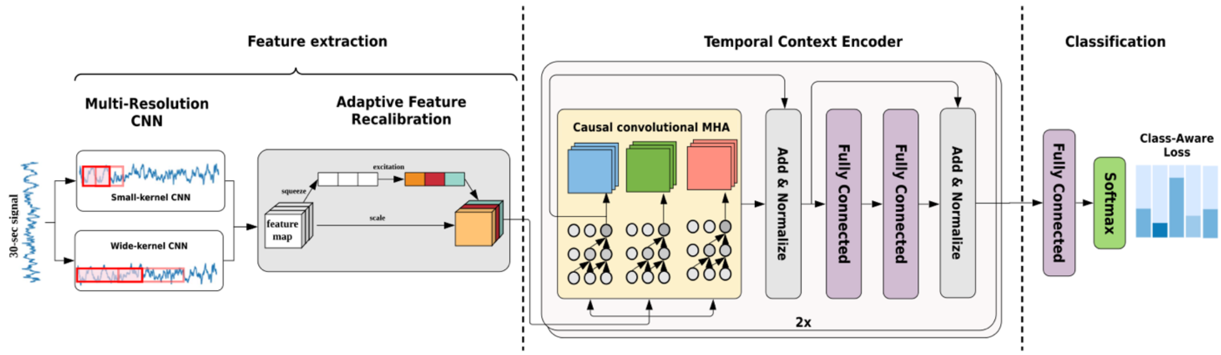

Figure 10: The framework proposed by the AttnSleep model for sleep stage classification.

Comparison of Our Model to Others

Table 5 shows a comparison of our approaches to sleep stage classification with the models described in the previous subsection. We compare the overall accuracy, the F1 scores, and the average training times of the models. As we can see, our simple model is very close in terms of accuracy to other highly complicated, and computationally expensive models. The average training time for our model is also significantly lower than the other models.

Comparing the performance of different models can be challenging due to the variability of various factors. It is important to recognize that the computational devices we are using to train our model are not the same as those used in the other models, and we may have an advantage due to the progress in technology, and computers becoming more powerful over time. We also recognize that the datasets used in the models are different. This can introduce variations in the data that can affect the performance of the models.

Model	Overall Accuracy	F1*	Average Training Time
TinySleepNet	0.854	0.805	Unknown
DeepSleepNet	0.82	0.81	~ 3 hours
AttnSleep	0.844	0.781	~ 21 mins
Approach 1 – Image Classification	0.742	0.746	Less than 5 mins

Approach 2 – Numerical Features	0.79	0.772	Less than 5 minutes

*Table 5: Comparison of accuracy for the Sleep-EDF-20 dataset across different deep-learning-based models. *While other models use macro-F1 scores, our approaches are using weighted F1 scores.*

Conclusion

As we have seen in this paper, our model produces comparable accuracy and results as compared to the more complex models seen in several papers. This problem faced in the field of data modeling can potentially be attributed to a lack of true understanding of the underlying problem and the kind of data being worked with. Complex models not only require more training data to be successful, but they are also incredibly computationally expensive. For a problem like sleep stage scoring, there is a need for the models to be easy to train, deploy and use since it is a real-world problem that requires an efficient application. For the medical community to adopt a model for sleep stage scoring, there is a need for interpretability, which is lacking with complex models. Through this paper, we hope to show that a true understanding of the data, along with a simple architecture and hyperparameter tuning of the model can allow for a less expensive, transparent, and comparable model.

Repositories

Below are links to relevant GitHub repositories referenced in this paper, as well as our own.

Our Repo – https://github.com/abigailalbuquerque/SimpleAndComparableSleepStageScoring
TinySleepNet - https://github.com/akaraspt/tinysleepnet
DeepSleepNet - https://github.com/akaraspt/deepsleepnet
AttnSleep - https://github.com/emadeldeen24/AttnSleep