Automatic Actigraphy and Polysomnography Sleep Scoring using Deep Learning

by

Omar Barazanji

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Electrical Engineering, Master of Science

Auburn, Alabama
August 5, 2023

Artificial Intelligence, Machine Learning, Deep Learning, Health, Sleep

Approved by

Griffith, SueAnne Nichole, Chair, Professor of Electrical and Computer Engineering
Sadi, Mehdi Zahid, Co-Chair, Professor of Electrical and Computer Engineering
Guin, Ujjwal, Professor of Electrical and Computer Engineering

Abstract

The utilization of deep learning models for sleep scoring has become an increasingly promising area of research due to their potential to automate and enhance the accuracy of this crucial task. Sleep scoring involves categorizing a patient's polysomnography (PSG) data into different sleep stages, which plays a vital role in diagnosing sleep disorders and understanding an individual's sleep patterns.

In this study, two significant sources of data were employed: actigraphy and PSG recordings. Actigraphy, a non-invasive method, captures physical activity and light exposure, enabling sleep/wake prediction. PSG, on the other hand, incorporates various physiological signals, such as EEG, ECG, and EOG recordings, providing comprehensive insights into brain activity, cardiac activity, and eye movements during sleep [1].

To address the complexity of sleep scoring and improve accuracy, three deep learning architectures were chosen for evaluation: Convolutional – Long Short-Term Memory (CNN-LSTM), Extreme Gradient Boosting (XGBoost), and LSTM. These models were assessed on a dataset comprising 109 subjects for actigraphy sleep/wake prediction and 30 subjects for PSG sleep staging. Each subject's dataset consisted of five nights of sleep data, offering diverse samples.

The integration of actigraphy and PSG data proved to be a valuable strategy, providing a more comprehensive understanding of an individual's sleep architecture. By utilizing the power of deep learning models and incorporating multi-modal data, clinicians and researchers can significantly improve sleep disorder diagnosis and treatment. The potential for automating the sleep scoring process promises to enhance the efficiency of sleep studies, allowing healthcare professionals to focus on tailored treatment plans and better patient care.

As the availability of large-scale sleep datasets and computational resources continues to grow, the future of sleep scoring with deep learning models holds great promise. With ongoing research and advancements, these models have the potential to become indispensable tools in sleep medicine, empowering healthcare providers to optimize sleep health and overall well-being for their patients.

Acknowledgments

I would like to thank Auburn University's Electrical–Computer Engineering department for all the support I was given in pursuing my research.

The National Sleep Research Resource was supported by the U.S. National Institutes of Health, National Heart Lung and Blood Institute (R24 HL114473, 75N92019R002).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

PSG	Polysomnography
EEG	Electroencephalography
ECG	Electrocardiography
EOG	Electrooculography
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
XGBoost	Extreme Gradient Boosting
MLP	Multi-Layer Perceptron
RNN	Recurrent Neural Network
STFT	Short-Time Fourier Transform
NSRR	National Sleep Research Resource

Chapter 1: Introduction

1.1 Background

Sleep is a fundamental biological process that is crucial for maintaining overall health and well-being. It plays a vital role in various physiological functions, such as memory consolidation, immune system regulation, and emotional well-being. The sleep-wake cycle is composed of different stages, including rapid eye movement (REM) sleep and non-rapid eye movement (NREM) sleep, each characterized by distinct brain activity patterns and physiological changes. Disruptions in these sleep stages or the occurrence of sleep disorders can have significant impacts on an individual's quality of life and overall health.

Sleep disorders, such as sleep apnea, insomnia, restless legs syndrome, and narcolepsy, are prevalent conditions affecting a substantial portion of the population worldwide. These disorders are associated with daytime sleepiness, impaired cognitive function, mood disturbances, and an increased risk of chronic conditions, including cardiovascular disease, obesity, and diabetes. Accurate and timely diagnosis of sleep disorders is crucial for appropriate treatment and management, as it allows healthcare professionals to target the underlying causes and alleviate the associated symptoms effectively.

1.2 Motivation

Traditionally, sleep scoring, the process of identifying and categorizing different sleep stages, has been performed manually by trained clinicians through visual analysis of polysomnography (PSG) data, which includes electroencephalography (EEG), electrooculography (EOG), electromyography (EMG), and other physiological signals. However, manual scoring is a time-consuming and labor-intensive task that is subject to inter-rater variability, leading to inconsistencies in the results. These limitations can hinder the efficiency and accuracy of sleep disorder diagnosis and subsequent treatment decisions.

With recent advances in deep learning techniques and the availability of large-scale sleep datasets, there is an opportunity to leverage the power of artificial intelligence to automate the sleep scoring process. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated remarkable success in various domains, including computer vision, natural language processing, and speech recognition. The application of these models to sleep scoring holds the potential to overcome the limitations of manual scoring, providing faster and more consistent results.

Furthermore, the integration of actigraphy data, which captures physical activity and light exposure, with PSG data could enhance the accuracy of sleep scoring. Actigraphy offers a non-invasive and cost-effective method for monitoring sleep patterns in real-life settings, making it suitable for large-scale studies and long-term monitoring. Combining actigraphy with PSG data can provide a more comprehensive understanding of an individual's sleep architecture, facilitating a more accurate characterization of sleep disorders.

Therefore, this study aims to explore the capabilities of deep learning models in automating the sleep scoring process. By comparing the performance of different models, including CNN-

LSTM, XGBoost, and LSTM, and utilizing both actigraphy data for sleep/wake prediction and PSG data for sleep staging, this research seeks to identify the optimal model for each task and assess their accuracies. The findings will provide valuable insights into the potential of deep learning models for sleep scoring, contributing to the improvement of current methods and ultimately benefiting the diagnosis and treatment of sleep disorders.

1.3 Contributions

The primary objective of this research was to explore the capabilities of deep learning models in automating the sleep scoring process, specifically focusing on sleep/wake prediction using actigraphy data and sleep staging using polysomnography (PSG) data. The study aimed to identify the optimal model for each task and compare the accuracies of different deep learning architectures, including Convolutional – Long Short-Term Memory (CNN-LSTM), Extreme Gradient Boosting (XGBoost), and LSTM.

For sleep/wake prediction, the study found that the XGBoost model (SleepWakeNet-v1) achieved the highest accuracy of 92%. The model effectively utilized raw signal data from accelerometer sensors in the x, y, and z axes to predict sleep and wake states. This result highlights the superiority of XGBoost over LSTM for sleep/wake prediction using actigraphy data.

In contrast, for sleep staging, the Time Distributed CNN model (SleepScoreNet-v1) outperformed the other architectures, achieving an accuracy of 87%. The model effectively used spectrogram representations obtained from PSG data, combining Time Distributed CNN layers for parallel spatial feature extraction. This finding emphasizes the effectiveness of the Time Distributed CNN approach for sleep staging using polysomnography data.

The integration of actigraphy data with PSG data also proved to be a valuable strategy, providing a more comprehensive understanding of an individual's sleep patterns. The combination of advanced machine learning techniques and multi-modal data showcased the potential to revolutionize the diagnosis and treatment of sleep disorders, offering faster and more consistent results compared to manual scoring by trained clinicians.

Overall, the study's contributions lie in its comprehensive evaluation and comparison of deep learning models for sleep scoring. By identifying the optimal models for sleep/wake prediction and sleep staging tasks, this research provides valuable insights for healthcare professionals, researchers, and developers seeking to leverage artificial intelligence and machine learning to enhance sleep disorder diagnosis and improve patients' overall health and well-being. These contributions pave the way for further research and applications in the domain of sleep medicine, ultimately benefiting individuals by optimizing sleep health and well-being.

Chapter 2: Related Works

2.1 Machine Learning Algorithms for Sleep Scoring

Early approaches to automatic sleep scoring and sleep/wake classification involved machine learning algorithms, such as the Support Vector Machine (SVM). SVMs demonstrated the ability to achieve accurate classification results [9]. However, compared to more recent methods that leverage advancements in machine learning, SVMs have limitations in terms of training efficiency and tuning [3, 6, 8]. An alternative option for large datasets is the utilization of a multi-layer perceptron (MLP), which offers easier model calibration and can deliver improved results [2]. MLPs are the foundation of feedforward neural networks commonly used in machine learning applications. While a simple MLP model can provide a solution, its accuracies may be lower when compared to state-of-the-art sleep scoring approaches [3].

2.2 **Sleep Scoring** with **Recurrent Neural Networks (RNNs)**

Expanding on the capabilities of MLPs, researchers have explored the use of recurrent neural networks (RNNs) for sleep scoring [7]. RNNs have the ability to capture contextual dependencies by analyzing patterns in sequential data. This is particularly advantageous for time-series forecasting tasks, such as sleep scoring, where the context of previous inputs can significantly impact future predictions. By leveraging the sequential nature of sleep data, RNNs enable deeper contextual analysis and have shown promise in improving sleep scoring accuracy.

2.3 Sleep Scoring with Extreme Gradient Boosting (XGBoost)

Another approach that has garnered attention in sleep scoring is Extreme Gradient Boosting (XGBoost). XGBoost combines decision trees and gradient descent, similar to a neural network, to minimize the loss function and achieve high accuracies with reduced computational complexity. The XGBoost method has demonstrated its effectiveness in various domains and has the potential to enhance sleep scoring performance [6].

2.4 Sleep Scoring with Convolutional Neural Networks and Long Short-Term Memory (CNN-LSTM)

Recent studies have explored the integration of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks for sleep scoring tasks. This approach has shown significant promise, particularly when dealing with large multi-dimensional inputs, such as images or spectrograms [3]. CNNs excel at extracting relevant features and reducing input dimensionality through feature extraction, while LSTMs are adept at processing sequential data and capturing temporal dependencies. By combining the strengths of both CNNs and LSTMs, the CNN-LSTM architecture can achieve higher accuracies in sleep scoring compared to using either model individually.

2.5 Related Works Summary

In summary, various machine learning approaches have been explored for sleep scoring and sleep/wake classification tasks. Early methods utilizing SVMs provided accurate results, but recent advancements in machine learning have paved the way for improved approaches. MLPs offer versatility and ease of calibration, while RNNs enable deeper contextual analysis by capturing sequential dependencies. XGBoost, with its gradient boosting and decision tree ensemble technique, achieves high accuracies with reduced computational complexity. The integration of CNNs and LSTMs in the CNN-LSTM architecture has shown exceptional performance in capturing both spatial and temporal features, resulting in improved sleep scoring accuracies. These advancements in machine learning techniques hold great potential for automating sleep scoring and enhancing the diagnosis and treatment of sleep disorders.

Chapter 3: Datasets and Feature Preparation

3.1 Sleep Scoring Dataset

The dataset used for sleep scoring in this study was obtained from the "Dreem Open Datasets: Multi-Scored Sleep Datasets to compare Human and Automated sleep staging." It comprised a subset of 27 participants with 25 nights of sleep recordings. The participants were divided into two groups: 15 participants for training and 12 participants for testing.

Each participant's sleep recordings included 16 signals, including EEG, ECG, and EOG data. These signals provide valuable insights into brain activity, cardiac activity, and eye movements during sleep. The dataset also provided labels indicating the sleep stage for each 30-second epoch.

To prepare the data for model processing, each signal was divided into segments with 7500 timesteps. These segments were transformed into spectrograms using the short-time Fourier transform (STFT), which converts the signals from the time domain to the frequency domain. The resulting spectrograms served as input for a CNN model to process spatial features. Figure 3 illustrates an example of a spectrogram derived from the dataset.

3.2 Sleep/Wake Prediction Dataset

For sleep/wake prediction, the training data was obtained from the NSRR dataset titled "Urban Poor in India." This dataset included actigraphy data recorded from accelerometer sensors in the x, y, and z axes, capturing physical activity patterns. Binary labels of 1 and 0 were provided to indicate sleep and wake states, respectively.

To create a training and testing split, 15 participants with labeled actigraphy data, representing days of recorded activity, were selected for training. Two additional participants were reserved for testing the accuracy, F1 score, recall, and precision metrics of each model. Figure 2 presents a visualization of the actigraphy data from a testing sample, displaying the accelerometer readings and the corresponding sleep/wake labels.

3.3 Feature Preparation

The datasets used in this study required specific feature preparation techniques. The sleep scoring dataset transformed the signals into spectrograms, representing the data in a 2D format suitable for CNN processing. The actigraphy data in the sleep/wake prediction dataset was already in a suitable format, consisting of accelerometer readings in the x, y, and z axes.
The datasets, along with their respective prepared features, provide the foundation for training and evaluating the deep learning models in the subsequent chapters.

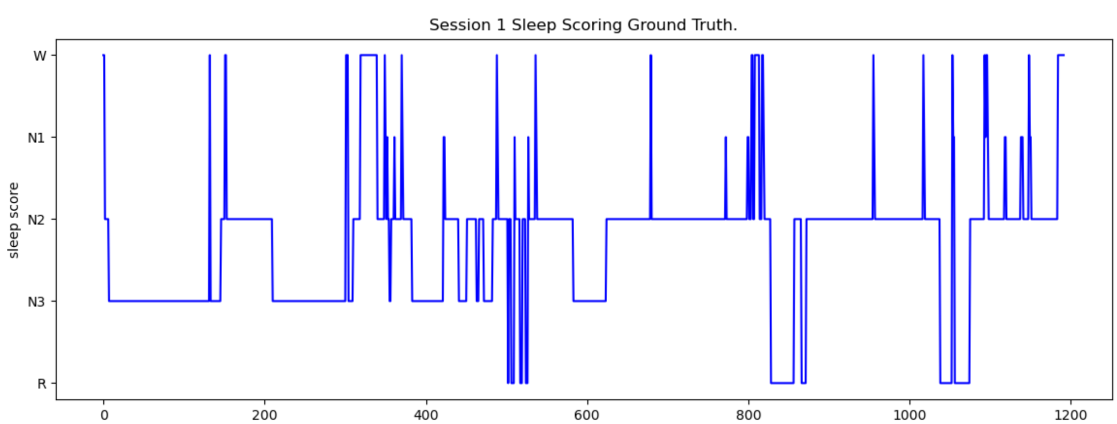

Figure 1 - Hypnogram visualization of session 1's ground truth sleep scores.

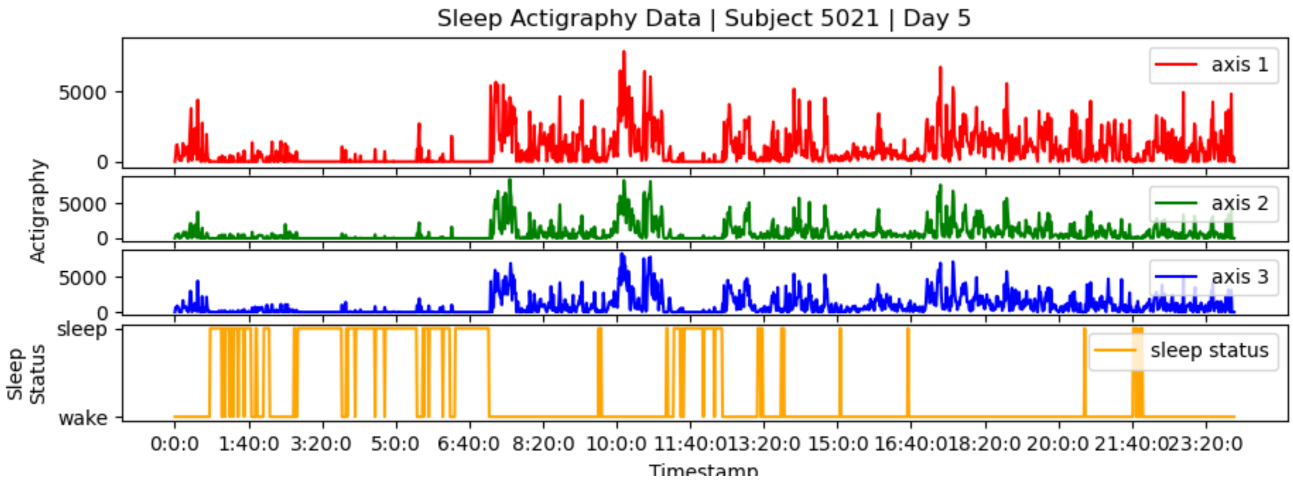

Figure 2 - Actigraphy data visualization on subject 5021, day 5. Axes 1-3 are x, y, and z accelerometer sensors.

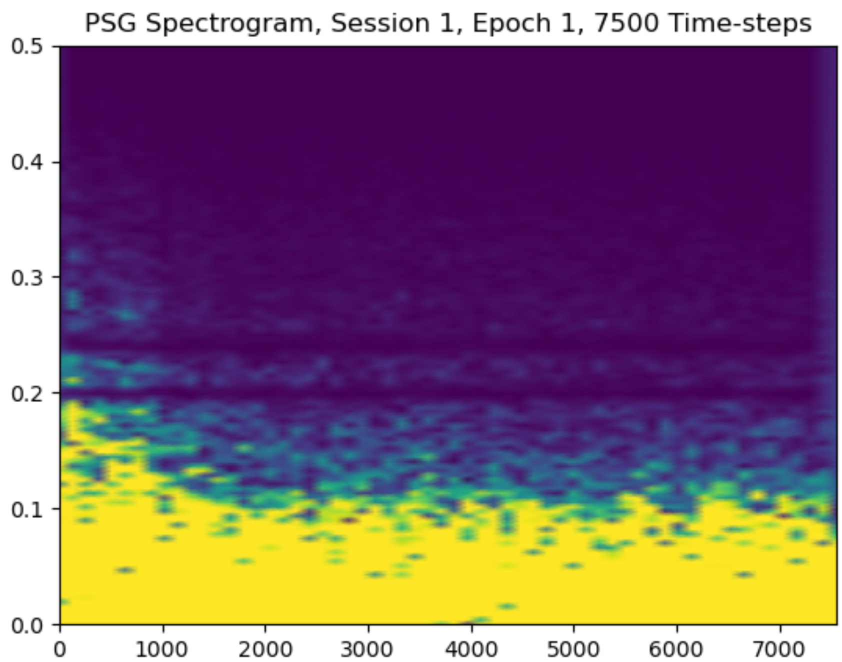

Figure 3 - PSG spectrogram visualization.

Chapter 4: Methods

4.1 Model Pipelines

For sleep scoring, four models are compared: LSTM, XGBoost, CNN, and CNN-LSTM. All models are trained and tested on the same samples. The evaluation metric for these models is the accuracy of the generated hypnogram compared to the ground truth sleep stages.

4.1.1 LSTM Model Pipeline

The LSTM model takes raw signal data as input for both sleep scoring and sleep/wake prediction tasks. For sleep scoring, the model utilizes 16 PSG signals, while for sleep/wake prediction, it uses 3-axis accelerometer signals (x, y, and z). The LSTM model is designed to capture temporal dependencies in the data. The architecture of the LSTM model remains the same for both tasks.

4.1.2 XGBoost Model Pipeline

Similar to the LSTM model, the XGBoost model also takes raw signal data without transforming it into a spectrogram. However, unlike the LSTM model, the input shape for the XGBoost model is flattened or 1D. The XGBoost model employs a decision tree ensemble technique and gradient descent to minimize the loss function. The model requires manual tuning of hyperparameters, such as the tree depth and learning rate.

4.1.3 CNN-LSTM Model Pipeline

The CNN-LSTM model preprocesses the signal data to create a 2D spectrogram representation. This is achieved using the Short-Time Fourier Transform (STFT) algorithm, which extracts features from the signal before feeding it into the model. The benefit of using a 2D input, similar to an image, is the ability to apply CNN layers for feature extraction before passing the data to LSTM cells to capture temporal features. The CNN layers enhance the model's understanding of the input by extracting 2D features before considering temporal information.

4.2 Sleep/Wake Prediction Models

The LSTM, XGBoost, and CNN-LSTM will be employed for sleep/wake prediction. The evaluation metric for these models is the accuracy of the sleep/wake predictions generated over the ground truth sleep/wake labels.

4.2.1 SleepWakeNet-V0 (Baseline LSTM)

SleepWakeNet-V0 represents the baseline LSTM model for sleep/wake classification. It comprises a single LSTM layer with three temporal cells, each representing the x, y, and z accelerometer signals. A Dense layer with one unit is added for binary classification. The model is trained using binary cross-entropy loss, and the learning rate is scaled with the Adam optimizer during training.

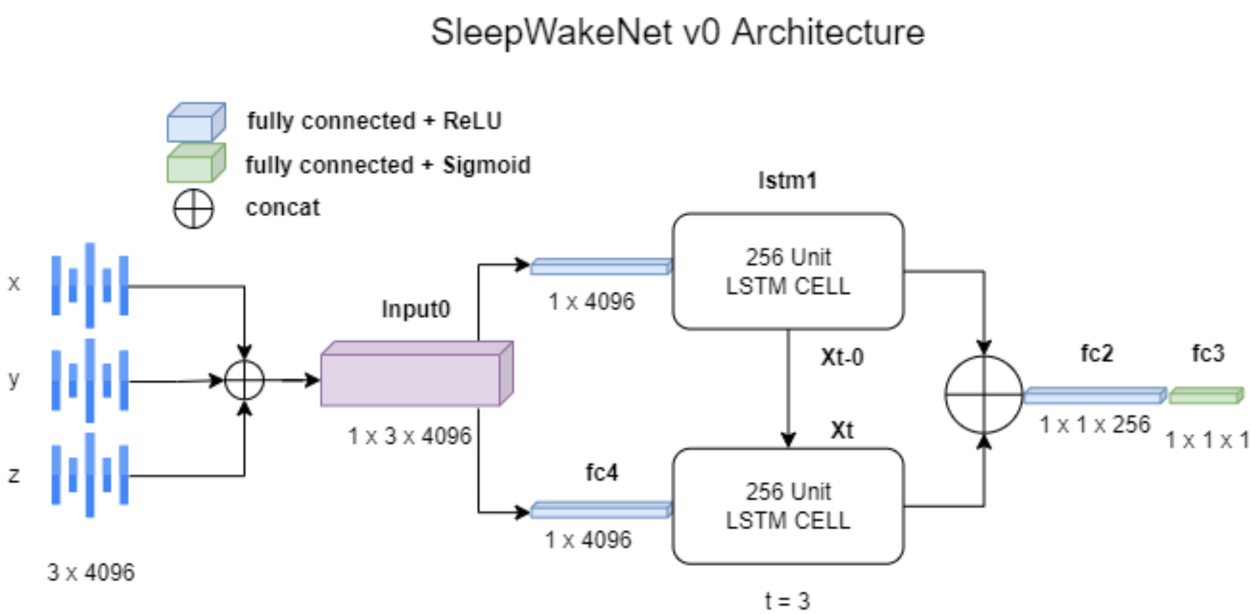

Figure 4 - SleepWakeNet-v0 architecture on sleep/wake prediction from 3 axis (x, y, z) accelerometer signals.

4.2.2 SleepWakeNet-V1 (XGBoost)

SleepWakeNet-V1 uses the XGBoost model for sleep/wake classification. This model is trained for 20 rounds and has a decision tree with a depth of 10 and learning rate of 0.7. The parameters for this model were determined by performing a parameter sweep on the number of rounds, tree depth, and learning rate.

4.2.3 SleepWakeNet-V2 (CNN-LSTM)

SleepWakeNet-V2 takes a time-distributed input to carry out three 3-layer CNNs on each spectrogram obtained from the x, y, and z accelerometer signals. The output of the CNN layers is then fed into a single LSTM layer with three temporal cells, each comprising 256 units. Finally, a Dense layer with one neuron is added for binary classification. The model is trained using binary cross-entropy loss, and the learning rate is scaled with the Adam optimizer during training.

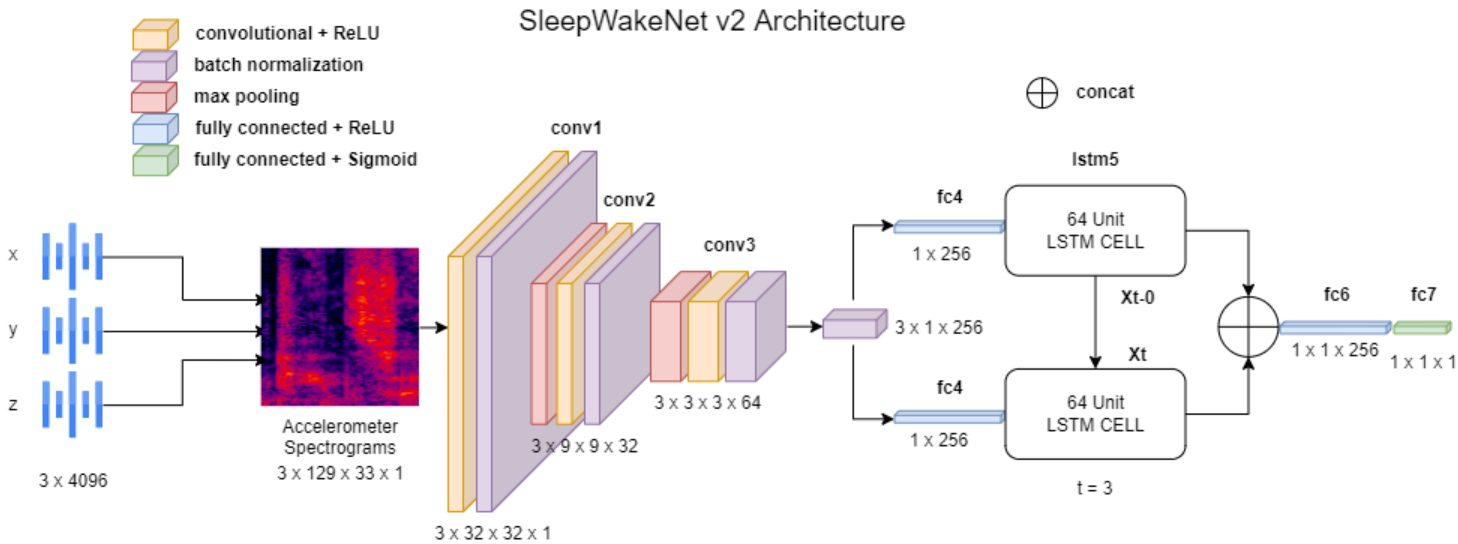

Figure 5 - SleepWakeNet-v2 architecture on sleep/wake prediction from 3 axis (x, y, z) accelerometer signals.

4.3 Sleep Stage Classification Models

The LSTM, CNN, and CNN-LSTM models were used for sleep scoring. The evaluation metric for these models is the accuracy of the hypnogram generated from testing data over the ground truth hypnogram.

4.3.1 SleepScoreNet-V0 (Baseline LSTM)

SleepScoreNet-V0 serves as the baseline architecture for sleep scoring. It takes 16 PSG signals as input and consists of a single LSTM layer with 16 temporal cells, each comprising 256 units. A Dense layer with five neurons is added for categorical classification. The model is trained using categorical cross-entropy loss, and the learning rate is scaled with the Adam optimizer during training.

4.3.2 SleepScoreNet-V1

SleepScoreNet-V1 employs a time-distributed input to process 16 3-layer CNNs on each PSG spectrogram. The output from the time-distributed CNN is flattened to 1 x 4096 and sent to a Dense layer with five neurons for categorical classification. The model is trained using categorical cross-entropy loss, and the learning rate is scaled with the Adam optimizer during training.

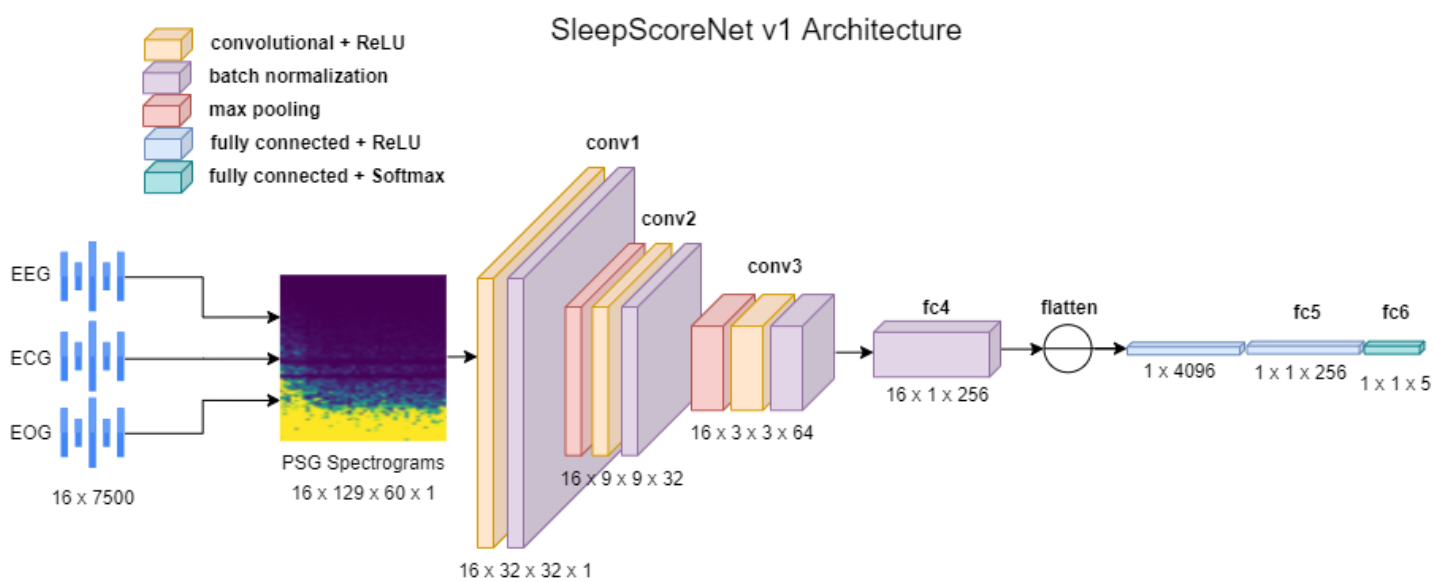

Figure 6 - SleepScoreNet-v1 sleep scoring architecture from 16 PSG signals.

4.3.3 SleepScoreNet-V2

SleepScoreNet-V2 utilizes a time-distributed input to process 16 3-layer CNNs on each PSG spectrogram. The output is then passed to a single LSTM layer with 16 temporal cells, each comprising 256 units. A Dense layer with five neurons is added for categorical classification.

The model is trained using categorical cross-entropy loss, and the learning rate is scaled with the Adam optimizer during training.

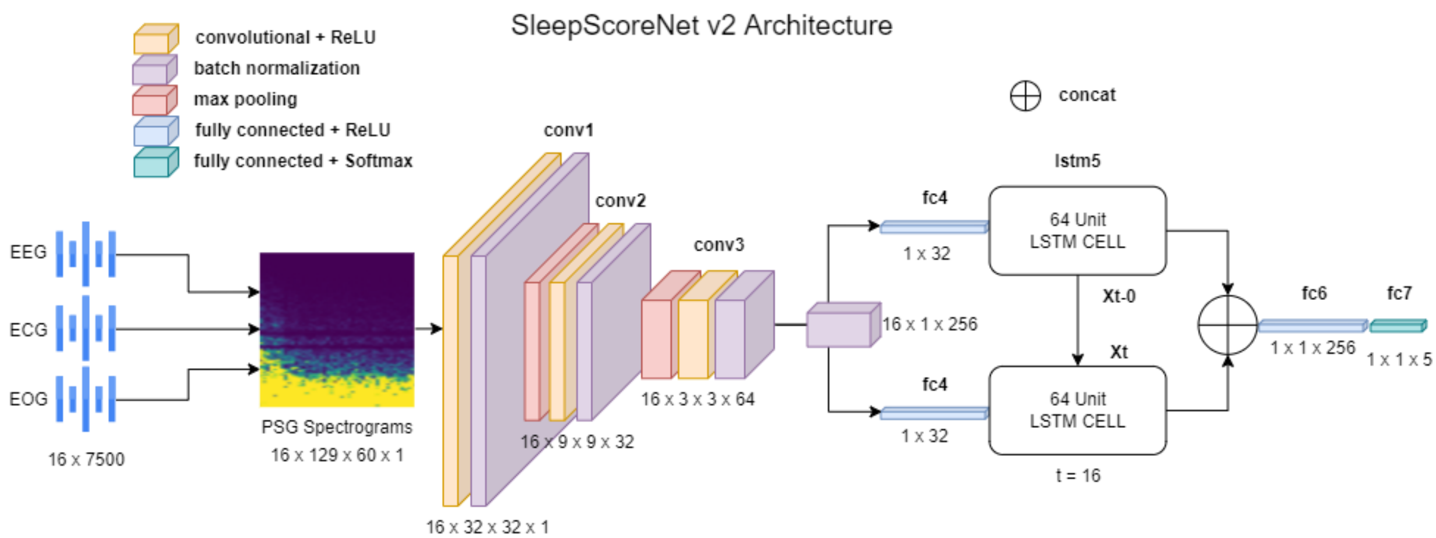

Figure 7 - SleepScoreNet-v2 sleep scoring architecture from 16 PSG signals.

By evaluating these additional model configurations, the study aims to compare their performances and identify the most effective approaches for sleep scoring and sleep/wake prediction tasks.

Chapter: 5 Evaluation Metrics for Model Performance

In this chapter, we discuss the evaluation metrics used to assess the performance of various models on the testing data. To gauge the effectiveness of each model's predictions, we calculate several key components, including True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These components form the foundation for computing essential metrics, such as Recall, Precision, F1, and Accuracy.

Each model will be evaluated with the same testing data per domain. For sleep/wake classification, each model will be scored on precision, recall, and F1 accuracy on the same testing data from the NSRR, which is 30% of the dataset and is not used during training. Table 1 shows the results for each sleep/wake classification model. For sleep scoring or sleep stage prediction, each model will be scored on 30% of the dataset from Dreem on classification of Wake, NREM1-3, and REM stages. Table 2 shows the results for each sleep scoring model.

5.1 True Positives, True Negatives, False Positives, and False Negatives

True Positives (TP) represent the number of positive instances that were correctly identified as positive by the model. On the other hand, True Negatives (TN) are the number of negative instances that were correctly identified as negative by the model. Conversely, False Positives (FP) denote the number of negative instances that were incorrectly predicted as positive, and False Negatives (FN) indicate the number of positive instances that were incorrectly predicted as negative.

5.2 Recall (Sensitivity or True Positive Rate)

Recall, also known as Sensitivity or True Positive Rate, quantifies the ability of the model to correctly identify positive instances. It is defined as the ratio of True Positives to the sum of True Positives and False Negatives:

$$Recall = \frac{TP}{TP + FN}$$

A higher Recall value indicates that the model is proficient at capturing positive instances, minimizing the occurrence of false negatives.

5.3 Precision (Positive Predictive Value)

Precision, often referred to as Positive Predictive Value, assesses the accuracy of positive predictions made by the model. It is calculated as the ratio of True Positives to the sum of True Positives and False Positives:

$$Precision = \frac{TP}{TP + FP}$$

A high Precision value suggests that the model has a low rate of false positives, and the positive predictions are reliable.

5.4 F1 Score (F1 Measure)

The F1 Score, also known as the F1 Measure, is the harmonic mean of Precision and Recall. It offers a balanced evaluation of the model's performance, considering both false positives and false negatives. The F1 Score is given by:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

This metric is particularly useful when there is an uneven class distribution, as it strikes a balance between Precision and Recall.

5.5 Accuracy

Accuracy measures the overall correctness of the model's predictions, encompassing both positive and negative instances. It is calculated as the ratio of the sum of True Positives and True Negatives to the total number of instances:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

A high Accuracy value indicates that the model is making correct predictions for both positive and negative instances.

In conclusion, by evaluating and comparing these metrics, we gain valuable insights into the performance of each model on the testing data. High Recall and Precision are essential for applications where false negatives and false positives carry significant consequences, respectively. The F1 Score provides a balanced evaluation, while Accuracy serves as a general measure of overall correctness. The comprehensive analysis of these metrics enables us to make well-informed decisions regarding the suitability of each model for the task at hand.

Model	**Sleep Wake Classification Model Metrics**			
	Precision	**Recall**	**F1**	**Accuracy**
SleepWakeNetv0	0.86270	0.80028	0.83032	0.88871
SleepWakeNetv1	0.86484	0.90080	0.88245	0.91664
SleepWakeNetv2	0.73799	0.68421	0.71008	0.81018

Table 1 – Sleep/wake classification accuracy scores.

Model	**Sleep Stage Classification Model Metrics**			
	Precision	**Recall**	**F1**	**Accuracy**
SleepScoreNetv0	0.31940	0.42133	0.34455	0.42133
SleepScoreNetv1	0.88701	0.87856	0.87998	0.87856
SleepScoreNetv2	0.87162	0.85111	0.85653	0.85111

Table 2 – Sleep stage classification accuracy scores.

Chapter 6: Discussion and Conclusion

The performance of different deep learning models for sleep scoring using actigraphy and polysomnography (PSG) data was evaluated. The study aimed to identify the optimal model for each task, comparing the accuracies of Convolutional – Long Short-Term Memory (CNN-LSTM), Extreme Gradient Boosting (XGBoost), and LSTM deep learning architectures.
For sleep/wake classification, the XGBoost model (SleepWakeNet-v1) achieved the highest accuracy of 92%. It utilized raw signal data from accelerometer sensors in the x, y, and z axes, effectively predicting sleep and wake states.

In contrast, for sleep scoring, the Time Distributed CNN model (SleepScoreNet-v1) outperformed the others, achieving an accuracy of 87% accuracy. The model used spectrogram representations obtained from PSG data, combining Time Distributed CNN layers for parallel spatial feature extraction.

The findings indicate that XGBoost is a promising approach for sleep/wake classification, while the Time Distributed CNN proves to be the optimal model for sleep staging. These deep learning models have the potential to automate the sleep scoring process, alleviating the time-consuming and labor-intensive manual scoring performed by trained clinicians.

The integration of actigraphy data with PSG data further enhances the accuracy of sleep scoring, providing a more comprehensive understanding of an individual's sleep architecture. The combination of advanced machine learning techniques and multi-modal data can revolutionize the diagnosis and treatment of sleep disorders, offering faster and more consistent results.

Overall, this study demonstrates the potential of deep learning models in the field of sleep scoring and opens new avenues for future research and applications in the domain of sleep medicine. The results provide valuable insights for healthcare professionals, researchers, and developers seeking to leverage artificial intelligence and machine learning to improve sleep disorder diagnosis and ultimately enhance patients' overall health and well-being.

References

[1] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. F. Quan, The AASM manual for the scoring of sleep and associated events : rules, terminology, and technical specifications. Istchester, IL: American Academy of Sleep Medicine, 2007.

[2] Guillot, A., Sauvet, F., During, E., Thorey, V. Dreem Open Datasets: Multi-Scored Sleep Datasets to compare Human and Automated sleep staging, 2019.

[3] Cho, J.H.; Choi, J.H.; Moon, J.E.; Lee, Y.J.; Lee, H.D.; Ha, T.K. Validation Study on Automated Sleep Stage Scoring Using a Deep Learning Algorithm. Medicina 2022, 58, 779.

[4] Zhang GQ, Cui L, Mueller R, Tao S, Kim M, Rueschman M, Mariani S, Mobley D, Redline S. The National Sleep Research Resource: towards a sleep data commons. J Am Med Inform Assoc. 2018 Oct 1;25(10):1351-1358. doi: 10.1093/jamia/ocy064. PMID: 29860441; PMCID: PMC6188513.

[5] Bessone P, Rao G, Schilbach F, Schofield H, Toma M. The Economic Consequences of Increasing Sleep Among the Urban Poor. Q J Econ. 2021 Apr 8;136(3):1887-1941. doi:10.1093/qje/qjab013. PMID: 34220361; PMCID: PMC8242594.

[6] A. Sano, W. Chen, D. Lopez-Martinez, S. Taylor and R. W. Picard, ”Multimodal Ambulatory Sleep Detection Using LSTM Recurrent Neural Networks,” in IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 4, pp. 1607-1617, July 2019, doi:10.1109/JBHI.2018.2867619.

[7] Sathyanarayana, Aartietal. “Sleep Quality Prediction From Wearable Data Using Deep Learning.” JMIR mHealth and uHealth vol. 4, 4e125.4Nov.2016, doi:10.2196/mhealth.6562.

[8] Chen, Tianqi, Carlos Guestrin. ’XGBoost’. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016. Web.

[9] Shantilal, K. D. Donohue and B. F. O'Hara, "SVM for automatic rodent sleep-wake classification," IEEE SoutheastCon 2008, 2008, pp. 581-586, doi:10.1109/SECON.2008.4494360.

[10] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. 'Attention Is All You Need'. arXiv, 2017.

Appendix

1. Source Code: https://github.com/omarzanji/AU_ECE_Health