Deep Learning-based EEG Analysis for Sleep Apnea Detection

Matthew Bonner[1], Benjamin Nikolai[2], Addison Gambhir[3], Quinn Glovier[4], Augustus Lamb[5],
School of Data Science - University of Virginia
Charlottesville, Virginia 22904
[1] e-mail: mkb2dxw@virginia.edu
[2] e-mail: pcf4ku@virginia.edu
[5] e-mail: ajg7pk@virginia.edu
[3] e-mail: qdg9xwb@virginia.edu
[4] e-mail: abl6ywp@virginia.edu

***Abstract*—Sleep apnea, a condition disrupting normal breathing patterns during sleep, poses significant health challenges. While the definitive diagnosis relies on polysomnography, this method's complexity underscores the need for innovative approaches to detecting the condition. Central to this diagnostic process is electroencephalography (EEG), which measures the brain's electrical activity and potentially provides detailed insights into sleep patterns and disturbances. Addressing this, the present study introduces a cutting-edge Temporal Convolutional Neural Network (TCNN) approach for the automatic detection of sleep apnea using single-lead EEG signals. TCNNs offer several advantages over traditional analytical methods by ensuring the preservation of temporal data sequence, which is essential for analyzing the complex, long-range dependencies observed in EEG recordings. This methodological innovation allows for enhanced computational efficiency and potentially greater accuracy in diagnosis. To validate model effectiveness, we employ a comprehensive evaluation framework, comparing it against established machine learning models on several key performance metrics, including accuracy, Area Under the ROC (AUC), F1-score, recall, and precision. Our objective is not only to examine the TCNN model's performance but also to investigate its utility in streamlining the diagnostic process for sleep apnea. By advancing the accuracy and efficiency of diagnosis, the present study aims to promote earlier treatment interventions, ultimately improving patient outcomes and reducing the health complications associated with sleep apnea.**

I. Introduction and Motivation

Sleep apnea is a complex sleep disorder that disrupts normal breathing patterns during sleep, characterized by recurrent pauses and resumptions in breathing. These disturbances, known as apneic events, can significantly impact the quality of sleep and overall health. Sleep apnea manifests in two primary forms: Obstructive Sleep Apnea (OSA) and Central Sleep Apnea (CSA). OSA is caused by a physical blockage in the airway, while CSA results from a failure in the brain's signaling mechanisms that control breathing. The number of apnea events per hour determines the severity of the disorder: 5-14 events per hour is categorized as mild sleep apnea, 15-29 events is categorized as moderate sleep apnea, and 30 or more events is categorized as severe sleep apnea [1]. This condition affects up to 7% of the population, and is a harbinger of numerous health risks, ranging from cognitive dysfunction to cardiovascular and cerebrovascular disorders [2].

The diagnosis of sleep apnea conventionally involves polysomnography (PSG), a comprehensive sleep study that utilizes a multitude of data channels to accurately assess the patient's condition. This procedure includes the use of electroencephalogram (EEG) and electrooculogram (EOG) for sleep staging, along with electromyogram (EMG), electrocardiogram (EKG), and various respiratory channels such as pulse oximetry (Sp02) [1], [3]. PSG is essential not only for diagnosing the type of sleep apnea but also for detecting other, co-occurring sleep disorders, particularly in patients who exhibit symptoms of OSA. During the diagnosis process, clinicians look for several key symptoms to confirm whether a patient has sleep apnea. These include snoring, difficulty breathing at night, and lack of quality sleep. Additional considerations involve the presence of non-sleep-related symptoms such as hypertension or mood disorders [1], [4].

Currently, traditional PSG can be difficult to come by for those potentially suffering from sleep apnea with wait times reaching as long as 60 months [2], [3], [5]. This extended delay, coupled with the invasive nature of PSG, underscores the pressing need for more efficient, less cumbersome diagnostic techniques. Recent advancements in machine learning, particularly in the use of single-channel EEG signals, have shown promise in addressing this shortfall. Technologies like Convolutional Neural Networks (CNNs) and Bi-directional Long Short-Term Memory (BI-LSTM) networks have demonstrated promising preliminary results, supporting their integration into portable sleep monitoring devices [2], [6]. These technological advances suggest a shift towards streamlined diagnostic methods that could facilitate more timely interventions and significantly reduce the long wait times associated with traditional PSG methods [3]. Moreover, many of those who currently live with sleep apnea

are undiagnosed [7], highlighting the need for new and more abundant diagnosis techniques. EEG signals have already proven to be a deep reservoir of useful data for investigating SA and other sleep-related issues [2], [8].

Sleep apnea not only affects the individual's health through diminished sleep quality, but it also poses more severe risks for cardiovascular complications in the most severe instances [9]. Thus, efficient and timely diagnosis and treatment are necessary. The present work seeks to not only aid sleep specialists in identifying and treating sleep apnea, but also to provide patients with faster relief and prevent further health complications. Ultimately, a future where the detection of sleep apnea is not only accurate and efficient but also patient-friendly and widely accessible, thereby enhancing patient outcomes and reducing the healthcare burden posed by this prevalent condition, might be achievable.

II. Data Description

For this research, the MIT-BIH Polysomnographic dataset was utilized, containing the results of a sleep study of 16 males aged 32-56. Performed at the Beth Israel Hospital Sleep Laboratory in Boston, Massachusetts, the study recorded the effects of Sleep Apnea on airway obstruction, and how a Constant Positive Airway Pressure (CPAP) machine mitigates the symptoms of the condition [10].

Broken down, the database consists of 18 recordings (two patients had 2 recordings) of approximately 80 hours of data in total, with the data separated into folders designating each recording. To illustrate the timeline for each patient's sleep activity, file contents include an ECG and single-lead EEG signals, a blood pressure signal, a nasal respiratory signal, a respiratory effort signal generated from inductance plethysmography, an EOG signal, an EMG signal, a stroke volume (SV) signal, and an earlobe oximeter (SO2) signal. Not all signal types were recorded for each patient, but, instead, they were chosen dependent on the condition of the patient and what was needed to generate a complete clinical picture of their sleep behaviors [10].

EEG Signal	Patient Count
C3-01	11
C4-A1	4
O2-A1	1

TABLE I: Count of Patients Studied Using Each Type of EEG Signal. NOTE: One patient with a C4-A1 signal was dropped during data cleaning

For EEG analysis, electrodes are placed on a patient's scalp to measure electrical fluctuations within the activity of neurons. While the activity of all patients were recorded using a single-channel of electrodes, the placement varied by individual. These placements were based upon the classic 10-20 electrode montage [11]. 11 of the recordings used a C3-O1 signal, four of the recordings used C4-A1, and one of the recordings used O2-A1 [10]. These codes represent the placement of the signal on the head during the time of the recording. The letter associated with each EEG code signified the location on the head of where electrical activity is being recorded; C represents the central area of the head (the top of the head), O represents the occipital area of the head (the back of the head), and A represents the auricular locations (the ears) [12]. We chose to train data on all channels available regardless of location as our aim is simply to detect Sleep Apnea as it occurs in the brain, and not location or strength.

In this task, the target variable is an indication of whether or not the subject underwent an SA event during the recording period. This is annotated within the dataset along side other diagnoses such as Central Hyponea or Obstructive Hyponea. We filtered this data to create a binary variable to indicate whether the pre-segmented 30-second slices-of-time included an apnea event.

III. Methodology

A. Mathematical Formulation

In this section, the problem is formulated as a time series classification or sequence modeling task. Time series classification involves generating outputs $y_1, y_2, ..., y_t$ for a given sequence of input time series $x_1, x_2, ..., x_t$, that satisfies the condition

$$f(x_1, x_2, ..., x_t) = \hat{y}_1, \hat{y}_2, ..., \hat{y}_T \tag{3.1}$$

Notably, f must be *causal*, meaning the output y_t is only dependent on the input sequence $x_1, x_2, ...x_{T=t}$ and not $x_{T+1}, x_{T+2}, ..., x_N$. This causal constraint is important to our architecture as we must ensure that predictions are made in accordance with the chronological progression of the input data. No leakage of future events x_{T+n} should be weighted against current events. Learning within this objective involves finding a network minimizing the discrepancy between actual and predicted classifications, considering sequences and outputs drawn from a predefined distribution [13], [14].

As described, the EEG signals were broken down into 30-second time intervals per patient. These intervals contain our input data $\vec{x_i}$, with a label classifying each apnea state, y_i. The goal is to predict $\hat{y_i}$ given a data for each of these time intervals.

$$\hat{y_1}, \hat{y_2}, ..., \hat{y_n} = f(\vec{x_1}, \vec{x_2}, ..., \vec{x_n}), \tag{3.2}$$

B. Temporal Convolutional Neural Networks (TCNN)

TCNNs can be employed within this formulation motivated by their ability to capture temporal relationships and patterns within sequential data [14]. This characteristic makes TCN's well-suited for sequence modeling tasks aimed at modeling temporal relationships especially within time series of differing lengths, allowing them to adaptively adjust to the temporal dynamics present in the data.

A TCNN can be defined as a series of *Dilated*, and *Causal* Convolutions, stacked inside multiple layers of *Residual Blocks* [15]. These *Causal* convolutions are made where an output at time t is convolved only with elements that occurred before it, as defined above. As this process is requires a deep

network for longer time sequences, *Dilated* convolutions are used to increase the visibility of the data between layers. A dilation factor l is applied to create a larger receptive field that skips over a certain number of steps in the input or kernel layers.

$$F *_l k(p) = \sum_{s+(lt)=p} F(s)k(t) \tag{3.3}$$

Where $*_l$ represent the convolution operator that is l dilated. This dilated causal convolution is what creates the *Residual Block*, of which we can create multiple layers as it can be merely defined as a transformations $\mathcal{F}$ which can be activated [14], [16].

C. Pre-Processing

The main TCNN model was trained on 12 patients' EEG data from the MIT-BIH Polysomnographic data set. Additionally, one patient's data was used for validation (patient 14) and the remaining two were utilized for our test set (patients 41 and 66). One patient (patient 37) was dropped entirely as it appeared that this patient's data were not reliable. No standard filters were applied to our data as the convolutional layers of our model are designed to learn the best filters for the problem. The mean and standard deviation of the EEG data across the training patients was determined and used to standardize the validation and test sets as well. This methodology permitted the keeping of all the data on the same scale while preventing any data leakage from the validation or test sets. As is standard practice, after hyper-parameter tuning we allowed the model to train on the 12 original train patients as well as patient 14 in order to maximize the use of limited data.

D. Model Architecture

Our proposed architecture follows the work of Ingolfsson et al. and their EEG-TCNet [17]. Our architecture consists of five blocks with three inital convolution layers followed by a TCNN with two layers, and finally a Fully Connected Layer with a $softmax$ activation.

In contrast to the aforementioned EEG-TCNet which utilizes Depthwise and Separable convolutions, the proposed approach exclusively employs 1-Dimensional Convolutions in Blocks 1, 2, and 3. This deliberate choice is motivated by the efficacy of 1D Convolutions in capturing temporal dependencies within EEG signals. While Depthwise and Separable convolutions offer advantages in reducing trainable parameters and capturing spatial relationships, an emphasis on 1D Convolutions stems from their suitability for analyzing single-channel EEG data, where the primary focus lies on discerning temporal patterns. This approach ensures that the model architecture aligns closely with the unique characteristics of single-channel EEG signals. Furthermore, to mitigate overfitting, techniques such as Batch Normalization, ELU Activation, and Dropout layers were employed. These measures serve to stabilize the learning process and enhance the model's generalization ability. Additionally, the incorporation of an average pooling layer effectively reduces the sampling rate of the signal, thereby reducing the amount of data needed in subsequent blocks. After the initial generation of learnable feature maps F_i, in the convolutional blocks, the integration of a TCNN in the fourth block enables an exploration of temporal relationships within the data. Finally, in the classification block, there is a Fully-Connected layer by a $softmax$ activation. These layers can be found in Table II.

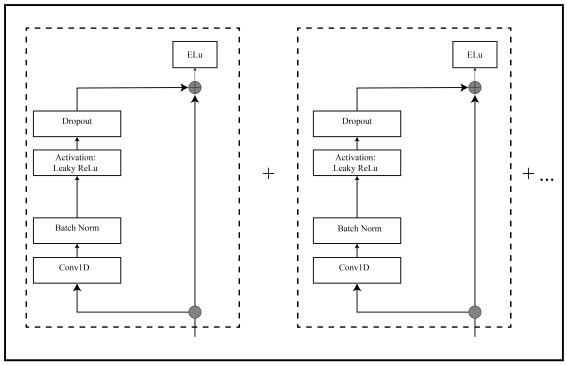

Fig. 1: Temporal Convolutional Neural Network (TCNN) Architecture. Each dashed box represents a residual block.

E. Training Procedure

We selected the Adam stochastic optimization algorithm [18] for the optimization of our model. This choice was predicated on Adam's advanced approach to regularization and adaptive learning rate capabilities. The defining equation of Adam is given by:

$$\theta_{t+1} = \theta_t - \eta_t \cdot \left(\frac{1}{\sqrt{\hat{v}_t} + \epsilon} + \hat{m}_t \right) \tag{3.4}$$

where θ_t denotes model parameters, η is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected estimates of gradient moments, ϵ a small scalar for numerical stability, and λ the weight decay. In addition to utilizing the Adam optimizer, class weights were assigned to counter-act any class imbalances in the dataset.

While previous works used scaled factors based around EEG attributes to determine feature map (F_1, F_2, F_3), kernel (k_1, k_2, k_3) and pooling layer $(Pool_1, Pool_2, Pool_3)$ sizes [17], to discover the best hyper-parameters for our model, a Hyperband tuner was utilized [19]. This permitted searching over a large space of potential hyper-parameters in a relatively short time. The tuner was configured to discover the best hyper-parameters to maximize the AUC of the selected validation patient (patient 14).

IV. Results

In this section, the model results after an intensive tuning process are presented. These results were obtained on test

Block	Layer	# Filters	Kernel Size	# Params	Output	Notes
1	Input				(T, C)	
	Conv1D	$F_1 = 50$	$k_1 = 300$	$(F_1 * k_1) + F_1$	$(T - k_1, F_1)$	Activation = ELU
	BatchNorm			$4 * F_1$		
	MaxPool1D				(T_2, F_1)	$Pool_1 = 12$ $T_2 = T - k_1//Pool_1$
	Dropout					$p = 0.3$
2	Conv1D	$F_2 = 80$	$k_2 = 50$	$(F_2 * k_2) + F_2$	$(T_2 - k_2, F_2)$	Activation = ELU
	BatchNorm			$4 * F_2$		
	MaxPool1D				(T_3, F_2)	$Pool_2 = 4$ $T_3 = T_2 - k_2//Pool_2$
	Dropout					$p = 0.3$
3	Conv1D	$F_3 = 50$	$k_3 = 20$	$(F_3 * k_3) + F_3$	$(T_3 - k_3, F_3)$	Activation = ELU
	BatchNorm			$4 * F_3$		
	MaxPool				(T_4, F_3)	$Pool_3 = 4$ $T_4 = T_3 - k_2//Pool_3$
	Dropout					
	Flatten				$(T_4 * k_3, 1)$	
TCN					$(T_4 * k_3, 10)$	
4	Dense			$(10 * 2) + 2$	(2)	Activation = $softmax$

TABLE II: Diagram of Model Architecture. T_1 is number of time samples within a given segment, C is number of channels, F_1 is initial number of filter maps, k_1 is initial kernel size, and $Pool_1$ is initial pooling size.

		Predicted	
		Non-Apnea	Apnea
True	Non-Apnea	903	87
	Apnea	102	127

TABLE III: Confusion Matrix

	AUC	Accuracy	Precision	Recall	F1
TCNN	0.75	0.84	0.59	0.55	0.57
LDA		0.82	0.66	0.90	0.77
SVM		0.94	0.64	0.94	0.76
k-NN		0.92	0.65	0.92	0.76
ANN		0.94	0.64	0.94	0.76
CNN	0.80	0.70	0.34	0.82	0.48

TABLE IV: Comparison of proposed model against similar studies' results.

patients (patients 41 and 66). The model achieved an AUC of 0.75 (on the ROC curve shown in Figure 2) and an accuracy of 85%. Table 3 shows the model's confusion matrix with a cutoff of 0.47.

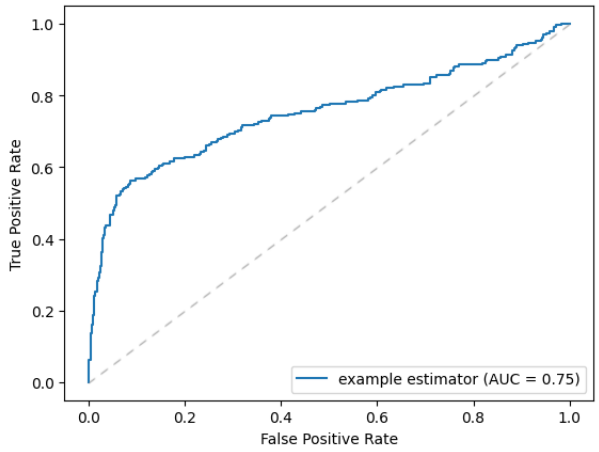

Fig. 2: ROC Curve for Test Patients. This shows an AUC of 0.75. The gray line represents the performance of a classifier if it made predictions at random.

In table IV, We compare these results against the models created and tested by Indrawati et al [20], [21]. There, they experimented on predicting sleep apnea events in the MIT-BIH dataset with various types of statistical models such as Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) and k-Nearest Neighbors (k-NN), along with an Artificial Neural Network (ANN). Additionally in table IV, we looked into the results gathered by Barnes et al. These researchers developed a Convoltional Neural Network (CNN) to classify sleep apnea using data from the Sleep Health Heart Study [2], [22]. This data set is not public and has over 2,000 patients compared to our data which had only 16 (prior to pre-processing). Early in our project's life, we attempted to gain access to the Sleep Health Heart Study data set but ultimately had to move forward with the MIT data.

V. Discussion

The TCNN model faced considerable challenges in achieving the desired level of performance, with various external factors contributing to these limitations. The following discussion highlights these challenges and suggests potential improvements for future research.

The model's ability to generalize was mainly tested by the diversity in EEG channel data within the patient dataset. Predominantly, the data utilized the C3-O1 electrode placement, with less frequent use of C4-A1 and O2-A1. This disparity necessitated a level of cross-channel generalization that proved difficult given the dataset's limited size, comprising EEG data from only 15 patients (2 of which were held out for testing). The model was thus tasked with identifying apnea patterns across disparate channels, which could be crucial for making accurate predictions across new, unseen EEG recordings. Future advancements should aim to refine

the model to better accommodate these variances. This may involve increasing the size and diversity of the dataset and integrating multiple-channel EEG data into our input for a larger search space.

Moreover, the training dataset, while substantial, may not have been sufficiently comprehensive to meet the requirements of the TCNN model, which was designed to learn over 200,000 parameters within its various feature maps and kernels. Unlike strategies employed in other studies, such as those mentioned by Ingolfsson et al., where separate models are generated for each patient [17], our approach trained a singular model using data from multiple patients. This could have impaired the model's capacity to recognize and adjust to the distinct relationships present in sleep apnea across different individuals' persoal EEG distributions. Efforts to enrich the dataset with a greater number of patient records are crucial for enhancing the model's generalization abilities across a broader demographic.

The constraints of the dataset and the pre-segmentation of EEG data into 30-second epochs present significant challenges to the TCNN's temporal resolution. This segmentation may have hindered the model's ability to detect nuanced anomalies indicative of apnea events. Considering an analysis approach that utilizes raw, continuous EEG data could enhance the model's temporal analysis for greater diagnostic accuracy. Furthermore, re-envisioning the problem as an image classification task, where each epoch is an 'image' to be classified, rather than a time sequence to model over could lead to a more effective use of the EEG data different features in future research.

These findings, particularly regarding the use of a singular model and its implications for EEG data interpretation, were similarly reflected in the challenges faced when applying the model to the ISRUC-Sleep dataset, where a high imbalance of non-apnea versus apnea events was discovered [23].

VI. Conclusions

These limitations in TCNN model development discussed emphasize that there is still much progress needed to transition neural networks from a research to a diagnostic tool. These models perform best when analyzing and classifying data that have recognizable EEG distributions, where they can serve as great tools for examining the nuances of individual patients. When forced to generalize, however, the model struggles to recognize sleep apnea across disparate sources. This is mainly due to differences in EEG patterns, frequency of events, length of events, or other varied characteristics of the data; regardless, this capability must be improved before introducing the model as a diagnostic tool.

Acknowledgments

We extend our sincere thanks to Dr. Jianxin Xie of the Xie Research Lab for her sponsorship and support. We are also grateful to our faculty mentor, Dr. John Van Horn, for his invaluable guidance and to the University of Virginia School of Data Science for their academic backing. We would also like to thank Lucia Quirke and Ingolfsson et al. as their previous work on this subject proved vital for our project. Special thanks to Rivanna at UVA for providing the essential computational resources necessary for our research. These contributions have been crucial to our project's success.

References

[1] C. Laratta, N. Ayas, M. Povitz, and S. Pendharkar, "Diagnosis and treatment of obstructive sleep apnea in adults," *Canadian Medical Association Journal*, vol. 189, no. 48, pp. E1481–E1488, 2017. [Online]. Available: https://doi.org/10.1503/cmaj.170296

[2] L. Barnes, K. Lee, A. Kempa-Liehr, and L. Hallum, "Detection of sleep apnea from single-channel electroencephalogram (EEG) using an explainable convolutional neural network (CNN)," *PLoS ONE*, vol. 17, no. 9, 2022. [Online]. Available: https://doi.org/10.1371/journal.pone.0272167

[3] J. Fleetham, N. Ayas, D. Bradley, M. Fitzpatrick, T. Oliver, D. Morrison, F. Ryan, F. Series, R. Skomro, and W. Tsai, "Canadian Thoracic Society 2011 guideline update: Diagnosis and treatment of sleep disordered breathing," *Canadian respiratory journal*, vol. 18, pp. 25–47, 2011. [Online]. Available: https://doi.org/10.1155/2011/506189

[4] G. Jean-Louis, F. Zizi, L. T. Clark, C. D. Brown, and S. I. McFarlane, "Obstructive sleep apnea and cardiovascular disease: role of the metabolic syndrome and its components," *J Clin Sleep Med*, vol. 4, no. 3, pp. 261–272, Jun 2008.

[5] W. W. Flemons, N. J. Douglas, S. T. Kuna, D. O. Rodenstein, and J. Wheatley, "Access to diagnosis and treatment of patients with suspected sleep apnea," *American Journal of Respiratory and Critical Care Medicine*, vol. 169, no. 6, pp. 668–672, 2004.

[6] Y. Wang, Z. Xiao, S. Fang, W. Li, J. Wang, and X. Zhao, "Bi-directional long short-term memory for automatic detection of sleep apnea events based on single channel eeg signal," *Computers in Biology and Medicine*, vol. 142, 2022.

[7] T. Young, L. Evans, L. Finn, and M. Palta, "Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women," *Sleep*, vol. 20, no. 9, pp. 705–706, 1997.

[8] D. Liu, Z. Pang, and S. R. Lloyd, "A neural network method for detection of obstructive sleep apnea and narcolepsy based on pupil size and eeg," *IEEE Transactions on Neural Networks*, vol. 19, no. 2, pp. 308–318, 2008.

[9] K. Hla, T. Young, E. Hagen, J. Stein, L. Finn, F. Nieto, and P. Peppard, "Coronary heart disease incidence in sleep disordered breathing: the Wisconsin Sleep Cohort Study," *Sleep*, vol. 38, no. 5, pp. 677–684, 2015. [Online]. Available: https://doi.org/10.5665/sleep.4654

[10] Massachusetts Institute of Technology/Beth Israel Hospital, "MIT-BIH Polysomnographic Database," https://doi.org/10.13026/C23K5S, 1999.

[11] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin Neurophysiol*, vol. 112, no. 4, pp. 713–719, Apr 2001.

[12] J. Britton, L. Frey, J. Hopp, P. Korb, M. Koubeissi, W. Lievens, E. Pestana-Knight, and E. St. Louis, *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. American Epilepsy Society, 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK390342/

[13] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," 2016.

[14] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[16] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2016.

[17] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, "Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020, pp. 2958–2965.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[19] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Efficient hyperparameter optimization and infinitely many armed bandits," *CoRR*, vol. abs/1603.06560, 2016. [Online]. Available: http://arxiv.org/abs/1603.06560

[20] A. N. Indrawati, N. Nuryani, A. S. Nugroho, and T. P. Utomo, "Obstructive Sleep Apnea Detection using Frequency Analysis of Electrocardiographic RR Interval and Machine Learning Algorithms," *J Biomed Phys Eng*, vol. 12, no. 6, pp. 627–636, Dec 2022.

[21] N. Chinchor, "Muc-4 evaluation metrics," in *Proceedings of the 4th Conference on Message Understanding*, ser. MUC4 '92. USA: Association for Computational Linguistics, 1992, p. 22–29. [Online]. Available: https://doi.org/10.3115/1072064.1072067

[22] S. F. Quan, B. V. Howard, C. Iber, J. P. Kiley, F. J. Nieto, G. T. O'Connor, D. M. Rapoport, S. Redline, J. Robbins, J. M. Samet, and W. Wahl, "The Sleep Heart Health Study: Design, Rationale, and Methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, 12 1997. [Online]. Available: https://doi.org/10.1093/sleep/20.12.1077

[23] S. Khalighi, T. Sousa, J. Santos, and U. Nunes, "ISRUC-Sleep: A comprehensive public dataset for sleep researchers," *Computer Methods and Programs in Biomedicine*, vol. 124, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.cmpb.2015.10.013