Automatic Sleep Stage Scoring on Raw Single-Channel EEG - A comparative analysis of CNN Architectures

Nirali Parekh*, Bhavi Dave†, Raj Shah‡ and Kriti Srivastava§
Department of Computer Engineering
Dwarkadas J. Sanghvi College of Engineering
Mumbai, India
*nirali25parekh@gmail.com, †bhavidave5@gmail.com, ‡rajshah2320@gmail.com, §kriti.srivastava@djsce.ac.in

***Abstract*—Medicine has long reached an overwhelming consensus on the importance of sleep in maintaining mental and physiological homeostasis, and the link that sleep disruption has with both disease and mortality. With the advent of the domain of HealthTech, Deep Learning approaches have generated State Of The Art performance in solving several problems in the medicinal arena. The study of sleep- Polysomnography- uses Electroencephalogram (EEG) readings, among other parameters, to gain a clearer picture of a patient's sleep patterns since different brain activities correspond to different stages of sleep. Monitoring and interpreting EEG signals and the body's reactions to the changes in these cycles can help identify disruptions in sleep patterns. Successfully classified sleep patterns can in turn help medical professionals with the prognosis of several pervasive sleep related diseases like sleep apnea and seizures. To address the pitfalls associated with the traditional manual review of EEG signals that help classify sleep stages, in this work, several Convolutional Neural Networks were trained and analysed to classify the five sleep stages (Wake, N1, N2, N3, N4 and REM by AASM's standard) using data from raw, single channel EEG signals. With PhysioNet's Sleep-EDF dataset, this comparative analysis of the performance of popular convolutional neural network architectures can serve as a benchmark to the problem of sleep stage classification using EEG signals. The analysis shows that CNN based methods are adept at extracting and generalizing temporal information, making it suitable for classifying EEG based data.**

***Index Terms*—Sleep stage scoring, EEG analysis, Deep Learning, Transfer Learning, Convolutional Neural Network**

I. Background and Motivation

Sleep plays a principal role in mental and physiological homeostasis. Sleep related diseases like insomnia and sleep apnea reduce the quality of life for the multitude of people who are affected. As connections between sleep disruption and both disease and mortality have become more firmly established in [1], accurate diagnosis of sleep disorders have become increasingly critical. Observing sleep cycles and observing body's reactions to the changes in these cycles, can help identify disruptions in sleep patterns. Typically, all-night polysomnographic (PSG) recording consisting of an electrooculogram (EOG), electroencephalogram (EEG) and electromyogram (EMG) are analyzed for determining the quality of sleep. This PSG is divided into segments of 30-second recordings, known as epochs, which are then be classified into different sleep stages by the experts by following established guidelines such as the Rechtschaffen and Kales (R & K) [2] and the American Academy of Sleep Medicine (AASM) [3]. This process is known as sleep stage scoring or sleep stage classification. In the AASM classification, there exists a wake stage, referred to as stage W, three Non-REM sleep stages named as N1, N2, and N3, with N3 reflecting slow wave sleep and one REM sleep stage, referred to as stage R. This procedure of sleep scoring is, however, manual and hence labor-intensive and time-consuming. Experts must visually inspect all epochs and label their sleep stages of entire PSG recordings. Thus, automatic sleep scoring for healthcare and well-being is in high demand.

Deep Learning's efficiency in handling large amounts of data and the power to learn hidden features automatically is what makes it popular in numerous domains, and has hence received considerable attention from the sleep research community. Deep Learning can aid physicians at hospitals and health systems in sleep scoring by providing them with real-time, automatic classification that they can alter and monitor based on their personal expertise. Many sleep scoring methods using Deep Learning for automatic classification of PSG data have been proposed. Some works have been discussed in the section II.

II. Literature Review

A. Review of the Existing Work

Various review works have been focussed on the use of EEG waves in the problem of sleep scoring techniques. In [17], the authors have reviewed 154 papers that apply DL to EEG, spanning different application domains such as epilepsy, sleep, brain–computer interfacing, and cognitive and affective monitoring. [18] is a review of automated sleep stage scoring systems since the year 2000. They analyse the systems that were developed for Electrocardiogram (ECG), Electroencephalogram (EEG), Electrooculogram (EOG), and a combination of signals.

[9] and [13] proposed DeepSleepNet and SleepEEGNet respectively, using CNNs for feature extraction and BiRNNs

TABLE I: Literature Review Analysis

Ref	Dataset Used	Architecture	Accuracy
[4]	MASS	SeqSleepNet - Hierarchical RNN	87.1%
[5]	Sleep-EDF	Transfer Learning Using CNN	84.3%
[6]	Sleep-EDF, MASS and SHHS	IITNet - transfer learning + bidirectional LSTM	Sleep-EDF: 83.9% MASS: 86.5% SHHS: 86.7%
[7]	Records from Massachusetts General Hospital Sleep Laboratory	SleepNet - RNN	85.76%
[8]	MASS	SVM	79.7%
[9]	Sleep-EDF and MASS	DeepSleepNet - CNN	MASS: 86.2%, Sleep-EDF: 82.0%
[10]	Sleep-EDF	Decision Tree	89.06%
[11]	Sleep-EDF	CNN + DNN with time-frequency image features	82.6%
[12]	Sleep-EDF	CNN + Temporal CNN + Conditional Random Field Layer	85.39%
[13]	Sleep-EDF	SleepEEGNet - CNN	84.26%
[14]	Records from Charite Clinic in Berlin	CNN + LSTM	40%
[15]	Sleep-EDF and MASS	CNN	Sleep-EDF: 82.3% MASS: 83.6 %
[16]	Sleep-EDF	bidirectional RNNs with attention + SVM	82.5%

for capturing temporal information. In [12] a CNN was similarly used, alongside Temporal Convolutional Neural Network (TCNN).

[7] deployed an annotation tool for sleep staging. SleepNet uses a RNN trained on the largest sleep physiology database, consisting of PSGs from over 10,000 patients from the Massachusetts General Hospital (MGH) Sleep Laboratory. The most successful model uses expert-defined features to represent 30 second intervals, annotating them with a RNN. [10] proposes the development of an Automatic Sleep Stage Classification (ASSC) system for detecting sleep stages using simple statistical features ideal for being implemented in an embedded device in real time. [5] presented a deep transfer learning approach to address the problem of insufficient data in many sleep studies so as to improve automatic sleep staging performance on small cohorts.

The authors of [14] perform sleep stage detection by using heartbeat signals, respiratory signals, and movement signals. They employ two different neural networks for classification: CNN and LSTM. In the paper [16], the authors used deep bidirectional RNNs with attention for single-channel sleep stage classification. The network works as a feature extractor to generate a high-level feature vector that is then given to an SVM for the classification.

A hierarchical recurrent neural network is proposed in [4] treated automatic sleep staging as a sequence-to-sequence classification problem for jointly classifying a sequence of multiple epochs at once. In the paper [6], inter- and intra-epoch temporal contexts were cosidered using raw single-channel EEG to effectively classify the time-series inputs. [11] proposes an efficient and simplified CNN that is capable of learning features at a multitude of temporal resolutions while capturing time shift-invariance property of EEG signals because of its 1-max pooling layer. A Mixed Neural Network (MNN) is proposed [8] that simultaneously aims to target the concerns of population heterogeneity and temporal pattern recognition. The authors of [15] used a novel CNN framework for sleep stage classification that simultaneously determined the classification label of the current epoch and the neighbouring epoch's prediction in the contextual output.

Table I summarizes the existing work discussed in this section.

B. Research Gaps

1) The review clearly demonstrates that most research uses only single-channel EEG signals to make classifications, and does not utilize all the available signal data including Fpz and Pz. The paper hypotheses that these signals contain valuable information about the temporal and spatial features of the waves and using all the signals can boost the accuracy.
2) The existing architectures are extremely bulky, requiring numerous layers and heavy preprocessing that inevitably increases the time required for training. Transfer learning can be invaluable in such a scenario.

III. Methodology

A. Dataset

In this study, two cohorts from the Physionet Sleep-EDF Extended [19] dataset contributed in 2018 are used for experimentation. The dataset contains PSG records and their corresponding sleep stages labeled by human sleep experts. These adopted cohorts [19], [20] present diverging health disorders which are:

1) Sleep-EDF-SC: This is the Sleep Cassette (SC) cohort of the Sleep-EDF Expanded dataset consisting of 153 recordings of 77 subjects were recorded. Two subsequent day-night PSG recordings were collected for each subject.
2) Sleep-EDF-ST: This is the Sleep Telemetry (ST) cohort of the Sleep-EDF Expanded dataset which was collected for studying the temazepam effects on sleep. It contains recordings of 22 subjects with mild difficulty falling asleep. Manual annotation was done similar to the Sleep-EDF-SC subset.

TABLE II: Demographic information of the cohorts of Sleep EDF dataset

Dataset	Avg. epochs	Category	No. of subjects	Age mean	Age min	Age max
Sleep Cassette	2650	male	36	59.3	26	97
		female	41	58.5	25	101
		total	77	58.9	25	101
Sleep Telemetry	2453	male	7	35.71	20	60
		female	15	50.85	18	79
		total	22	40.18	18	79

Table II summarizes the demographic information of datasets, including gender distributions and age characteristics. In each of the cohorts, each of the 30-second PSG epoch were manually labelled into one of eight categories W, N1, N2, N3, N4, REM, MOVEMENT, UNKNOWN by sleep experts according to the R &K standard [2]. Like some of the previous works [9], [11], [12], [15], N3 and N4 stages were merged into a single stage N3 and MOVEMENT and UNKNOWN categories were excluded to make it compliant to AASM standards [3]. The experiments were performed on the Fpz-Cz EEG and Pz-Oz EEG channels in this study.

B. Data Exploration

Using the Visbrain tool [21], the visualizations of the sample waveforms of different stages of sleep in an epoch of 30 seconds are shown in Figure 1.

C. Preprocessing

The recordings are given as EDF [22] files for each subject. One EDF file represents one PSG record. For the scope of this work, only two signals were needed from the original files: Fpz-Cz EEG, and Pz-Oz EEG signals. By using pyEDFlib library [23], the mentioned signals were extracted and processed. Signals from each subject were divided into time sequences of 3000 timesteps, which, with the frequency of 100 Hz, corresponds to 30 seconds epochs. The tabular dataset obtained for each record is shaped like a three dimensional vector as following: (n epochs, 3000 timesteps, 2 signals) where the 2 signals are Fpz-Cz and Pz-Oz.

The obtained dataset was highly unbalanced with a high number of instances i.e. epochs of Wake stage. To ensure a balanced and representative dataset across all the five classes, Wake, N1, N2, N3 and REM, undersampling has been performed ensuring that 1000 epochs for each class were utilized for training and 50 epochs per class were used for testing.

Due to CNN's commendable performance for processing and extracting the most important features of high dimensional data, this paper uses CNN architectures for experimentation as discussed in section IV Since the obtained dataset is tabular signal data, which isn't formatted as an image, transformation methods have been applied to apply CNN models. So, the sampled epochs obtained in the previous steps are first converted into signal images of grayscale images using such transformations. A couple sample images obtained are referenced in Figure 2. This preprocessing step helps to detect hidden features and topologies like homogeneity, periodicity, driftness and dispersion when dealing with time series like EEG signal data.

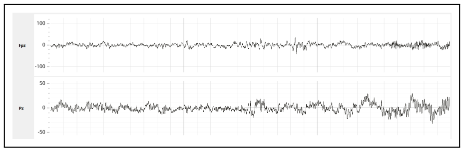
Thus, our image data is ready to be fed into the proposed Deep Learning architectures for training the models.

IV. Experimental Architectures

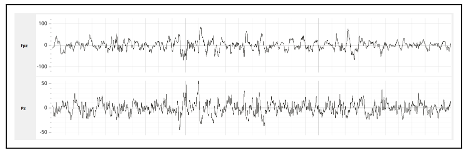
A convolutional neural network is most commonly used to gain relevant, abstract and location invariant features that are accurate depictions for recognition problems. CNNs greatly reduce the number of parameters because of sparse-connectivity and weight sharing, helping achieve a similar degree of complexity with a significantly smaller number of parameters. This paper performed experiments on several CNN architectures of varying complexity to analyse the effects of the different design features to the problem statement.The experiments were run with the pretrained models trained on the ImageNet dataset.The last layer was modified to reflect the five sleep stages. With the dataset being constant, a comparative analysis of the performance of popular CNN architectures can serve as a benchmark to the problem of sleep stage classification using EEG signals. The architectures chosen vary in depth, parameters, design, complexity and efficiency. A comparative analysis of this performance can help give insight to the aspects that will be most relevant to an ideal solution. The following section provides an overview to the key features of the various CNNs used.

A. AlexNet

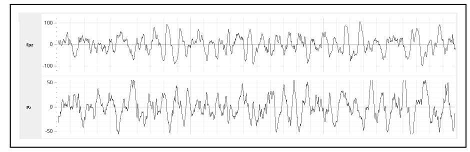
AlexNet [24] takes in images that are passed through 5 alternating convolution layers and 3 fully connected layers, with the last fully connected layer having neurons equivalent to the number of classes. The count of the total number of layers in the architecture does not include the max-pooling layers as



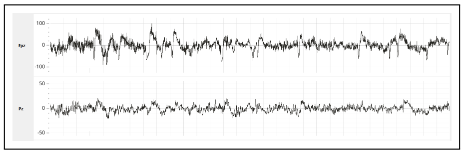
(a) N1 sleep stage



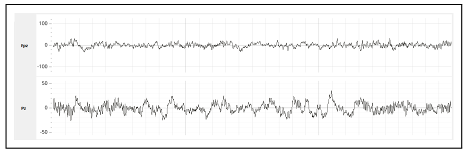
(b) N2 sleep stage



(c) N3 sleep stage

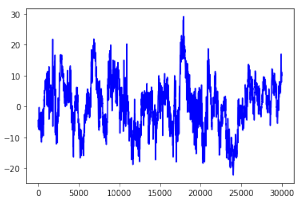


(d) Wake sleep stage

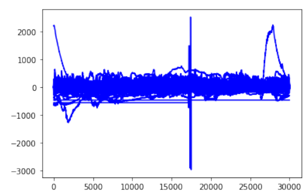


(e) REM sleep stage

Fig. 1: Raw EEG waveforms for each of the sleep stages



(a) Sleep Cassette sample



(b) Sleep Telemetry sample

Fig. 2: Samples of plotted signals of sleep stage 'REM' after raw data converted to graphs from Sleep Cassette and Sleep Telemetry cohorts

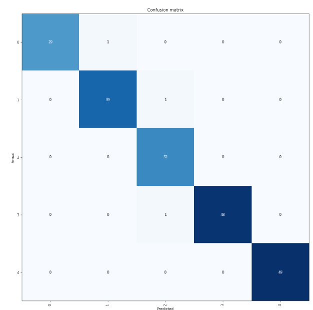
they do not carry weights. Thus, AlexNet has 8 layers. AlexNet was proposed in 2012 with a key contribution was using the ReLU activation function instead of tanh or sigmoid to achieve faster training times.

B. VGG

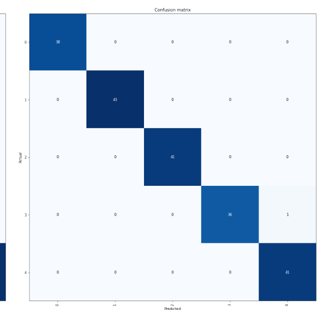
During the design of the VGGNet [25], it was found that alternating convolution & pooling layers were not required, and consequently VGG uses multiple Convolutional layers in sequence with pooling layers in between. The VGG architecture has several versions like the VGG-16 (a 16 layer network), VGG-19 (a 19 layer network) and so forth. VGG demonstrated a solution to the pertinent problem of increasing depth of a CNN - the small-size convolution filters in the architecture allowed VGG to have a large number of weight layers that increased depth and improved performance.

C. ResNet

As the depth of the VGG network was increased, the model performance began to deteriorate. It was hypothesised that the gradients were not able to flow well through the deeper network. [26]. In ResNet after every two layers the input given to the first layer along with the output obtained at the second layer. The information from the input that was earlier getting highly morphed and by the time it reached the output layers is now passed as a residue of the input once again



(a) VGG19 for Fpz-Cz



(b) MobileNetV2 for Pz-Oz

Fig. 3: Confusion Matrices for best performing models for Fpz-Cz (VGG19) and Pz-Oz (MobileNet-v2) channels of Sleep Cassette Cohort

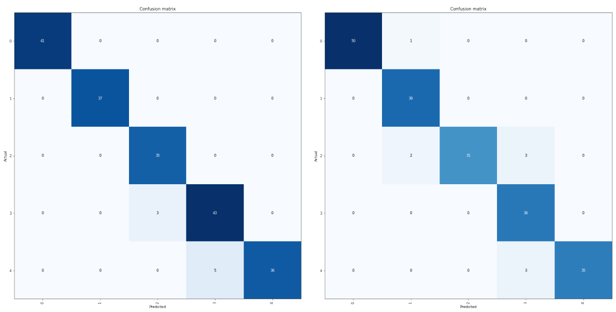
with the output. This helped the gradients to flow back better, improving training.

D. DenseNet

DenseNets [27] connect every layer to every other layer, implying that for L layers, there are L(L+1)/2 direct connections. Every layer uses the feature maps of all the preceding layers as inputs, and consequently its output feature maps are used as input for subsequent layers. The network is divided into densely connected blocks within which the feature map size remains the same. This facilitates both downsampling and feature concatenation. DenseNets require fewer parameters than a comparable size traditional CNN as the need to learn redundant feature maps is eliminated. It also facilitates the flow of information and gradient as each layer has direct access to the input and the gradient of the loss function.

E. Squeezenet

SqueezeNets [28] attempt to reduce the number of parameters in the network by using Fire modules which have a squeeze layer of 1x1 convolutions that can decrease parameters by restricting the number of input channels in every layer. SqueezeNets are therefore very low latency. They achieve



(a) ResNet50 for Fpz-Cz (b) MobileNetV2 for Pz-Oz

Fig. 4: Confusion Matrices for best performing models for Fpz-Cz (ResNet50) and Pz-Oz (MobileNet-v2) channels of Sleep Telemetry Cohort

AlexNet level accuracies while having 50 million fewer parameters.

F. MobileNet

MobileNets [29] utilise Depth convolutions and point convolutions, reducing the comparison and recognition time. They are designed to increase accuracy while simultaneously factoring the resource constrained environments of embedded devices or mobile phones, therefore getting their name. They also reduce the number of parameters and hence latency. To construct an even smaller and computationally cheap model, MobileNets also have useful model-shrinking parameters that help make the network thinner in a uniform manner at every layer.The architecture helps explore a reasonable accuracy, latency and size trade off.

V. Results and Discussions

A. Evaluation measures

After training, the model has been evaluated on the test set that it has never seen before. The model was trained on 80% of the data and tested on 20% of the complete data set.

1. Accuracy: It is the fraction of classifications the model got correct.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

2. Precision: Proportion of the positive identifications which are correct.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

3. Recall: Proportion of actual positives which were identified correctly.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

4. F1-score: It is a weighted average of Precision and Recall. It is calculated as the harmonic mean of the two.

$$F1- \text{ score } = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

B. Results

The architectures discussed in section IV were used to classify the data into 5 classes - the Wake stage, N1 stage, N2 stage, N3 stage and REM stage. The results are shown in Table III, using previously described measures. All metrics were calculated on test data. For every model and for each classification, learning accuracy reached around 94-95%. Comparing the models, for the Sleep-EDF-SC cohort, the best performing model for Fpz-Cz is VGG-19 and for Pz-Oz, it is MobileNet-v2. Similarly, for the Sleep-EDF-ST cohort, ResNet50 and MobileNet-v2 shows the best performance for Fpz-Cz and Pz-Oz channels respectively. Also, SqueezeNet 1.1 has the worst performance in comparison to the other models for both channels in SC and ST cohorts.

Figure 3 presents the confusion matrices processed by the best performing models using the Fpz-Cz and Pz-Oz channels respectively of the Sleep-EDF-Sleep Cassette Cohort. Similarly,Figure 4 shows the confusion matrices of the best models using the Fpz-Cz and Pz-Oz channels of the Sleep-EDF Sleep Telemetry cohort. The main diagonals in each confusion matrix denote the true positive (TP) values which indicate the number of stages scored correctly. It can be seen from the tables (the confusion matrices' parts) that TP values are significantly higher than other values in the same rows and columns, indicating notable performance of the models.

VI. Conclusion

The work has implemented and automated the sleep staging pipeline- when given raw, single channel EEG data of a patient i.e. Fpz-Cz and Pz-Oz EEG signals, the datapoint is classified into one of the five sleep stages (Wake, N1, N2, N3, N4 and REM) using several different CNN architectures. The preprocessing and sampling implemented helps avoid model bias, ensuring that the models evaluate sleep scoring reliably according to the new AASM standards.

The pipeline utilizes CNNs to extract time-invariant features within sleep stages from EEG epochs. The results also showed that the temporal information learned from the sequence residual learning helped improve the classification performance. The experiments clearly demonstrated that the models could learn features for sleep stage scoring from different raw single-channel EEGs. The paper also provides an exhaustive overview of the existing literature in the domain, allowing researchers to review the possible approaches of solving the problem. The work also benchmarked the performance of several popular CNN architectures on a consistent sleep stage classification dataset. A comparative analysis of the various CNN architectures demonstrates that a CNN based approach is a fitting solution for EEG data. This can serve as a step on tangible progress in the domain of sleep stage classification using EEG as well as AI in medicine.

References

[1] S. L. Worley, "The extraordinary importance of sleep: the detrimental effects of inadequate sleep on health and public safety drive an explosion of sleep research," *Pharmacy and Therapeutics*, vol. 43, no. 12, p. 758, 2018.

TABLE III: Result metrics of various experimental architectures

Dataset	Channel	Metric	Model					
			ResNet50	MobileNetV2	AlexNet	DenseNet121	VGG19	SqueezeNet 1.1
Sleep-EDF Sleep Cassette (SC)	Fpz-Cz	Accuracy	0.96	0.97	0.94	0.96	**0.98**	0.94
		Precision	0.97	0.98	0.95	0.97	**0.98**	0.96
		Recall	0.96	0.97	0.94	0.96	**0.98**	0.94
		F1-score	0.96	0.98	0.94	0.96	**0.98**	0.95
	Pz-Oz	Accuracy	0.98	**0.99**	0.98	0.97	0.98	0.96
		Precision	0.99	**1.00**	0.99	0.97	0.98	0.97
		Recall	0.98	**0.99**	0.98	0.98	0.99	0.96
		F1-score	0.99	**0.99**	0.98	0.97	0.98	0.96
Sleep-EDF Sleep Telemetry (ST)	Fpz-Cz	Accuracy	**0.96**	0.94	0.95	0.95	0.96	0.94
		Precision	**0.96**	0.93	0.95	0.95	0.95	0.93
		Recall	**0.95**	0.94	0.95	0.96	0.95	0.93
		F1-score	**0.95**	0.93	0.95	0.95	0.95	0.93
	Pz-Oz	Accuracy	0.94	**0.95**	0.95	0.94	0.94	0.91
		Precision	0.94	**0.96**	0.95	0.95	0.95	0.93
		Recall	0.94	**0.95**	0.95	0.94	0.93	0.91
		F1-score	0.94	**0.95**	0.95	0.94	0.94	0.91

[2] A. Rechtschaffen, “A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects,” *Brain information service*, 1968.

[3] C. Iber, “The aasm manual for the scoring of sleep and associated events: Rules,” *Terminology and Technical Specification*, 2007.

[4] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, “Seqsleepnet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 3, pp. 400–410, 2019.

[5] H. Phan, O. Y. Chén, P. Koch, Z. Lu, I. McLoughlin, A. Mertins, and M. De Vos, “Towards more accurate automatic sleep staging via deep transfer learning,” *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 6, pp. 1787–1798, 2020.

[6] H. Seo, S. Back, S. Lee, D. Park, T. Kim, and K. Lee, “Intra-and inter-epoch temporal context network (iitnet) using sub-epoch features for automatic sleep scoring on raw single-channel eeg,” *Biomedical Signal Processing and Control*, vol. 61, p. 102037, 2020.

[7] S. Biswal, J. Kulas, H. Sun, B. Goparaju, M. B. Westover, M. T. Bianchi, and J. Sun, “Sleepnet: automated sleep staging system via deep learning,” *arXiv preprint arXiv:1707.08262*, 2017.

[8] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, and Y. Guo, “Mixed neural network approach for temporal sleep stage classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 324–333, 2017.

[9] A. Supratak, H. Dong, C. Wu, and Y. Guo, “Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.

[10] K. A. Aboalayon and M. Faezipour, “Real time sleep detection system using new statistical features of the single eeg channel,” 2017.

[11] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, “Dnn filter bank improves 1-max pooling cnn for single-channel eeg automatic sleep stage classification,” in *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2018, pp. 453–456.

[12] E. Khalili and B. M. Asl, “Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel eeg,” *Computer Methods and Programs in Biomedicine*, vol. 204, p. 106063, 2021.

[13] S. Mousavi, F. Afghah, and U. R. Acharya, “Sleepeegnet: Automated sleep stage scoring with sequence to sequence deep learning approach,” *PloS one*, vol. 14, no. 5, p. e0216456, 2019.

[14] K. Stuburić, M. Gaiduk, and R. Seepold, “A deep learning approach to detect sleep stages,” *Procedia Computer Science*, vol. 176, pp. 2764–2772, 2020.

[15] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. De Vos, “Joint classification and prediction cnn framework for automatic sleep stage classification,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1285–1296, 2018.

[16] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, “Automatic sleep stage classification using single-channel eeg: Learning sequential features with attention-based recurrent neural networks,” in *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2018, pp. 1452–1455.

[17] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, “Deep learning-based electroencephalography analysis: a systematic review,” *Journal of Neural Engineering*, vol. 16, no. 5, p. 051001, aug 2019. [Online]. Available: https://doi.org/10.1088/1741-2552/ab260c

[18] O. Faust, H. Razaghi, R. Barika, E. J. Ciaccio, and U. R. Acharya, “A review of automated sleep stage scoring based on physiological signals for the new millennia,” *Computer methods and programs in biomedicine*, vol. 176, pp. 81–91, 2019.

[19] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye, “Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg,” *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185–1194, 2000.

[20] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals,” *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[21] E. Combrisson, R. Vallat, J.-B. Eichenlaub, C. O’Reilly, T. Lajnef, A. Guillot, P. M. Ruby, and K. Jerbi, “Sleep: an open-source python software for visualization, analysis, and staging of sleep data,” *Frontiers in neuroinformatics*, vol. 11, p. 60, 2017.

[22] B. Kemp, A. Varri, A. C. Rosa, K. D. Nielsen, and J. Gade, “A simple format for exchange of digitized polygraphic recordings,” *Electroencephalography and clinical neurophysiology*, vol. 82, no. 5, pp. 391–393, 1992.

[23] H. Nahrstaedt, “pyedflib,” https://github.com/holgern/pyedflib, 2020, accessed July 10, 2021.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[25] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[26] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[27] G. Huang, Z. Liu, and K. Q. Weinberger, “Densely connected convolutional networks,” *CoRR*, vol. abs/1608.06993, 2016. [Online]. Available: http://arxiv.org/abs/1608.06993

[28] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and $<$1mb model size," *CoRR*, vol. abs/1602.07360, 2016. [Online]. Available: http://arxiv.org/abs/1602.07360

[29] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017. [Online]. Available: http://arxiv.org/abs/1704.04861