Hisham ElMoaqet*, Mohammad Eid, Mutaz Ryalat, and Thomas Penzel

A CNN-BiLSTM Deep Learning Model for Automatic Scoring of EEG Signals

https://doi.org/10.1515/cdbme-2023-1161

Abstract: Recently, several automatic sleep stage classification methods have been proposed based on deep learning using convolutional (CNN) and recurrent (RNN) neural networks. However, the state of the art CNN methods are still complex which usually require significant time and considerable computational resources in order to set up and sufficiently train a deep CNN from scratch. This study eliminates the need to establish and train a deep CNN from scratch by leveraging a pre-trained deep architecture that has been previously trained from sufficient labeled data in a different context. A convolutional neural network (CNN) and a Bidrectional long short term memory network (BiLSTM) are integrated for automatic feature extraction and sleep stage scoring using only a single-channel EEG signal. To demonstrate the generalizability of our results, the proposed model was evaluated using PSG records of 81 patients that were collected in different environments, through different recording hardware, and annotated with different sleep experts. The use of a single EEG source and a one–to–one classification scheme in the proposed model can allow further development towards wearable systems and online in home monitoring applications.

Keywords: EEG Signal, CNN, BiLSTM, Deep Learning

1 Introduction and Background

Polysomnography (PSG) records are the gold standard used by clinicians to comprehensively evaluate and analyze sleep [1]. PSG data contain multivariate physiological signals such as electrocardiogram (ECG), electromyogram (EMG), electroencephalogram (EEG), and electrooculogram (EOG), in order to monitor different body functions and regions. Each 30-s time slot of PSG data defines a sleep epoch that can be classified into different sleep stages based on clinically defined rules and visual inspection. According to the American Academy of Sleep Medicine (AASM) [1], a sleep cycle can be classified into 5 stages as wakefulness (W), non-rapid eye movement (NREM) sleep, and rapid eye movement sleep (REM). The NREM sleep is further subdivided into three sleep stages referred to as N1, N2, and N3.

The significant advances in deep learning allowed researchers to improve the performance of sleep stage scoring systems by automatically learning feature representations from the PSG signals [2]. Towards this end, many studies used deep learning to develop algorithms for automatic scoring of sleep stages using a subset of PSG signals. This includes multichannel EEGs [3], EEG and EOG [4], as well as EEG, EOG, and EMG [5].

To overcome implementation issues associated with multi-channel scoring systems, many recent studies considered single channel EEG signals as inputs to their models [2, 6–11]. These methods have generally relied on establishing complex deep network from scratch which often requires significant time and considerable computational resources. Moreover, majority of these studies were designed to receive a sequence of EEG epochs as an input to these models in order to classify a target sleep epoch based on surrounding epochs (many–to–one classification) [6–8, 10] or to map an input sequence of multiple sleep epochs to the sequence of their corresponding target labels (many–to-many classification) [9].

This papers presents a novel deep learning model for automatic feature extraction and sleep stage scoring using only a single-channel EEG signal. The proposed pipeline is composed of a convolutional neural network (CNN) combined with a Bidrectional long short term memory network (BiLSTM) in order to automatically extract spatial as well as temporal features from time-frequency representations of successive EEG sleep epochs. The proposed approach eliminates the need to establish and train a deep CNN from scratch by using a pre-trained deep architecture that has been previously trained from sufficient labeled data in a different context. Furthermore, the proposed classification system receives a single EEG epoch as an input at a time and produces a single corresponding output label for the sleep stage, making it convenient of online and realtime sleep monitoring applications.

***Corresponding author: Hisham ElMoaqet,** Department of Mechatronics Engineering, German Jordanian University, 11118 Amman, Jordan, e-mail: hisham.elmoaqet@gju.edu.jo
Mohammad Eid, Department of Medical Engineering, Friedrich-Alexander-Universität Erlangen–Nürnberg, 91052 Erlangen, Germany
Mutaz Ryalat, Department of Mechatronics Engineering, German Jordanian University, 11118 Amman, Jordan
Thomas Penzel, Interdisciplinary Center of Sleep Medicine, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

Tab. 1: Number & Distribution of Sleep Epochs Across Different Sleep Stages

Number (%) of EEG Segments in Sleep Stages					Total
W	N1	N2	N3	REM	
17809 (24.57%)	10912 (15.05%)	30802 (42.49%)	5189 (7.160%)	7784 (10.74%)	72496

2 Materials and Methods

2.1 Data Sets and Preprocessing

The data sets include 81 PSG sleep studies that were performed at two sleep disorder centers in Germany and USA respectively. PSG data for 20 patients were recorded at the Interdisciplinary Center of Sleep Medicine in Charité-Universitätsmedizin Berlin in Berlin, Germany while the remaining 61 PSG data files were collected at the Sleep Disorders Center in the University of Michigan at Ann Arbor, Michigan in USA. These data sets have been collected in different environments and annotated with different sleep experts, which can be used to demonstrate the generalizability of the results in this study.

Sleep staging was carried out by expert clinicians according to recommendations of the AASM [1]. Accordingly, EEG signals were segmented at 30-s fragments such that each segment is labeled as in one of the five possible sleep stages. Finally, the Continuous Wavelet Transform (CWT) was applied to all EEG sleep segments to generate spectrogram images from the time-frequency representations of these segments. 72496 total labeled EEG segments were found across the total 81 subjects in this study. Table 1 provides a summary for the number and distribution of the EEG segments with respect to each of the 5 different sleep stages for the 81 patients in both data sets

2.2 Automatic Scoring System Architecture

The architecture for the proposed automatic sleep scoring system is shown in Figure 1. It considers GoogLeNet CNN as backbone model followed by two bidrectional long short term memory (BiLSTM) layers such that each of them is followed by a dropout layer to avoid overfitting. Finally, a fully-connected layer and a softmax output layer were added to enable classification between the five possible sleep stages.

2.3 Backbone CNN Model

The GoogleNet CNN is a 22 layers deep CNN arranged in a one tall stack with 9 inception modules (annotated by spins in Figure 1). In this study, the GoogLeNet model was first modified by removing the dropout layer, fully connected layer, and softmax layer as these were specific to original use of this model in classifying labels of the ImageNet data set [12]. The inception module is the basic block in the GoogLeNet. Each of the inception modules copies the corresponding input signal and feeds it to four different convolution layers with different kernel sizes in order to capture spatial patterns at different scales. The number of feature maps generated by each of the convolutional and pooling layers is shown before the kernel size in Figure 1. The 6 numbers listed in each of the inception blocks represent the number of feature maps output generated by each of the convolutional layers in the corresponding module. All convolutional layers use the ReLU activation function.

2.4 Bidirectional LSTM Memory Network (BiLSTM)

The BiLSTM network was added to extract temporal features in sleep epochs. It is composed of two BiLSTM layers. The first layer is composed of 2000 BiLSTM elements and the second layer is composed of 200 BiLSTM elements. Each BiLSTM layer is followed by a dropout regularization layer (probability of 50%) to avoid overfitting.

2.5 Training Setup

The ImageNet weights of the GoogleNet transferred layers were utilized as initial values, while all other added (trainable) layers were initialized with random weights. The proposed CNN-BiLSTM model was trained end-to-end using iterative optimization with the backpropagation algorithm to minimize the categorical cross entropy loss. To perform experiments, we used a workstation equipped with 2 Intel Xeon PE5 CPU and a GPU composed of 4 NVIDIA GeForce GTX 1080Ti graphic cards. All methods were implemented on MATLAB R2020a.

2.6 Evaluation Criteria

Evaluation metrics that were used to comprehensively evaluate the performance of the proposed CNN-BiLSTM model are per-class sensitivity (Sn_c), precision (Pr_c), F1-score ($F1_c$), specificity, and accuracy (ACC_c). Furthermore, we considered overall classification metrics including overall accuracy (ACC), macro-average F1 ($MF1$), overall sensitivity (Sn).

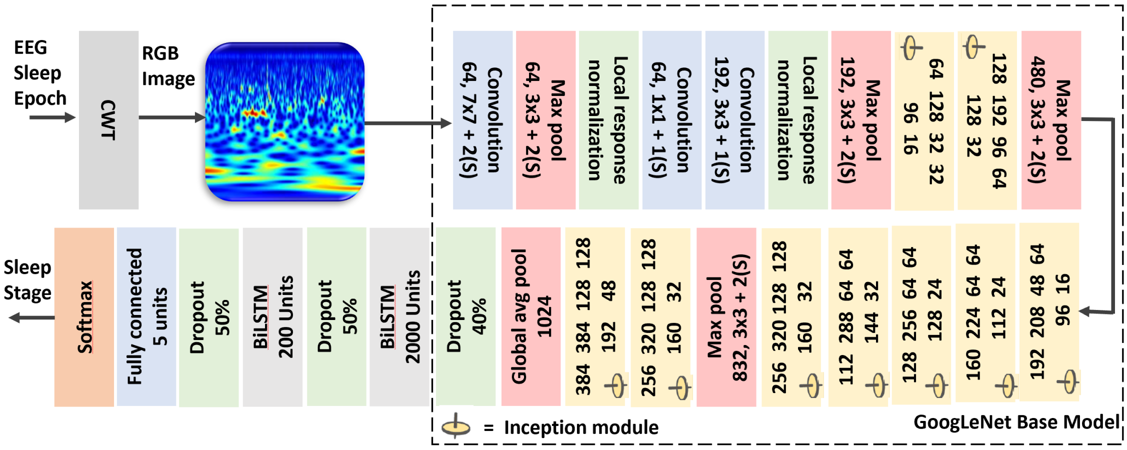

Fig. 1: Architecture of the Proposed CNN-BiLSM Automatic Sleep Scoring System

Tab. 2: 20-fold Per-Class Cross Validation Performance for the Proposed CNN-BiLSTM modeling approach. Results are reported as the mean (std) across the 20-fold results

Class	ACC_c	Sp_c	Sn_c	Pr_c	$F1_c$
W	92.5 (0.4)	96.4 (0.4)	80.5 (1.5)	87.9 (1.2)	84.0 (0.9)
N1	85.1 (0.7)	89.0 (0.7)	63.1 (2.1)	50.5 (1.9)	56.1 (1.9)
N2	85.7 (0.6)	89.4 (0.5)	80.6 (1.0)	84.9 (0.7)	82.7 (0.8)
N3	96.5 (0.2)	97.9 (0.2)	79.2 (2.2)	74.2 (1.9)	76.6 (1.4)
REM	96.1 (0.3)	97.9 (0.2)	81.5 (2.0)	82.0 (1.5)	81.7 (1.4)

3 Results

We considered the technique of oversampling to account for the class imbalance issue in the distribution of the EEG segments across the 5 different sleep stages (as illustrated in Table 1). The proposed model was evaluated using 20-fold cross validation to ensure generalizability and performance consistence over different testing folds. Table 2 reports 20-fold cross validation per-class performance of the proposed CNN-BilSTM model trained with oversampled spectrogram images while being evaluated on original spectrogram images in each of the 20 test folds. The proposed model shows a very high per-class sensitivity in detecting stages W, N2, N3, REM which is accompanied by generally high per-class precision. Stage N1 showed an acceptable but lower performance compared to other stages given that this stage is known to be the most challenging stage to detect [2]. Furthermore, the per-class performance results in Table 2 shows a small standard deviation in the performance across different test folds indicating that the model has an excellent potential to provide a consistent performance over unseen EEG data

It is worth to be mentioned here that using the BiLSTIM network allowed achieving excellent performance in classifying sleep stages which was not possible to achieve using a CNN only. In a previous study, we considered a CNN only to classify sleep-wake states (binary case) and the model achieved excellent performance [13]. However, considering the multi-class classification needed to model the sequential dependency between sleep stages in order to achieve a high classification performance.

The analysis that was carried out to report the performance of the proposed approach used complete data set. Thus, we considered the actual imbalanced class distributions between different sleep stages. Table 3 compares the performance of the proposed approach with the state-of-art deep learning methods that use a single-channel EEG as an input. One important observation to mention is that previous methods differ in the way that was used for handling the imbalance in sleep scoring data which may affect the different performance metrics in real cases. Nonetheless, Table 3 shows that the reported performance of the proposed approach is comparable to the stat of art deep learning methods. Interestingly, the proposed approach shows a relatively high performance compared to other studies in scoring stage N1.

Tab. 3: Performance Comparison Between The Proposed CNN-BiLSTM Method & Previous End-to-End Deep Learning Methods

Study	Subjects	Input Signal	Classification Scheme	Class Imbalance Handling	DL Model	Overall Performance			Per Class Sn (Sn_c)				
						ACC	$MF1$	Sn	W	N1	N2	N3	REM
[10]	42	Fpz-Cz	One to One	Not Reported	Squeezenet CNN	83.6	-	74.4	92.8	28.3	93.3	71.6	84.0
[10]	42	Fpz-Cz	Many to One	Not Reported	Squeezenet CNN	84.7	-	76.7	94.5	36.0	92.7	72.9	86.9
[7]	20	Fpz-Cz	Many to One	Oversampling	CNN+BiLSTM	82.0	76.9	78.7	83.4	50.1	81.7	94.2	83.9
[7]	20	Pz-Oz	Many to One	Oversampling	CNN+BiLSTM	79.8	73.1	-	-	-	-	-	-
[8]	20	Fpz-Cz	Many to One	Subsampling	CNN	81.9	73.8	73.9	-	-	-	-	-
[6]	20	Fpz-Cz	Many to One	Subsampling	CNN	75.0	70.0	73.6	70.0	60.0	73.0	91.0	74.0
[2]	5728	C4-A1	Many to One	Subsampling	CNN	87.0	78.0	77.2	91.0	35.0	89.0	85.0	86.0
[9]	20	Fpz-Cz	Many to Many	Original Sampling	CNN+BiLSTM	84.3	79.7	81.1	90.6	54.5	82.7	88.9	88.7
[11]	197	Fpz-Cz	Many to One	Not Reported	Parallel CNN-GRU	90.1	91.5	-	-	-	-	-	-
This Study	81	C4-M1	One to One	Oversampling	GoogLeNet CNN + BiLSTM	76.9	75.7	77.1	79.0	69.5	77.0	80.4	79.4

4 Conclusion and Future Work

This study proposed a new deep learning based method that integrates CNN and BiLSTM networks for automatic feature extraction and sleep stage scoring for EEG signals. In the proposed method, we used the CNN model for extracting time-invariant spatial features from the time-frequency representations for EEG sleep epochs and the BiLSTM model to extract temporal features in the sleep epochs. We mainly focused on developing a five-stage classification model to classify Wake stage, stage N1, stage N2, stage N3, and REM using data from a single EEG source. The proposed model was tested and evaluated using a patient data sets that were collected using different recording systems in two major sleep centers in USA and Germany respectively. Our results demonstrate an excellent potential for the proposed model to effectively extract representative features and score different sleep stages. Future work may include modifying the backbone CNN model to improve the classification performable as well as evaluating the performance on single-channel EEG data collected from wearable devices.

References

[1] R. Berry, R. Brooks, C. Gamaldo, *et al.*, “The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, version 2.6.0. american academy of sleep medicine, darien, illinois; 2020,” 2020.

[2] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, “A convolutional neural network for sleep stage scoring from raw single-channel eeg,” *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018.

[3] M. Sharma, P. Makwana, R. S. Chad, and U. R. Acharya, “A novel automated robust dual-channel eeg-based sleep scoring system using optimal half-band pair linear-phase biorthogonal wavelet filter bank,” *Applied Intelligence*, pp. 1–19, 2023.

[4] O. Yildirim, U. B. Baloglu, and U. R. Acharya, “A deep learning model for automated sleep stages classification using psg signals,” *International journal of environmental research and public health*, vol. 16, no. 4, p. 599, 2019.

[5] C.-E. Kuo, P.-Y. Liao, and Y.-S. Lin, “A self-attention-based ensemble convolution neural network approach for sleep stage classification with merged spectrogram,” in *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1262–1268, IEEE, 2021.

[6] O. Tsinalis, P. M. Matthews, and Y. Guo, “Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders,” *Annals of biomedical engineering*, vol. 44, no. 5, pp. 1587–1597, 2016.

[7] 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.

[8] 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.

[9] 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.

[10] P. Jadhav, G. Rajguru, D. Datta, and S. Mukhopadhyay, “Automatic sleep stage classification using time–frequency images of cwt and transfer learning using convolution neural network,” *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 494–504, 2020.

[11] I. A. Begawan, E. C. Djamal, D. Djajasasmita, F. Kasyidi, and F. Nugraha, “Sleep stage identification based on eeg signals using parallel convolutional neural network and recurrent neural network,” in *2022 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 39–44, IEEE, 2022.

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, IEEE, 2009.

[13] H. ElMoaqet, M. Eid, M. Ryalat, and T. Penzel, “An end-to-end deep leaning approach for sleep-wake classification using single channel eeg signals,” *Current Directions in Biomedical Engineering*, vol. 8, no. 2, pp. 801–804, 2022.