Automatic Sleep Stage Classification on EEG Signals Using Time-Frequency Representation

Paul Dequidt, Mathieu Seraphim, Alexis Lechervy, Ivan Igor Gaez, Luc Brun, Olivier Etard

Automatic sleep stage classification on EEG signals using time-frequency representation*

Paul Dequidt[1][0000−0002−8362−7735], Mathieu Seraphim[1][0000−0002−9367−1190], Alexis Lechervy[1][0000−0002−9441−0187], Ivan Igor Gaez[2][0000−0003−0529−1209], Luc Brun[1][0000−0002−1658−0527], and Olivier Etard[2][0000−0003−3661−0233]

[1] GREYC, UMR CNRS 6072, Normandie Univ, CNRS, UNICAEN, ENSICAEN, France
[2] COMETE, UMR-S 1075, Normandie Univ, CNRS, UNICAEN, France

Abstract. Sleep stage scoring based on electroencephalogram (EEG) signals is a repetitive task required for basic and clinical sleep studies. Sleep stages are defined on 30 seconds EEG-epochs from brainwave patterns present in specific frequency bands. Time-frequency representations such as spectrograms can be used as input for deep learning methods. In this paper we compare different spectrograms, encoding multiple EEG channels, as input for a deep network devoted to the recognition of image's visual patterns. We further investigate how contextual input enhance the classification by using EEG-epoch sequences of increasing lengths. We also propose a common evaluation framework to allow a fair comparison between state-of-art methods. Evaluations performed on a standard dataset using this unified protocol show that our method outperforms four state-of-art methods.

Keywords: Sleep scoring · time-frequency representation · computer vision · EEG · signal processing.

1 Introduction

Sleep is an important physiological process which can be monitored through polysomnography (PSG). A PSG involves multiple signals, such as electro-encephalogram (EEG), electro-oculogram (EOG) or electro-myogram (EMG). The American Academy of Sleep Medicine (AASM) edited guidelines [3] to classify sleep into different stages based on a 30 second time-frame called an epoch (EEG-epoch). The actual AASM standard identifies 5 stages : wakefulness (W), rapid eye movement (REM) and Non-REM sleep (N1, N2, N3). Sleep studies often need to score each EEG-epoch, which is a tedious task for a human expert. Therefore, sleep scoring could benefit from automation, especially in the case

* This study is co-funded by the Normandy County Council and the European Union (PredicAlert European Project - FEDER fund). Part of this work was performed using computing resources of CRIANN (Normandy, France). This work was performed using HPC resources from GENCI-IDRIS (Grant 2022-102446)

	Delta	Theta	Alpha	Beta_{low}	Beta_{high}	Gamma
Frequencies (Hz)	[0.5; 4[	[4; 8[	[8; 12[	[12; 22[	[22; 30[	[30; 45[

Table 1. Frequency bands for brain activity

of whole night recordings. Frontal EEG are often artefacted by ocular movements similar to the EOG signal. Moreover, the placement of multiple sensors for multi-modal analysis complicates the acquisition stage while the contribution of multimodality to the stage analysis is not clearly established.

Therefore, we want to investigate if we can simplify the acquisition by only using multiple EEG channels. The AASM groups different frequencies into frequency bands as detailed in Table 1. Each band corresponds to specific graphic elements in EEG signal used for sleep scoring. Representation of EEG signals as time-frequency images, or spectrograms, can show the variations in brain activity during sleep depth [12]. One important advantage of such representations is that it enables the use of deep networks devoted to pattern recognition and classification problems on images. Such networks have been subject to intensive investigations for more than 10 years.

2 State of the art

Sleep scoring methods can be analyzed through their input modalities, the computed features on each EEG-epoch, the way they take into account contextual information, or the type of method used in order to obtain the final classification.

Many authors use EEG, EOG and EMG as multimodal inputs for their networks [9, 10, 4, 7, 6, 18], resulting in a scoring aligned with the AASM standard. A classical heuristic [16, 5, 19, 20] consists in substracting the EOG signal to EEG acquisitions. The resulting signal provides interesting classification results but lacks interpretability for experts.

The number of EEG channels varies according to the methods and the datasets on which they are applied. For example, Qu et al. [13] only use one EEG signal while Jia et al. [7] use up to 20 EEG signals. Using more electrodes increases the number of input signals and should improve the scoring. However, increasing the number of electrodes increases the complexity of acquisition and analysis. In addition, electrodes become closer together on the skull as the number of electrodes increases, which enhances interference between signals from nearby electrodes. The optimal number of electrodes is still an open question.

The benefits of spectral analysis is recognized by sleep researchers as early as the 1980s [15]. However, spectral estimation based on Fourier transform assumes the signal is infinite, periodic and stationary. Therefore, used on EEG signals, which are finite, aperiodic and non-stationary, the spectrogram can be artefacted. Multitaper convolutions, or Tapers, have been proven to reduce this bias [12]. Tapers-based spectrograms have been used by Vilamala et al. [21] in conjunction with the VGG-16 convolutional neural network (CNN), on a 5 EEG-epoch sequence on a single EEG signal. VGG-16 takes (3,224,224) RGB images

as input. Vilamala et al. transform the spectrogram with a colormap; then use a VGG-16 pretrained on the ImageNet dataset, but with parameters in the fully connected layers randomized, and all weights unfreezed in their best performing experiments.

Manual sleep scoring often involves some form of contextual input. Many studies test the impact of using a temporal sequence of EEG-epochs as input of their network, instead of a single EEG epoch [10, 16, 5, 11]. The length of sequence differs greatly for one author to another. For example, Dong et al. [5] tried sequences from 1 to 6 EEG-epochs and found their best results between 2 and 4 EEG-epochs, while Phan et al. [10] found that sequences greater than 10 EEG-epochs had minimal impact on classification. Therefore, networks can often be divided into two sections: one that extracts features within each EEG-epoch (intra-epoch), and another that compares the features from neighboring EEG-epochs to get a better classification (inter-epoch).

The intra-epoch features can be handcrafted. For example, Dong et al. [5] and Sun et al. [18] extract handcrafted features from the power spectral density (PSD). Features can also be learned using CNNs [5, 9, 13, 16, 18] or recurrent neural networks (RNNs) [10].

Inter-epoch networks are mainly based on RNN structures, mostly Long-Short-Term-Memory (LSTM) layers, both bi-directional [16, 18, 19] and not [5, 20]. Phan et al. [10] also used RNNs, with bi-directional Gated Recurrent Units (GRU) at both the intra-epoch and inter-epoch levels. Contextual information within the inter-epoch section may also be analysed using non-recurrent networks. For example, Qu et al. [13] used a Transformer-like attention network to extract inter-epoch features, Jia et al.[7] combines a temporal convolution with an attention mechanism and Dong et al. [5] only used a softmax layer.

3 Method

3.1 Dataset preprocessing

Our network takes spectrograms extracted from 8 distinct and spatially diverse EEG signals, allowing us to include rich spatial information. To present this information in a visually significant way, we used two time-frequency representations of our signals: FFT and Tapers. We computed spectrograms using the Fast Fourier Transform (FFT) and the parameters provided by Phan et al. [10] : a 2-second Hamming window with 50% overlap. This gives us an image where 1 pixel encodes 0.5 Hz. Unlike Vilamala et al. [21], we cut the frequency axis at 45 Hz included, as higher frequencies do not carry relevant brainwave information. We then convert the amplitude spectrum to a logarithmic scale as done by Vilamala et al. After computing the spectrograms, we divided them into 30 seconds EEG-epochs. With C the number of EEG channels used as input, the resulting image for 1 EEG-epoch is a (C,30,90) tensor. For each EEG-epoch, we also computed the electrodes covariance matrices, as a way to convey spatial co-variation information, with F being the number of covariance matrices. These

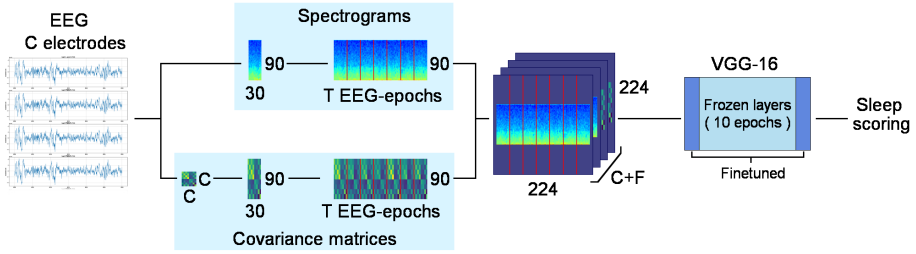

Fig. 1. The preprocessing and classification pipeline. The EEG signals is transformed into C spectrograms and F covariance matrices. Contextual information is added (here, T=7 and vertical lines have been added for visualization only). Spectrograms and covariance matrices are concatenated into a $(C+F,224,244)$ tensor, used as input for a finetuned VGG-16 network.

matrices have been computed on the native EEG-epoch signal, and on a filtered version for each frequency band described in Table 1, resulting in F=7. With C=8, these F (8,8) matrices have been reshaped through repetition of rows and concatenated into a $(F,30,90)$ tensor, then concatenated as supplementary channels to the $(C,30,90)$ spectrograms. The resulting $(C+F,30,90)$ tensor is then zero-padded to fit inside a $(C+F,224,224)$ shape, which is our adapted VGG-16 input layer, as shown in Figure 1.

To get the same spectrogram frequency sampling when computing the FFT and multitapers spectrograms, we used the heuristic described in Prerau et al. [12] where the number of tapers L is $L = \lfloor 2W \rfloor - 1$, with W being the half-time bandwith product and $\lfloor x \rfloor$ the floor function. To get comparable spectra between FFT and Tapers, we used L=3.

3.2 Contextual Information

In order to investigate the impact of contextual information for scoring, we used different EEG sequence lengths as input for the VGG input space, with T being the length of the sequence. Following the AASM guidelines, information above 3 to 5 EEG-epochs should not be relevant for scoring. Therefore, we tested T=1,3,5,7 EEG-epochs fitted inside the $(C+F,224,224)$ input tensor. In these samples, the EEG-epoch to classify has been centered in the input tensor, and we added past EEG-epochs on the left of the central epoch and future EEG-epochs on the right. We added future EEG-epochs as human experts also use them for manual scoring, especially during state transitions.

3.3 Finetuning a deep network to multi-electrode data

Manual sleep scoring is a visual process, therefore we use computer vision techniques to extract information from EEG signals. Extensive research has been done in the field of visual pattern recognition, therefore we chose a classification

network that is efficient in that regard. We used a VGG-16 CNN [17] with batch normalization layers. This network can be used for transfer learning, as it has been trained on ImageNet [14] natural images.

After loading the pretrained weights, we replaced the first layer by a $C + F$-channel deep convolutional layer to match the number of EEG channels and covariance matrices, and replaced the final dense layer to fit our 5-class classification problem. As VGG has been pretrained on color images, we need to finetune the first and last layer before training the whole network. We initially froze the weights of the network except the first and last layers, for 10 epochs, then unfroze the whole VGG for the rest of the training. We used checkpoints after each training epoch and early stopping to save the best network based on validation MF1 score.

4 Experiments

4.1 Dataset used

This study is focused on healthy subjects described by multiple EEG signals. We therefore use the Montreal Archive of Sleep Studies (MASS) dataset [8] which fits these criteria while remaining easily accessible.

The MASS dataset is divided in 5 studies (SS1 to SS5). The only study which involves healthy subjects, scored on 30-second epochs is the SS3 subset. This subset gathers one whole-night recording of 62 subjects using 20 EEG channels, which allows comparison between different brain region signals. On each EEG channel, a 60 Hz notch filter has been applied, as well as a low and high-pass filter with a cutoff frequency of 0.30 Hz and 100 Hz respectively. Due to the biological nature of the studied data, this dataset is imbalanced, with around 50.2% of epochs being in the N2 sleep stage, as shown in Table 2. To correct the class imbalance, we repeat samples during training so that each class has the same number of samples. For validation and testing, we used unbalanced data as these subsets are supposed to represent real case studies, meaning unbalanced classes.

As stated in section 3.1, we have elected to study 8 EEG signals in particular, recorded from a variety of locations on the skull: F3, F4, C3, C4, T3, T4, O1 and O2. We will refer to each electrode couple by their location (F for {F3;F4}). Unless stated otherwise in section 4.5 of the ablation study, each result is presented with C=8 electrodes (FCTO).

	Awake	REM	N1	N2	N3	Total
Number of EEG-epochs	6.442	10.581	4.839	29.802	7.653	59.317
Percentage	10.86%	17.84%	8.16	50.24%	12.90%	100%

Table 2. Class imbalance on the MASS SS3 dataset

4.2 Folds and metrics

We divided the SS3 dataset using the 31 folds proposed by Seo et al. [16], and available on their Github repository. Each fold has 50 training subjects, 10 validation subjects and 2 test subjects. Since there are 62 subjects, the set of test folds covers the whole datatest set without overlap. We used a Tree-structured Parzen Estimator (TPE) approach [2] as implemented in Optuna [1] to determine the best hyperparameters for the first fold. Namely, the learning rate, momentum, weight decay and learning rate decay. Due to time constraints, we applied the TPE only on the first fold, and used the resulting fine-tuned set of hyperparameters for training and validation on the 31 folds. The resulting model for each fold is then evaluated on the test set of the fold, leading to one value of each metric per fold. We monitored the main metrics used in the literature : macro-averaged F1 (MF1), macro-averaged accuracy (MaccroAccuracy) and Cohen's kappa. The MASS dataset has a strong class imbalance, therefore overall accuracy and F1 score become biased metrics. Consequently, MF1 and Macro-accuracy are more indicative of relative per-class accuracy. In our results, we will present the mean and standard deviation of each metric based on the 31 fold-based predictions.

4.3 Time-frequency representations

We first test the impact of the spectral representation by testing FFT against Tapers. We make sure both representations have the same spectral resolution by using L=3 as number of Tapers, which gives a frequency resolution of 0.5Hz, identical to a 2-second window FFT. The comparison of FFT and Tapers scoring is provided in Table 3. Tapers gets better results on all three statistics while the gap between both representations remains within the standard deviation. We will therefore use Tapers spectrograms in the following tests.

	MF1	MacroAccuracy	Kappa
FFT	77.79 ± 3.80	80.68 ± 3.84	0.759 ± 0.051
Tapers	**78.53 ± 3.77**	**81.06 ± 3.39**	**0.766 ± 0.057**

Table 3. Performances reached for FFT and Tapers

4.4 Increasing context information

Using Tapers as time-frequency representation, we tested the effect of contextual inputs by gradually increasing the length of the sequence T. Results are shown in Table 4. We observe an improvement of all the metrics and a decreasing standard deviation as the length of the EEG sequences grows. Reaching better performances from 1 to 3 EEG-epochs underlines the importance of contextual information, which is congruent with the AASM standard. Our best performances is reached for 7 epochs, which does not align with Dong et al. [5],

as their best results were reached between 2 and 4 EEG-epochs. These results suggest that even longer sequences can enhance classification. Especially, reducing the standard deviation seems to be relevant when using the fold provided by Seo et al. [16]. Consequently, we used T=7 for the remaining tests.

T EEG epochs	MF1	MacroAccuracy	Kappa
1	78.53 ± 3.77	81.06 ± 3.39	0.766 ± 0.057
3	80.00 ± 3.65	81.83 ± 3.70	0.782 ± 0.052
5	80.02 ± 3.80	81.89 ± 3.40	0.7869 ± 0.049
7	**81.79 ± 2.95**	**82.96 ± 2.88**	**0.809 ± 0.038**

Table 4. Results from 1 to 7 EEG epochs in the VGG space

4.5 Removing spatial information

We want to study the effect reducing the number of EEG channels. We tested 5 decreasing sets of electrodes, from only left electrodes {F3,C3,T3,O1} (FCTO_left) or right {F4,C4,T4,O2} (FCTO_right), then tested on FCO_right, FO_right and F_right to see the influence of the number of electrodes. Comparing (FCTO_left and FCTO_right), we observe a slight decrease in performance compared to the full FCTO set, while still giving good performances. The right side performs better, so we successively removed the T, C and O electrodes. We observe a rise of the standard deviation, but also good performances for the FCO_right subset. This aligns with the AASM standard, which recommends using at least FCO from one side for human expert scoring. These results also suggest that the temporal (T) electrode does not seem relevant for this task, as removing it seems to give slightly better results.

4.6 Comparison with State-of-the-Art Methods

In order to compare our results to the state of the art, we ran 3 recent methods [7, 16, 19]. Jia et al. [7] is a method which gives good results on MASS, Supratak et al. [19] is often cited as the state-of-art baseline, and Seo et al. [16] uses the folds we based our method on. All methods used the MASS dataset and have

Electrodes	MF1	MaccroAccuracy	Kappa
FCTO_left	79.82 ± 3.81	81.38 ± 3.58	0.788 ± 0.046
FCTO_right	80.71 ± 3.01	82.07 ± 3.25	0.798 ± 0.039
FCO_right	**81.49 ± 3.21**	**83.11 ± 3.37**	**0.802 ± 0.043**
FO_right	80.60 ± 3.44	82.67 ± 3.58	0.790 ± 0.049
F_right	78.77 ± 3.62	81.01 ± 3.19	0.776 ± 0.049

Table 5. Reducing the number of electrodes

their code publicly available on Github. To get a fair comparison, we ran their code using the same folds and metrics that we used, as presented in section 4.2.

All three studies originally used 31 folds on MASS, but both Supratak et al. [19] and Jia et al. [7] have a different fold composition than Seo et al. [16]. Their fold involves 60 subjects per fold as a training set, and the remaining 2 as a validation set. To the best of our knowledge, their published results have been obtained on their validation set only. This is not best practice but can be understood, as neither their code nor their papers show signs of hyperparameter optimization. In order to allow a robust comparison between methods, we retrained Supratak et al. and Jia et al. on Seo et al.'s folds.

All three studies computed their metrics by concatenating the predictions of each fold into a single array of predictions, and comparing those predictions with a similarly obtained array of targets. This way, they compute their metrics on all EEG-epochs of all 62 subjects without omission or repetition. However, this technique is debatable, as it groups together results obtained from 31 different networks (one per fold) without taking into account that different folds have different number of EEG-epochs during prediction. Therefore, computing their metrics on a concatenated prediction array creates an implicit weighting relative to each fold's prediction set size. Moreover, they used overall accuracy instead of Maccro-Accuracy, and none of them give the standard deviation. Consequently, we computed the Macro-Accuracy, MF1 and Kappa score for each fold, then computed the averaged and standard deviation.

These differences may explain the discrepancy between the results published and the results we obtained by running their code using our folds and methodology, as seen in Table 6. We also tested Vilamala et al. [21] on the MASS dataset, as they are also using Tapers and transfer-learning on VGG-16. While training Vilamala's method, we froze the convolution layers beforehand as it gave this method better results.

With this shared protocol, our method outperforms all four methods, as shown in Table 6.

5 Discussion

Our results suggest that automatic sleep scoring could benefit from using multitapers spectrograms as time-frequency representation. Using a high number of epochs as contextual input gave higher results, and seems to reduce the standard

Methods	MF1	MaccroAccuracy	Kappa
Seo [16]	77.36 ± 4.76	77.17 ± 4.08	0.774 ± 0.052
Supratak [19]	79.67 ± 4.49	79.99 ± 4.26	0.792 ± 0.047
Jia [7]	76.03 ± 4.01	76.35 ± 4.53	0.751 ± 0.056
Vilamala [21]	72.87 ± 5.72	73.24 ± 5.78	0.666 ± 0.072
Our method	**81.79 ± 2.95**	**82.96 ± 2.88**	**0.809 ± 0.038**

Table 6. Comparison with SOA methods

deviation. Our results are congruent with the fact that some state transitions can be influenced by the previous epochs in the AASM standard, and unlike Dong et al. [5], we still got improvement with a sequence length greater than 6 EEG-epochs. We showed in Table 5 than halving the number of electrode could maintain strong classification results, thus questioning on the redundancy between left and right side. However, left-side EEGs are often artefacted by the cardiac activity, which may explain why we reach slightly better performances when using only right-side electrodes compared to left-side electrodes. Using only EEG, we got comparable results than Supratak et al. [19] which used EEG and EOG signals, and better result than Jia et al.[7] which use EEG, EOG and EMG. This underline that EEG alone can give robust results for classification, thus leading to simpler acquisition protocols.

6 Conclusion

In this paper we compared two types of time-frequency spectrograms for sleep scoring. Using a fine-tuned deep visual network, we outperforms state-of-the-art results. We did an ablation study to determine the number of contextual EEG-epochs needed, and study how the number of electrodes could impact classification results. Our results seems relevant with the AASM standard and EEG expertise regarding the number of EEG-epochs and when comparing left and right side for electrodes. Our results suggest that acquisition protocol could be reduced to a lesser number of modalities and sparser electrodes. Finally, we propose a common methodology for training and method comparison, using the same folds as Seo et al. [16], which allows hyperparameter research.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2623–2631 (2019)
2. Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B.: Algorithms for hyper-parameter optimization. Advances in neural information processing systems **24** (2011)
3. Berry, R.B., Brooks, R., Gamaldo, C., Harding, S.M., Lloyd, R.M., Quan, S.F., Troester, M.T., Vaughn, B.V.: Aasm scoring manual updates for 2017 (version 2.4) (2017)
4. Chambon, S., Galtier, M., Arnal, P.J., Wainrib, G., Gramfort, A.: A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. IEEE Transactions on Neural Systems and Rehabilitation Engineering **26**(4), 17683810 (Mar 2018)
5. Dong, H., Supratak, A., Pan, W., Wu, C., Matthews, P.M., Guo, Y.: Mixed neural network approach for temporal sleep stage classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering **26**(2), 324–333 (2018)
6. Jia, Z., Lin, Y., Wang, J., Ning, X., He, Y., Zhou, R., Zhou, Y., Lehman, L.w.H.: Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering **29**, 1977–1986 (2021)

7. Jia, Z., Lin, Y., Wang, J., Zhou, R., Ning, X., He, Y., Zhao, Y.: Graphsleepnet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification. In: IJCAI. pp. 1324–1330 (2020)
8. O'reilly, C., Gosselin, N., Carrier, J., Nielsen, T.: Montreal archive of sleep studies: an open-access resource for instrument benchmarking and exploratory research. Journal of sleep research **23**(6), 628–635 (2014)
9. Phan, H., Andreotti, F., Cooray, N., Chén, O., de Vos, M.: Joint classification and prediction cnn framework for automatic sleep stage classification. IEEE Transactions on Biomedical Engineering **66**, 1285–1296 (05 2019)
10. Phan, H., Andreotti, F., Cooray, N., Chén, O.Y., De Vos, M.: Seqsleepnet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. IEEE Transactions on Neural Systems and Rehabilitation Engineering **27**(3), 400–410 (2019)
11. Phan, H., Mikkelsen, K., Chén, O.Y., Koch, P., Mertins, A., De Vos, M.: Sleeptransformer: Automatic sleep staging with interpretability and uncertainty quantification. IEEE Transactions on Biomedical Engineering **69**(8), 2456–2467 (2022)
12. Prerau, M.J., Brown, R.E., Bianchi, M.T., Ellenbogen, J.M., Purdon, P.L.: Sleep neurophysiological dynamics through the lens of multitaper spectral analysis. Physiology **32**(1), 60–92 (2017)
13. Qu, W., Wang, Z., Hong, H., Chi, Z., Feng, D.D., Grunstein, R., Gordon, C.: A residual based attention model for eeg based sleep staging. IEEE journal of biomedical and health informatics **24**(10), 2833–2843 (2020)
14. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision **115**(3), 211–252 (2015)
15. Salinsky, M., Goins, S., Sutula, T., Roscoe, D., Weber, S.: Comparison of sleep staging by polygraph and color density spectral array. Sleep **11**(2), 131–8 (1988)
16. Seo, H., Back, S., Lee, S., Park, D., Kim, T., Lee, K.: 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 **61**, 102037 (2020)
17. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
18. Sun, C., Chen, C., Li, W., Fan, J., Chen, W.: A hierarchical neural network for sleep stage classification based on comprehensive feature learning and multi-flow sequence learning. IEEE Journal of Biomedical and Health Informatics **24**(5), 1351–1366 (2020)
19. Supratak, A., Dong, H., Wu, C., Guo, Y.: Deepsleepnet: a model for automatic sleep stage scoring based on raw single-channel eeg. IEEE Transactions on Neural Systems and Rehabilitation Engineering **25**(11), 1998–2008 (Nov 2017)
20. Supratak, A., Guo, Y.: Tinysleepnet: An efficient deep learning model for sleep stage scoring based on raw single-channel eeg. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC). pp. 641–644 (2020)
21. Vilamala, A., Madsen, K.H., Hansen, L.K.: Deep convolutional neural networks for interpretable analysis of eeg sleep stage scoring. In: 2017 IEEE 27th international workshop on machine learning for signal processing (MLSP). pp. 1–6. IEEE (2017)