Conference Paper/Konferans Bildirisi

GAZİ
JOURNAL OF ENGINEERING SCIENCES

Automatic Sleep Staging by EEG and 2D-Convolutional Neural Network

Ibrahim Kaya*[a]

Submitted: 27.12.2021 Revised: 05.07.2022 Accepted: 27.10.2022 doi:10.30855/gmbd.0705037

Keywords: Automatic sleep staging, EEG, Deep learning, Convolutional Neural Network (CNN)

[a,*] İzmir Katip Celebi University, Faculty of Engineering and Architecture, Dept. of Biomedical Engineering 35620 - İzmir, Türkiye
Orcid: 0000-0003-0802-4376
e mail: ibrahimkaya21@yahoo.com

*Corresponding author: ibrahimkaya21@yahoo.com

ABSTRACT

Sleep disorders have high prevalence and cause various health problems. For the diagnostics of these disorders and assessment of the sleep quality, many physiological data are collected using polysomnogram (PSG) method. The most important PSG data is the EEG recorded from the brain during sleep. Analysis of hours of sleep EEG data by experts is an onerous task which requires high attention. Recently, many automatic sleep staging classifiers using EEG are developed in order to prevent human error, and to provide a quick objective analysis. They use machine learning techniques and predict the sleep stage of each EEG epoch. Compared to traditional machine learning, deep learning which requires no hand-crafted feature extraction was able to classify sleep stages better. 1D Convolutional Neural Networks (CNN) are the main methods used in automatic sleep staging recently. In this research a simple 2D-CNN based automatic sleep staging feasibility is investigated. It has been found that a 2D CNN can classify the sleep stages by accuracy of 92.55% and with a Cohen's kappa of 0.82.

EEG ve 2 Boyutlu Evrişimsel Sinir Ağları Aracılığıyla Otomatik Uyku Evre Tayini

ÖZ

Uyku bozuklukları toplumda oldukça yaygın görülmekle birlikte çeşitli sağlık sorunlarına neden olmaktadır. Bu bozuklukların teşhis edilmesi ve uyku kalitesinin belirlenmesi için Polisomnogram metodu ile birçok fizyolojik veri toplanılır. En önemli veri uyku halinde beyinden kaydedilen EEG verisidir. Saatler süren uykuya ait EEG verilerinin uzmanlar tarafından analiz edilmesi yüksek dikkat isteyen çok zahmetli bir iştir. Son zamanlarda insan hatalarını önlemek ve hızlı nesnel bir analiz gerçekleştirmek amacıyla EEG sinyallerini kullanan otomatik uyku evre sınıflandırıları geliştirilmiştir. Bu sınıflandırıcılar makine öğrenmesi yöntemlerini kullanır ve her bir EEG kesitine dair uyku evresini tahmin eder. Geleneksel makine öğrenmesi yöntemlerine kıyasla elle hiçbir öznitelik çıkarımı gerektirmeyen derin öğrenme uyku evre sınıflandırmasında daha başarılı olabilmiştir. Son zamanlarda, tek boyutlu evrişimsel sinir ağları otomatik uyku evre sınıflandırmasında ana yöntem olmuştur. Bu araştırmada iki boyutlu basit bir evrişimsel sinir ağlarına dayalı otomatik uyku evre sınıflandırılmasının uygulanabilirliği incelenmiştir. iki boyutlu evrişimsel sinir ağlarının %92.5 doğruluk ve 0.82 Cohen Kappa değeri ile sınıflandırmabildiği bulunmuştur.

Anahtar Kelimeler: Otomatik uyku evresi sınıflandırma, EEG, Derin öğrenme, Evrişimli sinir ağları

To cite this article: I. Kaya, "Automatic Sleep Staging by EEG and 2D-Convolutional Neural Network," *Gazi Journal of Engineering Sciences,* vol. 8, no. 3, pp. 491-498, 2022. doi:10.30855/gmbd.0705037

1. Introduction

Sleep is a vital process for human health and function. Correct and early diagnosis of sleep problems is critical. Sleep staging is used to diagnose sleep disorders. It is done by collecting diagnostic data from subjects during sleep and evaluation by sleep experts. Overnight polysomnogram (PSG) is a gold standard method to investigate the quality of sleep and rate the sleep stages [1]. In PSG, many sensors are connected to the subject and various data such as EEG, EOG, EMG, ECG, respiratory efforts, airflows, and blood oxygenation are collected [2]. EEG is the most widely used reliable data related with the brain activity. Human sleep consists of repetitions of six different stages which are; wakefulness (W), 4 Non Rapid Eye Movement (REM) stages (S1-S4), and REM sleep as described by Rechtschaffen and Kales (R&K) [3-4]. Typical sleep stage EEG rhythms for Wake, S1 to S4 and REM sleep are shown in Figure 1. Each colored plot is a typical 30second EEG signal of particular sleep stage. A night sleep usually involves 90 minutes cycles of N2, N3, REM and N1 [5]. This cycle repeats 4-5 times in one night sleep.

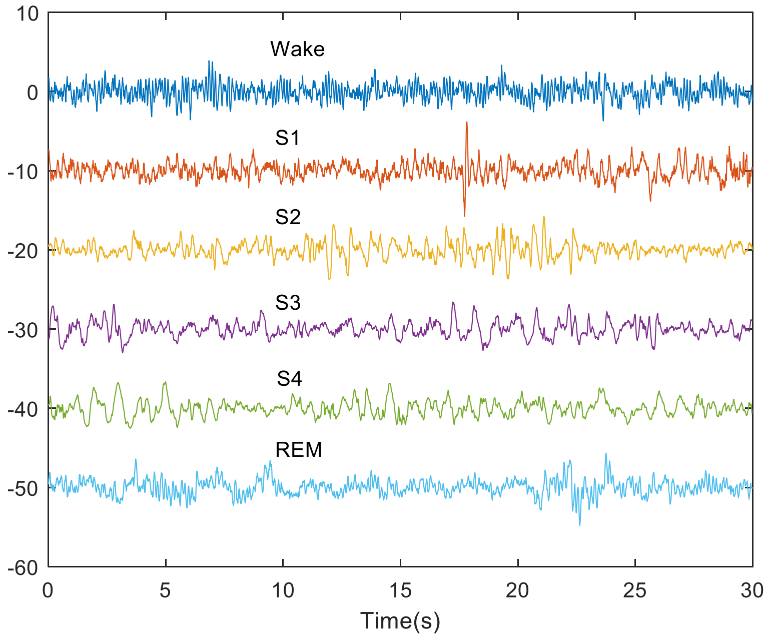

Figure 1. Six different sleep stages according to R&K.

In the sleep staging various features such as K-complexes, spindles, alpha bursts, amplitudes, frequency content, EMG level, eye movements are used [6]. In a typical PSG, EEG data is segmented with 30s epochs and each epoch is classified as one of the 6 stages of sleep. However, the American Academy of Sleep Medicine (AASM) has another standard suggesting grouping sleep epochs into 5 distinct sleep stages [7]. Sleep staging used to be done by inspection of each epoch by a trained sleep expert, however since PSG data contains many hours of information it is very time consuming and prone to mistakes. Therefore, some sort of automatic sleep staging is required. While manual feature extraction based machine learning methods have reasonable performances in the automatic sleep staging by EEG, there exist some limitations such as prior knowledge and inability to generalizing to broader datasets and different subjects [8]. Thus, deep learning methods become a better alternative, since they do not require the hand crafted feature extraction and can be generalized to new different data easily.

In the sleep staging various features such as K-complexes, spindles, alpha bursts, amplitudes, frequency content, EMG level, eye movements are used [6]. In a typical PSG, EEG data is segmented with 30s epochs and each epoch is classified as one of the 6 stages of sleep. However, the American Academy of Sleep Medicine (AASM) has another standard suggesting grouping sleep epochs into 5 distinct sleep stages [7]. Sleep staging used to be done by inspection of each epoch by a trained sleep expert, however since PSG data contains many hours of information it is very time consuming and prone to mistakes. Therefore, some sort of automatic sleep staging is required. While manual feature extraction based machine learning methods have reasonable performances in the automatic sleep staging by EEG, there exist some limitations such as prior knowledge and inability to generalizing to broader datasets and

different subjects [8]. Thus, deep learning methods become a better alternative, since they do not require the hand crafted feature extraction and can be generalized to new different data easily.

2. Materials and Method

2.1. Dataset

The data is obtained from the sleep-edfx expanded database on physionet [12]. The dataset contains 197 whole-night PSG sleep recordings, however only the first 20 subject recordings are used, because there is a clear subject age difference between the first 20 subjects and the rest. The selected dataset consists of two PSG files for each of 20 subjects with ages ranging from 25 to 34. Each subject recordings took place during two consequtive day and night periods, except subject 13 who had only one night data. Also each PSG file is associated with a hypnogram file which basically labels the data epochs as Wake, REM, Stage 1, Stage 2,Stage 3, Stage 4, M (Movement time) and ? (not scored) according to R&K sleep staging method. In PSG data there are four electrode recordings from two EEG channels Fpz-Cz and Pz-Oz, one EOG and one EMG channels. In addition, there are Resporonasal, EMGSubmenta, Tempbody, and Eventmarker in the data. In this study, two channels of EEG data from Fpz-Cz and Pz-Oz electrodes are exploited. The sampling rate is 100Hz and one sleep EEG epoch contains 3000 time series data points corresponding to 30s. Epochs labelled as M and ? are excluded from the dataset because they are artifacts. The sleep data stages are converted from 6 stage R&K standard to conventional AASM staging standard of 5 sleep stages by combining the stages 3 and stages 4 as non-REM3 (N3). In AASM there are Wake ,three non-REM N1, N2, N3 stages and a REM stage. The distribution of sleep stages in the dataset is not even and this is a major problem with the proper training of the network.

The hold out method is used to separate the test and train data. The first 14 subjects, Subjects S0 to S14 (excluding S13) are included in the training set and remaining 5 subjects (S15-S19) are grouped as the test dataset. Following the training, the model is tested with the data that has never been used in training.

2.2. Method

2.2.1 Image formation

In MATLAB environment, a single dimensional time series EEG signal is converted into a 2D image to fit the input format and to train 2D CNN model as shown in Figure-2. EEG sleep epochs of 30s or 3000 data points from Fpz-Cz channel are reshaped into 50x60x1. Similarly another image color layer is created by subtracting the 30s, 3000 data points of Pz-Oz from the Fpz-Cz channel data and assigned as the second color of the 50x60x2 image. This subtraction which a basic spatial filtering, cancelled the common background activity between Fpz-Cz and Pz-Oz electrodes.

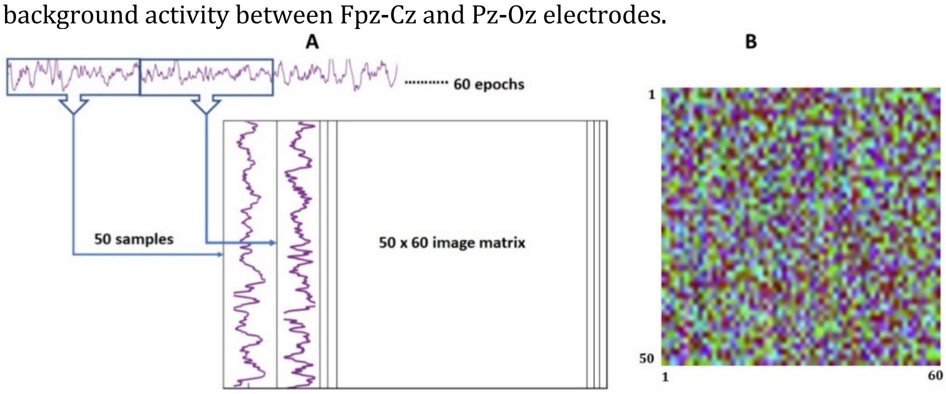

Figure 2. (A) 2D image from EEG data. (B) A 50 x 60 pixel image from one EEG time series data. Signal values are set as the image pixel colors. Additional color is added to the image using the subtraction of Pz-Oz and Fpz-Cz EEG data as the second color.

2.2.2 Two-Dimensional convolution neural network (2D-CNN)

In CNN architecture, first a 2D convolution layer is used to extract the features from the images. It applies sliding convolutional filters to 2D input. Batch normalization layer is used between convolutional layers and ReLU to speed up training of the convolutional neural network and reduce the sensitivity to network initialization. ReLU layer is used as the activation function between layers in order to account for nonlinearity of the network. It replaces all the negative pixel values with zeroes. A 2-D max pooling layer is inserted after ReLU layers to perform downsampling of the input images. It divides the input into rectangular pooling regions then computes the maximum of each region. Therefore the number of parameters and computational load in the network is reduced and overfitting can be controlled. In the full connected layer, the input is multiplied by a weight matrix and then bias vector is added. Fully connected layer uses all of the features from the previous layers across the image to recognize the larger patterns. These layers are repeated four times to increase the complexity of the model. Finally softmax layer together with a classification layer assigns the sleep stage with the highest probability as the output of the classifier. The overall diagram of the 2D CNN model is given in Figure-3. The number of convolutional level can be reduced further to minimize the complexity and the number of parameters. The filter sizes, paddings, strides can be optimized.

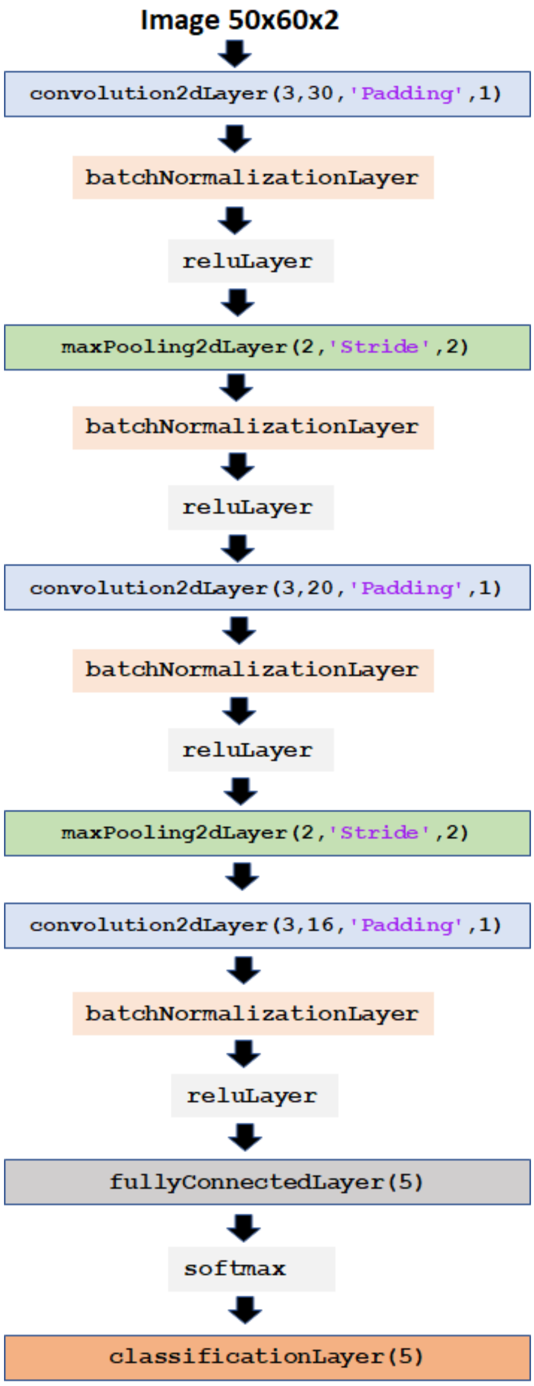

Figure 3. 2D CNN model for sleep staging.

3. Research Findings

The results are evaluated by the accuracy, precision, recall, F1-score, and Cohen's Kappa (K) values that are computed for the test group, namely Subjects 15-19. These are the main performance metrics in literature for machine learning and particularly automatic sleep staging [13]. The formulas for calculation of the above mentioned metrics are given in equations (1) (2) (3) (4) (5) in order [13]. Accuracy Precision and Recall are computed using the True Positives (TP), True Negatives (TN), False Positives (FP) and False Negative (FN) values from the confusion matrix. F1 score is computed with Precision and Recall. Cohen's Kappa is a measure of interraters performances, how well is the classification result in agreement with the correct labels. Inter-rater agreement of standard EEG scorings between international sleep centers is above 0.70. In (5) N is the number of sleep epochs, x values are the cell values in the confusion matrix for each row i and column j.

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

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

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

$$F1-score = \frac{2\,Precison\ Recall}{Precision+Recall} \tag{4}$$

$$K = \frac{\frac{\sum_{i=1}^{n} x_{ii}}{N} - \frac{\sum_{i=1}^{n} x_{ii}\left(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}\right)}{N^2}}{1 - \frac{\sum_{i=1}^{n} x_{ii}\left(\sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} x_{ji}\right)}{N^2}} \tag{5}$$

In Table 1, each cell shows the number of epochs from the labelled by the expert as the row label and classified as column label by the 2D CNN model. The data is not uniformly distributed. Wake state dominates the EEG data. The probabilities of each sleep state is given in Table 2. N2 follows the dominant Wake state by probability of 17%. High F1 score for Wake emphasizes the classification performance of the method, however N1 is not correctly classified. The low F1 in the classification of N1 is observed in other studies as well [13].

Table 1. The confusion chart for the test data is given. The true sleep stage distribution is given as sleep stage epoch counts and the corresponding predicted sleep stage epoch counts.

		Predicted				
		Wake	**N1**	**N2**	**N3**	**REM**
True	**Wake**	17633	60	72	9	269
	N1	175	106	86	1	252
	N2	152	60	3832	262	353
	N3	3	3	183	1539	0
	REM	231	122	285	1	1595

Table 2. Classification performances for each sleep stage are given as stage, probability, recall , precision and F1 Scores.

Stage	**Prob.**	**Recall**	**Precision**	**F1-score**
Wake	0.66	0.98	0.97	0.97
N1	0.02	0.17	0.30	0.22
N2	0.17	0.82	0.86	0.84
N3	0.06	0.89	0.85	0.87
REM	0.08	0.71	0.65	0.68
Macro	1.00	0.72	0.73	0.72

The classification performances for the test dataset are given in Table 3. The mean accuracy of the method using the test dataset is 90.55%. The interrater Cohen's Kappa value K is found between 0.74

and 0.89, higher than the acceptable 0.7 level. Subject 15 performance values are the highest, 94.63% accuracy and 0.89 Cohen's Kappa.

Table 3. Accuracy and Cohen's kappa values for test subjects.

Subject	Accuracy (%)	Cohen's Kappa (K)
S15	94.63	0.89
S16	89.63	0.80
S17	86.14	0.74
S18	92.66	0.84
S19	89.59	0.82
Mean	90.55	0.82

4. Discussion

The results of this study are compared with the state of the art methods in the literature given in Table 4. It can be seen that, while this approach has high classification accuracy compared to the other methods, it lacks the F1 score due to the data imbalance problem. One of the problems with the low performance of the other researches is the subjective determination of the sleep classes by experts. Xu et al. found that sleep dataset annotation may be inaccurate due to high workload of the sleep experts [14]. Therefore this also deteriorates the automatic classification performance. Another point is the high misclassification of the N1 stage, this is attributed to the extremely small percentage of the N1 stage in the dataset and the similarity between N1 and REM stages [14]. Although Salamatian & Khadem developed a 1D CNN and obtained great accuracy, however the confusion matrix that is provided is from a very small data [21]. Deep learning method (1D CNN) presented in Yildirim et al. 2019 study achieved 90.98 % accuracy however they used EOG channel in addition to EEG [22].

Table 4. Results of the deep learning studies in this field.

Study	Method	Macro F-1	Cohen's kappa	Accuracy %
Xu et al. [14]	1D-CNN	81.18	0.80	85.53
Tsinalis et al. [15]	1D-CNN	84		78.9
Cai et al. [16]	Dual input 1D-CNN	-	0.80	87.21
Phan et al. [11]	Joint 2D-CNN		0.75	81.9
Supratak et al. [17]	CNN + Bidirectional LSTM	76.9	0.76	82
Fu et al. [18]	CNN + Bidirectional LSTM	82.14	0.77	83.78
Mousavi et al. [19]	1D-CNN	79.66	0.79	84.26
Khalili & Asl [20]	TCNN + CRF	79.29	0.80	85.39
Salamatian & Khadem [21]	1D-CNN	-	-	94.09
Zhou et al. [13]	1D-CNN	-	0,81	86.1
Jia et al. [23]	Dilated CNN	0.83	-	-
Nie et al. [24]	CNN+RNN	0.81	-	-
This study	2D-CNN	72.00	0.82	90.55

5. Conclusion

This study demonstrates the efficient use of a simple 2D CNN in automatic sleep staging. The overall sleep stage epochs and the corresponding prediction states that are given in Table-1 show that Wake and N3 stages are accurately predicted with the method. The performances of each sleep stage are given in Table-2 shows this performance in terms of performance metrics. Except for N1 stage the performance metrics highlight the accuracies of the method. It achieved performance comparable to the state of the art methods, however there are differences in the data and preprocessing steps. In future studies, first I am interested in optimizing the 2D image matrix generation Then I want to evaluate the efficacy of the method on the in-bed data only, and broader sleep-edfx dataset or alternative datasets. I am also interested in incorporating multiple PSG data such as EOG EMG into the deep learning 2D CNN algorithm. Moreover, I want to tackle the problem of data imbalance which prevents the proper classification of N1 stage. Another direction could be comparison of the possible electrode channels and spatial filtering approaches before applying CNN architecture.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

References

[1] W. H. Spriggs, *Essentials of Polysomnography: A Training Guide and Reference for Sleep Technicians,* Burlington, MA, USA: Jones & Bartlett Learning, 2014.

[2] H. Schulz, “Rethinking sleep analysis,” *Journal of Clinical Sleep Medicine*, vol. 4, no. 2, pp. 99–103, 2008. doi:10.5664/jcsm.27124

[3] T. Hori, Y. Sugita, E. Koga et al., Proposed supplements and amendments to “A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects’, the Rechtscha_en & Kales (1968) standard,” *Psychiatry and Clinical Neurosciences*, no. 55, pp. 305–310. 2001.

[4] D. W. Carley and S. S. Farabi, “Physiology of sleep,” *Diabetes Spectrum*, vol. 29, no. 1, pp. 5–9, 2016. doi:10.2337/diaspect.29.1.5

[5] G. Z. Özen, R. Sultanov, Y. Özen and Z. Y. Güneş, “A Convolutional Neural Network Based on Raw Single Channel EEG for Automatic Sleep Staging,” *Sakarya University Journal of Computer and Information Sciences*, vol. 3, no. 2, pp. 149-158, 2020. doi:10.35377/saucis.03.02.731628

[6] K. Šušmákov, "Human sleep and sleep EEG," *Measurement science review*, vol. 4, no. 2, pp. 59-74, 2004.

[7] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus and B. V. Vaughn, “The AASM manual for the scoring of sleep and associated events,” *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, no. 176, 2012.

[8] Z. Mousavi, T. Yousefi Rezaii, S. Sheykhivand et al., "Deep convolutional neural network for classification of sleep stages from single-channel EEG signals," *Journal of neuroscience methods,* vol. 324, 2019. doi:10.1016/j.jneumeth.2019.108312

[9] A. Craik, Y. He, J. L. Contreras-Vidal, “Deep learning for electroencephalogram (EEG) classification tasks: a review,” *Journal of neural engineering*, vol. 16, no. 3, 2019. Art. doi:10.1088/1741-2552/ab0ab5

[10] M. Bojarski, D. Del Testa, D. Dworakowski et al., “End to end learning for self-driving cars, ” *arXiv preprint arXiv:1604.07316*, 2016. doi:10.48550/arXiv.1604.07316

[11] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén and M. De Vos, "Joint Classification and Prediction CNN Framework for Automatic Sleep Stage Classification," *in IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1285-1296, 2019, doi:10.1109/TBME.2018.2872652.

[12] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. C. Kamphuisen and J. J. L. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG," in *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185-1194, Sept. 2000, doi: 10.1109/10.867928.

[13] D. Zhou, J. Wang, G. Hu, et al. “Singlechannelnet: A model for automatic sleep stage classification with raw single-channel EEG,” *Biomedical Signal Processing and Control*, vol. 75, 2022.

[14] K. Xu, S. Xia, G. Li, “Automatic Classification of Sleep Stages Based on Raw Single-Channel EEG,” *In Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, October-16, 2020, Cham: Springer, 2020. pp. 356-368. doi:10.1007/978-3-030-60639-8_30

[15] O. Tsinalis, P. M. Matthews, Y. Guo et al. “Automatic sleep stage scoring with single-channel EEG using convolutional neural networks,” *arXiv preprint arXiv:1610.01683,* 2016. doi:10.48550/arXiv.1610.01683

[16] Q. Cai, Z. Gao, J. An, S. Gao and C. Grebogi, "A Graph-Temporal Fused Dual-Input Convolutional Neural Network for Detecting Sleep Stages from EEG Signals," *in IEEE Transactions on Circuits and Systems II: Express Briefs,* vol. 68, no. 2, pp. 777-781, 2021. doi:10.1109/TCSII.2020.3014514.

[17] A. Supratak, H. Dong, C. Wu and Y. Guo, "DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG," *in IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998-2008, 2017. doi:10.1109/TNSRE.2017.2721116.

[18] M. Fu, Y. Wang, Z. Chen et al., “Deep Learning in Automatic Sleep Staging With a Single Channel Electroencephalography,” *Front. Physiol*, vol. 12, no. 628502. doi:10.3389/fphys.2021.628502

[19] S. Mousavi, F. Afghah, U. R. Acharya, "SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PloS one*, vol. 14, no. 5, 2019. doi:10.1371/journal.pone.0216456

[20] 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, 2021. doi:10.1016/j.cmpb.2021.106063

[21] A. Salamatian and A. Khadem, "Automatic sleep stage classification using 1D convolutional neural network," *Frontiers in Biomedical Technologies*. vol. 7, no. 3, pp. 142-150, 2020. doi:10.18502/fbt.v7i3.4616

[22] 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, pp. 599, 2019. doi:10.3390/ijerph16040599

[23] Z. Jia, Y. Lin, J. Wang et al., "SalientSleepNet: Multimodal salient wave detection network for sleep staging," *arXiv preprint arXiv:2105.13864,* 2021. doi:10.48550/arXiv.2105.13864

[24] H. Nie, S. Tu and L. Xu, "RecSleepNet: An Automatic Sleep Staging Model Based on Feature Reconstruction," *in 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM),* pp. 1458-1461, 2021. doi:10.1109/BIBM52615.2021.9669687.

** Bu makale International Conference on Engineering Technologies-ICENTE 21 isimli konferansta sunulmuş bildirinin genişletilmiş halidir.*