ORIGINAL ARTICLE

Estimating daytime sleepiness with previous night electroencephalography, electrooculography, and electromyography spectrograms in patients with suspected sleep apnea using a convolutional neural network

Sami Nikkonen[1,2,*,], Henri Korkalainen[1,2,], Samu Kainulainen[1,2,], Sami Myllymaa[1,2,], Akseli Leino[1,2,], Laura Kalevo[1,2], Arie Oksenberg[3], Timo Leppänen[1,2,] and Juha Töyräs[1,2,4,]

[1]Department of Applied Physics, University of Eastern Finland, Kuopio, Finland, [2]Diagnostic Imaging Center, Kuopio University Hospital, Kuopio, Finland, [3]Sleep Disorders Unit, Loewenstein Hospital—Rehabilitation Center, Raanana, Israel, [4]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia

*Corresponding author. Sami Nikkonen, Department of Applied Physics, University of Eastern Finland, P.O Box 1627 (Canthia), 70211 Kuopio. Email: sami.nikkonen@uef.fi.

Abstract

A common symptom of obstructive sleep apnea (OSA) is excessive daytime sleepiness (EDS). The gold standard test for EDS is the multiple sleep latency test (MSLT). However, due to its high cost, MSLT is not routinely conducted for OSA patients and EDS is instead evaluated using sleep questionnaires. This is problematic however, since sleep questionnaires are subjective and correlate poorly with the MSLT. Therefore, new objective tools are needed for reliable evaluation of EDS. The aim of this study was to test our hypothesis that EDS can be estimated with neural network analysis of previous night polysomnographic signals. We trained a convolutional neural network (CNN) classifier using electroencephalography, electrooculography, and chin electromyography signals from 2,014 patients with suspected OSA. The CNN was trained to classify the patients into four sleepiness categories based on their mean sleep latency (MSL); severe (MSL < 5min), moderate (5 ≤ MSL < 10), mild (10 ≤ MSL < 15), and normal (MSL ≥ 15). The CNN classified patients to the four sleepiness categories with an overall accuracy of 60.6% and Cohen's kappa value of 0.464. In two-group classification scheme with sleepy (MSL < 10 min) and non-sleepy (MSL ≥ 10) patients, the CNN achieved an accuracy of 77.2%, with sensitivity of 76.5%, and specificity of 77.9%. Our results show that previous night's polysomnographic signals can be used for objective estimation of EDS with at least moderate accuracy. Since the diagnosis of OSA is currently confirmed by polysomnography, the classifier could be used simultaneously to get an objective estimate of the daytime sleepiness with minimal extra workload.

Statement of Significance

Daytime sleepiness is a common symptom of obstructive sleep apnea (OSA), but it is somewhat ignored in sleep apnea diagnostics and treatment planning since the multiple sleep latency test is not routinely conducted for sleep apnea patients. The convolutional neural network classifier developed in this study enables the estimation of objective daytime sleepiness for OSA patients using signals recorded during polysomnography. Therefore, a reasonably accurate sleepiness estimation can be acquired without the need to conduct any additional tests. The only currently available alternatives are subjective sleep questionnaires, such as Epworth Sleepiness Scale, which the developed classifier slightly outperforms. The accuracy of the classifier could be further improved in the future with broader training material.

Introduction

Obstructive sleep apnea (OSA) is a common sleep disorder affecting approximately half of the adult population [1, 2]. A major symptom of OSA is excessive daytime sleepiness (EDS). Although EDS is not directly lethal, it has a significant deteriorating impact on the quality of life causing depression and cognitive impairment [3–5]. In addition, EDS is a major cause of motor vehicle accidents and sick leaves making it a substantial economic burden [6].

The gold-standard test for EDS is the multiple sleep latency test (MSLT) [7]. The MSLT is an objective, full-day trial performed in a sleep laboratory where sleep latency is measured multiple times and the average of these latencies, that is, mean sleep latency (MSL) is used to assess EDS [7]. The subjects are clinically classified into four sleepiness categories based on their MSL: severe (MSL < 5 min), moderate (5 ≤ MSL < 10), mild (10 ≤ MSL < 15), and normal (MSL ≥ 15) [8, 9]. Alternatively, a single MSL threshold of 8 or 10 min is often used to differentiate between normal patients and patients suffering from EDS [10, 11]. However, as MSLT is time-consuming and expensive, it is not routinely performed for OSA patients and EDS is instead evaluated using simpler tests, such as sleep questionnaires [12, 13]. Sleep questionnaires are problematic however, since they are dependent on the patients' interpretation of the rating system and therefore only offer an estimation on subjective sleepiness. For example, the results of the most common subjective test, the Epworth Sleepiness Scale (ESS), do not correlate well with MSLT and have been proven to be insufficient in estimating daytime sleepiness [12, 14–16]. Due to these shortcomings, simpler and easier objective tools are needed for evaluation of EDS especially for OSA patients.

Machine learning has been proven to be a powerful tool in medical signal analysis and has also shown promise in automatic diagnostics of OSA [17–19]. For example, artificial neural networks have been used for automated sleep staging using electroencephalography (EEG) [20, 21]. Based on the promising previous research, we hypothesized that previous night EEG could be used to estimate the daytime sleepiness of an OSA patient. Therefore, the aim of this study was to develop an objective, neural network method for estimation of EDS in patients with suspected OSA. We test our hypothesis by training a convolutional neural network (CNN) that estimates the results of the MSLT based on the previous night's EEG, electrooculography (EOG), and electromyography (EMG) signals.

We chose to use a convolutional neural network, which is a type of deep neural network inspired by the human visual cortex and developed specifically for visual machine learning tasks [22]. CNNs also have less parameters and are faster to train than equally sized multilayer perceptron networks which is important with large inputs such as high resolution images. Like in regular multilayer perceptron networks, layers of a CNN have neurons, which receive inputs, calculate a weighed sum from them according to the learnable weights, pass them through an activation function and generate an output. However, in CNNs, the layers are not fully connected and instead only a small part of the input layer is being operated on by the convolution kernel at a time. This kernel is then moved over the whole input layer generating the full output.

Methods

We developed a convolutional neural network (CNN) classifier to automatically estimate the MSLT result using EEG, EOG, and chin EMG signals recorded during in-lab polysomnography (PSG) the previous night. The CNN classifier was trained to classify the patients with suspected OSA into four sleepiness categories based on their MSL; severe (MSL < 5 min), moderate (5 ≤ MSL < 10), mild (10 ≤ MSL < 15), and normal (MSL ≥ 15). Additionally, we classified patients to EDS and normal groups using an MSL < 10 min as the threshold for EDS.

Dataset

The patient population consisted of 2,014 patients with suspected OSA who had undergone in-lab PSG and a next day MSLT (Table 1). The recordings were conducted during 2001–2011 in the Sleep Disorders Unit, Loewenstein Hospital—Rehabilitation Center, Raanana, Israel and analyzed using the prevailing American Academy of Sleep Medicine (AASM) guidelines [23, 24]. According to the clinical protocol at the Loewenstein Hospital, patients were referred to the MSLT because they had complained of daytime sleepiness during the clinical interview. No preliminary sleep questionnaires were performed. Ethical permission was obtained from the Ethical Committee of Loewenstein Hospital (Permission number: 0006-17-LOE). The MSLTs were conducted using four-nap protocol in uninterrupted conditions with 2 h intervals between each nap attempt [25]. The sleep onset was determined from the first stage of sleep. If no sleep occurred, the nap attempt was terminated at 20 min and the sleep latency was determined to be 20 min for that nap attempt. A total of four nap attempts were conducted and the MSL was calculated as the mean of these four readings.

The EEG electrodes were placed according to the international 10–20 system [26]. Two EEG channels, C4-A1 and PZ-A1, EOG channel (ROC-A1), and chin EMG were used as an input to the CNN. AASM recommends C4-A1, F4-A1, and O2-A1 EEG channels together with EOG and EMG channels for sleep staging [24]. However, as our dataset included frontal and occipital EEG channels only for a very limited number of patients, we chose to use C4 and PZ channels because they were most frequently recorded among the patients and thus the highest possible number of patients could be included in the study. We also chose to include the EOG and chin EMG channels since according to our preliminary testing, slightly better results were obtained with all four channels compared to using EEG only.

Signal processing

The raw signals sampled at 256 Hz frequency were exported from REMbrandt Manager System (MedCare Co, Amsterdam, the Netherlands) and imported to MATLAB 2018b (MathWorks Inc., Natick, Massachusetts, USA), which was used to conduct all preprocessing tasks. The signals were truncated so that only the time between the lights off mark and lights on mark was included and normalized using z-score normalization, that is, subtracting the signal mean and dividing by standard deviation resulting in a signal with zero mean and a standard deviation of one. The normalization was done to unify the greatly varying signal amplitudes between different patients.

Table 1. Subject characteristics

	Mean	Range	SD
Age (years)	50.9	18.0–88.0	13.8
BMI (kg/m^2)	30.8	13.8–63.7	6.4
AHI 1/h	30.0	0.3–148.1	28.9
MSL (min)	10.2	0.5–20.0	5.1
Recording duration (h)	7.2	6.0–8.7	0.4
	Number	Percentage	
Total number of patients	2,014		
Male patients	1,492	74.1	
Female patients	522	25.9	
EDS category			
Normal	368	18.3	
Mild	649	32.2	
Moderate	580	28.8	
Severe	417	20.7	
OSA category			
Normal	401	19.9	
Mild	438	21.8	
Moderate	422	21.0	
Severe	753	37.4	

Number and percentage for categorical variables and mean, range and standard deviation for continuous variables.
EDS, excessive daytime sleepiness; OSA, obstructive sleep apnea; BMI, body mass index; AHI, apnea–hypopnea index; MSL, mean sleep latency; SD, standard deviation.

The signals were divided into 512 epochs with 50% overlap. No padding at the start or end was used. As a result, each epoch length was 2/513rds of the time between lights off and lights on marks. Welch's power spectral density (PSD) estimate [27] was then calculated for each epoch using 8 windows with 50% overlap. The PSD estimate was calculated for a frequency range of 0.3–30.3Hz using 512 data points. This frequency range was chosen since it contains the common diagnostic bands (Delta, Theta, Alpha, and Beta) and, for example, AASM recommends filtering out frequencies outside 0.3–35Hz when scoring sleep [28, 29]. The PSD estimates were converted to dB scale ($x_{db} = 10 \log_{10} x$) and arranged into a 512 × 512 spectrogram image where one column corresponds to one epoch. The same procedure was repeated for all four channels and for each patient. Finally, all spectrograms were arranged into a 2,014 × 512 × 512 × 4 matrix where the first dimension represents patients, the second and third dimensions represent the spectrograms, and the fourth dimension represents the four signal channels. Example figure of the spectrograms are presented in Figure 1.

Neural network

The CNN was trained in Python 3.7.3 with Tensorflow 1.14.0 using Keras 2.2.4. The CNN consisted of four convolutional blocks and one fully connected block (Figure 2). Each convolutional block consisted of two 2D-convolution layers followed by a max pooling layer with a pool size of 2 × 2 and a stride of 2–2. All convolution layers used 3 × 3 convolution kernels, stride of 1–1 and a tanh activation function. The number of output filters of the convolutional layers was 12 in the first block, 18 in the second block, 24 in the third block, and 30 in the fourth block. The last block consisted of a dropout layer with a 0.3 dropout
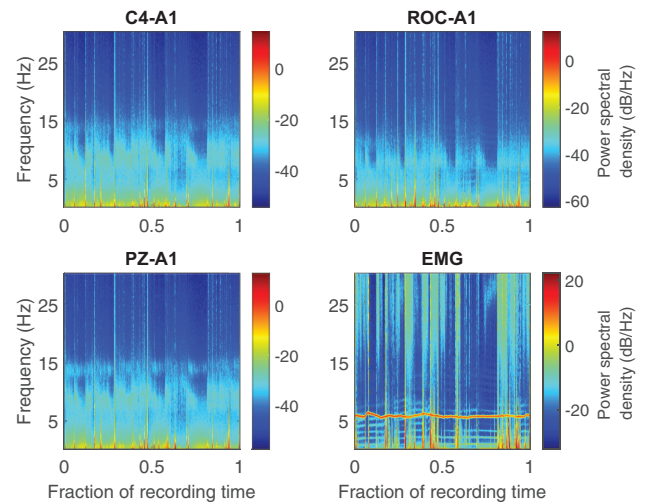

Figure 1. Example of the spectrograms given to the convolutional neural network as an input.
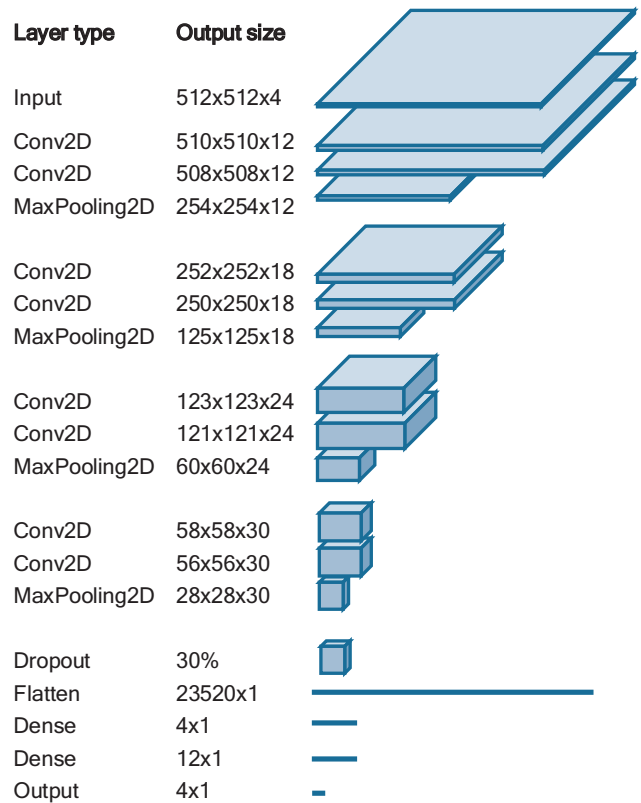

Figure 2. Structure of the convolutional neural network.

followed by a flattening layer and two fully connected layers with layer sizes of 4 and 12 and a ReLU activation. The last layer, that is, the output layer, was a fully connected layer with a size of 4 and a softmax activation. The network was trained with the Adam optimizer using a learning rate of 0.0001. Different neural network structures were also tested but they resulted in worse performance (see Table S2). The network with the lowest mean validation set loss was selected from the tested networks. We used class weighting during training to mitigate the effect of
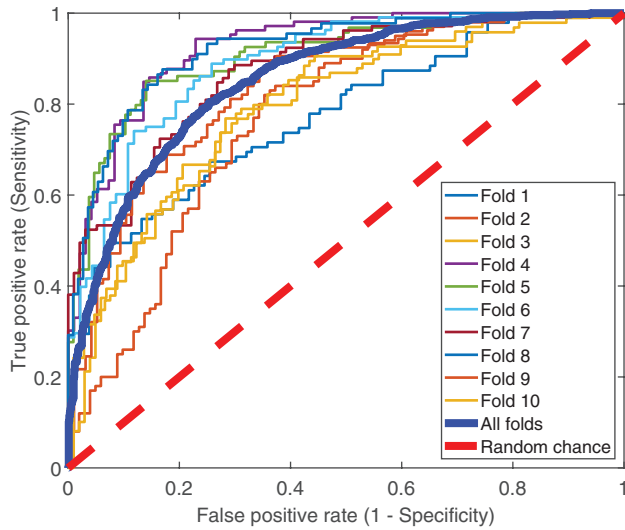
Figure 3. ROC curves for the classifier in each fold and across all folds.

Figure 4. Confusion matrix showing the classification accuracy of the convolutional neural network classifier across all folds.

imbalanced classes. Each class weight was set to be inversely proportional to the number of patients in the class.

We used 10-fold cross-validation to test the performance of the classifier. The patient population was randomly divided into 10 subpopulations, each consisting of 201 or 202 patients. The CNN was trained 10 times such that each subpopulation was used once as a test set, and 9 times in the training set. During each fold, 10% of the training set was further used as the validation set to assess the performance during training and to avoid overfitting. The training accuracy was monitored during training using sparse categorical cross-entropy as the loss function. The training was stopped after the validation set loss did not decrease for 100 continuous epochs after which the model with the lowest validation loss was selected as the model for that fold.

To further interpret the model, we performed an occlusion test to estimate the relative importance of different parts of the spectrogram. A 32 × 32 mask, that sets the spectrogram values under the mask to zero, was used to occlude part of the spectrogram and these occluded spectrograms were given as an input to the trained classifier. The process was repeated by moving the mask over the whole spectrogram with no overlap resulting in a total of 256 occlusions. The accuracy of the classifier was then calculated for each occlusion.

Results

By using a single, 10-min, threshold for EDS classification, the classifier achieved an accuracy of 77.2% in differentiating sleepy and non-sleepy patients with suspected OSA. Sensitivity and specificity of the classifier were 76.5% and 77.9%, respectively. The receiver operating characteristic (ROC) curves for the classifier in each fold and across all folds are presented in Figure 3. The area under ROC curve (AUC) for the classifier across all folds was 0.853. The classifier achieved a positive predictive value of 78.0% and negative predictive value of 76.5%. Cohen's kappa [30] value for the binary classification was 0.544 and F1-score was 0.772.

When classifying patients to the four sleepiness categories, the CNN achieved an overall accuracy of 60.6%. Cohen's kappa

[30] value for the classifier was 0.464. The training, validation, and test set accuracies varied slightly between the folds. Mean training, validation, and test accuracies were 70.7%, 61.1%, and 60.6% with standard deviations of 4.5%, 7.3%, and 8.3%, respectively (see Table S1 for full cross-validation statistics). Confusion matrix showing the patient classification across all folds is presented in Figure 4. The CNN performed best in the moderate sleepiness category with an accuracy of 66.9% and worst in the normal category with an accuracy of 52.0%.

To assess which group of patients is most likely to be classified correctly, the classification accuracy was compared in age, sex, BMI, and AHI subgroups (Table 1). The classification accuracy varied slightly between the subgroups. Patients with severe OSA were slightly more likely to be classified correctly than patients with lesser severity of OSA. Patients with higher BMI or age were also classified slightly more accurately than patients with low BMI or age.

The results of the occlusion test when classifying the patients to the four sleepiness categories are presented in Figure 5. The accuracy varied greatly between the occlusions. Occluding the lower frequencies (0–15Hz) had slightly more detrimental effect on the accuracy of the classifier than occluding the higher frequencies.

Discussion

We developed a CNN classifier that estimates daytime sleepiness based on polysomnographic (EEG, EOG, and chin EMG) signals recorded the night before MSLT. We found that the classifier was able to estimate sleepiness with moderate accuracy. The classifier classified patients to all sleepiness categories relatively evenly with no apparent bias for any sleepiness category (Figure 3).

In detecting EDS, the sensitivity (76.5%) and specificity (77.9%) were good with reasonably high positive predictive value (78.0%) and negative predictive value (76.5%). In comparison, similar sensitivities (70% and 80%) and negative predictive values (75% and 76%) have been reported with ESS using cohort-optimized cutoff values (16 and 12 points) [31, 32]. However, the
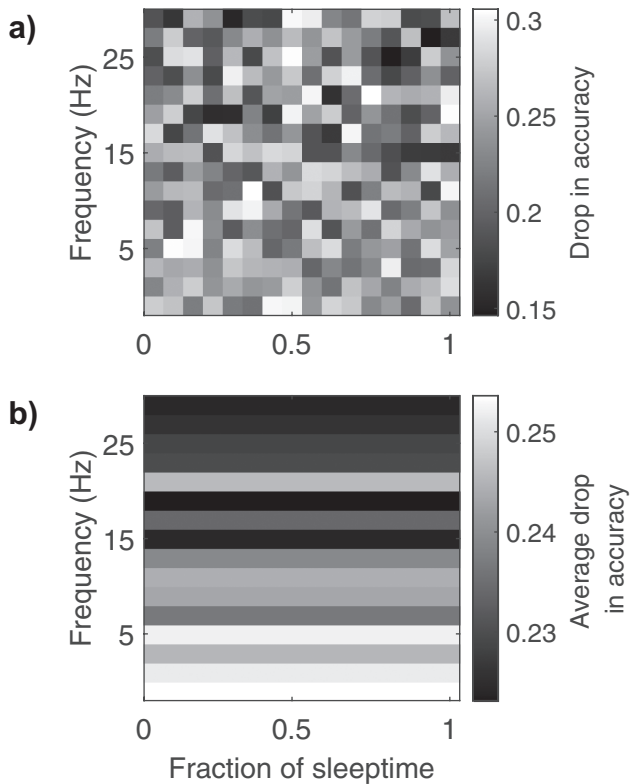
Figure 5. Occlusion plots for the convolutional neural network classifier when classifying patients to the four sleepiness categories. All 32 × 32 occlusions (A) showing the difference in classification accuracy when the corresponding area of the input spectrograms are occluded. Time average of the occlusions (B) showing the average drop in accuracy for each frequency. Brighter color corresponds to a larger drop in accuracy, that is, the occluded area is more important, and darker color corresponds to a smaller drop in accuracy.

Table 2. Classification accuracy in subgroups across all folds when classifying patients to the four sleepiness categories

Subgroup	Number of patients in subgroup	Classification accuracy (%)
Males	1,492	61.5
Females	522	57.9
AHI < 5	401	52.1
5 ≤ AHI < 15	438	59.4
15 ≤ AHI < 30	422	57.5
AHI ≥ 30	753	67.5
BMI < 25	355	56.7
25 ≤ BMI < 30	660	60.0
30 ≤ BMI < 35	598	62.7
BMI ≥ 35	421	61.5
age < 40	430	54.2
40 ≤ age < 50	406	60.1
50 ≤ age < 60	668	62.7
age ≥ 60	510	63.5

AHI, apnea–hypopnea index; BMI, body mass index.

The occlusion test showed that there does not seem to be a well-defined, specific region in the spectrogram image that is most important for sleepiness classification (Figure 5, A). However, occluding the frequencies less than 15Hz seemed to have a more detrimental effect on the accuracy of the classifier (Figure 5, B). This makes sense since most of the power in the spectrogram is at the lower frequencies. In addition, delta waves, associated with deep sleep are at this low frequency range (0.5–4Hz) [28]. As the amount of slow wave sleep is important in recovery during the night and greatly affects sleepiness, it could be that the amount of slow wave sleep detected from the spectrogram is a major component of the classifier function. However, since the accuracy suffered at least slightly when any part of the image was occluded, it seems that the whole night and all frequencies are at least somewhat important for the classifier.

This study has certain limitations. The use of Pz electrode might limit generalizability of the classifier since it is not a commonly used electrode placement in PSG montage. However omitting this electrode would have lowered classifier accuracy (see supplement). Although the accuracy of the classifier was moderate, it still leaves room for improvement. One complicating factor in estimating the MSLT result is that the patient's sleepiness is not entirely dependent on the previous night. Some of the patients might have been sleep deprived for a long time while others might only be sleepy because of poor sleep during the previous night polysomnography. While they both might be classified to the severe sleepiness category, their EEG and EMG spectrograms are likely significantly different. That is, all information on the patient's sleepiness is not available in the single night polysomnographic recording, which limits the performance of the classifier. Another limiting factor is the patient population. Although the patient population was relatively large, using even larger population would have likely improved the results. Larger population would also have allowed a larger test set and thus enabled a more robust validation of the developed classifier. Another limiting factor of the patient population is that the baseline ESS test was not conducted and therefore we could not compare classifier accuracy to ESS or assess the subjective and objective sleepiness in the

specificities (55% and 69%) were considerably lower compared to our classifier along with lower positive predictive values (61% and 74%) [31, 32]. High specificity (76%) has also been reported using ESS, but with low sensitivity (64%) [33]. Simultaneous high sensitivity and specificity has been difficult to achieve with ESS even when using cohort-optimized cutoff values [31–33].

Based on the present results, the CNN classifier developed in this study seems to be able to estimate sleepiness slightly better than ESS [31–33]. However, it is important to note that ESS is better suited as a measure of chronic, long-time sleepiness rather than the acute sleepiness. Therefore, ESS is still a valuable tool in sleepiness estimation. The classifier could be used as a simple estimator of the patients' sleepiness since it is easy to implement and does not require long and laborious full-day test (i.e. MSLT) while still providing an objective estimate of the patient's acute daytime sleepiness. In addition, ESS could be used in conjunction with the classifier to provide information on the chronic situation.

In the subgroup analysis, older patients, patients with severe OSA or patients with high BMI were classified slightly more accurately than younger patients, patients with lower severity of OSA or patients with low BMI (Table 2). Since OSA severity generally increases with age and obesity [34], it could be that the sleepiness of these patients is mainly caused by the sleep apnea, which might be more clearly detectable from the spectrograms.

same population. In addition, no information on medications or comorbidities was available for this study population. This can be considered a study limitation as both of these could have an effect on the patients' sleepiness. It is also important to note that all of the patients in this study were suspected OSA patients complaining from daytime sleepiness during clinical interview. This results in a biased population only consisting of patients with subjective sleepiness while including no patients who were objectively sleepy but not subjectively sleepy. Thus, the network might behave differently if applied to a different population such as a healthy population or to a population with different sleep disorders.

In conclusion, objective estimation of daytime sleepiness using polysomnographic signals shows promising results. The developed CNN classifier could be applied for OSA patients that undergo polysomnography to get an objective EDS evaluation with minimal workload.

Supplementary material

Supplementary material is available at *SLEEP* online.

Disclosure Statement

References

1. Heinzer R, *et al*. Prevalence of sleep-disordered breathing in the general population: the HypnoLaus study. *Lancet Respir Med*. 2015;**3**:310–318.
2. Benjafield AV, *et al*. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;**7**:687–698.
3. Regestein Q, *et al*. Sleep debt and depression in female college students. *Psychiatry Res*. 2010;**176**:34–39.
4. Ohayon MM, *et al*. Daytime sleepiness and cognitive impairment in the elderly population. *Arch Intern Med*. 2002;**162**:201–208.
5. Slater G, *et al*. Excessive daytime sleepiness in sleep disorders. *J Thorac Dis*. 2012;**4**:608–616.
6. Knauert M, *et al*. Clinical consequences and economic costs of untreated obstructive sleep apnea syndrome. *World J Otorhinolaryngol Head Neck Surg*. 2015;**1**:17–27.
7. Carskadon MA, *et al*. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep*. 1986;**9**:519–524.
8. American Academy of Sleep Medicine. *The International Classification of Sleep Disorders, Revised Daignostic and Coding Manual*. Chicago, IL: American Academy of Sleep Medicine, 2001. doi:10.1145/3274694.3274742
9. Thorpy MJ. The clinical use of the multiple sleep latency test. The standards of practice committee of the American sleep disorders association. *Sleep*. 1992;**15**:268–276.
10. Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. *Chest*. 2014;**146**: 1387–1394.
11. Arand D, *et al*. The clinical use of the MSLT and MWT. *Sleep*. 2005;**28**:123–144.
12. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth Sleepiness Scale. *Chest*. 1993;**103**:30–36.
13. Buysse DJ, *et al*. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res*. 1988;**28**:193–213. doi:10.1016/0165-1781(89)90047-4
14. Fong SYY, *et al*. Comparing MSLT ansd ESS in the measurement of excessive daytime sleepiness in obstructive sleep apnoea syndrome. *J Psychosom Res*. 2005;**58**:55–60.
15. Park YK, *et al*. Clinical and polysomnographic characteristics of patients with excessive daytime sleepiness. *Sleep Med Res*. 2018;**9**:32–38.
16. Olson LG, *et al*. Correlations among Epworth Sleepiness Scale scores, multiple sleep latency tests and psychological symptoms. *J Sleep Res*. 1998;**7**:248–253.
17. Miotto R, *et al*. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2017;**19**:1236–1246.
18. Uddin MB, *et al*. Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: a systematic review. *Physiol Meas*. 2018;**39**:03TR01.
19. Nikkonen S, *et al*. Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea. *Sci Rep*. 2019;**9**:13200.

20. Boostani R, *et al*. A comparative review on sleep stage classification methods in patients and healthy individuals. *Comput Methods Programs Biomed*. 2017;**140**:77–91.

21. Sun H, *et al*. Large-scale automated sleep staging. *Sleep*. 2017;**40**. doi:10.1093/sleep/zsx139

22. Zhang W, *et al*. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Appl Opt*. 1990;**29**:4790–4797.

23. American Academy of Sleep Medicine. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. *Sleep*. 1999;**22**:662–689.

24. American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. **Vol 1**. Westchester, IL: American Academy of Sleep Medicine, 2007.

25. Littner MR, *et al*. Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep*. 2005;**28**:113–121.

26. Jasper HH. The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol*. 1958;**10**:371–375.

27. Welch P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans audio Electroacoust*. 1967;**15**:70–73.

28. Bronzino J. *Medical Devices and Systems, The Biomedical Engineering Handbook,* 3rd ed. Boca Raton, FL: CRC Press, 2006.

29. American Academy of Sleep Medicine. *AASM Manual for the Scoring of Sleep and Associated Events*. Darien, IL: American Academy of Sleep Medicine, 2017.

30. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;**20**:37–46.

31. Trimmel K, *et al*. Wanted: a better cut-off value for the Epworth Sleepiness Scale. *Wien Klin Wochenschr*. 2018;**130**:349–355.

32. Cai SJ, *et al*. Correlation of Epworth Sleepiness Scale with multiple sleep latency test and its diagnostic accuracy in assessing excessive daytime sleepiness in patients with obstructive sleep apnea hypopnea syndrome. *Chin Med J (Engl)*. 2013;**126**:3245–3250.

33. Chung KF. Use of the Epworth Sleepiness Scale in Chinese patients with obstructive sleep apnea and normal hospital employees. *J Psychosom Res*. 2000;**49**:367–372.

34. Jordan AS, *et al*. Adult obstructive sleep apnoea. *Lancet*. 2014;**383**:736–747.