

<https://helda.helsinki.fi>

Deep Learning Enables Accurate Automatic Sleep Staging Based on Ambulatory Forehead EEG

Leino, Akseli

2022

Leino , A , Korkalainen , H , Kalevo , L , Nikkonen , S , Kainulainen , S , Ryan , A , Duce , B , Sipila , K , Ahlberg , J , Sahlman , J , Miettinen , T , Westeren-Punnonen , S , Mervaala , E , Toyras , J , Myllymaa , S , Leppanen , T & Myllymaa , K 2022 , ' Deep Learning Enables Accurate Automatic Sleep Staging Based on Ambulatory Forehead EEG ' , IEEE Access , vol. 10 , pp. 26554-26566 . <https://doi.org/10.1109/ACCESS.2022.3154899>

<http://hdl.handle.net/10138/342641>

<https://doi.org/10.1109/ACCESS.2022.3154899>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Received February 11, 2022, accepted February 21, 2022, date of publication February 25, 2022, date of current version March 14, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3154899

Deep Learning Enables Accurate Automatic Sleep Staging Based on Ambulatory Forehead EEG

AKSELI LEINO^{1,2}, HENRI KORKALAINEN^{1,2}, LAURA KALEVO^{1,2}, SAMI NIKKONEN^{1,2},
SAMU KAINULAINEN^{1,2}, ALEXANDER RYAN³, BRETT DUCE^{3,4}, KIRSI SIPILÄ^{5,6},
JARI AHLBERG⁷, JOHANNA SAHLMAN⁸, TOMI MIETTINEN⁹,
SUSANNA WESTEREN-PUNNONEN², ESA MERVAALA^{2,10}, JUHA TÖYRÄS^{1,11,12},
SAMI MYLLYMAA^{1,2}, TIMO LEPPÄNEN^{1,2,12}, AND KATJA MYLLYMAA^{1,2}

¹Department of Applied Physics, University of Eastern Finland, 70211 Kuopio, Finland

²Diagnostic Imaging Center, Kuopio University Hospital, 70210 Kuopio, Finland

³Sleep Disorders Centre, Department of Respiratory and Sleep Medicine, Princess Alexandra Hospital, Brisbane, QLD 4102, Australia

⁴Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD 4000, Australia

⁵Research Unit of Oral Health Sciences, University of Oulu, 90570 Oulu, Finland

⁶Medical Research Center, Oulu University Hospital, University of Oulu, 90570 Oulu, Finland

⁷Department of Oral and Maxillofacial Diseases, University of Helsinki, 00100 Helsinki, Finland

⁸Department of Otorhinolaryngology, Kuopio University Hospital, 70210 Kuopio, Finland

⁹Department of Oncology, University of Helsinki and Helsinki University Hospital, 00029 Helsinki, Finland

¹⁰School of Medicine, Institute of Clinical Medicine, University of Eastern Finland, 70211 Kuopio, Finland

¹¹Science Service Center, Kuopio University Hospital, 70210 Kuopio, Finland

¹²School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia

Corresponding author: Akseli Leino (akseli.leino@uef.fi)

This work was supported in part by the EU Commission Horizon 2020 Framework Program under Grant Agreement 965417; in part by the NordForsk (NordSleep Project 90458-06111) through Business Finland under Grant 5133/31/2018; in part by the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding under Grant 5041767, Grant 5041794, Grant 5041797, Grant 5041798, Grant 5041776, and Grant 5041803; in part by the Academy of Finland under Grant 323536; in part by The Research Foundation of the Pulmonary Diseases; in part by the Finnish Cultural Foundation North Savo Regional Fund; in part by the Finnish Cultural Foundation; in part by the Päivikki and Sakari Sohlberg Foundation; in part by the Finnish Anti-Tuberculosis Association; in part by the Tampere Tuberculosis Foundation; and in part by the Kuopio Area Respiratory Foundation.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee of the Northern Savo Hospital District under Application Nos. 215/2015 and 126/2016 and the Institutional Human Research Ethics Committee of the Princess Alexandra Hospital under Application No. HREC/16/QPAH/668.

ABSTRACT We have previously developed an ambulatory electrode set (AES) for the measurement of electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG). The AES has been proven to be suitable for manual sleep staging and self-application in in-home polysomnography (PSG). To further facilitate the diagnostics of various sleep disorders, this study aimed to utilize a deep learning-based automated sleep staging approach for EEG signals acquired with the AES. The present neural network architecture comprises a combination of convolutional and recurrent neural networks previously shown to achieve excellent sleep scoring accuracy with a single standard EEG channel (F4-M1). In this study, the model was re-trained and tested with 135 EEG signals recorded with AES. The recordings were conducted for subjects suspected of sleep apnea or sleep bruxism. The performance of the deep learning model was evaluated with 10-fold cross-validation using manual scoring of the AES signals as a reference. The accuracy of the neural network sleep staging was 79.7% ($\kappa = 0.729$) for five sleep stages (W, N1, N2, N3, and R), 84.1% ($\kappa = 0.773$) for four sleep stages (W, light sleep, deep sleep, R), and 89.1% ($\kappa = 0.801$) for three sleep stages (W, NREM, R). The utilized neural network was able to accurately determine sleep stages based on EEG channels measured with the AES. The accuracy is comparable to the inter-scorer agreement of standard EEG scorings between international sleep centers. The automatic AES-based sleep staging could potentially improve the availability of PSG studies by facilitating the arrangement of self-administrated in-home PSGs.

INDEX TERMS Ambulatory EEG, automatic scoring, deep learning, electroencephalography, sleep staging.

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

I. INTRODUCTION

Current guidelines defined by the American Academy of Sleep Medicine (AASM) divide sleep into five stages: wake (W), N1, N2, N3, and rapid eye movement (R) [1]. The stages are identified based on electroencephalography (EEG), electrooculography (EOG), and chin electromyography (EMG) signals recorded during polysomnography (PSG). PSG is usually conducted in a sleep laboratory (type I PSG), as the standard 10-20 system EEG electrodes (Fig. 1c) require pre-application by a trained expert and standard type II PSG EEG electrodes are too complex to be fully self-administrated in a home environment.

To overcome this shortcoming of the PSG, various types of headbands and electrode sets have been developed for EEG measurement [2]–[4]. Moreover, we have previously introduced an ambulatory electrode set (AES) enabling the recording of the EEG, EOG, and EMG during sleep [5]. The Ag/AgCl electrodes in the AES are screen-printed on a flexible polyethylene terephthalate (PET) film that attaches easily to the skin with a self-adhesive hydrogel membrane and medical foam. The design of the AES and illustrated instructions help to achieve consistent placement of the electrodes. The AES has 1.5 mm touch-proof safety socket connectors suitable to be used with most modern EEG amplifiers. The EEG and EOG electrodes are located around the face, near the hairline, as presented in the Fig. 1b. The EEG measured with AES has been shown to be suitable for manual sleep staging [6] and the success rate and technical quality of the EEG, EOG, and EMG channels in self-applied AES recordings have been shown to be comparable to conventional type II in-home PSGs [5], [7].

In addition to the complex measurement setup used in a standard type I PSG, the sleep staging process requires time-consuming manual analysis of the recordings by experienced sleep technicians. However, recent advances in deep learning have brought a surge of automated sleep staging applications [8]–[11]. These deep learning applications have achieved excellent sleep staging accuracies and compare well with the inter-scorer agreement between experienced manual scorers from international sleep centers [12]–[14].

Recently, we introduced a deep learning model for the automatic identification of sleep stages [15]. The model achieved an accuracy of 82.9% ($\kappa = 0.77$) by utilizing only a single EEG channel (F4-M1) extracted from type I PSGs. To further enhance the diagnostics of sleep disorders and facilitate the arrangement of type II PSG studies, we aim to utilize a similar model [15] for sleep staging based on signals acquired from AES recordings. Building on the recent results on deep learning-based automatic scoring [15] and the feasibility of the AES for manual sleep scoring [6], we hypothesize that deep learning can be utilized for accurate automatic AES-based sleep staging. Furthermore, electrode malfunctions and other EEG artifacts appear as high amplitude variations in the signals. Thus, we present a novel method for eliminating electrode-originated noise in the neural network

input. We hypothesize that the automatic scoring accuracy can be increased by using the variance of the concurrent EEG epochs as a simple metric for selecting the input for the deep learning model. In addition, we investigate which AES channel derivations provide the most accurate results to advance the development of the AES. Successful implementation of automatic AES-based sleep staging could potentially enhance the clinical usability of the AES and improve the availability of PSG studies by facilitating the measurement of self-administrated type II PSG recordings.

II. MATERIALS AND METHODS

A. PATIENTS AND DATA

Related to the development and clinical validation of the AES, we have tested it in several small cohorts, some of these being currently under analysis and some already published [5], [7]. Three separate datasets were utilized in this study (Table 1). The first dataset consists of 80 home sleep apnea tests (HSAT) supplemented with AES that were conducted at Kuopio University Hospital (Kuopio, Finland) and City of Helsinki Unit of Specialized Oral Care in the Metropolitan Area and Kirkkonummi (Helsinki, Finland) for 50 subjects with suspected sleep bruxism [16] in 2015–2017 (bruxism dataset). The second dataset consists of 33 in-lab polygraph recordings supplemented with AES that were conducted at Kuopio University Hospital (Kuopio, Finland) for suspected sleep apnea patients in 2016–2019 (sleep apnea dataset 1). The third dataset consists of 22 in-lab PSGs supplemented with AES that were conducted for suspected sleep apnea patients at the Princess Alexandra Hospital (Brisbane, Australia) in 2019 (sleep apnea dataset 2). The collection and processing of the datasets were given a favorable statement by The Research Ethics Committee of the Northern Savo Hospital District (215/2015 and 126/2016) and The Institutional Human Research Ethics Committee of the Princess Alexandra Hospital (HREC/16/QPAH/668). Informed consent was acquired from all patients prior to the recording.

In all datasets, the AES signals were recorded with a Nox A1 PSG monitor (Nox Medical, Reykjavík, Iceland). The AES data comprised four EEG (Af8-T9, Fp2-T9, Fp1-T10, and Af7-T10), two EOG (F8-T9 and F7-T10), and four EMG (S1-SF, S2-SF, MassL, MassR) channels (Fig. 1a). EEG, EOG, and chin EMG channels are used for manual scoring of sleep stages, whereas MassL and MassR are for the identification of sleep bruxism episodes. In addition, each electrode of the AES was also recorded against the common ground to enable the acquisition of non-standard channel derivations. The sleep staging was initially performed manually by four experienced scorers according to the latest AASM guidelines [1] based on EEG, EOG, and EMG signals recorded with the AES. Two individual scorers took part in the scoring of the bruxism dataset in Finland, sleep apnea dataset 1 was scored by one scorer in Iceland, and sleep apnea dataset 2 was scored by one scorer in Australia.

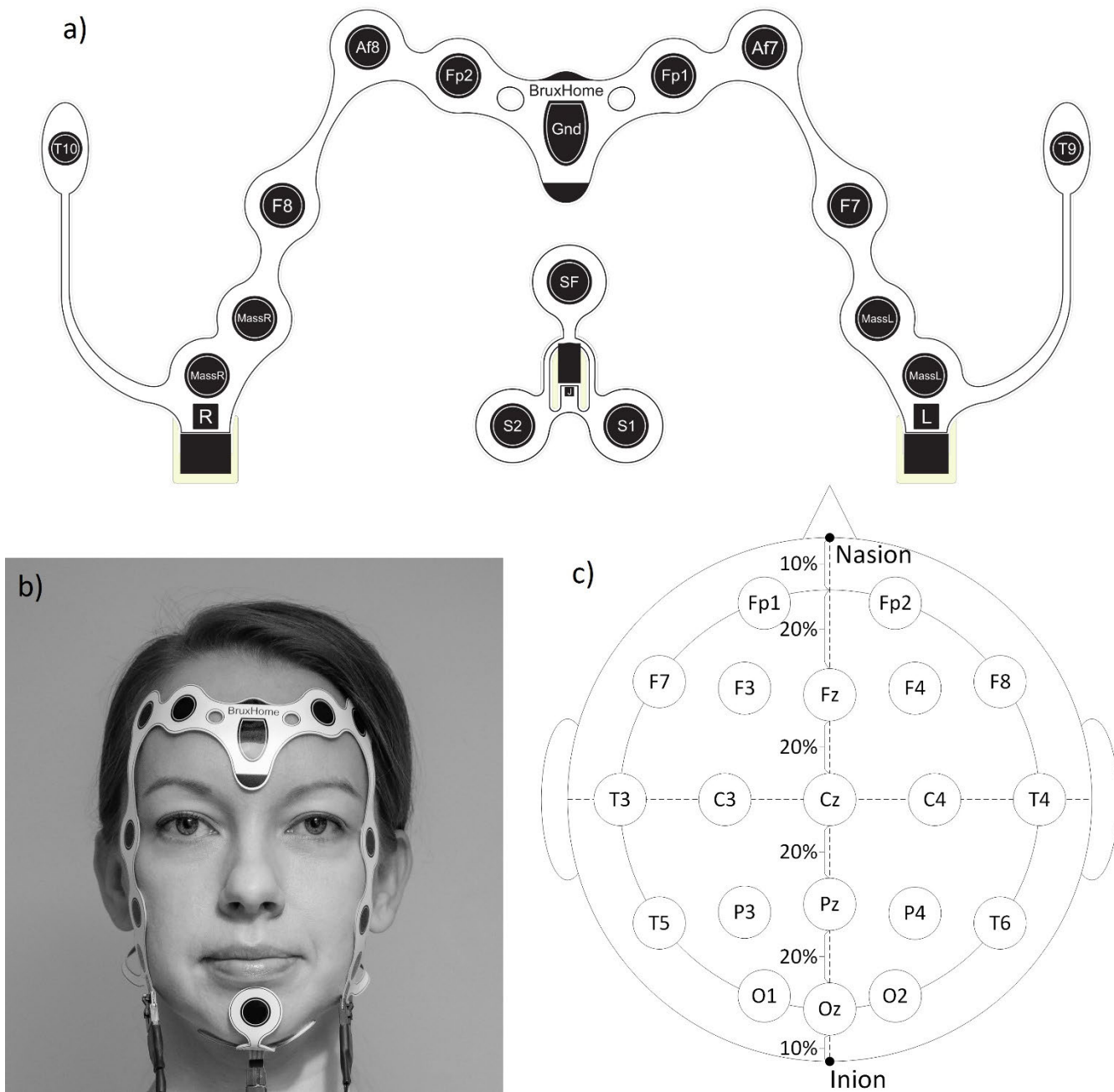


FIGURE 1. (a) The electrode locations in the ambulatory electrode set (AES), (b) the AES worn by a volunteering researcher, and (c) standard 10–20 system electrode locations for comparison.

TABLE 1. Patient demographics of the datasets.

	Bruxism dataset	Sleep apnea dataset 1	Sleep apnea dataset 2
Number of patients	50	33	22
Number of recordings	80	33	22
Age (years)	41.0 (29.0, 47.0)	44.0 (35.0, 56.5)	55.5 (40.3, 64.0)
BMI (kg/m ²)	23.9 (22.1, 26.8)*	27.8 (25.0, 31.8)	33.5 (29.7, 39.0)
AHI (h ⁻¹)	N/A	12.7 (7.5, 25.8)	15.7 (7.3, 36.6)

The values are presented as median (interquartile range).

*Missing data from six patients

BMI, body mass index; AHI, apnea-hypopnea index

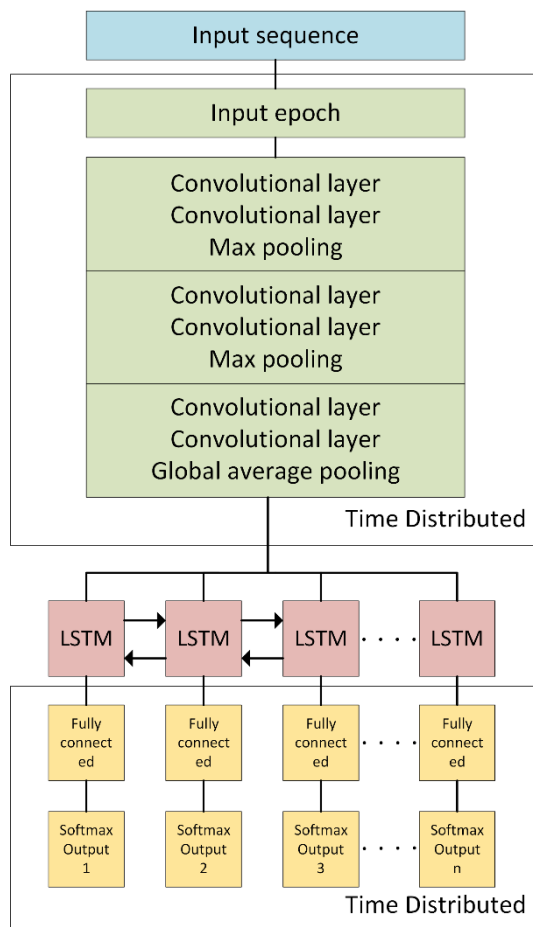


FIGURE 2. The architecture of the utilized neural network. The input sequence consists of 100 epochs that are fed separately to a convolutional neural network with a time distributed layer. The global average pooled feature maps of individual epochs of the sequence are then fed to a long short-term memory (LSTM) layer. The output of the LSTM is then transformed into a sequence of sleep stages with time distributed fully connected layer with a softmax activation function.

B. NEURAL NETWORK ARCHITECTURE

We developed a deep learning model in a previous study, which achieved a sleep staging accuracy of 82.9% ($\kappa = 0.77$) based on a single standard EEG channel (F4-M1) using a dataset comprising 891 PSG recordings [15]. The same neural network architecture was used without modifications in the present study. The architecture comprises a combination of a convolutional neural network (CNN) and a recurrent neural network (RNN) (Fig. 2). The CNN is used for epoch-wise feature extraction and the RNN for the learning of the sequential characteristics of the sleep stages. The input for the neural network is a sequence of a hundred consecutive 30 s EEG epochs and the output is a sequence of the corresponding sleep stages. The sequence length was set to one hundred as a compromise between memory preservation, data augmentation, and sufficiently long sequences for capturing sleep cycles.

The CNN comprised three convolutional blocks with each including two convolutional layers and a pooling layer. The first two blocks end with a max pooling layer with a size of 2 and the last block ends with a global average pooling layer. In the first block, the number of kernels was 128 and the kernel size was 21 data points. The first convolutional layer had a stride size of 5 to reduce the amount of data, and all later convolutional layers had a stride size of 1. In the latter two blocks, the number of kernels was 256, and the kernel size was 5. Each convolutional layer was followed by a batch normalization layer. After the convolutional neural network, a Gaussian dropout was applied. The data was then fed into a bidirectional long short-term memory (LSTM) neural network. The LSTM had 256 units, a dropout value of 0.3 and a recurrent dropout value of 0.5. The last layer of the neural network was a fully connected dense layer with softmax activation, which can be interpreted as a probability for each sleep stage.

C. DATA PREPARATION

The EEG and EMG signals were zero-phase filtered with Butterworth bandpass filter (0.3–32 Hz) and downsampled from 200 Hz to 64 Hz as the deep learning model was originally developed for 64 Hz signal [15]. To unify the scale of the signals, each signal was z-score normalized. After the signal processing, the signals were divided into 30 s epochs and the epochs were set to sequences of 100 consecutive epochs. In the training and validation sets, these sequences were generated with 75% overlap between adjacent sequences to multiply the amount of the training data whereas test set sequences had no overlap. In accordance with previous studies [10], [15], [17], we included a maximum of 30 minutes of wake EEG signal before and after sleep whenever excess signal was recorded. Furthermore, to mimic AASM guidelines of switching to contralateral channels in case of poor signal quality, we also created new combination channels. The combination channels were constructed by calculating the variance of concurrent EEG epochs of the opposite EEG channels (e.g., Fp1-T10 and Fp2-T9) and adding the epoch with lower variance to the combination channel. With this method, we can ideally select the side with less high amplitude electrode-based noise from the opposite EEG channels. These combination channels are later referenced as Fp1/Fp2 and Af7/Af8 combination channels.

D. TRAINING OF THE NEURAL NETWORK

The 10-fold cross-validation was used to test the performance of the network to reduce the bias caused by the relatively small dataset. The three datasets were divided into 10 equal-sized folds so that each fold had a constant number of recordings from each dataset. One of the folds was used for testing, one for validation, and the remaining eight folds were used for training. Ten different models were then trained so that each fold was used once in the validation set and once in the test set. Each model was trained for a maximum

TABLE 2. The neural network 5-stage sleep scoring accuracies.

Input channel	Accuracy, % (κ)			
	All datasets	Bruxism dataset	Sleep apnea dataset 1	Sleep apnea dataset 2
Fp1/Fp2 combination channel	79.7 (0.729)	79.6 (0.727)	80.2 (0.738)	80.5 (0.727)
Fp2	77.5 (0.698)	78.3 (0.708)	77.6 (0.703)	74.2 (0.638)
Fp1	79.0 (0.719)	79.1 (0.720)	80.6 (0.743)	77.1 (0.678)
Fp1 and Fp2	78.8 (0.716)	78.3 (0.708)	79.7 (0.731)	79.9 (0.717)
Fp2-GND	77.8 (0.703)	79.0 (0.718)	77.3 (0.699)	74.3 (0.641)
Af7/Af8 combination	77.8 (0.703)	77.5 (0.699)	78.6 (0.717)	78.4 (0.699)
All EEG and EOG channels (Fp1, Fp2, Af7, Af8, F7, F8)	76.8 (0.688)	77.7 (0.700)	77.5 (0.702)	72.3 (0.609)
Fp1/Fp2 combination and S2-Sf EMG channel	76.8 (0.690)	77.6 (0.700)	77.6 (0.704)	73.0 (0.620)

The accuracies are calculated across all test sets in 10-fold cross validation for models trained with different inputs. The results are reported as a percent agreement between neural network scoring and manual scoring, followed by Cohen's kappa (κ). The channels are referenced to contralateral T9 or T10 electrodes unless otherwise stated.

of 200 cycles or until the value of the validation loss function had not decreased during 20 consecutive cycles. All calculations were conducted on a server with 32-core AMD Ryzen Threadripper 2990WX, 128GB of RAM, and NVIDIA GeForce RTX 2080. With our server setup, the training of a single model takes approximately 2–4 hours and evaluation of a single patient data with a trained model takes only a few seconds.

E. DATA ANALYSIS

The performance of the deep learning model was evaluated as the test set accuracies across all the 10 folds. Furthermore, the inter-scoring agreement between automatic scoring and manual scoring was evaluated with Cohen's kappa coefficient (κ). Performance was calculated for complete test sets (all datasets included) and independently for each dataset by extracting the corresponding patients from the test sets. The confusion matrices were drawn for 5-stage, 4-stage, and 3-stage scorings across all folds including patients from all datasets. The same model was used for the 4-stage and 3-stage scorings and confusion matrices were derived by combining N1 and N2 to light sleep (4-stage), and N1, N2, and N3 to NREM (3-stage).

To further assess the usability of the AES in clinical use, total sleep time (TST) and wake after sleep onset (WASO) were calculated for the automatic and manual scorings. The results are reported as medians, interquartile ranges, and Bland-Altman plots with mean difference and 95% confidence interval. Furthermore, the time spent in each sleep stage was calculated for each dataset for manual and automatic scoring. The statistical significance between manually and automatically acquired sleep parameters was evaluated with Wilcoxon signed-rank test using $p < 0.001$ as the limit for statistical significance.

III. RESULTS

The performances of the neural network-based sleep staging with different channel derivations are presented in Table 2. The highest accuracy was achieved with the Fp1/Fp2

combination channel, outperforming the Fp2, Fp1, and Fp2-GND channels by 2.2%, 0.7%, and 1.9%, respectively. The lowest accuracy was acquired either with Fp1/Fp2 combination channel accompanied with S2-Sf EMG channel or when using all EEG and EOG channels. However, when considering all datasets, the overall accuracies with every tested input were close to each other, and the difference between the highest and the lowest accuracy was only 2.9%.

The accuracies of the individual datasets show the largest differences in the sleep apnea dataset 2, where the highest accuracy with Fp1/Fp2 combination channel was 80.5% and the lowest accuracy with all EEG and EOG channels was 72.3%. The most consistent performance throughout the three datasets was achieved with Fp1/Fp2 combination channel. However, adding S2-SF EMG channel to the input with Fp1/Fp2 combination channel slightly worsened the accuracy, especially in the sleep apnea 2 dataset.

The confusion matrices of the most accurate Fp1/Fp2 combination channel are presented in Fig. 3a-c for the 5-stage, 4-stage, and 3-stage predictions. The 5-stage neural network model achieved high accuracy in the W, N2, N3, and R stages. However, it struggled to identify the N1 stage. When the N1 and N2 stages were combined into light sleep, the overall accuracy increased from 79.7% to 84.1% ($\kappa = 0.773$). Scoring to only 3 stages (W, NREM, R) further increased the accuracy to 89.1% ($\kappa = 0.801$). Furthermore, the sensitivity, specificity, and F1-score for each sleep stage with 5-stage scoring based on Fp1/Fp2 combination channel in all datasets are presented in the Table 4. In general, sensitivity and specificity are high in all sleep stages except N1.

Optimal sequence length was also tested with the best performing Fp1/Fp2 combination channel. The accuracies for sequence lengths of 25, 50, 75, 100, and 125 were 78.5%, 79.0%, 79.6%, 79.7%, and 79.4%, respectively. Thus, optimal sequence length was 100 epochs also used in the previous study by Korkalainen *et al.* [15].

The total sleep time (TST, Fig. 4a, Table 3), wake after sleep onset (WASO, Fig. 4b, Table 3), and time spent in

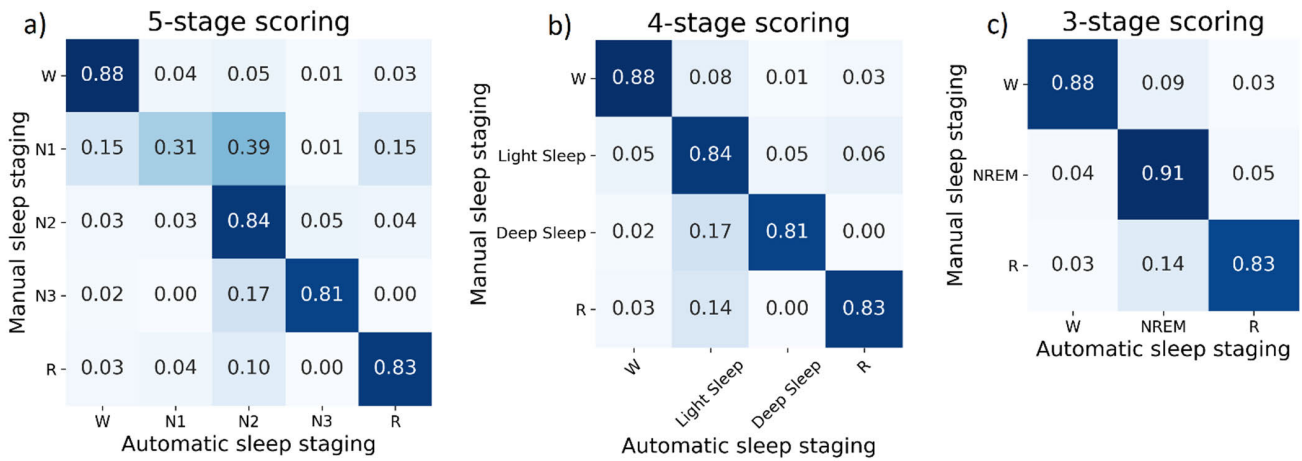


FIGURE 3. Confusion matrix for the (a) 5-stage, (b) 4-stage (wake (W), light sleep (N1 + N2), deep sleep (N3), rapid eye movement (R)), and (c) 3-stage (W, non-rapid eye movement (NREM), R) neural network sleep scoring based on Fp1/Fp2 combination channel recorded with ambulatory electrode set (AES). The results are calculated across all test sets in 10-fold cross-validation including patients from all datasets ($n = 135$).

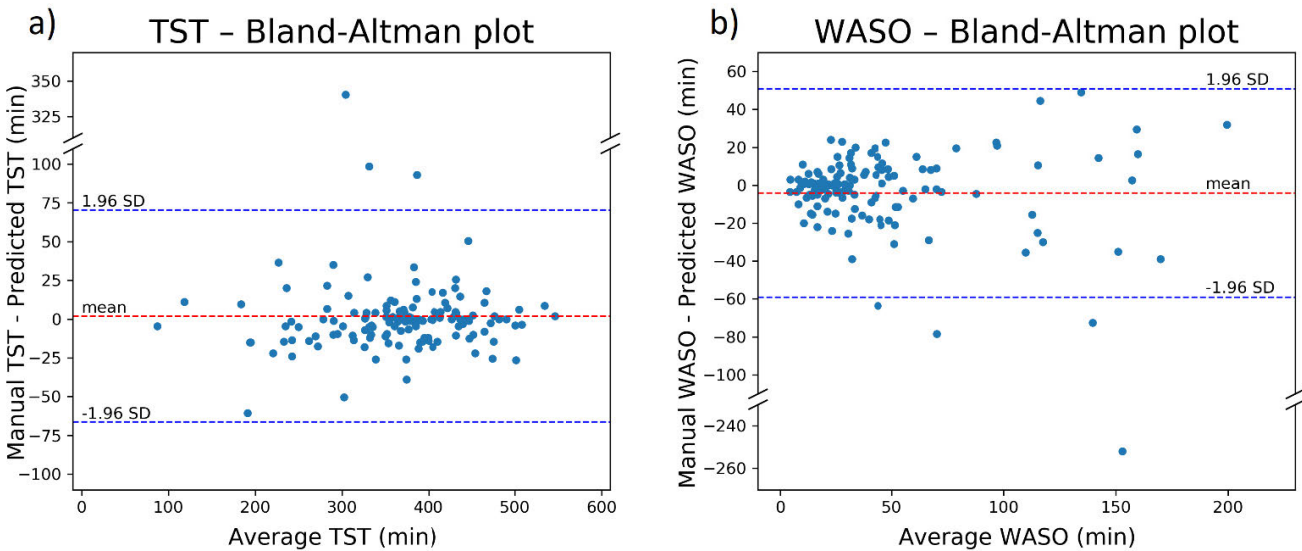


FIGURE 4. Bland-Altman plots of the (a) total sleep time (TST) and (b) wake after sleep onset (WASO) calculated from the difference of manual and automatic scoring. The results are calculated with all test sets in 10-fold cross-validation including patients from all datasets ($n = 135$).

different sleep stages (Table 3) were calculated from the manual scoring and the predictions made with the Fp1/Fp2 combination channel. TST and WASO show only minor differences between medians of manual and automated scoring (Table 3). Furthermore, the number of each sleep stage and unscored epochs based on manual scoring are presented in Table 5. The deep learning model slightly overestimates the time spent in N2 and underestimates the time spent in N1 (Table 3), which can also be seen from the 5-stage confusion matrix (Fig. 3a). Based on the Bland-Altman plots, the constant bias of the TST and WASO is minimal (Fig. 4).

To better illustrate the performance of the automatic scoring algorithm, hypnograms of the automatic and manual

scoring of a subject with median scoring accuracy were plotted (Fig. 5). By visual inspection, the automatic scoring manages to identify the ultradian rhythm well. The manual scoring, however, has more rapid sleep stage transitions. Furthermore, a Sankey diagram of the scored sleep stages between manual and automatic scoring was plotted (Fig. 6). From the Sankey diagram and confusion matrix (Fig. 3a) sleep stage pairs that rarely get mixed up (e.g., REM-N3) and pairs that sometimes cause confusion (e.g., N1-N2 or N2-N3) can be identified.

IV. DISCUSSION

In this study, we utilized a previously developed deep learning model [15] for sleep staging based on frontal EEG, EOG,

TABLE 3. Sleep parameters and time spent in different sleep stages based on the manual and automatic scorings.

	Input channel	All datasets	Bruxism dataset	Sleep apnea dataset 1	Sleep apnea dataset 2
Manual scoring	TST (min)	378.0 (326.0, 429.0)	394.8 (358.9, 439.3)	369.0 (307.5, 433.0)	300.3 (246.4, 327.5)
	WASO (min)	61.0 (34.5, 102.5)	44.5 (24.8, 71.0)	79.0 (47.0, 104.0)	143.0 (85.3, 293.5)
	W (min)	88.0 (73.3, 109.8)	78.0 (70.0, 94.3)	96.0 (82.5, 111.5)	131.0 (102.1, 201.0)
	N1 (min)	34.0 (21.5, 55.5)	35.8 (24.4, 54.1)	53.0 (27.0, 70.5)	13.0 (7.5, 23.3)
	N2 (min)	171.0 (146.3, 197.0)	182.0 (157.3, 208.3)	155.0 (112.5, 189.0)	157.3 (144.8, 176.4)
	N3 (min)	81.5 (56.5, 105.3)	89.0 (68.8, 113.0)	63.5 (29.0, 88.0)	59.0 (29.9, 82.5)
	R (min)	79.0 (60.0, 99.8)	78.5 (65.3, 96.7)	94.5 (64.5, 103.5)	36.8 (27.4, 83.4)
Automatic scoring	TST (min)	374.0 (326.0, 419.3)	396.3 (357.4, 430.5)	364.0 (299.5, 414.5)	307.5 (242.0, 338.5)
	WASO (min)	62.0 (41.0, 115.5)	49.5 (33.0, 89.8)	69.0 (54.0, 116.0)	129.0 (71.3, 267.5)
	W (min)	85.5 (74.3, 109.3)	79.8 (69.9, 93.1)	87.5 (79.0, 110.0)	116.0 (90.1, 187.9)
	N1 (min)	21.5 (11.8, 29.8)	20.8 (13.8, 29.1)	24.0 (15.5, 37.0)	10.3 (4.1, 22.3)
	N2 (min)	183.0 (157.5, 220.0)	193.3 (158.5, 225.0)	176.0 (151.5, 219.5)	168.0 (144.6, 189.4)
	N3 (min)	74.5 (47.0, 106.3)	90.3 (61.1, 117.5)	61.0 (28.0, 97.0)	47.5 (17.6, 89.1)
	R (min)	85.0 (59.3, 105.3)	84.5 (65.9, 106.3)	89.0 (69.5, 102.5)	46.0 (31.4, 88.6)

The results are reported as median (interquartile range). Statistically significant ($p < 0.001$) difference between manual and automatic sleep scorings is marked in bold. Statistical significance was calculated with Wilcoxon signed-rank test.

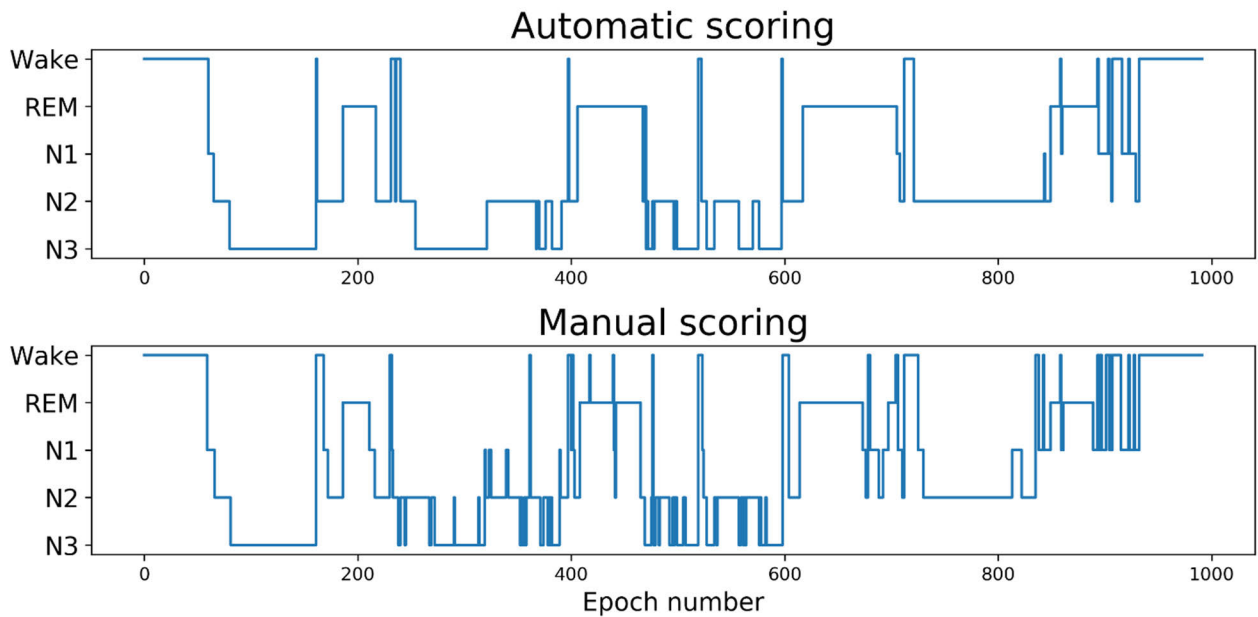


FIGURE 5. An example hypnogram of manual and automatic scoring for a subject with median accuracy.

TABLE 4. Sensitivity, specificity, and F1-score of different sleep stages for 5-stage automatic scoring using all datasets.

	Sensitivity	Specificity	F1-score
Wake	0.86	0.87	0.86
N1	0.52	0.31	0.39
N2	0.77	0.84	0.81
N3	0.86	0.81	0.84
REM	0.80	0.82	0.82
Total	0.79	0.80	0.79

TABLE 5. The number of each sleep stage in the datasets based on manual scoring.

	All datasets	Bruxism	Sleep apnea 1	Sleep apnea 2
Wake	26 777	13 445	6 836	6 496
REM	21 026	12 816	5 913	2 297
N1	10 862	6 657	3 478	727
N2	45 467	28 555	9 958	6 954
N3	21 803	14 844	4 264	2 695
Total scored	125 935	76 317	30 449	19 169
Unscored	3 861	642	2 206	1 013

and EMG data measured with an ambulatory electrode set (AES). When considering all datasets, the highest automatic scoring accuracy 79.7% ($\kappa = 0.729$) was achieved with Fp1/Fp2 combination EEG channel. The accuracy was only

3.2% lower than that achieved with the same deep learning model trained with a nearly 7 times larger PSG dataset [15]. The accuracy is also comparable with the inter-scorer agreement of standard EEG between individual scorers, in which kappa values have been reported to be between 0.58–0.76

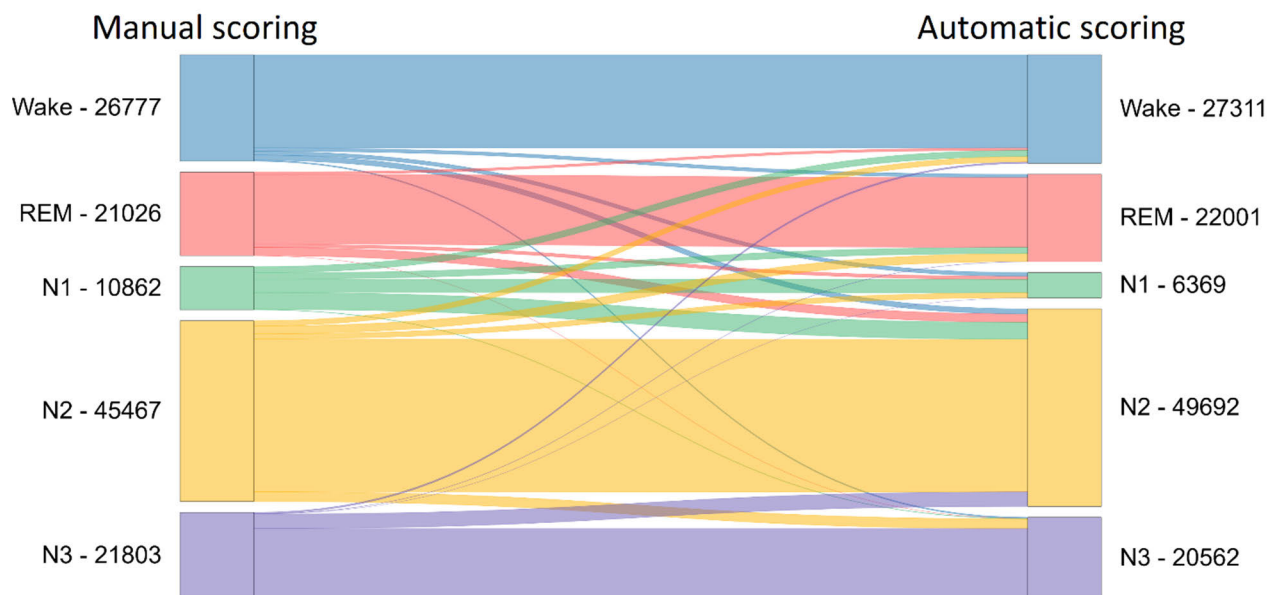


FIGURE 6. A Sankey diagram of the number of sleep stage epochs in manually and automatically scored recordings.

[12]–[14], [18]. The accuracy also exceeded the 76.1% manual inter-scorer agreement previously reported for the AES recordings [6]. Furthermore, the overall accuracy was similar in each of the three datasets, which indicates that the neural network generalizes well to different datasets with distinct recording environments (home, sleep lab), and different independent scorers in multiple international centers.

The most accurate EEG derivation was the Fp1/Fp2 combination channel. This confirms the hypothesis that the automatic scoring accuracy can be slightly increased by using the variance of EEG epochs as a simple metric to choose the optimal input for the deep learning model. The rationale behind this metric is to mimic the AASM scoring rules that recommend switching the scoring from the F4-M1, C4-M1, a O2-M1 channels to F3-M2, C3-M2, and O1-M2 channels in the case of an electrode malfunction [1]. Electrode malfunctions and other EEG artifacts appear as high amplitude variations in the signal. Furthermore, typical electrode-related artifacts appear only in single channels with the AES [5]. Thus, calculating the variance of the epochs from the opposite EEG channels can be used to select the epoch from the opposite channels with less noise induced by poor electrode contact.

In addition, the scoring accuracy of the neural network was the most consistent across the different datasets when using the combination channels. The differences between datasets were negligible when using combination channels, the accuracies being within 0.9% with the Fp1/Fp2 combination channel and 1.1% with the Af8/Af7 combination channel. With single bipolar channels, the differences between datasets were slightly larger. This, however, could be

explained by electrode-based artifacts or other types of noise in single channels.

The largest difference between different inputs was seen in sleep apnea dataset 2, in which the scoring accuracy increased by 6.4% and 3.4% when comparing Fp2 and Fp1 channels to the Fp1/Fp2 combination channel, respectively. Even though the combination channels improved the accuracy compared to single channels, adding more EEG, EOG, or EMG channel derivations did not improve the accuracy. However, this is not the first time this kind of behavior has been reported [19], [20]. Several different deep learning models have been shown to have only minor benefit or even performance degradation from additional input signals [19], [20]. This may be caused by the fact that adding possibly redundant input increases the noise of the deep learning model's input, making it more difficult for the model to extract relevant information from the signals.

When considering all datasets, the highest accuracy (88%) was achieved when scoring wake. Furthermore, N2, N3, and REM were predicted with high accuracy (81–84%). The lowest accuracy (31%) was acquired when predicting the N1 stage. This was expected, as the N1 stage is also the most difficult sleep stage to score manually even for experienced scorers. The inter-scorer agreement between experienced scorers for N1 has been reported to be only $\kappa = 0.31$ – 0.46 [12], [13], [18] based on standard EEG. In addition, the dataset was slightly imbalanced (Table 5), and the N1 stage is the least prevalent stage in all datasets, which further complicates the feature learning of the N1. However, balancing the dataset with sample weights had negative effect on the overall performance based on preliminary tests. This indicates that despite

the imbalance, there was enough information of each sleep stage for reliable feature learning, and the low accuracy of N1 is mainly caused by inconsistent manual scoring.

Based on the confusion matrix of the 5-stage scoring (Fig. 3a), a slight bias towards the N2 stage can be seen. This most probably results from the N2 stage being the most prevalent sleep stage in the datasets (Table 2). An unequal number of different sleep stages causes the loss function to get its lowest value when uncertain epochs are scored as the most prevalent one. This leads to a higher number of false positive N2 identifications when compared with other sleep stages. The bias could be corrected by using sample weights inversely proportional to the number of the corresponding sleep stage during the training. This would, however, have a slight negative impact on the overall accuracy of the neural network.

The medians and Bland-Altman plots for TST and WASO (Fig. 4, Table 3) show only minor differences between the automatic and manual scorings. Furthermore, the 95% confidence intervals of the difference between the manual and estimated parameters are fairly low. Only one extreme outlier is present in the dataset, which is caused by significant ECG artifacts in the EEG channels. The ECG artifacts cause the deep learning model to score significantly more wake than the manual scorer. Similar misinterpretations could be easily avoided in clinical circumstances with a visual inspection before feeding the signal to the neural network.

The achieved 79.7% ($\kappa = 0.73$) scoring accuracy compares favorably to the recent advances in the deep learning-based automatic sleep scoring presented in the literature. Based on standard EEG recordings, several deep learning models with different architectures have achieved an accuracy of 82.9–86.2% ($\kappa = 0.77$ –0.80) [10], [15], [17], [21], [22] with single-channel inputs. Furthermore, a random forest classifier has been used to achieve 72.98% accuracy based on a single frontal EEG channel [23]. There have also been other studies using automatic sleep staging methods on signals collected with different ambulatory EEG systems. Based on signal measured with ear-EEG electrodes, random forest classifier has been used to achieve scoring accuracies of $\kappa = 0.45$ –0.65 [24] and $\kappa = 0.73$ [25]. A fabric headband with forehead electrodes developed by Cognionics Inc. (San Diego, CA, USA) has been recently used to achieve 74.0% deep learning scoring accuracy based on two-channel data [26]. Furthermore, a commercially available device, Dreem headband (Dreem Inc., Paris, France), has been used to achieve 83.5% ($\kappa = 0.748$) accuracies with a deep learning model that was trained with consensus hypnograms of five experienced scorers [4]. However, the deep learning models of these studies have not been published in detail. The high accuracy based on consensus manual scoring highlights the importance of the scoring quality and using data from multiple different scorers to avoid unintentional and unwanted bias in the learning process.

A minor limitation in this study is the size of the dataset ($n = 135$). To compensate for this, we multiplied the number

of the EEG sequences in the training and validation sets by overlapping consecutive sequences. Furthermore, we utilized 10-fold cross-validation to minimize the bias that could be caused by a small test. Despite the relatively small dataset, the model achieved good accuracy, and further increasing the size or number of the AES datasets could increase the accuracy of the automatic sleep staging even further. Furthermore, the neural network was trained with three different datasets analyzed in three different international sleep centers. The automatic scoring accuracy would arguably be higher if the used dataset were scored by one person. Considering this, the achieved accuracy is excellent which indicates that the algorithm is able to find the relevant signal features despite differences in the scorer habits and recording environments. In any case, the algorithm should be thoroughly validated on large AES datasets preferably scored by multiple sleep experts before advancing to routine clinical use.

Another factor that might limit the scoring accuracy is that the AES has been designed to be used without skin abrasion to allow easy self-application. This leads to higher electrode-skin impedances when compared to a standard EEG setup with cup electrodes accompanied with skin abrasion procedure. The high electrode-skin impedance has been reported to lead to noticeable sweat artifact [27], [28] in the delta wave bandwidth (0.5–4.0 Hz), which is suggested to partly originate from the imbalanced impedances of the electrodes [29]. However, with further research and product development, the signal quality of the AES could be enhanced to improve its resistance to artifacts [27], [28]. This has the possibility of simultaneously enhancing the sleep scoring accuracy of the utilized deep learning model.

The transition from type I PSGs to type II in-home measurements has been shown to improve the availability and reduce the waiting times of the sleep studies in diagnostics of obstructive sleep apnea [30]. Furthermore, the direct costs of obstructive sleep apnea management have been shown to decrease because of the absence of hospitalization [30]. In addition to financial benefits, type II PSGs with accurate automatic scoring algorithm could potentially produce better quality sleep data, as the sleep efficiency has been shown to be better and total sleep time longer compared to type I PSGs since the patients can sleep in a familiar and comfortable home environment instead of a sleep laboratory [31], [32]. Self-applicable EEG electrodes with automatic scoring could also enable easy and cost-effective multi-night measurements, further improving the data quality by enabling the consideration of the first-night effect and night-to-night variability [2], [33], [34]. Therefore, type I PSG could be replaced in some cases with self-administrated type II PSG with AES to free up human resources, given due consideration on the needs of the patient. In addition to facilitating the type II PSGs, the AES could be useful when conducting type I PSG recordings in cumbersome situations, such as for patients with reduced mobility in stroke units where routine screening for sleep apnea is uncommon [35]. Furthermore, the AES together with the proposed automatic sleep staging

model could also be used as a supplement in HSAT to improve its diagnostic accuracy without a significant increase in the costs or manual labor.

Our study also included subjects with possible sleep bruxism. Our previous study confirmed that home PSG with the AES, recording at least 3 nights, can help to better capture the oral behavior in a natural sleeping environment [7]. For such extensive clinical studies to be possible, new wearable technologies and automated analytical methods are needed. The neural network-based sleep staging together with self-applicable AES described in this study could also offer new possibilities for the diagnosis of sleep bruxism. Our study also showed that a very simple measurement solution might be sufficient to perform reliable automatic analysis. This will allow simplification of the current layout of the AES.

In conclusion, the previously developed deep learning model [15] was successfully utilized for automatic sleep scoring based on EEG recorded with an AES. The automatic scoring reached an accuracy comparable to inter-scoring agreement of standard EEG recordings [12]–[14], [18]. Therefore, the automatic scoring algorithm that can be used with a self-applicable EEG electrode set could significantly facilitate the arrangement of type II in-home PSGs.

ACKNOWLEDGMENT

The authors would like to thank Sigríður Sigurðardóttir and Erna Sif Arnardóttir from the Reykjavik University Sleep Institute for their valuable contribution in the data analysis.

REFERENCES

- [1] R. B. Berry, *The AASM Manual for the Scoring of Sleep and Associated Events, Version 2.5*. Darien, IL, USA: American Academy of Sleep Medicine, 2018.
- [2] D. J. Levendowski, L. Ferini-Strambi, C. Gamaldo, M. Cetel, R. Rosenberg, and P. R. Westbrook, "The accuracy, night-to-night variability, and stability of frontopolar sleep electroencephalography biomarkers," *J. Clin. Sleep Med.*, vol. 13, no. 6, pp. 791–803, Jun. 2017.
- [3] M. Younes, M. Soiferman, W. Thompson, and E. Giannouli, "Performance of a new portable wireless sleep monitor," *J. Clin. Sleep Med.*, vol. 13, no. 2, pp. 245–258, Feb. 2017.
- [4] P. J. Arnal, V. Thorey, E. Debellemanniere, M. E. Ballard, A. B. Hernandez, A. Guillot, H. Jourde, M. Harris, M. Guillard, P. Van Beers, M. Chennaoui, and F. Sauvet, "The dream headband compared to polysomnography for electroencephalographic signal acquisition and sleep staging," *Sleep*, vol. 43, no. 11, pp. 1–13, Nov. 2020.
- [5] T. Miettinen, K. Myllymaa, S. Westeren-Punnonen, J. Ahlberg, T. Hukkanen, J. Töyräs, R. Lappalainen, E. Mervaala, K. Sipilä, and S. Myllymaa, "Success rate and technical quality of home polysomnography with self-applicable electrode set in subjects with possible sleep bruxism," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 4, pp. 1124–1132, Jul. 2018.
- [6] S. Myllymaa, A. Muraja-Murro, S. Westeren-Punnonen, T. Hukkanen, R. Lappalainen, E. Mervaala, J. Töyräs, K. Sipilä, and K. Myllymaa, "Assessment of the suitability of using a forehead EEG electrode set and chin EMG electrodes for sleep staging in polysomnography," *J. Sleep Res.*, vol. 25, no. 6, pp. 636–645, Dec. 2016.
- [7] T. Miettinen, K. Myllymaa, T. Hukkanen, J. Töyräs, K. Sipilä, and S. Myllymaa, "Home polysomnography reveals a first-night effect in patients with low sleep bruxism activity," *J. Clin. Sleep Med.*, vol. 14, no. 8, pp. 1377–1386, Aug. 2018.
- [8] S. Biswal, H. Sun, B. Goparaju, M. B. Westover, J. Sun, and M. T. Bianchi, "Expert-level sleep scoring with deep neural networks," *J. Amer. Med. Assoc.*, vol. 25, no. 2, pp. 1643–1650, 2018.
- [9] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 400–410, Mar. 2019.
- [10] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.
- [11] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, Oct. 2019, Art. no. 051001.
- [12] S. Deng, X. Zhang, Y. Zhang, H. Gao, E. I.-C. Chang, Y. Fan, and Y. Xu, "Interrater agreement between American and Chinese sleep centers according to the 2014 AASM standard," *Sleep Breathing*, vol. 23, no. 2, pp. 719–728, Jun. 2019.
- [13] U. J. Magalang, N.-H. Chen, P. A. Cistulli, A. C. Fedson, T. Gíslason, D. Hillman, T. Penzel, R. Tamisier, S. Tufik, G. Phillips, A. I. Pack, and f. the SAGIC Investigators, "Agreement in the scoring of respiratory events and sleep among international sleep centers," *Sleep*, vol. 36, no. 4, pp. 591–596, Apr. 2013.
- [14] X. Zhang, X. Dong, J. W. Kantelhardt, J. Li, L. Zhao, C. Garcia, M. Glos, T. Penzel, and F. Han, "Process and outcome for international reliability in sleep scoring," *Sleep Breathing*, vol. 19, no. 1, pp. 191–195, Mar. 2015.
- [15] H. Korkalainen, J. Aakko, S. Nikkonen, S. Kainulainen, A. Leino, B. Duce, I. O. Afara, S. Myllymaa, J. Töyräs, and T. Leppänen, "Accurate deep learning-based sleep staging in a clinical population with suspected obstructive sleep apnea," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 7, pp. 2073–2081, Jul. 2020.
- [16] F. Lobbezoo, J. Ahlberg, A. G. Glaros, T. Kato, K. Koyano, G. J. Lavigne, R. de Leeuw, D. Manfredini, P. Svensson, and E. Winocur, "Bruxism defined and graded: An international consensus," *J. Oral Rehabil.*, vol. 40, no. 1, pp. 2–4, Jan. 2013.
- [17] 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, May 2019, Art. no. e0216456.
- [18] H. Danker-Hopfe, P. Anderer, J. Zeithofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics, B. Saletu, A. Schmidt, and G. Dorffner, "Interrater reliability for sleep scoring according to the rechtschaffen & Kales and the new AASM standard," *J. Sleep Res.*, vol. 18, no. 1, pp. 74–84, Mar. 2009.
- [19] H. Phan, O. Y. Chen, M. C. Tran, P. Koch, A. Mertins, and M. De Vos, "XSleepNet: Multi-view sequential model for automatic sleep staging," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Mar. 31, 2021, doi: 10.1109/TPAMI.2021.3070057.
- [20] H. Phan, O. Y. Chen, P. Koch, Z. Lu, I. McLoughlin, A. Mertins, and M. De Vos, "Towards more accurate automatic sleep staging via deep transfer learning," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 6, pp. 1787–1798, Jun. 2021.
- [21] M. Fu, Y. Wang, Z. Chen, J. Li, F. Xu, X. Liu, and F. Hou, "Deep learning in automatic sleep staging with a single channel electroencephalography," *Frontiers Physiol.*, vol. 12, pp. 1–11, Mar. 2021.
- [22] W. Qu, Z. Wang, H. Hong, Z. Chi, D. D. Feng, R. Grunstein, and C. Gordon, "A residual based attention model for EEG based sleep staging," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2833–2843, Oct. 2020.
- [23] A.-M. Tautan, A. C. Rossi, R. de Francisco, and B. Ionescu, "Automatic sleep stage detection using a single channel frontal EEG," in *Proc. E-Health Bioeng. Conf. (EHB)*, Nov. 2019, pp. 2019–2022.
- [24] K. B. Mikkelsen, D. B. Villadsen, M. Otto, and P. Kidmose, "Automatic sleep staging using ear-EEG," *Biomed. Eng. OnLine*, vol. 16, no. 1, pp. 1–15, Dec. 2017.
- [25] K. B. Mikkelsen, J. K. Ebajemito, M. A. Bonmati-Carrion, N. Santhi, V. L. Revell, G. Atzori, C. D. Monica, S. Debener, D. Dijk, A. Sterr, and M. Vos, "Machine-learning-derived sleep-wake staging from around-the-ear electroencephalogram outperforms manual scoring and actigraphy," *J. Sleep Res.*, vol. 28, no. 2, p. e12786, Apr. 2019.
- [26] A. A. Casciola, S. K. Carlucci, B. A. Kent, A. M. Punch, M. A. Muszynski, D. Zhou, A. Kazemi, M. S. Mirian, J. Valerio, M. J. McKeown, and H. B. Nygaard, "A deep learning strategy for automatic sleep staging based on two-channel eeg headband data," *Sensors*, vol. 21, no. 10, pp. 1–17, 2021.
- [27] L. Kalevo, T. Miettinen, A. Leino, S. Kainulainen, H. Korkalainen, K. Myllymaa, J. Töyräs, T. Leppänen, T. Laitinen, and S. Myllymaa, "Effect of sweating on electrode-skin contact impedances and artifacts in EEG recordings with various screen-printed Ag/AgCl electrodes," *IEEE Access*, vol. 8, pp. 50934–50943, 2020.

- [28] L. Kalevo, T. Miettinen, A. Leino, S. Kainulainen, K. Myllymaa, J. Toyras, T. Leppanen, and S. Myllymaa, "Improved sweat artifact tolerance of screen-printed EEG electrodes by material selection-comparison of electrochemical properties in artificial sweat," *IEEE Access*, vol. 7, pp. 133237–133247, 2019.
- [29] E. S. Kappenman and S. J. Luck, "The effects of electrode impedance on data quality and statistical significance in ERP recordings," *Psychophysiology*, vol. 23, no. 1, pp. 1–7, Mar. 2010.
- [30] A. Safadi, T. Etzioni, D. Fliss, G. Pillar, and C. Shapira, "The effect of the transition to home monitoring for the diagnosis of OSAS on test availability, waiting time, patients' satisfaction, and outcome in a large health provider system," *Sleep Disorders*, vol. 2014, pp. 1–6, Apr. 2014.
- [31] M. Bruyneel, C. Sanida, G. Art, W. Libert, L. Cuvelier, M. Paesmans, R. Sergysels, and V. Ninane, "Sleep efficiency during sleep studies: Results of a prospective study comparing home-based and in-hospital polysomnography," *J. Sleep Res.*, vol. 20, no. 1, pp. 201–206, Mar. 2011.
- [32] C. Iber, S. Redline, A. M. K. Gilpin, S. F. Quan, L. Zhang, D. J. Gottlieb, D. Rapoport, H. E. Resnick, M. Sanders, and P. Smith, "Polysomnography performed in the unattended home versus the attended laboratory setting—Sleep heart health study methodology," *Sleep*, vol. 27, no. 3, pp. 536–540, May 2004.
- [33] J. Newell, O. Mairesse, P. Verbanck, and D. Neu, "Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples," *Psychiatry Res.*, vol. 200, nos. 2–3, pp. 795–801, 2012.
- [34] O. Le Bon, L. Staner, G. Hoffmann, M. Dramaix, I. S. Sebastian, J. R. Murphy, M. Kentos, I. Pelc, and P. Linkowski, "The first-night effect may last more than one night," *J. Psychiatric Res.*, vol. 35, no. 3, pp. 165–172, May 2001.
- [35] D. L. Brown, X. Jiang, C. Li, E. Case, C. B. Sozener, R. D. Chervin, and L. D. Lisabeth, "Sleep apnea screening is uncommon after stroke," *Sleep Med.*, vol. 59, pp. 90–93, Jul. 2019.



LAURA KALEVO received the M.Sc. (Tech.) degree in electrical engineering from the Tampere University of Technology, Tampere, Finland, in 2007. She is currently pursuing the Ph.D. degree.

She joined the Sleep Technology and Analytics Group, Kuopio University Hospital, and the University of Eastern Finland, Kuopio, Finland, in 2018. From 2007 to 2018, she worked as a Hardware Designer and the Project Manager in engineering companies. She currently works as a Junior Researcher with the Diagnostic Imaging Center, Kuopio University Hospital, and the Department of Applied Physics, University of Eastern Finland. Her current research interests include the development EEG electrodes and the development of an easy-to-use ambulatory electrode set for home polysomnography.



SAMI NIKKONEN received the B.Sc., M.Sc., and Ph.D. degrees in applied physics from the University of Eastern Finland, Kuopio, Finland, in 2016, 2017, and 2020, respectively.

After his Ph.D. degree, he has been working as a Senior Researcher with the University of Eastern Finland and the Kuopio University Hospital. His research interests include sleep, diagnostics of sleep disorders, data analysis, and machine learning applications.



AKSELE LEINO received the M.Sc. (Tech.) degree in electrical engineering from the Tampere University of Technology, Tampere, Finland, in 2017. He is currently pursuing the Ph.D. degree with the Department of Applied Physics, University of Eastern Finland, Kuopio, Finland.

He has been a Junior Researcher with the Diagnostic Imaging Center, Kuopio University Hospital, and the Department of Applied Physics, University of Eastern Finland, since 2017. His

research interests include development of sleep apnea diagnostics with deep learning methods and investigating the relationship between stroke and sleep apnea.



SAMU KAINULAINEN was born in Joensuu, Finland, in September 1992. He received the B.Sc., M.Sc., and Ph.D. degrees in medical physics from the University of Eastern Finland, in 2017, 2018, and 2020, respectively.

He is currently working as a Postdoctoral Researcher with the Sleep Technology and Analytics Research Group and the Project Manager in a sleep laboratory construction project. His research interests include cardiorespiratory signal analysis in the scope of sleep disorders and utilization of multimodal diagnostic information in medical diagnostics. He is a member of international delegate reserve in humanitarian aid for Finnish Red Cross and actively participates in teaching duties at the Department of Applied Physics, University of Eastern Finland.



HENRI KORKALAINEN received the B.Sc., M.Sc., and Ph.D. degrees in medical physics from the University of Eastern Finland, in 2017, 2018, and 2020, respectively.

He joined the Sleep Technology and Analytics Research Group, in 2018. He is currently working as a Postdoctoral Researcher with the Department of Applied Physics, University of Eastern Finland, and the Diagnostic Imaging Center, Kuopio University Hospital. His research interests include the

development of deep learning methods in medical diagnostics, mainly in the field of sleep medicine, and the simplification of current diagnostic recordings with the development of artificial intelligence applied to easily applicable and wearable sensors.

ALEXANDER RYAN received the B.H.Sc. and GDipClinPhysiol degrees from Griffith University, Australia, in 2015 and 2016, respectively. He is currently a Sleep Scientist at a nationally recognized sleep disorders center, and since 2018, he has also been a Research Officer with the Department of Respiratory and Sleep Medicine, Princess Alexandra Hospital, Brisbane, Australia. His current research interests include EEG asynchrony within OSA patients, and novel methods for diagnosis and severity assessment of obstructive sleep apnoea.



BRETT DUCE was born in Ipswich, Australia, in 1972. He received the B.Sc. degree in physiology and pharmacology from The University of Queensland, in 1996. He is currently pursuing the Ph.D. degree with the Institute for Biomedical and Health Innovation, Queensland University of Technology, Brisbane, Australia.

Currently, he is also the Scientific Director of the Sleep Disorders Centre, Princess Alexandra Hospital, Brisbane. His research interests include snoring, sleep apnea, sleep diagnostics, and non-invasive ventilation for chronic respiratory failure.



KIRSI SIPILÄ was born in Puolanka, Finland, in 1962. She received the D.D.S. and Ph.D. degrees in dentistry from the University of Oulu, in 1986 and 2002, respectively.

She is currently a Professor of prosthetic dentistry and stomatognathic physiology with the University of Oulu. Her research interests include facial pain and temporomandibular disorders—their epidemiology, diagnostics, and treatment.

JARI AHLBERG was born in 1961. He received the D.D.S. and Ph.D. degrees from the University of Helsinki, Finland, in 1988 and 1997, respectively, and the specialist degree in clinical dentistry. In 2004, he became an Associate Professor at the Faculty of Medicine, University of Oulu. He was the Chief Dental Officer at the Finnish Broadcasting Company for ten years. He worked as a Resident Doctor at the Department of Oral and Maxillofacial Diseases, Helsinki University Central Hospital, from 2000 to 2003. After a three-year period as a Senior Lecturer at the former Institute of Dentistry, Helsinki, he served as a Specialist Dentist and a Supervisor for dentists in specialist training at the Unit of Specialized Oral Care at Helsinki and the metropolitan area. He has been an Associate Professor, since 2006, and a Senior Scientist with the Department of Oral and Maxillofacial Diseases, University of Helsinki. Currently, he is the Chief Dentist with the Head and Neck Center, Hospital District of Helsinki and Uusimaa, responsible for undergraduate teaching. He has supervised several research projects regarding occupational stress, orofacial pain, biopsychosocial symptoms, health effects of irregular shift work, sleep disorders, and especially their complex relationships with bruxism. He has supervised six doctoral theses among the field, four on-going. Together with his colleagues, he has developed the Finnish version of the dual axis research diagnostic criteria for temporomandibular disorders. His groups most recent research focuses on genetic and environmental effects on bruxism, sleep disorders and sleep disordered breathing, and cardiovascular events. He has developed international collaboration with world leading centers among this field of investigation. In peer-reviewed journals, he has published 80 articles.



JOHANNA SAHLMAN was born in 1975. She graduated from the Medical School, University of Kuopio, in 2001, and the Ph.D. degree from the University of Eastern Finland, in 2011, with a thesis related to low-grade inflammation and weight loss in sleep apnea.

She completed the training in ear-, nose-, and throat-diseases and rhinology at the University of Eastern Finland, in 2007 and 2020, respectively.

As a young scientist, she attended the Marie Curie Project “The Training in Sleep Research and Sleep Medicine” provided by the European Sleep Research Society. Currently, she works as a Consultant Rhinologist and the Head of the Rhinology Team, Kuopio University Hospital. Her current research interests include clinical rhinology, especially chronic rhinosinusitis, and polyposis. In sleep apnea research, she has provided a link to everyday clinical practice with sleep apnea patients. She is a member of the Sleep Disorders Working Group, Kuopio University Hospital.



TOMI MIITTINEN was born in Vehmersalmi, Finland, in 1987. He received the M.Sc. degree in medical physics from the University of Eastern Finland, Kuopio, Finland, in 2015, and the Ph.D. degree, in 2019, with a thesis related to the suitability of an ambulatory electrode set for the home polysomnography-based assessment of sleep bruxism.

Since 2017, he has been employed by the Cancer Center, Hospital District of Helsinki and Uusimaa, Helsinki, Finland, as a specializing Medical Physicist.

SUSANNA WESTEREN-PUNNONEN received the M.D. degree from the University of Turku, Finland, in 1990, and the specialist degree in clinical neurophysiology from the University of Kuopio, Finland, in 1999.

Currently, she is working as the Head of the Department of Clinical Neurophysiology at the Diagnostic Imaging Center, Kuopio University Hospital. Recently, she has been involved in research projects concerning the relationship between stroke and sleep apnea, the development of novel methods for diagnostics of sleep disorders, and EEG recordings in acute medicine.

ESA MERVAALA received the M.D. and Ph.D. degrees in medicine from the University of Kuopio, in 1985 and 1987, respectively. Currently, he is a Professor of clinical neurophysiology at the School of Medicine, University of Eastern Finland, and the Chief Physician of the Department of Clinical Neurophysiology at the Diagnostic Imaging Center, Kuopio University Hospital. His research interests include among other topics implementation of EEG services into sleep and acute medicine.



JUHA TÖYRÄS was born in Rovaniemi, Finland, in 1975. He received the M.Sc. and Ph.D. degrees in medical physics from the University of Kuopio, in 1999 and 2001, respectively.

He is currently the Director of research and innovation with the Kuopio University Hospital and a Professor of medical physics and engineering with the Department of Applied Physics, University of Eastern Finland, Kuopio, Finland.

He is the Co-Head of the Biophysics of Bone and Cartilage Research Group and the Sleep Technology and Analytics Research Group at the Department of Applied Physics, University of Eastern Finland. His current research interests include development of quantitative arthroscopy and computed tomography methods for diagnostics of osteoarthritis, as well as development of artificial intelligence solutions for diagnostics and severity estimation of obstructive sleep apnea.



SAMI MYLLYMAA was born in Rauma, Finland, in 1975. He received the M.Sc. degree in electrical engineering from the Tampere University of Technology, Finland, in 2001, and the Ph.D. degree in biomedical engineering from the University of Eastern Finland, Kuopio, Finland, in 2010.

He is currently acting as a Senior Inspector at the Medical Devices Unit of the Finnish Medicines Agency Fimea. He is also acting as a Docent and the Co-Head of the Sleep Technology and Analytics Research Group, Department of Applied Physics, University of Eastern Finland, and a Researcher at the Diagnostic Imaging Center, Kuopio University Hospital. His current research interests include the development of novel wearable sensors and measurement techniques for personalized diagnostics and finding ways to improve the prognostics of sleep disorders.



TIMO LEPPÄNEN received the B.Sc., M.Sc., and Ph.D. degrees in applied physics from the University of Eastern Finland, in 2014, 2015, and 2016, respectively.

He started his studies at the University of Eastern Finland, in 2011, after two years of service with the Finnish Defense Forces. Since 2017, he has been a Postdoctoral Researcher and the Co-Head of the Sleep Technology and Analytics Research Group with the Department of Applied

Physics, University of Eastern Finland. He was appointed as a Senior Researcher, in 2018, and received the title of Docent, in 2020. Since March 2021, he has also been affiliated to the School of ITEE, The University of Queensland, as an Adjunct Fellow. His research interests include development of artificial intelligence solutions for diagnostics and severity estimation of sleep apnea, phenotyping, and progression of obstructive sleep apnea; sleep apnea related psychomotor vigilance and daytime sleepiness; and development of wearable sensor for home sleep apnea testing.

Dr. Leppänen is a member of the European Sleep Research Society, the Finnish Sleep Research Society, the World Sleep Society, the IEEE Engineering in Medicine and Biology Society, the Finnish Society for Medical Physics and Medical Engineering, and the Finnish Society of Clinical Neurophysiology.



KATJA MYLLYMAA was born in Haapajärvi, Finland, in 1981. She received the M.Sc. degree in medical physics from the University of Kuopio, Kuopio, Finland, in 2003, and the Ph.D. degree in medical physics from the University of Eastern Finland, Kuopio, in 2010.

Currently, she is working as a Medical Physicist at the Diagnostic Imaging Center, Kuopio University Hospital, and holds a title of Docent in medical physics at the University of Eastern Finland.

Recently, she has been very active in investigating the relationship between stroke and sleep apnea and development of novel methods for diagnostics of sleep disorders.

...