



# Home-Use and Real-Time Sleep-Staging System Based on Eye Masks and Mobile Devices with a Deep Learning Model

Tsung-Hao Hsieh<sup>1,2</sup> · Meng-Hsuan Liu<sup>1,3</sup> · Chin-En Kuo<sup>4</sup> · Yung-Hung Wang<sup>5</sup> · Sheng-Fu Liang<sup>1,3</sup>

Received: 7 November 2020 / Accepted: 17 August 2021 / Published online: 4 September 2021  
© Taiwanese Society of Biomedical Engineering 2021

## Abstract

**Purpose** Sleep is an important human activity. Comfortable sensing and accurate analysis in sleep monitoring is beneficial to many healthcare and medical applications. From 2020, owing to the COVID-19 pandemic that spreads between people when they come into close physical contact with one another, the willingness to go to hospital for receiving care has reduced; care-at-home is the trend in modern healthcare. Therefore, a home-use and real-time sleep-staging system is developed in this paper.

**Methods** We developed and implemented a real-time sleep staging system that integrates a wearable eye mask for high-quality electroencephalogram/electrooculogram measurement and a mobile device with MobileNETV2 deep learning model for sleep-stage identification. In the experiments, 25 all-night recordings were acquired, 17 of which were used for training, and the remaining eight were used for testing.

**Results** The averaged scoring agreements for the wake, light sleep, deep sleep, and rapid eye movement stages were 85.20%, 87.17%, 82.87%, and 89.30%, respectively, for our system compared with the manual scoring of PSG recordings. In addition, the mean absolute errors of four objective sleep measurements, including sleep efficiency, total sleep time, sleep onset time, and wake after sleep onset time were 1.68%, 7.56 min, 5.50 min, and 3.94 min, respectively. No significant differences were observed between the proposed system and manual PSG scoring in terms of the percentage of each stage and the objective sleep measurements.

**Conclusion** These experimental results demonstrate that our system provides high scoring agreements in sleep staging and unbiased sleep measurements owing to the use of EEG and EOG signals and powerful mobile computing based on deep learning networks. These results also suggest that our system is applicable for home-use real-time sleep monitoring.

**Keywords** Automatic sleep-staging method · EEG · EOG · Home use · Eye mask · Real time · Mobile platform · MobileNetV2

---

✉ Sheng-Fu Liang  
sfliang@mail.ncku.edu.tw

Tsung-Hao Hsieh  
z10712038@email.ncku.edu.tw

Meng-Hsuan Liu  
p76071014@email.ncku.edu.tw

Chin-En Kuo  
cekuo@nchu.edu.tw

Yung-Hung Wang  
yunghung@ncu.edu.tw

<sup>1</sup> Dept. of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan

<sup>2</sup> Dept. of Psychology, National Cheng Kung University, Tainan, Taiwan

<sup>3</sup> Institute of Medical Informatics, National Cheng Kung University, Tainan, Taiwan

<sup>4</sup> Dept. of Applied Mathematics, National Chung Hsing University, Taichung, Taiwan

<sup>5</sup> Dept. of Mechanical Engineering, National Central University, Taoyuan, Taiwan

## 1 Introduction

Sleep is an important human activity. Monitoring and recognizing sleep stages has many healthcare and medical applications, such as long-term sleep quality evaluation, sleep environment control, and sleep disorder diagnosis [1, 2]. Comfortable sensing [compared with that afforded by a polysomnogram (PSG)] with automatic sleep staging is required for home-use sleep monitoring. Because of the COVID-19 pandemic, which spreads among people when they come into close physical contact with one another, the willingness to go to a hospital for care has reduced. Care-at-home has been considered a practical solution to prevent infection and reduce the burden on healthcare systems.

A conventional PSG is a clinically approved sleep-monitoring device. A PSG can be understood as a multivariate system that records different biological signals, such as electroencephalograms (EEGs), electrooculograms (EOGs), electromyograms (EMGs), and electrocardiograms (ECGs). The recorded data are divided into 30-s intervals called epochs. Then, one or more experts classify each epoch into one of five stages [N1, N2, N3, rapid eye movement (REM), or wake] by quantitatively and qualitatively examining the signals of the PSG in the time and frequency domains. Because visual sleep scoring by experts is time consuming and subjective, various automatic sleep scoring methods have been developed [3, 4]. However, the excessive number of wired connections in a conventional PSG disturbs a user's sleep, and therefore, automatic sleep-staging methods based on a single channel EEG/EOG have been developed to reduce the number of wires and disturbance to a user [5–13].

For facilitating convenient sensing at home, various approaches without sleep EEGs have been proposed to describe sleep behavior/states, such as body activity [14, 15], ECGs [16–18], voice [19], and breathing [20], that can be measured by a self-applicable device for home use. The characteristics of these methods are usually part of the standard PSG sleep protocol, and owing to a lack of EEG information, physiological details and descriptions of the sleep for diagnosis cannot be provided.

In this study, a real-time sleep-monitoring system that integrates a wearable eye mask [21] for EEG/EOG measurement and a mobile device for stage identification is proposed. The eye mask includes an embedded module that records EEGs and EOGs with high signal quality and calculates the features. A mobile device is used to receive the features through Bluetooth Low Energy (BLE) and analyzes them for sleep-stage identification with MobileNetV2. A novel performance evaluation strategy that considers sleep-staging agreements and sleep measures was proposed for parameter selection. The results of sleep staging and fundamental sleep measures obtained using our system were compared with the manual scorings obtained through PSG data for performance evaluation.

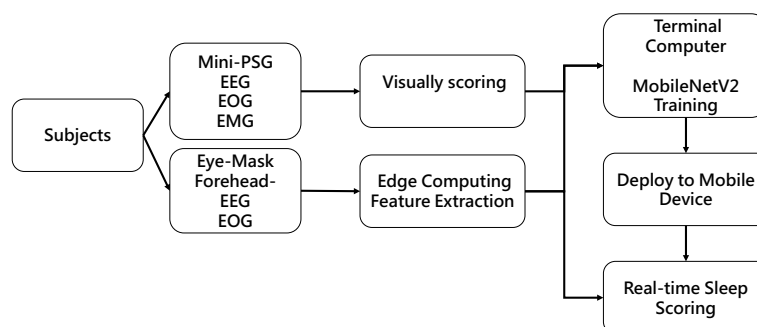
## 2 Methods and Materials

Figure 1 shows the development flowchart of this study. In the experiments, the proposed wearable eye mask and a PSG were mounted on the subjects for simultaneous recording. These data were used for training and validation. During the model training step, signal preprocessing and feature extraction were executed on a personal computer. Then, signal preprocessing and feature extraction were executed on a system embedded in the wearable eye mask. The mobile platform received the features sent from the eye mask through Bluetooth and used the trained stage classification model for real-time sleep scoring.

### 2.1 Hardware Design

To achieve a home-use solution, we proposed a portable wireless physiological measurement module. This module can be used as either a standard PSG measurement device or a wearable device in combination with an eye mask frame for convenient and comfortable sensing. The module employs edge computing for sleep-related feature extraction on the embedded system to achieve real-time sleep interpretation and extend the application of the system to home-care use.

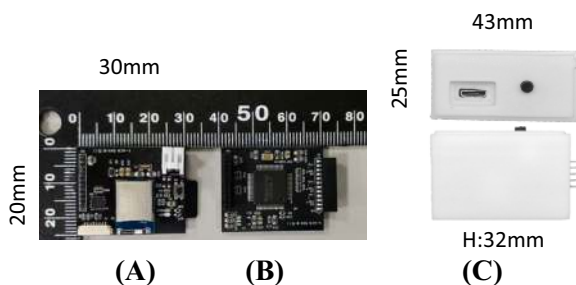
**Fig. 1** System development flowchart



### 2.1.1 Portable Wireless Physiological Measurement Module

The developed sensing module can be used for microvolt-level physiological signal recordings, such as EEG, EOG, EMG, or ECG, in a single-end/differential mode. The data are stored in an SD memory card and/or transferred in real time to the application device through BLE. The module integrates an nRF52840 microchip unit and an ADS1299 analog front-end as shown in Fig. 2. The detailed specifications are listed in Table 1. The main requirements of the module considered for applicability are listed as follows:

- Light weight:** The weight of the module with the battery is 19 ( $\pm 1$ ) g (the sensing module weighs 9 g and the 500 mAh battery weighs 10 g). The light design significantly reduces the interference in sleep quality.
- Continuous 24-h recording:** This proposed module can record and transfer raw sleep data for up to 30 h continuously. In general, a sleep-monitoring device should be able to record at least 9 h continuously for overnight measurement. In addition, operating in the on-line sleep-monitoring mode (with edge computing and wireless feature transmission), the system can record sleep continuously for 9.5 h and perform staging in home-use applications.
- Edge computing:** To achieve real-time sleep scoring in mobile platforms and to reduce the power consumption of the measurement module, signal preprocessing and feature extraction are implemented in the sensing module. This approach prevents frequent BLE transmissions/connections, effectively reduces power consumption, and preserves personal privacy without releasing raw data. The features are regarded as encrypted information, and a matching model (e.g., the classification model) is required to translate the actual situation of the user.



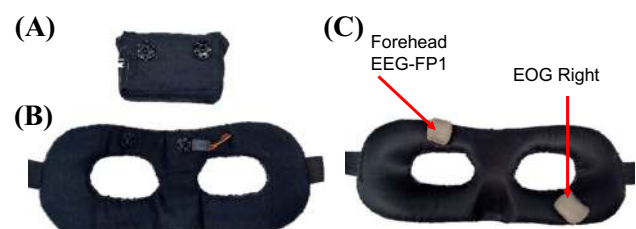
**Fig. 2** Circuits of the designed portable wireless physiological measurement module: **A** Top board circuit with MCU, USB plugin, SD card, and other parts; **B** Bottom board circuit with ADC and signal input pin; **C** Packaging with battery for wearable application

**Table 1** Specifications of the designed portable wireless physiological measurement module

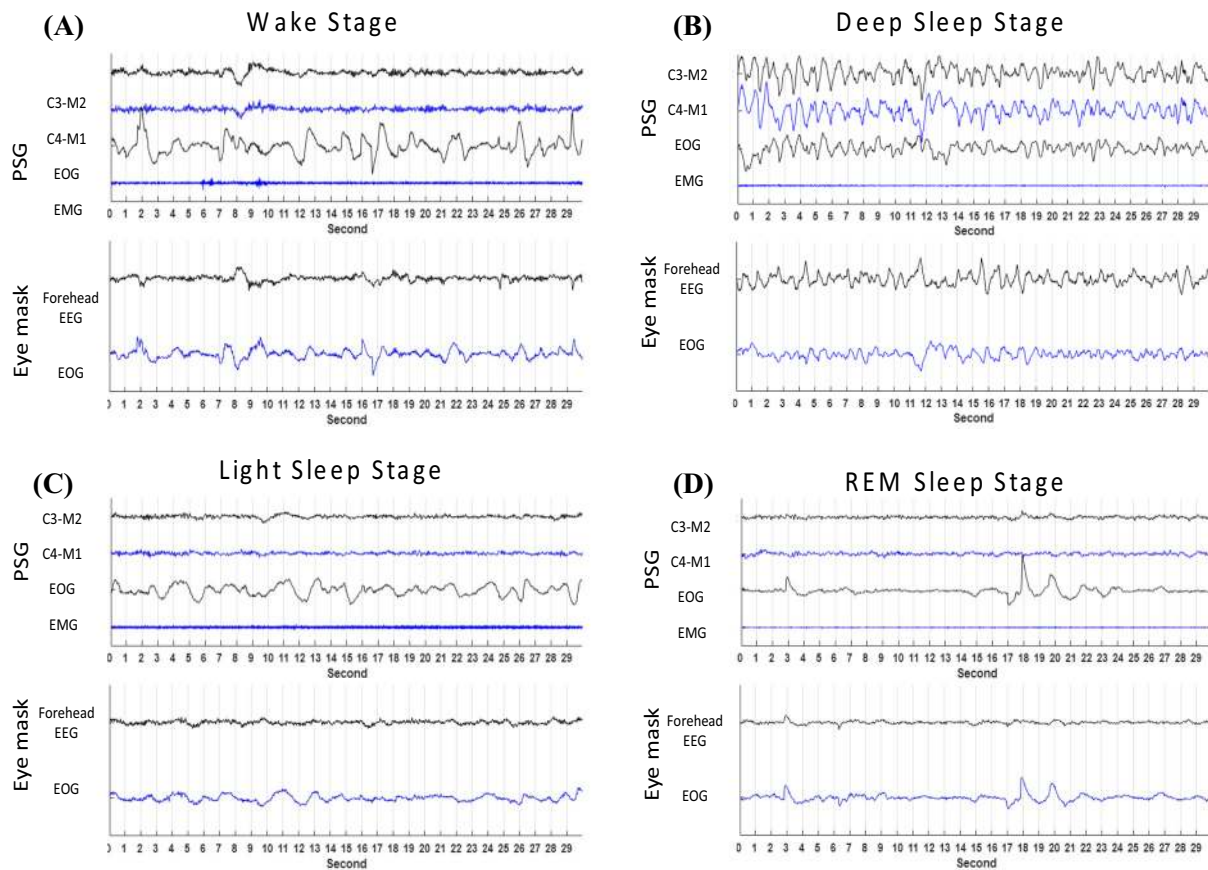
Item	PID	Function
MCU	nRF52840	Frequency: 2.4 GHz RAM: 256 kB Microcontroller: ARM Cortex-M4F Bluetooth: 5.0
ADC	ADS1299	4 channels Resolution: 24 Bits Sample rate: 250 Hz ~ 16 kHz Input range (V): 0–5.25 Gain: 1–24
Accelerometer	ADXL362	3-axis Sensor range: $\pm 8$ g Resolution: 12 Bits
Memory Card		With FAT32 page system
Total power cost		About 16.38 mAh

### 2.1.2 Wearable Eye Mask Design

The core design concept for comfortable sleep sensing is to integrate a sleep eye mask with EEG and EOG electrodes. To reduce pressure on the eyeballs, we use sponges located between the sensor and the eye mask to make the sensor fit the shape of the orbit. Our eye mask design and soft-fabric electrodes protect users from interference due to the electrode wire and allow them to wear the eye mask conveniently, as shown in Fig. 3. This design has been proven to be more comfortable than a traditional PSG [21]. Two improvements were made in this study. First, we selected softer materials to make the mask lighter (from 63 to  $>42$  g). Second, instead of a single bipolar EOG signal recorded to calculate features, the forehead EEG and EOG-R singles were recorded in single-end mode to obtain EEG-related and EOG-related characteristics during sleep. The difference between the signals of these two channels was then calculated to obtain the EOG signal.



**Fig. 3** **A** Packaged sensing module, which is connected to the eye mask with metal snap buttons; **B** Outside and **C** inside of the eye mask, which can measure the forehead EEG and EOG-R signals; The entire eye mask wearable device weighs 74 g ( $\pm 1$ )



**Fig. 4** Sleep recordings corresponding to different sleep stages obtained simultaneously by a PSG and the eye mask

Figure 4 shows the sleep measurement results obtained by using our sensing module (Fig. 2) to function as a mini-PSG (to record 2 EEG channels, 1 EOG channel, and 1 EMG channel) and the eye mask (Fig. 3) to perform simultaneous recordings. In the wake stage (Fig. 4A), the alpha rhythm (8–12 Hz) is observed in the EEG channels of both devices. The alpha rhythm is also observed in the EOG signal of the eye mask. In the light sleep stage (N1 stage, Fig. 4B), the theta rhythm (4–7 Hz) is the major component in the EEG channels of both devices. In the deep sleep stage, characteristics of high-amplitude, slow-wave activity (1–3 Hz) are observed in all channels of the two devices. In the REM stage, the EOG signals indicate REM with large amplitude. These recordings indicate that the eye mask can fix the dry electrodes to capture the EEG and EOG characteristics during different sleep stages. However, because the forehead EEG was recorded instead of typical C3-M2/C4-M1 channels, the EEG signal of the eye mask was not identical to the EEG signal of the PSG; therefore, a specific automatic sleep scoring model is required for the eye mask to perform home-use sleep monitoring.

## 2.2 Subjects and Recording

These measurements were approved by the internal review board of National Cheng Kung University. A total of 25 overnight PSG and wearable eye mask sleep recordings were obtained simultaneously from 25 subjects (12 men and 13 women, aged  $23.2 \pm 1.8$  years, college students of NCKU) in this study. All subjects had no prior history of smoking or drug or alcohol abuse or neurological, psychiatric, or sleep disorders. No outside interference was observed during data collection, and no medications were used to induce sleep. To maintain the subjects' sleeping habits and body clocks as much as possible, we provided an independent and user-controlled sleeping environment (temperature: 25–28 °C, light, and airflow), and the scheduled bedtime was at their leisure. Subjects arrived at the laboratory at approximately 10:00 PM and were given instructions for the experiment. Before execution the experiments, they did not take any tea, coffee, alcohol, etc.... Then, two sleep monitors (a mini-PSG and the eye mask) were set up, which took 30–40 min. The mini-PSG recordings (utilizing the sensing module in Fig. 2) included two EEG (C3-M2 and C4-M1) channels, one EOG (EOG\_R—EOG\_L) channel, and a chin EMG

channel. The eye mask recordings contained a forehead EEG (FP1) and an EOG (EOG\_R) channel, as shown in Fig. 3. The sampling rate of these two devices was 250 Hz, and the data were transmitted to the terminal computer in real time and were saved to the on-board SD card. All 25 PSG sleep recordings (EEG, EOG and EMG) were visually scored by a sleep specialist using the AASM guidelines in 30-s intervals (named an epoch) as the gold standard to develop the automatic scoring model for the recordings obtained from the eye mask. To efficiently and effectively construct and evaluate our method, we sorted the recordings based on the sleep efficiency (SE) obtained from the manual scoring of the mini-PSG. The recordings of every third subject from the sorted list of the recordings based on SE were used for verification (8 subjects), and the remaining data from 17 subjects were used for model construction.

Table 2 lists the statistics of sleep measures obtained from the 25 subjects. From the 24,029 epochs, the average percentages of the wake, light sleep, deep sleep, and REM stages were 10.78%, 53.6%, 17.38%, and 18.23%, respectively. The SE ranged between 66 and 96%, and the SE of two subjects was < 80% (72.42% and 66.7%, respectively).

The average total sleep time was > 7 h, ranging 325–524 min, and two subjects slept < 7 h (379 and 325 min, the same subjects with poor SE).

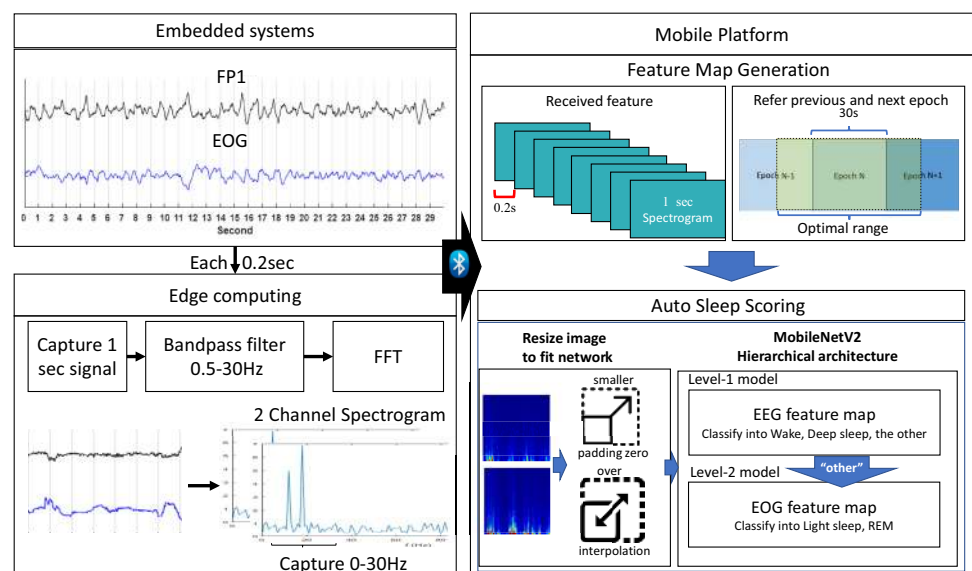
### 2.3 Edge Computing for Feature Extraction

To effectively reduce power consumption as well as preserve personal privacy, the embedded physiological sensing module in the eye mask provides a signal recording mode and a sleep scoring mode. In the scoring mode, an ARM processor in the embedded sensing module performs signal processing and feature extraction to reduce the data size of the BLE package and avoid the leaking of private information. Figure 5 shows the flowchart of the real-time sleep analysis procedure, wherein edge computing is employed by the embedded physiological sensing module in the eye mask to extract features and mobile computing is employed in a mobile device for sleep-stage identification. The embedded sensing module in the eye mask records the forehead EEG and EOG signals at a sampling rate of 250 Hz. Before feature extraction, the EEG and EOG raw data were filtered with a cutoff frequency of 0.5–30 Hz by using a 30-order

**Table 2** Statistics of sleep measures obtained from the 25 subjects

Sleep stage	Wake	Light	Deep	REM
Avg(%)	10.78	53.60	17.38	18.23
S.D	7.26	8.16	3.89	3.88
Avg(mins)	51.3	255.1	82.7	86.7
Sleep parameters	SE (%)	TST (min)	SOT (min)	WASOT (min)
Avg	89.22	429.0	29.6	17.3
S.D	7.26	61.4	19.7	22.2

**Fig. 5** Flowchart of the real-time sleep analysis procedure, wherein edge computing is employed by the embedded physiological sensing module in the eye mask to extract features and mobile computing is employed in a mobile device with MobileNETV2 for sleep-stage identification





FIR low-pass filter and a 125-order FIR high-pass filter to remove artifacts and maintain the data characteristics for sleep analysis. Every 0.2 s of data and the previous 0.8 s of data were combined as 1 s of data to perform a 256-point fast Fourier transform calculation. The resultant spectrograms of the EEG and EOG signals were sent from the eye mask through BLE to the mobile platform to generate the feature map and identify sleep stages by using MobileNetV2.

## 2.4 Mobile Computing for Feature Map Generation and Sleep Staging

For home-use healthcare applications, mobile devices, such as smart phones and pads, provide acceptable computing, communication, display, and management functions. Taking advantage of the advances in powerful deep learning networks, we used MobileNetV2, a convolutional neural network (CNN)-based next-generation portable computer vision network developed by Google Inc., to identify the sleep stages based on the features calculated and transmitted from the eye mask. The network program is small enough to be deployed on a mobile device. This API contained a new architecture called “linear bottlenecks” and simplifies the layer connection through a shortcut. The shortcut helps improve the encoding of the model’s intermediate inputs and outputs, whereas the inner layer allows the model to transform from lower-level concepts, such as pixels, to higher-level descriptors, such as image categories [22].

The mobile-based sleep scoring method proposed in this study generated two MobileNetV2 models to compose a hierarchical classification process. The level-1 model first classifies the epoch into three classes—wake, deep sleep stage (N3), and other stages (N1, N2, and REM)—based on their EEG features. Next, the other stages are further classified as light sleep (N1 and N2) and REM through the level-2 model, based on their EOG features. Therefore, any unknown epoch is identified as one of the four stages: wake, light sleep, deep sleep, and REM.

Because MobileNetV2 is a pretrained deep learning network by using the ImageNet dataset, we generated feature maps (images) based on the spectrograms of the EEG and EOG signals to fit the input structure of MobileNetV2. The completeness and coverage of essential information in the feature map and the suitable objective to achieve for the model are two major factors that influence the performance of model tuning and optimization. For sleep scoring, the following are considered.

### 2.4.1 Temporal Context for Feature Maps

For manual sleep scoring, in addition to the signal characteristics of the current epoch to be identified, the characteristics of nearby epochs are observed by experts to consider the

temporal context of sleep cycles. Therefore, in this study, the input feature maps were not limited to the spectrogram of the current epoch. Some portion of the previous and subsequent epochs were also included. The received spectrograms 15 s before and after the current 30-s epoch were included in the current 30-s spectrogram, to obtain a 60-s spectrogram as the feature maps (EEG and EOG). Moreover, the regular input image size of MobileNetV2 is  $224 \times 224$  pixels, so the feature maps were resized into  $224 \times 224$  pixels before they were fed into the MobileNetV2 network. In the experiments, the performances of including 20 s of previous and subsequent data in the current 30-s spectrogram were also compared. Similarly, the feature maps were resized to  $224 \times 224$  pixels. If the generated feature map was smaller than  $224 \times 224$  pixels, the empty portions were designated as zero. If the generated feature maps were larger than  $224 \times 224$ , the maps were resampled.

### 2.4.2 Objective Selection for the Classifier

Most of the automatic sleep scoring method compares the epoch-by-epoch agreements between the model outputs and manual scorings to adjust parameters or select models. However, for sleep diagnosis, beyond sleep staging, sleep indices calculated from the hypnograms are also important for evaluating a subject’s objective sleep quality. Considering or evaluating only the epoch-based agreement may cause overfitting and not ensure that accurate diagnostic sleep measurements are reported. Therefore, in addition to the average agreement of epoch staging, objective sleep measurements such as SE, sleep latency, wake after sleep onset time (WASOT), and total sleep time (TST) in model optimization were proposed and evaluated in this paper.

## 3 Results

In total, 25 overnight PSG and wearable eye mask sleep recordings were obtained simultaneously from 25 subjects (12 men and 13 women, aged  $23.2 \pm 1.8$  years). The SE ranged between 66 and 96%, and the recordings were sorted based on the SE. From the sorted list, every third recording from the subjects were used for verification (8 subjects), and the remaining data from 17 subjects were used to establish and fine-tune the MobileNetV2 model. The spectrogram 15 s before and after the current 30-s epoch was included with the current 30-s spectrogram to obtain a 60-s spectrogram as the feature image (resized to  $224 \times 224$  pixels) to be fed into MobileNetV2. The length of neighbor data to be included was determined by averaging the ranks of agreements corresponding to the wake, light sleep, deep sleep, and REM sleep stages, as well as the accuracies corresponding to objective sleep

measurements (SE), sleep latency, TST, sleep onset time (SOT), and WASOT). The feature map with the 60-s spectrogram (including 15 s before and after the current 30-s epoch) helps achieve a good balance between sleep staging and accuracy in sleep measurements.

The confusion matrix of the four-stage epoch classification obtained using the data from eight test subjects and sensed by the eye mask and the classifier on the smart phone (Google Pixel 3a) is presented in Table 3. For the sleep stages, the overall agreement between the scores provided by the expert and those obtained by the proposed system was 86.72% ( $\pm 1.85\%$ ); all interscorer agreements were higher than 82% [3]. The sensitivities of the wake, light sleep, deep sleep, and REM stages were 85.20%, 87.17%, 82.87%, and 89.30%, respectively. The agreements between the predicted scores and the scores provided by experts for all stages were good and exceeded 80%. In addition, the mean average error absolute error was also calculated to measure the agreements between the results of the expert and the proposed method with respect to various sleep measurements. The mean absolute errors (MAEs) with respect to four objective sleep measures—SE error, TST error, SOT error, and WASOT error—were 1.68% ( $\pm 2.59\%$ ), 7.56 ( $\pm 7.88$ ) min, 5.50 ( $\pm 9.07$ ) min, and 3.94 ( $\pm 2.35$ ) min, respectively.

For healthy adults, a deep sleep stage constitutes approximately 20% of the TST and the REM stage constitutes approximately 25% of the TST. Therefore, the report addresses percentage of TST that each sleep stage occupies is essential for sleep diagnosis. Figure 6 shows the comparisons of the subject-by-subject percentage of the TST each sleep stage occupies, as estimated by our system and manual PSG scoring. The MAE between the results estimated by our system and those estimated through manual PSG scoring corresponding to the four sleep stages—wake, light sleep, deep sleep, and REM—were 1.68% ( $\pm 1.92\%$ ),

2.94% ( $\pm 1.66\%$ ), 1.99% ( $\pm 1.54\%$ ), and 2.10% ( $\pm 2.10\%$ ), respectively.

In the statistics analysis, no significant differences were observed between the proposed system and manual PSG scoring in terms of the percentage of each stage and the objective sleep measurements ( $p$ -values of the wake, light sleep, deep sleep, REM, SE, TST, SOT, and WASOT were 0.9, 0.8, 0.33, 0.26, 0.9, 0.96, 0.7, and 0.32, respectively). Figure 7 shows the hypnograms of two test subjects, including the PSG manual scoring results and the results of our system. These experimental results demonstrate the applicability of our system for home-use sleep monitoring.

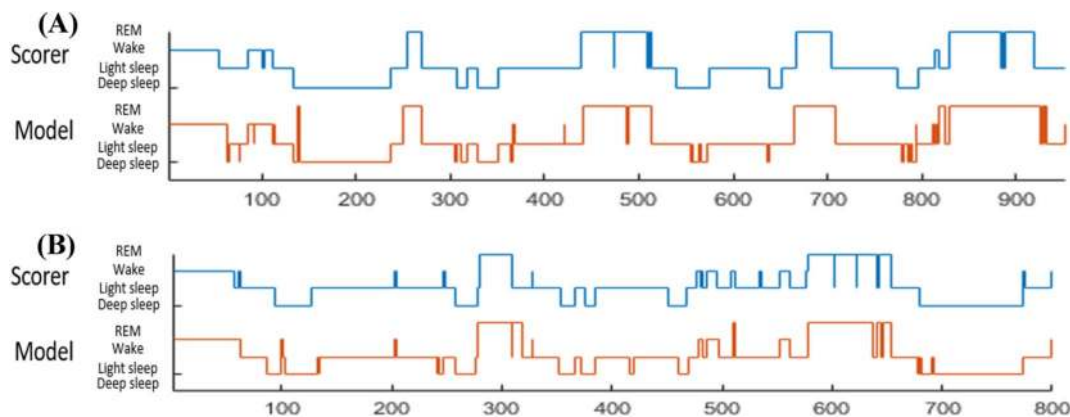
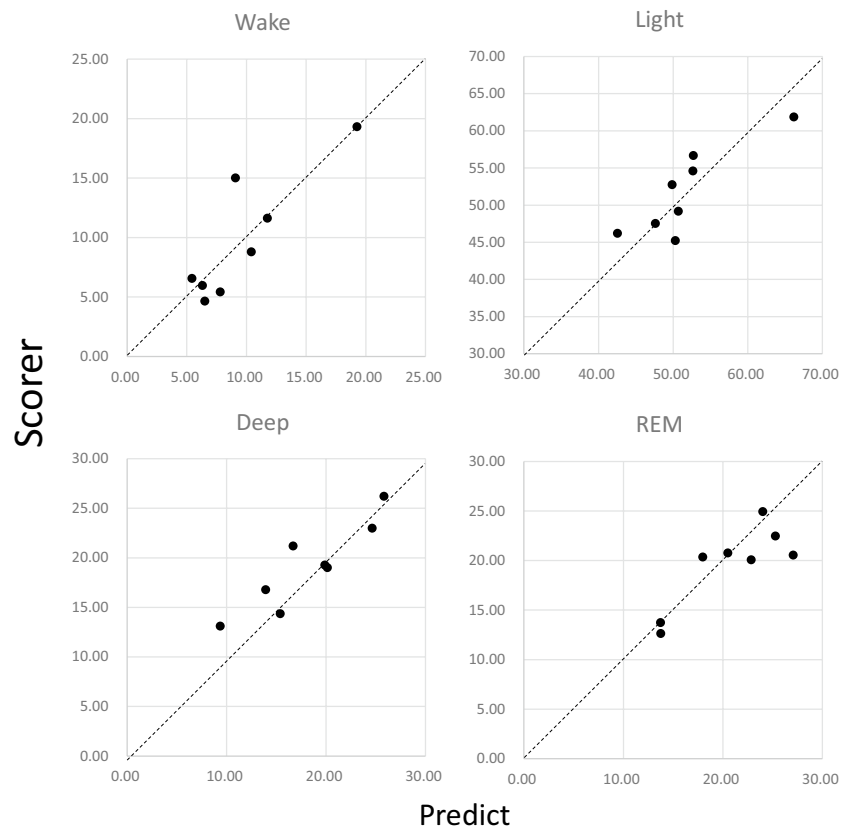
## 4 Discussion and Conclusion

From 2020, owing to the COVID-19 pandemic that spreads between people when they come into close physical contact with one another, the willingness to go to hospital for receiving care has reduced; care-at-home is the trend in modern healthcare. In this study, a home-use and real-time sleep-monitoring system that integrates a comfortable eye mask and a mobile device was developed. The wearable eye mask [21] obtains high-quality EEG and EOG signals, uses edge computing for essential feature calculation, and facilitates real-time data transmission through BLE. A mobile device was used to receive the calculated features, generate the feature maps, and analyze the feature maps with Mobile-NETV2 for sleep-stage identification. The averaged scoring agreements between our proposed system and the manual scoring of PSG recordings for the wake, light sleep, deep sleep, and REM stages were 85.20%, 87.17%, 82.87%, and 89.30%, respectively. In addition, the MAEs with respect to the objective sleep measurements—SE, TST, SOT, and WASOT—were 1.68%, 7.56 min, 5.50 min, and 3.94 min,

**Table 3** Confusion matrices between the mobile scoring method and the visual scorings obtained with eight test subjects with respect to sleep stages and sleep measurements

Sleep stage	Predict				
	Wake	Light	Deep	REM	ACC
Scorer					
Wake	85.20% (616)	13.96% (99)	0.28% (2)	0.83% (6)	85.20%
Light	2.49% (97)	87.17% (3389)	4.22% (164)	6.12% (238)	87.17%
Deep	0.35% (5)	16.70% (236)	82.27% (1171)	0.07% (1)	82.87%
REM	0.55% (8)	10.15% (148)	0.00% (0)	89.30% (1302)	89.30%
ACC					86.72%
Error of sleep measurement	SE (%)	TST (min)	SOT (min)	WASOT (min)	
Errors					
Average	2.10	1.68	7.56	5.50	
STD	2.10	2.59	7.88	9.07	
p-value	0.26	0.90	0.96	0.70	

**Fig. 6** Comparisons of subject-by-subject percentage of TST each sleep stage occupies, as estimated by the proposed system and manual PSG scoring



**Fig. 7** Hypnograms of two test subjects, including those of the mini-PSG manual scoring results and results of our system

respectively. No significant differences were observed between the proposed system and manual PSG scoring in terms of the percentage of each stage and the objective sleep measurements. The experimental results demonstrate the applicability of the proposed home-use and real-time sleep-monitoring system.

Various automatic sleep-staging methods based on a single EEG channel have been developed for home-used sleep monitoring. These methods were developed with the objective of using less recording wires to reduce sleep

disturbance [7, 9–13, 23]. However, most of these methods use the EEG signals from PSG recordings instead of home-based wearable devices, and although commercial wearable devices, such as wristbands [15, 24], are easy to use, they may not provide accurate sleep hypnograms and unbiased objective sleep measurements [25]. Our system provides high scoring agreements in sleep staging and unbiased sleep measurements owing to the use of EEG and EOG signals and powerful mobile computing based on deep learning networks.



To enhance the performance of sleep staging, the temporal context is considered for feature map generation. As [23, 26] suggested, a trade-off exists between performance and length. As reported, the improvement in performance was marginal, and an overly long context extension may affect the detection of some sleep stages [11, 12, 23, 26]. We investigated the benefits of context extension from 0 to 20 s in 5-s steps. We found that a context extension of 5 s improved the performance of stage scoring, and the improvement tended to be smooth when using 15-s or 20-s context extensions. The experimental result showed that the feature map with the 60-s spectrogram (including 15 s before and after the current 30-s epoch) aids in achieving a good balance between sleep staging and accuracy in sleep measurements. Our proposed system also helps overcome the limitations of computing capability, communication speed, and power consumption for the mobile platform.

The major limitation of this study was the young average age (under 30 years old) of the respondents. Elders need to be included and analyzed in future research. Older people find it more difficult to sleep for as long as possible and tend to wake up several times throughout the night. In addition, these insomnia symptoms of elders are more likely to be caused by sleep disordered breathing, such as sleep apnea syndrome. Although the current dataset includes subjects with low sleep quality, most of symptoms are difficulty in falling asleep. Our future studies should include sleep disordered breathing-related insomnia symptoms in the dataset.

On-line sleep monitoring is required for applications such as memory consolidation [27], efficient and effective napping [28, 29], and sleep environment enhancement [30]. These applications usually provide stimuli or change conditions at specific sleep stages or when specific events occur. The proposed eye mask integrated with a smart phone meets such requirements; moreover, it is portable and can be used by someone on their own, which is critical for day-to-day living. In the future, more applications or advanced sleep research can be developed by using our system.

**Acknowledgements** This work was supported by the Ministry of Science and Technology of Taiwan under Grant MOST 108-2634-F-006-012, 108-2321-B-006-024-MY2, 109-2634-F-006-013, 109-2221-E-035-066 -MY3, 109-2627-H-006-005-02 and carried out with AI Biomedical Research Center (AIBMRC) at National Cheng Kung University.

## References

- Ahmad, S., Batkin, I., Kelly, O., Dajani, H. R., Bolic, M., & Groza, V. (2013). Multiparameter physiological analysis in obstructive sleep apnea simulated with Mueller maneuver. *IEEE Transactions on Instrumentation and Measurement*, 62(10), 2751–2762.
- Behar, J., Roebuck, A., Domingos, J. S., Geder, E., & Clifford, G. D. (2013). A review of current sleep screening applications for smartphones. *Physiological Measurement*, 34(7), R29.
- Rosenberg, R. S., & Van Hout, S. (2013). The American Academy of Sleep Medicine inter-scorer reliability program: Sleep stage scoring. *Journal of Clinical Sleep Medicine*, 9(1), 81–87.
- Liang, S.-F., Kuo, C.-E., Hu, Y.-H., & Cheng, Y.-S. (2012). A rule-based automatic sleep staging method. *Journal of Neuroscience Methods*, 205(1), 169–176.
- Shambroom, J. R., Fábregas, S. E., & Johnstone, J. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research*, 21(2), 221–230.
- Hsu, Y.-L., Yang, Y.-T., Wang, J.-S., & Hsu, C.-Y. (2013). Automatic sleep stage recurrent neural classifier using energy features of EEG signals. *Neurocomputing*, 104, 105–114.
- Liang, S.-F., Kuo, C.-E., Hu, Y.-H., Pan, Y.-H., & Wang, Y.-H. (2012). Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models. *IEEE Transactions on Instrumentation and Measurement*, 61(6), 1649–1657.
- Hassan, A. R., & Bhuiyan, M. I. H. (2016). A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. *Journal of Neuroscience Methods*, 271, 107–118.
- Berthomier, C., et al. (2007). Automatic analysis of single-channel sleep EEG: Validation in healthy individuals. *Sleep*, 30(11), 1587–1595.
- Koley, B., & Dey, D. (2012). An ensemble system for automatic sleep stage classification using single channel EEG signal. *Computers in Biology and Medicine*, 42(12), 1186–1195.
- Tsinalis, O., Matthews, P. M., Guo, Y., & Zafeiriou, S. (2016). Automatic sleep stage scoring with single-channel EEG using convolutional neural networks. <https://arxiv.org/abs/1610.01683>
- Supratak, A., Dong, H., Wu, C., & Guo, Y. (2017). DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11), 1998–2008.
- Bresch, E., Großekathöfer, U., & Garcia-Molina, G. (2018). Recurrent deep neural networks for real-time sleep stage classification from single channel EEG. *Frontiers in Computational Neuroscience*, 12, 85.
- Sazonov, E., Sazonova, N., Schuckers, S., Neuman, M., & CS Group. (2004). Activity-based sleep-wake identification in infants. *Physiological Measurement*, 25(5), 1291.
- Kuo, C.-E., Liu, Y.-C., Chang, D.-W., Young, C.-P., Shaw, F.-Z., & Liang, S.-F. (2016). Development and evaluation of a wearable device for sleep quality assessment. *IEEE Transactions on Biomedical Engineering*, 64(7), 1547–1557.
- Camm, A. J. et al. (1996) Heart rate variability: Standards of measurement, physiological interpretation and clinical use. In: *Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology*.
- Mack, D., Kell, S., Alwan, M., Turner, B. & Felder, R. (2003). Non-invasive analysis of physiological signals (naps): A vibration sensor that passively detects heart and respiration rates as part of a sensor suite for medical monitoring. In *Summer Bioengineering Conference*, pp. 25–29.
- Mack, D. C., Patrie, J. T., Suratt, P. M., Felder, R. A., & Alwan, M. (2008). Development and preliminary validation of heart rate and breathing rate detection using a passive, ballistocardiography-based sleep monitoring system. *IEEE Transactions on Information Technology in Biomedicine*, 13(1), 111–120.
- Al-Mardini, M., Aloul, F., Sagahyoon, A., & Al-Husseini, L. (2014). Classifying obstructive sleep apnea using smartphones. *Journal of Biomedical Informatics*, 52, 251–259.
- Pereira, C. B., Yu, X., Czaplak, M., Rossaint, R., Blazek, V., & Leonhardt, S. (2015). Remote monitoring of breathing dynamics

- using infrared thermography. *Biomedical Optics Express*, 6(11), 4378–4394.
21. Liang, S.-F., et al. (2015). Development of an EOG-based automatic sleep-monitoring eye mask. *IEEE Transactions on Instrumentation and Measurement*, 64(11), 2977–2985.
  22. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L. -C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520.
  23. Sors, A., Bonnet, S., Mirek, S., Vercueil, L., & Payen, J.-F. (2018). A convolutional neural network for sleep stage scoring from raw single-channel EEG. *Biomedical Signal Processing and Control*, 42, 107–114.
  24. Diaz, K. M., et al. (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International Journal of Cardiology*, 185, 138–140.
  25. Haghayegh, S., Khoshnevis, S., Smolensky, M. H., Diller, K. R., & Castriotta, R. J. (2019). Accuracy of wristband Fitbit models in assessing sleep: Systematic review and meta-analysis. *Journal of Medical Internet Research*, 21(11), e16273.
  26. Chambon, S., Galtier, M. N., Arnal, P. J., Wainrib, G., & Gramfort, A. (2018). A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(4), 758–769.
  27. Leminen, M. M., et al. (2017). Enhanced memory consolidation via automatic sound stimulation during non-REM sleep. *Sleep*, 40(3), zsx003.
  28. Hilditch, C. J., Dorrian, J., & Banks, S. (2017). A review of short naps and sleep inertia: Do naps of 30 min or less really avoid sleep inertia and slow-wave sleep? *Sleep Medicine*, 32, 176–190.
  29. Liang, S.-F., Shih, Y.-H., Hu, Y.-H., & Kuo, C.-E. (2020). A method for napping time recommendation using electrical brain activity. *IEEE Transactions on Cognitive and Developmental Systems*, 12, 645.
  30. Hsieh, T. -H., Liu, Y. S. Liang, S. -F. Yasutake, M. & Ishibashi, A. (2019). The tent-type clean unit system platform for air cleaning and non-contact sleep assessment. In *Proceedings of the 2019 3rd International Conference on Computational Biology and Bioinformatics*, pp. 47–51.