Standardized Image-Based Polysomnography Database and Deep Learning Algorithm for Sleep Stage Classification

Jaemin Jeong[1o], Wonhyuck Yoon[2o], Jeong-Gun Lee[1], Dongyoung Kim[2], Yunhee Woo[2], Dong-Kyu Kim[2,3*], Hyun-Woo Shin[2,4*]

Chuncheon[1,3], Seoul[2,4], Repubilc of Korea

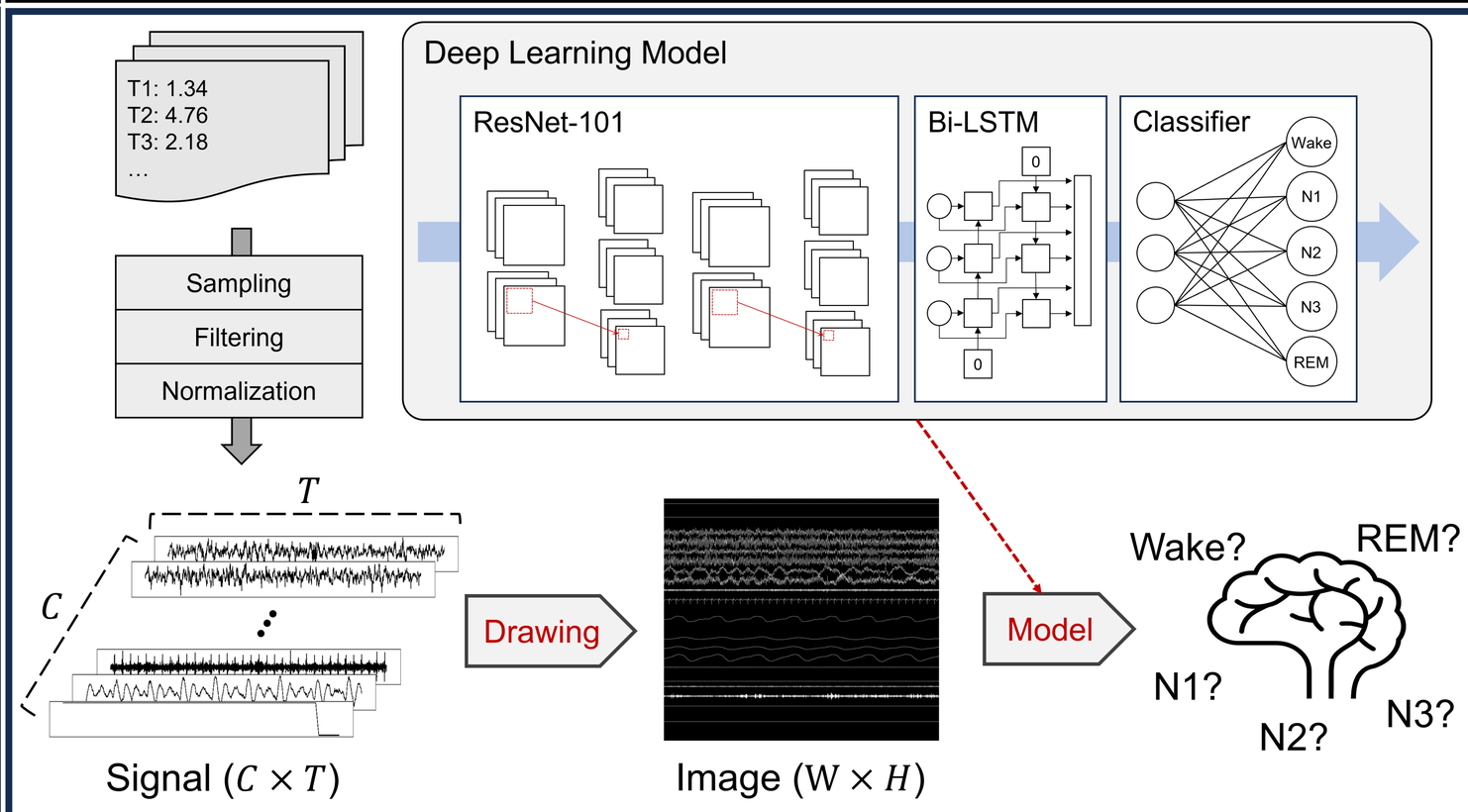
[1]Hallym University, [2]OUaR LaB, Inc, [3]Hallym University Hospital

[4]Seoul National University College of Medicine and Seoul National University Hospital

Introduction

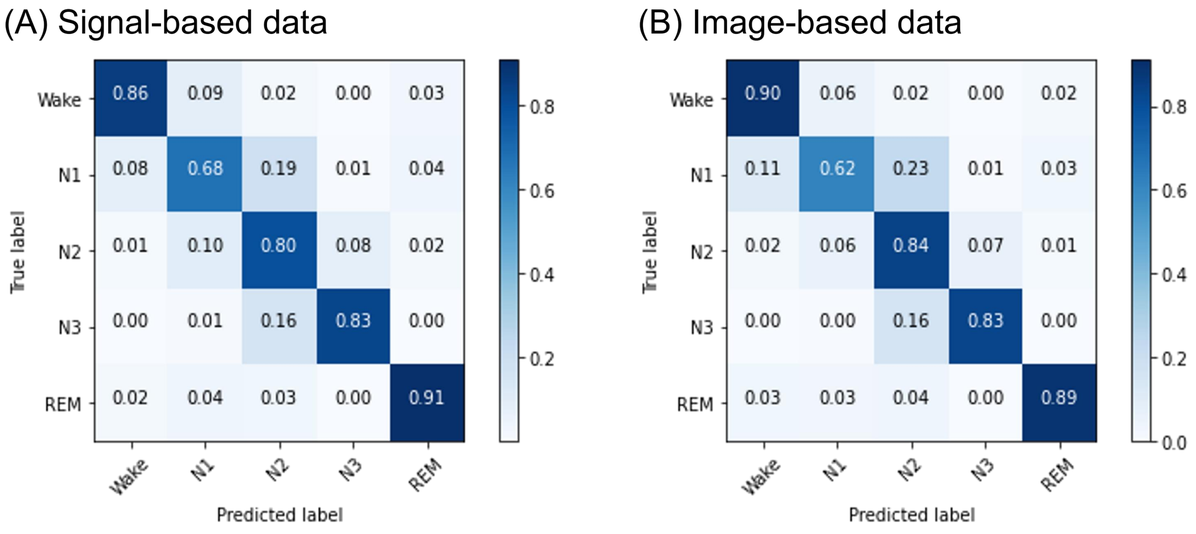
- Polysomnography (PSG) scoring is time-consuming, subjective, and often ambiguous. Automated sleep scoring using deep learning has been widely investigated using public and nonpublic PSG datasets consisting of signal-type data.
- However, clinically effective methods do not yet exist due to performance limitations seen during external validation. In this study, we tried to construct a standardized image-based PSG dataset because constructing PSG data with standardized image files can unify the heterogeneity of raw signal data obtained from various PSG devices and various sleep laboratory environments in each hospital.
- All individually exported European data format files containing raw signals were converted into images with an annotation file in all PSG datasets. Each annotation file contained the demographics, diagnoses, and sleep statistics. We constructed 10,253 image-based PSG datasets using a standardized format. Among these, 7,745 patient data were used to develop our model. We found that the model using the image dataset performed similarly to the signal-based dataset for the same subject regardless of adjusting the input sampling rate or number of channels. In addition, the overall accuracy was greater than 80%, even with severe obstructive sleep apnea.
- Moreover, for the first time, we show using an explainable artificial intelligence method in sleep medicine as visualized key inference regions using class activation maps.

Method



- Framework for automatic sleep-stage classification: After performing preprocessing such as sampling, filtering, and normalization on the input raw signal dataset, a standardized image dataset was created.
- Next, to test the five-class sleep-stage scoring, the standardized image dataset was added to the deep learning model, which combines the bidirectional long short-term memory networks (Bi-LSTM) layer and fully connected layer with the convolutional neural network layer.

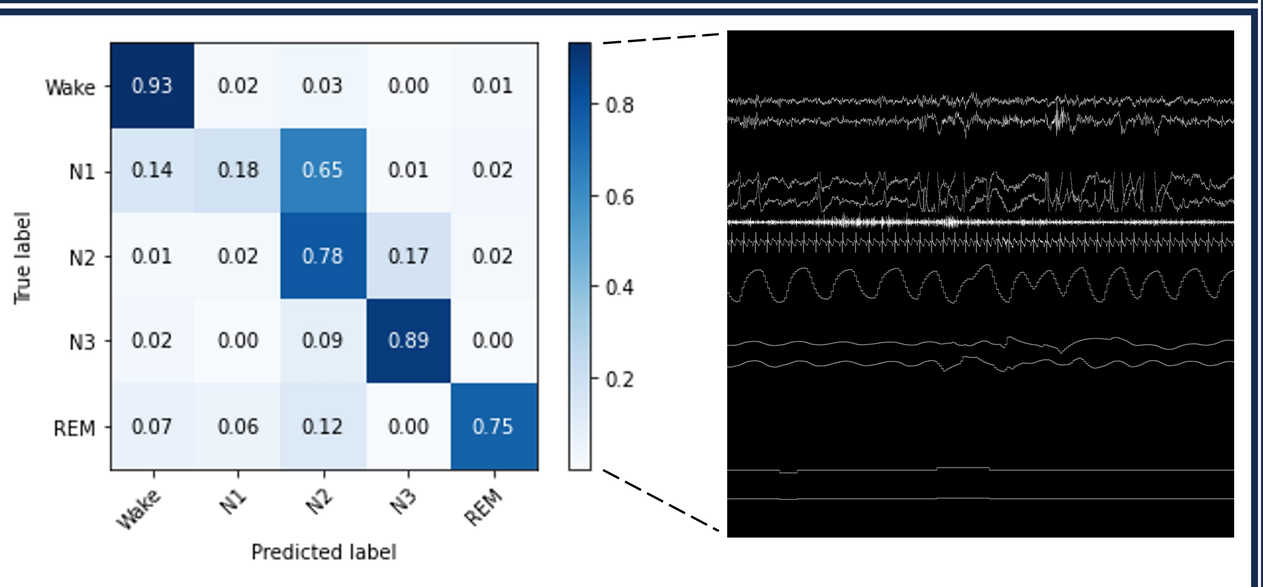
Performance



	Accuracy (%)	Macro F1-score (%)	Weighted F1-score (%)
DeepSleepNet	81.48	80.89	81.62
Our	82.91	82.90	82.76
* Normal	86.84	83.60	86.54
* Mild	86.62	83.37	86.30
* Moderate	84.44	81.98	84.11
* Severe	80.74	80.68	80.66

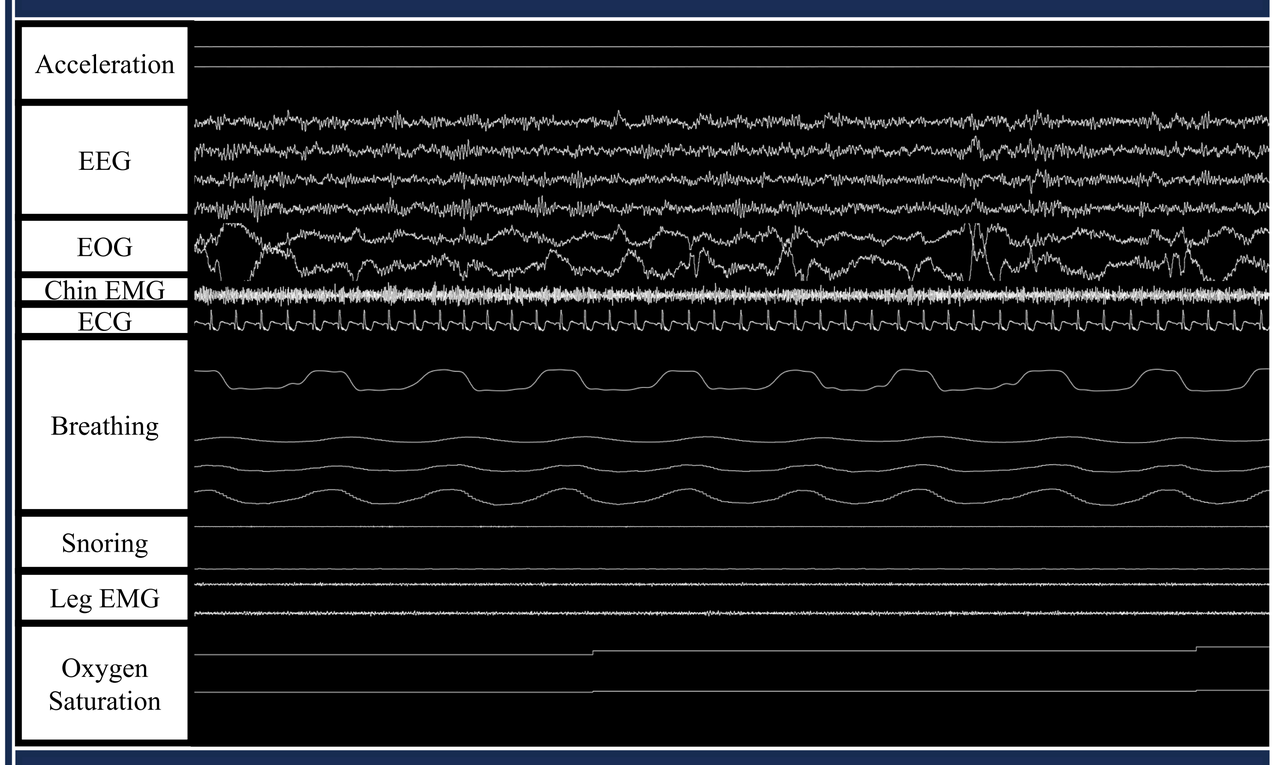
- To identify whether the image-based PSG dataset could be useful for improving performance, we compared the model performance of the image-based and signal-based datasets.
- The performance of the model obtained using the image-based PSG dataset was similar to the performance of the model obtained using the signal-based PSG dataset.
- For the model performance of all the tested epochs, a confusion matrix was generated.

External Validation

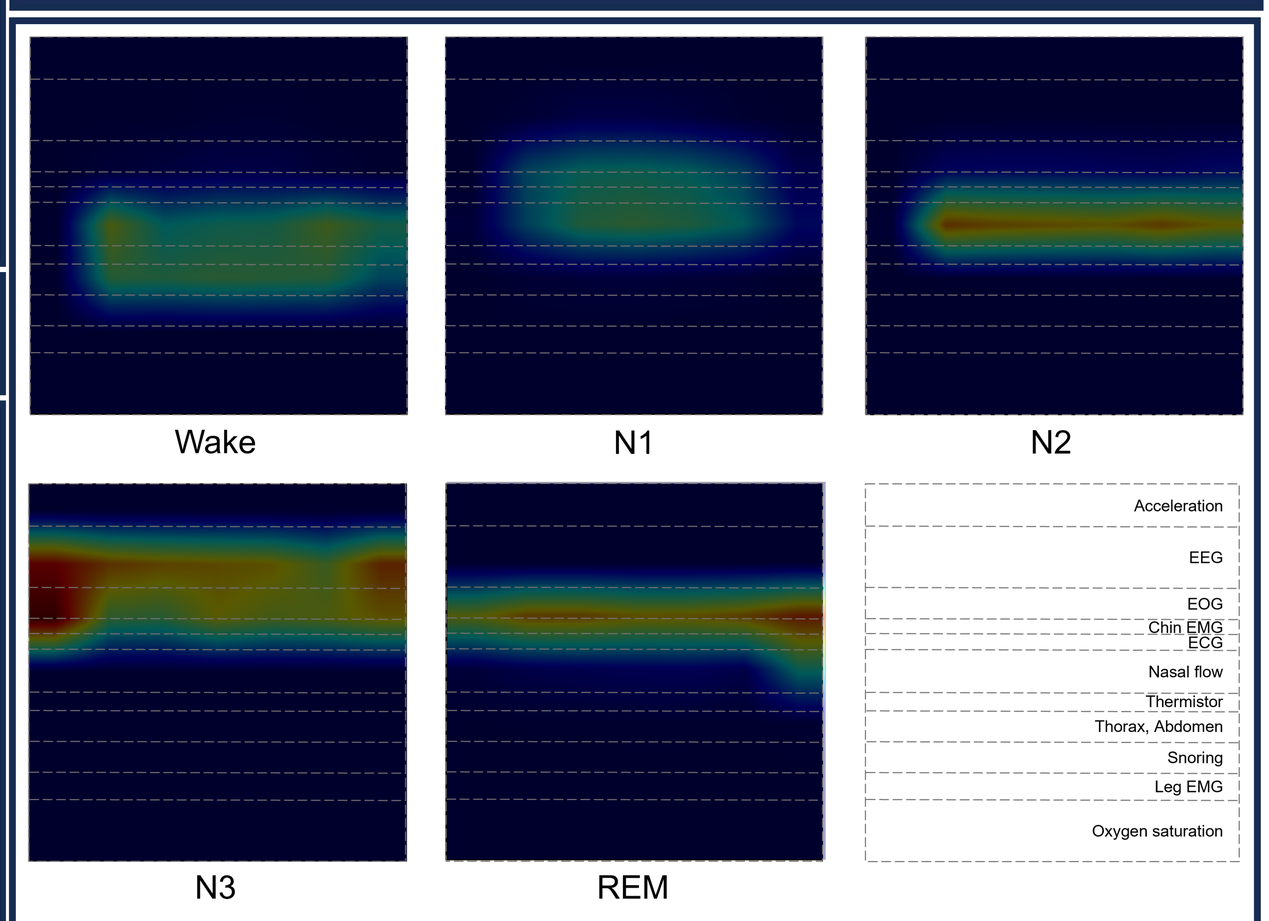


- To evaluate the generalization ability of our image data format, we conducted an external validation with an open-access dataset. We utilized 2,652 PSG data from the Sleep Heart Health Study (SHHS) Visit 2 dataset as the external data and converted them to our image data format.

Data Format



Visualization



- To determine which area was the deciding component for classification in the model using the image-based PSG dataset, we used Eigen-CAM. The final convolution layer contains spatial information indicating discriminative regions to make classifications and generates a spatial heatmap from the activations of the previous convolutional layer.
- We visualized the discriminative region for sleep staging by averaging each Eigen-CAM, which consisted of 10,000 images, to demonstrate where the model usually focused on each class.

Website: https://sleepai.kr/