

Narcolepsy Diagnosis With Sleep Stage Features Using PSG Recordings

Jiquan Wang¹, Sha Zhao¹, Member, IEEE, Yangxuan Zhou, Haiteng Jiang, Zhenghe Yu, Tao Li, Shijian Li, and Gang Pan², Senior Member, IEEE

Abstract—Narcolepsy is a sleep disorder affecting millions of people worldwide and causes serious public health problems. It is hard for doctors to correctly and objectively diagnose narcolepsy. Polysomnography (PSG) recordings, a gold standard for sleep monitoring and quality measurement, can provide abundant and objective cues for the narcolepsy diagnosis. There have been some studies on automatic narcolepsy diagnosis using PSG recordings. However, the sleep stage information, an important cue for narcolepsy diagnosis, has not been fully utilized. For example, some studies have not considered the sleep stage information to diagnose narcolepsy. Although some studies consider the sleep stage information, the stages are manually scored by experts, which is time-consuming and subjective. And the framework using sleep stages scored automatically for narcolepsy diagnosis is designed in a two-phase learning manner, where sleep staging in the first phase and diagnosis in the second phase,

causing cumulative error and degrading the performance. To address these challenges, we propose a novel end-to-end framework for automatic narcolepsy diagnosis using PSG recordings. In particular, adopting the idea of multi-task learning, we take the sleep staging as our auxiliary task, and then combine the sleep stage related features with narcolepsy related features for our primary task of narcolepsy diagnosis. We collected a dataset of PSG recordings from 77 participants and evaluated our framework on it. Both of the sleep stage features and the end-to-end fashion contribute to diagnosis performance. Moreover, we do a comprehensive analysis on the relationship between sleep stages and narcolepsy, correlation of different channels, predictive ability of different sensing data, and diagnosis results in subject level.

Index Terms—Narcolepsy diagnosis, sleep staging, PSG recordings, multi-task learning.

I. INTRODUCTION

SLEEP plays a critical role in promoting mental and physical health [1], [2]. Problems with the quality, timing and amount of sleep severely interfere with normal physical, mental, social and emotional functioning. Such problems are brought about by sleep disorders which affect millions of people worldwide and cause serious public health problems [3]. There are about 50 to 70 millions people in America suffering from a chronic sleep or wakefulness disorder [4], such as narcolepsy, insomnia, restless legs syndrome, and sleep apnea. Among the disorders, narcolepsy, characterized by excessive daytime sleepiness and brief episodes of involuntary sleep, may severely interfere with work or social commitments in daily life [1]. Patients suffer from the sudden onset and irresistible urges to sleep. Meantime, about 70% of patients affected also experience episodes of sudden loss of muscle strength, known as cataplexy [5]. Moreover, narcolepsy tends to happen among relatively young people, for which 15 and 36 years of age is the peak time periods [6]. It is extremely harmful to young people's physical and mental health and even leads to a variety of complications, such as depression, mania, bipolar disorder and schizophrenia. Given the fact that narcolepsy has great harm, it is critical to diagnose narcolepsy, so as to timely protect mental and physical health.

Early diagnosis of narcolepsy is typically based on the presented symptoms. In clinical practice, doctors usually determine subjectively whether one has narcolepsy by asking the patient through direct inquiries or questionnaires. In this way,

Manuscript received 24 March 2023; revised 16 June 2023 and 18 August 2023; accepted 29 August 2023. Date of publication 6 September 2023; date of current version 14 September 2023. This work was supported in part by the STI 2030 Major Projects under Grant 2021ZD0200400. (Corresponding author: Sha Zhao.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Zhejiang University.

Jiquan Wang, Sha Zhao, and Gang Pan are with the College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310013, China, and also with the State Key Laboratory of Brain-Machine Intelligence, Zhejiang University, Hangzhou, Zhejiang 311121, China (e-mail: wangjiquan@zju.edu.cn; szhao@zju.edu.cn; gpan@zju.edu.cn).

Yangxuan Zhou is with the School of Software Technology, Zhejiang University, Ningbo, Zhejiang 315103, China, and also with the State Key Laboratory of Brain-Machine Intelligence, Zhejiang University, Hangzhou, Zhejiang 311121, China (e-mail: 22251325@zju.edu.cn).

Haiteng Jiang and Tao Li are with the Department of Neurobiology, Affiliated Mental Health Center, Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310063, China, also with the MOE Frontier Science Center for Brain Science and Brain-Machine Integration, Hangzhou, Zhejiang 310063, China, and also with the State Key Laboratory of Brain-Machine Intelligence, Zhejiang University, Hangzhou, Zhejiang 311121, China (e-mail: h.jiang@zju.edu.cn; litaozjusc@zju.edu.cn).

Zhenghe Yu is with the Department of Sleep Medicine, Affiliated Mental Health Center, Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310063, China (e-mail: yuzhcoo@sina.com).

Shijian Li is with the College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310013, China, also with the Qiantang Science and Technology Innovation Center, Hangzhou, Zhejiang 310018, China, and also with the State Key Laboratory of Brain-Machine Intelligence, Zhejiang University, Hangzhou, Zhejiang 311121, China (e-mail: shijianli@zju.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3312396

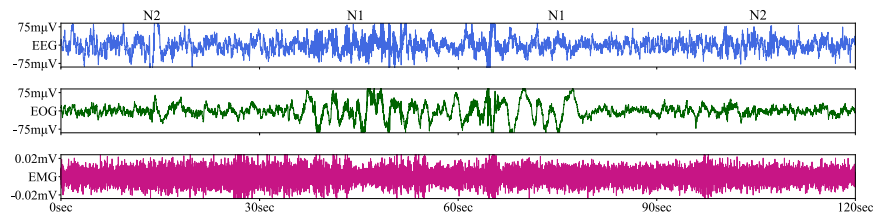


Fig. 1. One real example of recorded PSG signals.

misdiagnosis may be caused, because narcolepsy and other sleep disorders have some similar symptoms and patients may not describe their symptoms accurately and objectively enough. In fact, since people with narcolepsy are often misdiagnosed with other conditions, such as psychiatric disorders or emotional problems, it can take years for someone to get the proper diagnosis [6]. Due to the difficulty in diagnosing narcolepsy, a comprehensive, objective, and high-quality manner is urgently needed to help diagnose narcolepsy. With the development of biomedical engineering and sleep medicine, polysomnography (PSG) in hospitals or sleep centers has become the most effective way to understand the sleep status of subjects. The PSG consists of electroencephalogram (EEG), electrooculogram (EOG), electromyography (EMG), and other physiological signals (e.g., electrocardiogram (ECG), Nasal pressure, and body position). PSG recordings are typically segmented into epochs of 30-second duration, each of which is manually assigned a sleep stage by an expert or technician. This process of sleep staging follows the rule of the American Academy of Sleep Medicine (AASM) sleep standard [7], which defines five different sleep stages: Wake (W), rapid eye movement (REM), and three types of non-REM sleep (N1, N2, N3). A real example of PSG signals in our used dataset is given in Fig. 1. Here, EEG, EOG and EMG signals in each sleep stage are presented. Given the richness and objectivity of sensing recordings, PSG has been considered the gold standard for sleep monitoring [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. Some studies [20], [21], [22], [23] have adopted traditional machine learning methods for automatic sleep staging from PSG. Some studies [11], [14], [24], [25], [26], [27], [28] have proposed deep learning models to predict sleep stages from PSG. It provides abundant and objective cues for narcolepsy diagnosis.

There have been some studies using PSG signals for sleep disorder diagnosis [29], [30], [31], including narcolepsy. Some studies [30], [32] extracted hand-crafted features from PSG recordings, and then fed them into traditional machine learning classifiers (e.g., random forest) for identifying narcolepsy. The sequential relationship within each epoch and between epochs is missed, which is important due to the sequential nature of sleep. With the development of deep learning, the deep neural network is used for narcolepsy diagnosis. However, there are still some limitations that the sleep stage information, an important cue for narcolepsy diagnosis, has not been fully utilized. For example, some studies have not considered the sleep stage information to diagnose narcolepsy. The difference of sleep stage label could be potential biomarkers to classify narcolepsy [33], and the PSG signals belonging to different sleep stages have different performance on narcolepsy diagnosis [31]. However, some studies ignore such

information, directly using PSG recordings for narcolepsy diagnosis [30], [34]. Although the sleep stage information is considered for diagnosis by some studies, the sleep stages are manually scored by experts, which is time-consuming and requires incredible amount of human labor. In some previous studies, the sleep stage labels are first manually assigned, and then combined with PSG recordings for narcolepsy diagnosis [31], [33], [35]. Besides, although some studies scored sleep stages automatically, the methods for narcolepsy diagnosis are designed in a two-phase learning manner, and then they are combined with PSG signals for diagnosis in the second phase [36]. The sleep stages scored in the first phase contain incorrect labels which could not be well optimized in the second phase, causing cumulative error and degrading the performance for the disorder diagnosis.

In order to address the limitations mentioned above, we propose a novel end-to-end framework for narcolepsy diagnosis from PSG signals. We automatically score the sleep stages, and then take advantages of them for narcolepsy diagnosis, by adopting the idea of multi-task learning [37]. To evaluate the framework, we collected a dataset of PSG recordings in our cooperated hospital, consisting of 50 narcolepsy patients and 27 people without disabilities. For convenience, we will later call narcolepsy patients as “patients”, and people without disabilities as “normals”. Compared with other approaches, our framework achieves the state-of-the-art performance. Our contributions are as follows:

- Considering that PSG recordings are the gold standard for sleep monitoring, we collected a dataset of PSG recordings in the cooperated hospital from 50 narcolepsy and 27 healthy people to analyze the relationship between sleep stages and narcolepsy and evaluate our method. In the future, we will release the dataset.
- We design a novel end-to-end framework for automatically diagnosing narcolepsy from PSG recordings by adopting the idea of multi-task learning and setting sleep staging as auxiliary task. Experimental results show that both of the sleep stage related features and the end-to-end fashion significantly contribute to the performance of narcolepsy diagnosis.
- We do a comprehensive analysis, including the relationship between sleep stages and narcolepsy, correlation of different channels (e.g., EEG channel), predictive ability of different sensing data (EEG, EOG, EMG, etc.), and diagnosis results in subject level.

II. DATASET

We first collect a dataset, named as SSND (sleep-stage narcolepsy diagnosis), in our cooperated hospital. All experiments

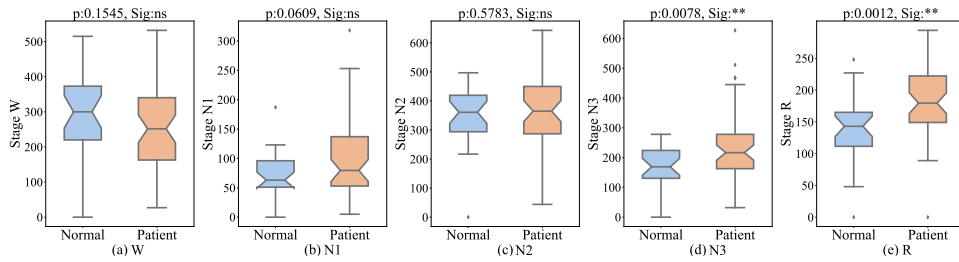


Fig. 2. Significance test on the number of epochs in each sleep stage for normals and patients.

TABLE I
SUBJECTS STATISTICS OF SSND DATASET

Dataset	Total	Female	Male	Age($\mu \pm \sigma$)	Max Age	Min Age
All	77	42	35	22.65 \pm 7.58	49	11
Patient	50	28	22	20.16 \pm 8.22	49	11
Normal	27	14	13	27.26 \pm 2.55	32	22

reported in this work were conducted on this dataset. We first introduce the collection process in details.

A. Dataset Collection

We recruited 77 subjects with age from 11 to 49 (22.65, \pm 7.58), consisting of 42 females and 35 males, to collect PSG recordings in Affiliated Mental Health Center & Hangzhou Seventh People’s Hospital, Zhejiang University School of Medicine. The research was performed at the Zhejiang University with Institutional Review Board approval and written consent was obtained from all the subjects or their caregivers. An overview statistics of subjects are shown in Tab. I.

In the collection procedure, each participant was asked to be in a special ward in the hospital. Before collection, we first need to place the multiple sensors to each participant’s body. The technician put more than 20 wired attachments, including the pulse oximeter, pressure transducer, thermocouple, and electrodes on different positions of the subject’s body (such as head, eyes, nose, chin, and leg). After that, each subject lies in the bed and falls asleep gradually. The wired attachments begin to collect physical signal from different parts of the subject’s body. The PSG recordings were collected according to the AASM sleep standard [7]. During the collection process, EEG, EOG, ECG, Chin EMG and Leg EMG signals were sampled at 512Hz which can capture the fine-grained information for these signals. For each subject, we collected her/his PSG recordings for one whole night, from about 21:00 to 5:00 the next morning, about 8 hours in total. All signals were stored using standard EDF+ data formats with .edf extension. The recordings were segmented into epochs of 30 seconds, and then each epoch was manually labeled as a sleep stage by sleep experts or technician according to AASM [7], including Wake, N1, N2, N3, REM, MOVEMENT, and UNKNOWN.

The sensors for the PSG include EEG (Electroencephalogram), EOG (Electrooculogram), chin EMG (chin Electromyogram), ECG (Electrocardiogram), Nasal Pressure, Therm (Nasal Thermistance), Thor, Abdo (Abdominal Movement), SpO₂ (Oxygen saturation), Snore, PositionSen (Body position), Leg EMG (Leg Electromyogram), IPAP (Inspiratory Positive Airway Pressure), EPAP (Expiratory Positive Airway Pressure), and EtCO₂ (End Tidal Carbon Dioxide).

To ensure a fair comparison, we initially performed pre-processing on the datasets, and subsequently evaluated all the methods using the same prepared datasets. Some signals such as EEG, EOG, Chin EMG, and ECG were band-pass filtered and notch filtered. In subsequent experiments, we removed the epochs annotated as MOVEMENT or UNKNOWN.

B. Dataset Analysis

In order to give a better understanding of our SSND dataset, we analyze it from different perspectives.

1) *Sleep Stage Analysis*: We first analyze sleep stage of our SSND dataset by calculating the number of epochs in total, the number of epochs in each sleep stage, and the proportion of the epochs in each stage for all the subjects, patients and normals, shown in Tab. II. It is worth noting that patients have a lower proportion (22.03% vs 28.46%) of epochs in Wake stage than normals, which shows patients are more likely to fall asleep in bed than normals. This statistical result is consistent with the basic characteristics of narcolepsy. Meanwhile, *for patients, both the proportion of epochs in N3 and REM are bigger than those of normals* (21.08% v.s. 16.81%, 16.15% v.s. 13.59%). The statistical results of REM stage are consistent with the previous discovery that patients with narcolepsy typically have higher REM sleep density than normals [38].

To further analyze the relationship between sleep stage distribution and narcolepsy, we conducted a significance test on the number of epochs in each sleep stage and whether one subject is a patient or normal, shown in Fig. 2. Here, the “p” value is an indicator of the difference between the patients and the normals on each stage. The “Sig” is an indicator of significance. From Fig. 2 we can see, p values in Wake, N1, N2, N3 and REM stages are respectively 0.1545, 0.0609, 0.5783, 0.0078 and 0.0012. “Sig:ns” denotes $p \geq 0.05$, indicating there no significant difference between patients and normals. “Sig:**” denotes $p \leq 0.01$, which indicates there a significant difference between patients and normals. Obviously, *compared with other stages, the differences in N3 and REM stage between patients and normals are more significant* ($p=0.0078$ in N3 stage and $p=0.0012$ in REM stage). This result indicates that patients with narcolepsy are more likely to enter the N3 and REM stages than normals.

2) *Hypnogram Analysis*: To further analyze the relationship between sleep stage and narcolepsy, we compare two examples of hypnograms manually scored by a sleep expert from two whole-night PSG recordings of a patient and a normal in Fig. 3. Hypnogram is a graph that represents the stages of sleep as a function of time. Hypnograms are usually obtained by scoring the recordings from EEG, EOG and EMG. From Fig. 3(a) we can see that transitions of sleep stages happen frequently in a patient with narcolepsy. On the contrary,

TABLE II
STATISTICS OF SLEEP EPOCHS

Dataset	All Epochs	Wake	N1	N2	N3	REM
All	84,546	20,402 (24.13%)	6,927(8.19%)	27,630(32.68%)	16,642(19.68%)	12,945(15.31%)
Patients	56,895	12,532(22.03%)	4,975(8.74%)	18,208(32.00%)	11,993(21.08%)	9,187(16.15%)
Normals	27,651	7870(28.46%)	1,952(7.06%)	9,422(34.07%)	4,649(16.81%)	3,758(13.59%)

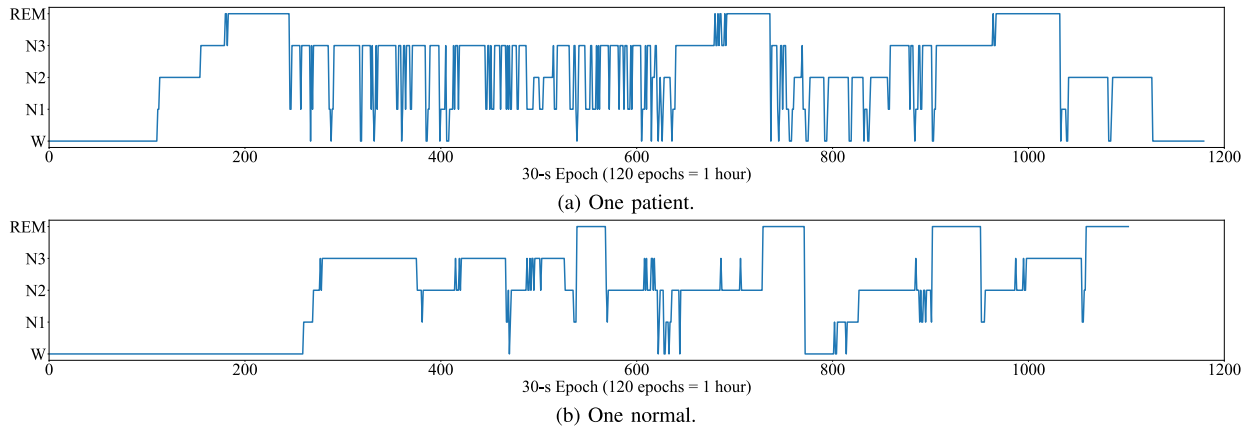


Fig. 3. The hypogram of one whole-night recording from (a) one patient and (b) one normal.

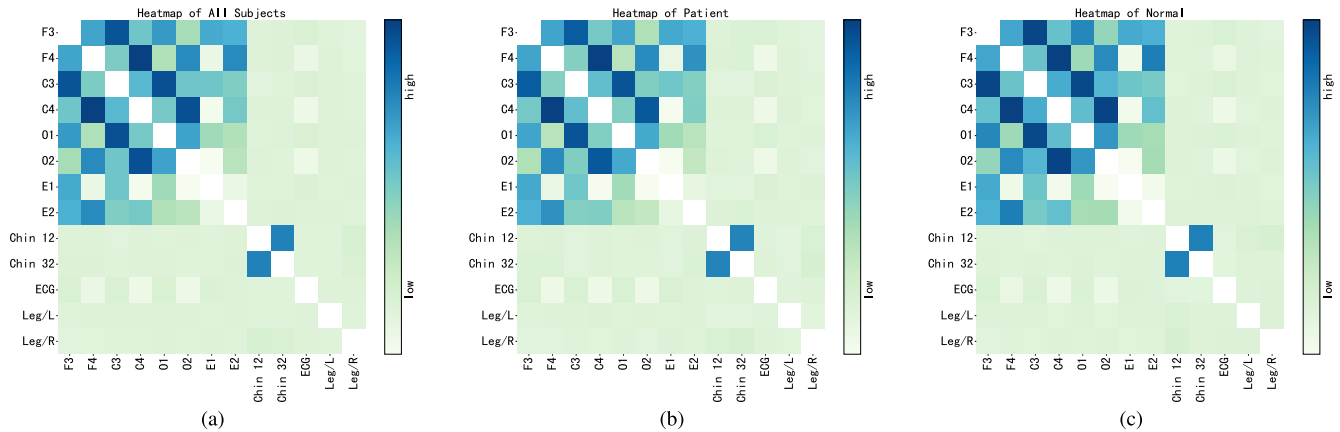


Fig. 4. Pearson correlation coefficient between different channels.

a normal can have more stationary transitions of sleep stages than a patient, shown in Fig. 3(b). Above findings illustrate that *narcolepsy can bring brief arousals in sleep and promote sleep fragmentation that typically disturbs sleep architecture*. It further proves that the known sleep stages can help diagnose narcolepsy. Therefore, we try to introduce a sleep staging task as an auxiliary task [39] in our deep learning model for narcolepsy diagnosis, which helps compete the primary task of narcolepsy diagnosis and improve the performance.

3) *Correlation Analysis of Different Channels*: In pervious work, EEG, EOG, EMG and ECG have frequently been given higher importance compared to other signals. Here, we investigate the correlation between different modalities, by calculating the Pearson correlation coefficient between different signals from 13 important channels of EEG, EOG, EMG, and ECG. The heatmaps of Pearson correlation coefficient are shown in Fig. 4. Firstly, the heatmaps of all the subjects, patients and normals are similar in our dataset, which illustrates that overall results of Pearson correlation coefficient on patients and normals are coincident. Then, the values of Pearson correlation coefficient between 6 EEG channels are high, especially the value between F4 and C4. It illustrates that single-channel EEG may achieve the performance similar

to that of the fusion of 6 EEG channels. It is worth noting that the value of Pearson correlation coefficient between Chin1-Chin2 EMG and Chin3-Chin2 EMG is high, which shows two Chin EMG channels are similar and single-channel Chin EMG may represent information of two-channels Chin EMG.

III. METHODS

A. Problem Formulation

Our model is designed in an end-to-end fashion, which processes a sequence of sleep epochs and outputs a narcolepsy prediction with a sequence of predicted sleep stages. We denote $x \in \mathbb{R}^{n \times C}$ as a sleep epoch, where n is the number of sampling points in a sleep epoch and C is the number of channels. The input sequence of sleep epochs is defined as $X = \{x_1, x_2, x_3, \dots, x_L\}$, where $x_i \in \mathbb{R}^{n \times C}$ denotes a sleep epoch, L is length of the sequence and $i \in [1, 2, 3, \dots, L]$.

For automatic sleep staging, we denote the number of sleep stages as N , and $N = 5$ (Wake, N1, N2, N3, REM), according to the AASM sleep standard [7]. We define $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \hat{y}_3, \dots, \hat{y}_L\}$ as the sequence of sleep stages corresponding to $X = \{x_1, x_2, x_3, \dots, x_L\}$, where $\hat{y}_i \in \{0, 1\}^N$ is the one-hot encoding of ground-truth sleep stage of x_i .

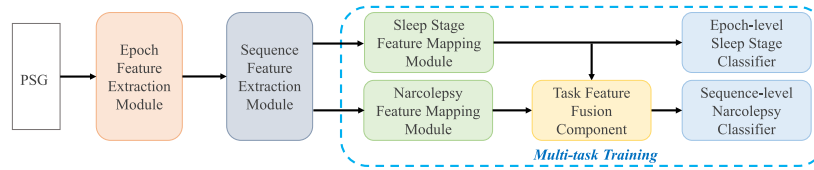


Fig. 5. Our framework overview.

For automatic narcolepsy diagnosis, we denote the number of narcolepsy diagnosis as M , and $M = 2$ (patient and normal). $\hat{z} \in \{0, 1\}^M$ is defined as an one-hot encoding of ground-truth narcolepsy diagnosis. For a sequence of sleep epochs $X = \{x_1, x_2, x_3, \dots, x_L\}$, $\hat{z} \in \{0, 1\}^M$ is ground-truth narcolepsy diagnosis.

Therefore, our sleep staging task is defined as learning a mapping function F that maps a sequence of sleep epochs X into the corresponding sequence of sleep stages \hat{Y} and a narcolepsy diagnosis \hat{z} .

B. Overview

In order to diagnose narcolepsy from PSG recordings, we design an end-to-end framework, which captures sequential relationship within each epoch and between epochs, automatically scores sleep stages and combines the scored stages with PSG recordings for narcolepsy diagnosis. Specifically, we adopt the idea of multi-task learning, and take the sleep staging as our auxiliary task which contributes to the performance improvement of our primary task: narcolepsy diagnosis. In the auxiliary task, we automatically score the sleep stages and simultaneously learn the sleep stage features that are then combined with narcolepsy features extracted from PSG recordings for the primary task.

Our deep learning model is illustrated in Fig. 5. The model consists of seven modules: (1) Epoch Feature Extraction Module, (2) Sequence Feature Extraction Module, (3) Sleep Stage Feature Mapping Module, (4) Narcolepsy Feature Mapping Module, (5) Task Feature Fusion Component, (6) Epoch-level Sleep Stage Classifier, (7) Sequence-level Narcolepsy Classifier. We first design Epoch Feature Extraction Module to extract the local features within each epoch of raw signals from PSG. Then, the epoch features are input to Sequence Feature Extraction Module. Next, we design two task-guided feature mapping modules, Sleep Stage Feature Mapping Module and Narcolepsy Feature Mapping Module. Sleep Stage Feature Mapping Module is used to map features for sleep staging and Narcolepsy Feature Mapping Module is used to map features for narcolepsy diagnosis. The sequence features are fed into Sleep Stage Feature Mapping Module and Narcolepsy Feature Mapping Module to obtain sleep stage features and narcolepsy features, respectively. Then, sleep stage features are fed into Epoch-level Sleep Stage classifier to predict sleep stages and are also fed into Task Feature Fusion Component with narcolepsy features to obtain fused narcolepsy features. Finally, fused narcolepsy features are fed into Sequence-level Narcolepsy Classifier to diagnose narcolepsy.

C. Feature Extraction Module

1) *Epoch Feature Extraction Module*: Local salient wave features are critical in sleep staging for sleep experts [7]. For

automatic sleep staging, extracting features from local salient waveforms within each epoch can help classify sleep staging in epoch level. In addition, existing studies on sleep disorder [30], [31] show that local salient wave features from each epoch are also helpful for diseases diagnosis, such as narcolepsy. Therefore, in our deep learning network, we design an Epoch Feature Extraction Module to extract local feature within each epoch. Epoch Feature Extraction Module consists of Convolutional Neural Network (CNN), Batch Normalization [40], and GELU [41] activation function. Existing studies on sleep staging from PSG [8], [9], [10], [11] have proved that CNN is able to capture the local features of significant waveforms. Therefore, we utilize CNN to extract local features from salient waveforms within each epoch.

We feed sleep sequence $X = \{x_1, x_2, x_3, \dots, x_L\}$ into Epoch Feature Extraction Module. The process is as follows:

$$X_{j+1} = \text{Max Pooling}_j(G(\text{BN}(\text{Conv}_j(X_j))), j \in [0, 1, 2, 3] \quad (1)$$

$$X^{\text{epoch}} = \text{Avg Pooling}(X_4) \quad (2)$$

where X_j is the j -th features (X_0 is X), Conv_j is the j -th convolution layer of Epoch Feature Extraction Module, BN is Batch Normalization, G is GELU activation function, Max Pooling_j is the j -th max pooling layer, and Avg Pooling is an average pooling layer. Finally, Epoch Feature Extraction Module outputs the epoch features $X^{\text{epoch}} = \{x_1^{\text{epoch}}, x_2^{\text{epoch}}, x_3^{\text{epoch}}, \dots, x_L^{\text{epoch}}\}$, where $x_i^{\text{epoch}} \in \mathbb{R}^d$ and d is the feature dimension.

2) *Sequence Feature Extraction Module*: Transition patterns of sleep stages between epochs play an critical role in sleep staging [7]. Therefore, modeling the relationship between sleep epochs in sequence is helpful for sleep staging. In addition, for narcolepsy diagnosis, extracting global context features from the sequence of sleep epoch can avoid being limited to the local characteristics of the waveform within an epoch. In other words, modeling the sleep sequence can expand the receptive field of model to learn global characteristics of the waveform, which can improve the performance of narcolepsy diagnosis. Due to effectiveness of modeling global relationship, we propose a Sequence Feature Extraction Module to extract context features between epochs in a sleep sequence. In pervious work on automatic sleep staging [12], [13], [14], Transformer or multi-head attention is used to model global temporal context and achieves a high performance. Inspired by these studies, we use a Transformer Encoder as a Sequence Feature Extraction Module, which can encode global context features through multi-head attention.

The Transformer layer, just like standard Transformer [42], adopts scaled dot-product attention, which is defined as

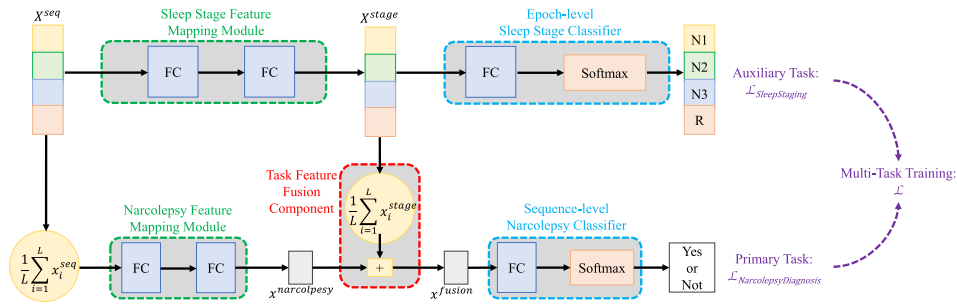


Fig. 6. Illustration of multi-task process (here only $L = 4$ for visual purposes). Here, FC is fully-connected layer, and $+$ is element-wise addition.

follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

where matrices Q , K , and V consist of queries, keys, and values, respectively, and d_k is the dimension of keys. We feed epoch features $X^{epoch} = \{x_1^{epoch}, x_2^{epoch}, x_3^{epoch}, \dots, x_L^{epoch}\}$ into Sequence Feature Extraction Module. The process is as follows:

$$X^{seq} = Transformer(X^{epoch}) \quad (4)$$

where $Transformer$ is standard Transformer Encoder, $X^{seq} = \{x_1^{seq}, x_2^{seq}, x_3^{seq}, \dots, x_L^{seq}\}$ is sequence context features, and $x_i^{seq} \in \mathbb{R}^d$.

D. Narcolepsy Diagnosis With Sleep Staging Features

As Fig. 2 and Fig. 3 in Section II-B show, there are significant differences in the proportion and transition of sleep stages between patients and normals. Some existing studies [31], [33] have proved that known sleep stage information can improve the performance of narcolepsy diagnosis. Therefore, we try to take advantages of sleep staging for narcolepsy diagnosis. Here, we adopt the idea of multi-task learning, where we take sleep staging as the auxiliary task to automatically extract sleep stage features for narcolepsy diagnosis. For mapping sequence context features into sleep stage features and narcolepsy features, we design two task-guided feature mapping modules, which are Sleep Stage Feature Mapping Module and Narcolepsy Feature Mapping Module. The two feature mapping modules mainly consist of multilayer perceptrons. The Sleep Stage Feature Mapping Module maps sequence features into sleep stage features to predict sleep stages by the Epoch-level Sleep Stage Classifier. The Narcolepsy Feature Mapping Module maps sequence features into narcolepsy features to diagnose narcolepsy. Then, to combine sleep stage features and narcolepsy features for narcolepsy diagnosis, we design a Task Feature Fusion Component to fuse them together, so as to obtain fused narcolepsy features. Finally, the fused narcolepsy features are fed into the Sequence-level Narcolepsy Classifier to diagnose narcolepsy.

1) *Auxiliary Task: Automatic Sleep Staging*: We take the automatic sleep staging as the auxiliary task of narcolepsy diagnosis. Our sleep staging process is shown in Fig. 6. For sequence context features $X^{seq} = \{x_1^{seq}, x_2^{seq}, x_3^{seq}, \dots, x_L^{seq}\}$, the process of sleep stage feature mapping is as follows:

$$X^{stage} = MLP(X^{seq}) \quad (5)$$

where MLP is multilayer perceptrons consist of two fully-connected layers, $X^{stage} = \{x_1^{stage}, x_2^{stage}, x_3^{stage}, \dots, x_L^{stage}\}$ is sleep stage features, $x_i^{stage} \in \mathbb{R}^{d'}$ and d' is task-guided feature dimension.

After mapping sequence context features X^{seq} into sleep stage features X^{stage} , we feed X^{stage} into Epoch-level Sleep Stage Classifier, which consists of fully-connected layer and a softmax function, to obtain $Y = \{y_1, y_2, y_3, \dots, y_L\}$, where $y_i \in \mathbb{R}^N$ is the predicted probability in N sleep stage classes of the i -th epoch. We use the cross-entropy (CE) function as sleep staging loss function:

$$\mathcal{L}_{SleepStaging} = - \sum_{i=1}^L \sum_{j=1}^N \hat{y}_{i,j} \log(y_{i,j}) \quad (6)$$

where $y_{i,j} \in \mathbb{R}$, the j -th element of y_i , denotes the probability that the i -th epoch is predicted to the j -th sleep stage class, and $\hat{y}_{i,j} \in \{0, 1\}$, the j -th element of \hat{y}_i , denotes the probability that the i -th epoch actually belongs to the j -th class.

2) *Primary Task: Narcolepsy Diagnosis*: Our narcolepsy diagnosis process is shown in Fig. 6. Considering that the task of narcolepsy diagnosis is sequence-level, we calculate the average feature of sequence context features X^{seq} before feeding it into MLP. The process of narcolepsy feature mapping is as follow:

$$x^{narcolepsy} = MLP\left(\frac{1}{L} \sum_{i=1}^L x_i^{seq}\right) \quad (7)$$

where x_i^{seq} is i -th feature of sequence context features X^{seq} , MLP is multilayer perceptrons consisting of two fully-connected layers, $x^{narcolepsy} \in \mathbb{R}^{d'}$ is narcolepsy feature.

For making sleep staging task as the auxiliary task of narcolepsy diagnosis, we design a Task Feature Fusion Component to fuse sleep stage feature and narcolepsy feature together. The Task Feature Fusion Component is as follows:

$$x^{fusion} = x^{narcolepsy} + \frac{1}{L} \sum_{i=1}^L x_i^{stage} \quad (8)$$

where x_i^{stage} is the i -th feature of sleep stage features X^{stage} and $x^{fusion} \in \mathbb{R}^{d'}$ is the fused narcolepsy feature. Then we feed x^{fusion} into Sequence-level Narcolepsy Classifier, which consists of fully-connected layer and a softmax function, to obtain $z \in \mathbb{R}^M$. z is the predicted probability in M Narcolepsy classes of the sleep sequence. We use the cross-

TABLE III
THE DETAILS OF DATA SPLITTING OF SUBJECT-WISE
6-FOLD CROSS-VALIDATION

Fold	Subjects	Patients	Normals
1	13	9	4
2	13	9	4
3	13	8	5
4	13	8	5
5	13	8	5
6	12	8	4
Sum	77	50	27

entropy (CE) function as narcolepsy diagnosis loss function:

$$\mathcal{L}_{NarcolepsyDiagnosis} = - \sum_{j=1}^M \hat{z}_j \log(z_j) \quad (9)$$

where $z_j \in \mathbb{R}$, the j -th element of z , denotes the probability that the sequence is predicted to the j -th narcolepsy diagnosis class, and $\hat{z}_j \in \{0, 1\}$, the j -th element of \hat{z} , denotes the probability that the sequence actually belongs to the j -th narcolepsy diagnosis class.

E. Joint Training

In the training procedure, sleep staging and narcolepsy diagnosis are jointly trained together by the same one objective function, which consists of two different parts, staging loss and diagnosing loss, described in Equation 10:

$$\mathcal{L} = \lambda \times \mathcal{L}_{SleepStaging} + (1 - \lambda) \times \mathcal{L}_{NarcolepsyDiagnosis} \quad (10)$$

where λ is the coefficient of two loss functions for sleep staging and narcolepsy diagnosis.

IV. EXPERIMENT

A. Performance Measurement and Implementation

We use ACC (accuracy) and F_1 -score (F_1) to measure model performance. In particular, given that the task of sleep staging is a multi-class classification problem, we replace F_1 score with Macro F_1 score. In other word, we calculated the F_1 scores in a class-wise manner, and reported the mean value to get Macro F_1 score.

Inspired by most of existing methods on automatic sleep staging, we adopted a subject-wise 6-fold cross-validation policy by dividing the subjects in the dataset into 6 groups. In each fold, five groups were used for training, and the left one for testing, ensuring that the data from the same one subject never appear in the training set and testing set simultaneously. In addition, we ensured that each fold has the same number of subjects and includes patients and normals. The details of data splitting are shown in Tab. III.

We implemented our deep learning model based on the PyTorch [43]. We evaluated EEG, EOG, ECG, EMG (including chin EMG and Leg EMG) and Nasal Pressure from PSG recordings on our deep learning model. The model was trained using the Adam optimizer with default settings and the learning rate was set to $1e-4$. The mini-batch size was set to 32 and dropout [44] rate was set to 0.1. We adopted early stopping [45] policy in the training process. If the model does not achieve a better performance any more for ten consecutive

TABLE IV
PERFORMANCE COMPARISON WITH OTHER METHODS
USING SINGLE-CHANNEL EEG (F4-M1)

Methods	Narcolepsy Diagnosis		Sleep Staging	
	Accuracy	F_1	Accuracy	Macro- F_1
SVM	61.66	69.46	65.85	60.14
RF	61.78	68.55	67.21	62.46
CNN	72.82	80.72	79.32	72.09
CNN+RNN	74.81	81.81	79.91	71.87
Transformer	76.34	83.59	80.14	72.46
Our method	78.94	85.45	81.24	74.85

epochs, the training ends. The Transformer block of Sequence Feature Extraction Module has 8 heads and 512 hidden states. We set the length of sleep epoch sequence as $L = 20$, feature dimension as $d = 512$, task-guided feature dimension as $d = 128$ and the coefficient of two loss functions as $\lambda = 0.5$. Before being fed into deep learning model, EEG, EOG, ECG, EMG and Nasal Pressure signals were resampled to 100Hz. We trained the model on the machine with Intel Core i9 10900K CPU and eight NVIDIA RTX 3080 GPUs.

B. Compared Methods

In our experiment, we compared our proposed method with the following approaches on sleep staging and narcolepsy diagnosis. For fair comparison, all the approaches were evaluated on the same dataset, and adopted subject-wise training policy:

SVM (Support Vector Machine) [46] uses a Gaussian kernel function for automatic sleep staging and narcolepsy diagnosis.

RF (Random Forests) [47] is an ensemble learning method.

CNN (Convolutional Neural Network) is used as the feature extractor of raw PSG recordings for automatic sleep staging and narcolepsy diagnosis.

CNN + RNN, where CNN is used to extract local features within each epoch and RNN is used to extract context features from an epoch sequence.

Transformer is used as the feature extractor of PSG recordings for automatic sleep staging and narcolepsy diagnosis.

C. Overall Results

We first compared our model with other approaches for sleep staging and narcolepsy diagnosis on single-channel EEG (F4-M1). Previous studies have proved that using EEG achieves good performance [8], [36]. Here, all the approaches were evaluated using EEG signals for sleep staging and narcolepsy diagnosis. As we can see from Tab. IV, our method achieves the best performance. SVM and RF perform the worst, about 16% lower in accuracy than our method (65.85% v.s. 67.21% v.s. 81.24% on sleep staging and 61.66% v.s. 61.78% v.s. 78.94% on narcolepsy diagnosis). It indicates the sequential relationship in EEG signals is important for the diagnosis, which the traditional machine learning methods cannot model yet. CNN performs worse than our method on sleep staging (79.32% v.s. 81.24% in accuracy and 72.09% v.s. 74.85% in Macro- F_1) and narcolepsy diagnosis (72.82% v.s. 78.94% in accuracy and 80.72% v.s. 85.45% in F_1), indicating fully CNN without context features extractor cannot well model sequential relationship between epochs, which helps for the disorder diagnosis. CNN+RNN, where CNN is used as

TABLE V

ANALYSIS OF AUXILIARY TASK IN SINGLE-CHANNEL EEG (F4-M1)

Methods	Narcolepsy Diagnosis		Sleep Staging	
	Accuracy	F ₁	Accuracy	Macro-F ₁
Our model	78.94	85.45	81.24	74.85
Single-task method	75.97(-2.97)	82.69(-2.76)	—	—
No-fusion method	77.10(-1.84)	84.55(-0.9)	80.84(-0.40)	75.04(+0.19)
Two phase method	77.11(-1.83)	82.56(-2.89)	80.69(-0.55)	74.58(-0.27)

epoch feature extractor and RNN is used as sequence feature extractor, performs worse than our method on sleep staging (79.91% v.s. 81.24% in accuracy and 71.87% v.s. 74.95% in Macro-F₁) and on narcolepsy diagnosis (74.81% v.s. 78.94% in accuracy and 81.81% v.s. 85.45% in F₁). Transformer, using fully Transformer to capture local and global features from EEG signals, performs worse about 1.7% accuracy and 1.9% F₁ than our method on narcolepsy diagnosis (76.34% v.s. 78.94% in accuracy and 83.59% v.s. 85.45% in F₁). The Transformer cannot well extract local features within each epoch from EEG signals. Compared with other approaches, our deep learning model, using CNN as epoch feature extractor and Transformer as sequence feature extractor, utilizing sleep staging as the auxiliary task, can well model local and global features from EEG signals and make full use of sleep stage information to improve the performance of narcolepsy diagnosis.

V. ANALYSIS AND DISCUSSION

A. Analysis of Sleep Staging Task

To investigate the effectiveness of the auxiliary task of sleep staging, the Task Feature Fusion Component and the end-to-end manner, we compared our model with the three following methods:

Single-Task Method: We set single-task method as a baseline method, where we ablate the Sleep Feature Mapping Module, the Task Feature Fusion Component and the Epoch-level Sleep Stage Classifier from our model.

No-Fusion Method: We set no-fusion method as another baseline method, where we only ablate the Task Feature Fusion Component. This model can be used to classify sleep stages and narcolepsy, but the sleep stage features and narcolepsy features are not fused together for narcolepsy diagnosis.

Two-Phase Method: In the two-phase method, the sleep staging is automatically scored in the first phase, and the narcolepsy is diagnosed in the second step. The two tasks are trained separately.

For fair comparison, we set the same hyperparameters for these models as our model. The results of ablation experiments are shown in Tab. V. From Tab. V we can see that single-task method performs the worst, 2.97% lower in accuracy and 2.76% lower in F₁ than our model, on narcolepsy diagnosis (75.97% v.s. 78.94% in accuracy and 82.69% v.s. 85.46% in F₁). It is reasonable that the single-task method without sleep staging task can not well extract features and learn the transition rules about sleep stages, which can help classify narcolepsy. No-fusion method performs close to our model on sleep staging (80.84% v.s. 81.24% in accuracy and 75.04% v.s. 74.85% in Macro-F₁). Obviously, ablating Task Feature Fusion Component has no significant impact on performance of sleep

TABLE VI

THE RESULTS OF USING SINGLE CHANNEL AND MULTIPLE CHANNELS

Channels	Narcolepsy Diagnosis		Sleep Staging	
	Accuracy	F ₁	Accuracy	Macro-F ₁
F4-M1	78.94	85.45	81.24	74.85
C4-M1	77.10	83.48	80.45	74.31
F3-M2	76.44	83.13	78.67	72.33
C3-M2	76.76	82.54	78.28	71.76
O2-M1	76.96	83.42	79.09	72.75
O1-M1	76.92	83.55	79.22	73.02
F4-M1 + C4-M1	76.59	82.89	82.08	76.85
All EEG Channels	76.79	83.48	80.44	74.17
Chin1-Chin2	69.98	78.37	53.35	40.63
Chin3-Chin2	70.88	79.97	52.77	42.18
LegL	73.23	80.15	47.28	28.86
LegR	73.60	80.40	46.98	28.44
Chin1-Chin2 + Chin3-Chin2	70.57	79.76	52.58	42.36
All EMG Channels	74.18	80.57	48.74	35.85

staging. However, on narcolepsy diagnosis, no-fusion method performs 1.84% lower in accuracy and 0.9% lower in F₁ than our model (77.10% v.s. 78.94% in accuracy and 84.55% v.s. 85.45% in F₁). *It further indicates that the sleep stage features can improve the performance of narcolepsy diagnosis.* Two phase method performs close to our model on sleep staging (80.69% v.s. 81.24% in accuracy and 74.58% v.s. 74.85% in Macro-F₁). However, for narcolepsy diagnosis, it performs 1.83% lower in accuracy and 2.89% lower in F₁ than our model (77.11% v.s. 78.94% in accuracy and 82.56% v.s. 85.45% in F₁). These indicate that two phase method works well in sleep staging. However, the sleep stages scored in the first phase contain incorrect labels which could not be well optimized in the second phase, causing cumulative error and leading the poor performance for the disorder diagnosis. *All the results prove the importance of setting sleep staging as the auxiliary task for narcolepsy diagnosis.*

B. Analysis of Highly Correlated Channels

As shown in Fig. 4, some channels are highly correlated, such as the six EEG channels and the two chin EMG channels. In EEG, F4-M1 and C4-M1 channels are highly correlated. In EMG, the correlation coefficient between two chin EMG is high. Channels that exhibit high correlation with each other can lead to information redundancy. Among them, we could choose only one channel to feed into our deep learning model to achieve a high performance. Here, we tested the model performance when using single channel and using highly correlated channels, respectively. Specifically, we evaluated our model on single channel (F4-M1, C4-M1, F3-M2, C3-M2, O2-M1, O1-M1 in EEG and Chin1-Chin2, Chin3-Chin2, LegL, LegR in EMG), two highly correlated channels (F4-M1 + C4-M1 and Chin1-Chin2 + Chin3-Chin2), all EEG channels and all EMG channels, shown in Tab. VI.

For EEG, when only using F4-M1, our model achieves the best performance on narcolepsy diagnosis (78.94% in accuracy and 85.45% in F₁). Compared with F4-M1, our model using C4-M1 performs a little worse on sleep staging (81.24% v.s. 80.45% in accuracy and 74.85% v.s. 74.31% in Macro-F₁) and narcolepsy diagnosis (78.94% v.s. 77.10% in accuracy and 85.45% v.s. 83.48% in F₁). In addition, our model using other single-channel EEG performs much worse than F4-M1

TABLE VII
THE RESULTS OF USING SINGLE MODALITIES AND MULTI MODALITIES

	Modalities	Narcolepsy Diagnosis		Sleep Staging	
		Accuracy	F ₁	Accuracy	Macro-F ₁
Single Modality	EEG	78.94	85.45	81.24	74.85
	EOG	76.63	84.06	81.11	74.28
	EMG	74.18	80.57	48.74	35.85
	ECG	75.54	83.09	44.85	33.05
	Nasal Pressure	71.30	79.67	55.21	45.63
Multiple Modalities	EEG+EOG	79.71	84.47	82.39	76.37
	EEG+EMG	75.23	83.49	80.16	74.08
	EEG+ECG	72.31	81.95	76.95	69.93
	EEG+Nasal Pressure	71.95	79.57	65.86	56.76
	All Modalities	72.43	81.71	73.41	64.79

and C4-M1 on sleep staging and narcolepsy diagnosis. These results indicate that the F4-M1 channel is the most powerful on narcolepsy diagnosis.

Our model combining F4-M1 and C4-M1 channels performs best on sleep staging (82.08% in accuracy and 76.85% in Macro-F₁), but performs worse than single F4-M1 channel on narcolepsy diagnosis (76.59% v.s. 78.94% in accuracy and 82.89% v.s. 85.45% in F₁). Furthermore, our model in all EEG channels performs worse than single F4-M1 channel on narcolepsy diagnosis (76.79% v.s. 78.94% in accuracy and 83.49% v.s. 85.45% in F₁). The results suggest that combining highly correlated channels or combining all the EEG channels could not improve the performance of narcolepsy diagnosis.

For EMG, our model using single Chin1-Chin2 channel, single Chin3-Chin2 channel, and the two channels achieves a relatively close performance on narcolepsy diagnosis (69.98% v.s. 70.88% v.s. 70.57% in accuracy and 78.37% v.s. 79.97% v.s. 79.76% in F₁). Similarly, combining highly correlated EMG channels could not contribute to narcolepsy diagnosis. Besides, our model using LegL and LegR performs worse than Chin3-Chin2 on sleep staging (47.28% v.s. 46.98% v.s. 52.77% in accuracy and 28.86% v.s. 28.44% v.s. 42.18% in Macro-F₁) but performs better than Chin3-Chin2 on narcolepsy diagnosis (73.23% v.s. 73.60% v.s. 70.88% in accuracy and 80.15% v.s. 80.40% v.s. 79.97% in F₁). In Tab. 4, the values of Pearson correlation coefficient between Chin EMG and Leg EMG are low. It indicates that Leg EMG contains different information from Chin EMG, which performs worse on sleep staging but performs better on narcolepsy diagnosis than Chin EMG. It is worth noting that our model in all EMG channels performs better than Chin3-Chin2 on narcolepsy diagnosis (74.18% v.s. 70.88% in accuracy and 80.57% v.s. 79.97% in F₁). It further indicates that Leg EMG can help provide effective features for narcolepsy diagnosis.

C. Analysis of Multiple Modalities

In this experiment, we investigated the predictive abilities of different modalities and their combination for narcolepsy diagnosis, including EEG, EOG, EMG, ECG, Nasal Pressure, EEG+EOG, EEG+EMG, EEG+ECG, and EEG+Nasal Pressure. In single-modality experiments, the EEG here refers to the single-channel EEG of F4-M1 channel which achieves the best performance in narcolepsy diagnosis.

Table VII shows the performance comparison. As we can see from Tab. VII, when using single modality of EEG, our method achieves the best performance (81.24% in accuracy

and 74.85% in Macro-F₁) on sleep staging and the best performance (78.94% in accuracy and 85.45% in F₁) on narcolepsy diagnosis compared with other single-modality results, indicating that EEG is the most predictive for sleep staging and narcolepsy diagnosis in PSG recordings. Using EOG also has a good performance. The sleep staging results are close to EEG (81.11% v.s. 81.24% in accuracy, 74.28% v.s. 74.85% in Macro-F₁), but the narcolepsy diagnosis results are lower than EEG (76.63% v.s. 78.94% in accuracy and 84.06% v.s. 85.46% in F₁). According to AASM sleep standard [7], EOG is also an important standard for experts to assign sleep stages. In addition, EMG, ECG and Nasal pressure are not so helpful for sleep staging for which all the accuracy are lower than 60%, but they are relatively useful for narcolepsy diagnosis for which all the accuracy values are higher than 70%.

When using the combined modalities, on sleep staging, our model using EEG+EOG performs the best, about 1.4% higher than EEG (82.64% v.s. 81.24% in accuracy and 76.98% v.s. 74.85% in Macro-F₁), which indicates that there is additional information that can be used to identify sleep stages in EOG compared with EEG. Other performance of our model on multiple modalities (EEG+EMG, EEG+ECG and EEG+Nasal Pressure) performs worse than EEG on sleep staging (80.16% v.s. 76.95% v.s. 65.86% v.s. 81.24% in accuracy and 74.08% v.s. 69.93% v.s. 56.76% v.s. 74.85% in Macro-F₁). Specifically, our model on EEG+ECG and EEG+Nasal Pressure performs quite poor on sleep staging compared with EEG+EMG. It may be because that ECG and Nasal Pressure may bring about too much irrelevant information for sleep staging and cause significant performance decline. For narcolepsy diagnosis, obviously, our model using the multiple modalities of EEG+EOG achieves the best performance, about 0.8% higher than EEG in accuracy (79.71% v.s. 78.94%), which combining EEG and EOG could bring about useful features for narcolepsy diagnosis. Our model using multiple modalities (EEG+EMG, EEG+ECG and EEG+Nasal Pressure) performs worse than EEG on narcolepsy diagnosis (75.23% v.s. 72.31% v.s. 71.95% v.s. 78.94% in accuracy and 83.49% v.s. 81.95% v.s. 79.57% v.s. 85.45% in F₁). Similarly, combining EMG, ECG or Nasal Pressure with EEG may bring about additional noise for narcolepsy diagnosis.

D. Subject-Level Case Study

In order to give a better understanding of the narcolepsy diagnosis, we selected one patient from our dataset to illustrate her/his hypnogram through the whole night. Here, we present

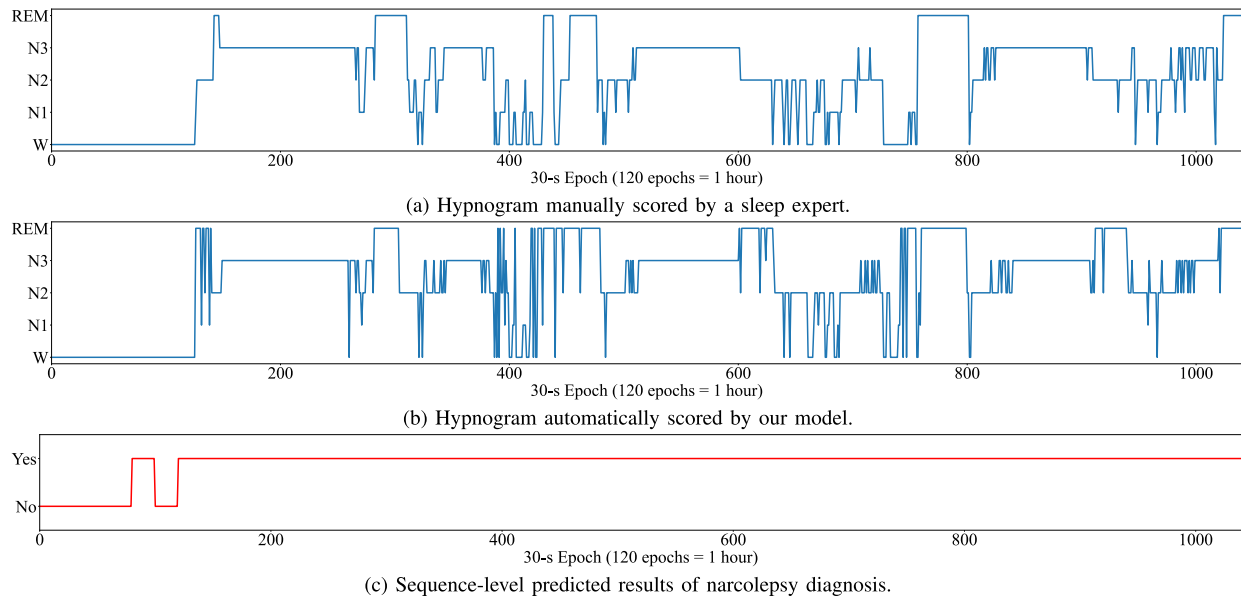


Fig. 7. Case study on one patient. In (c), “Yes” denotes that this sequence is predicted as narcolepsy. “No” denotes that this sequence is not predicted as narcolepsy.

the groundtruth hypnogram and the hypnogram automatically scored by our model of the patient, respectively, shown in Fig. 7(a) and (b). Meanwhile, we present the narcolepsy diagnosis results obtained by our model, shown in Fig. 7(c). It can be seen from Fig. 7(a) and (b), the sleep stages of most epochs of this patient are correctly scored by our model, and only a few epochs are misclassified. It is difficult to correctly score the sleep stages with rapid sleep transitions. The sequential relationship among such sleep fragments is hard to model. As we can see from Fig. 7(c), we can correctly diagnose the narcolepsy for most sequences by our model. In our model, when input a sequence of 20 epochs, it will output a diagnosis result. In this way, for each subject, there are multiple diagnosis results. Therefore, we could determine the diagnosis result in subject-level. Specifically, for each subject, we take all of her/his diagnosis results into account, and if more than 50% of the results determine the subject with narcolepsy, we determine that she/he is with narcolepsy. In this way, our model achieves a 100% in accuracy in subject-level narcolepsy diagnosis. It suggests that we could improve the robustness of our model by taking more diagnosis results in sequence-level.

E. Limitations

We must acknowledge the limitations of the dataset used in this work. First, due to the difficulty in recruiting a large number of patients, the total number of subjects in our dataset was relatively small, 77 in total. All the subjects in our dataset are from China, and the conclusions we obtained were mainly for a Chinese population. Second, types of sleep disorders in our dataset were limited to narcolepsy. There were many other sleeping disorders, such as insomnia, restless legs syndrome, and sleep apnea, that were endangering people’s health. We cannot research on these sleep disorder in our dataset. Finally, there are many challenges on classifying

narcolepsy into fine-grained categories, including type 1 narcolepsy, type 2 narcolepsy, and unspecified narcolepsy. The labels provided for narcolepsy were limited to narcolepsy and normal, without fine-grained categories. In the future, we will continue to study sleep disorders and try to address these challenges.

VI. CONCLUSION

In clinic, it is difficult for doctors correctly and objectively to diagnose narcolepsy. In this paper, we address the problem of diagnosing narcolepsy automatically and objectively using PSG signals. We collected a dataset of PSG recordings from 77 participants. We propose a novel end-to-end framework for narcolepsy diagnosis, which embeds the sequential relationship within each epoch and between epochs in PSG signals, automatically scores the sleep staging, and combines the sleep stage related features with narcolepsy features together for narcolepsy diagnosis. In particular, we adopt the idea of multi-task learning, where we take the sleep staging as the auxiliary task, and take the narcolepsy diagnosis as the primary task. The framework was evaluated on the collected dataset, and the results show that both of the sleep stage features and the end-to-end fashion help diagnose narcolepsy. Moreover, we do a comprehensive analysis on the PSG recordings, including the importance of sleep staging for the diagnosis, highly correlated channels, and the predictive ability of different modality (e.g., EEG, EOG, EMG, and ECG).

REFERENCES

- [1] M. R. Irwin, “Why sleep is important for health: A psychoneuroimmunology perspective,” *Annu. Rev. Psychol.*, vol. 66, pp. 143–172, 2015.
- [2] B. Zhai, I. Perez-Pozuelo, E. A. D. Clifton, J. Palotti, and Y. Guan, “Making sense of sleep: Multimodal sleep stage classification in a large, diverse population using movement and cardiac sensing,” *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 1–33, Jun. 2020.

- [3] V. Chattu, M. Manzar, S. Kumary, D. Burman, D. Spence, and S. Pandi-Perumal, "The global problem of insufficient sleep and its serious public health implications," *Healthcare*, vol. 7, no. 1, p. 1, Dec. 2018.
- [4] B. M. Altevogt et al., *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. Washington, DC, USA: National Academies Press, 2006.
- [5] J. G. Peacock and R. M. Benca, "Narcolepsy: Clinical features, co-morbidities & treatment," *Indian J. Med. Res.*, vol. 131, no. 2, pp. 338–349, 2010.
- [6] (2017). *Narcolepsy Information Page*. [Online]. Available: <https://rarediseases.org/rare-diseases/narcolepsy/>
- [7] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. F. Quan, *The American Academy of Sleep Medicine (AASM) Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*, vol. 1. Westchester, IL, USA: American Academy of Sleep Medicine, 2007.
- [8] 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.
- [9] H. Seo, S. Back, S. Lee, D. Park, T. Kim, and K. Lee, "Intra- and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG," *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 102037.
- [10] A. Supratak and Y. Guo, "TinySleepNet: An efficient deep learning model for sleep stage scoring based on raw single-channel EEG," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 641–644.
- [11] M. Perslev, M. H. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-Time: A fully convolutional network for time series segmentation applied to sleep staging," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 4415–4426.
- [12] W. Qu et al., "A residual based attention model for EEG based sleep staging," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 10, pp. 2833–2843, Oct. 2020.
- [13] E. Eldele et al., "An attention-based deep learning approach for sleep stage classification with single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 809–818, 2021.
- [14] J. Phyo, W. Ko, E. Jeon, and H.-I. Suk, "Enhancing contextual encoding with stage-confusion and stage-transition estimation for EEG-based sleep staging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 1301–1305.
- [15] H. Phan, K. Mikkelsen, O. Y. Chén, P. Koch, A. Mertins, and M. De Vos, "SleepTransformer: Automatic sleep staging with interpretability and uncertainty quantification," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 8, pp. 2456–2467, Aug. 2022.
- [16] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, 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.
- [17] H. Phan, O. Y. Chén, 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.*, vol. 44, no. 9, pp. 5903–5915, Sep. 2022.
- [18] Z. Jia, Y. Lin, J. Wang, X. Wang, P. Xie, and Y. Zhang, "SalientSleepNet: Multimodal salient wave detection network for sleep staging," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, Aug. 2021, pp. 2614–2620.
- [19] Z. Jia et al., "Multi-view spatial-temporal graph convolutional networks with domain generalization for sleep stage classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1977–1986, 2021.
- [20] O. Tsinalis, P. M. Matthews, and Y. Guo, "Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders," *Ann. Biomed. Eng.*, vol. 44, no. 5, pp. 1587–1597, May 2016.
- [21] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, Jun. 2018.
- [22] X. Li, L. Cui, S. Tao, J. Chen, X. Zhang, and G.-Q. Zhang, "HyCLASSS: A hybrid classifier for automatic sleep stage scoring," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 2, pp. 375–385, Mar. 2018.
- [23] H. Ghimatgar, K. Kazemi, M. S. Helfroush, and A. Aarabi, "An automatic single-channel EEG-based sleep stage scoring method based on hidden Markov model," *J. Neurosci. Methods*, vol. 324, Aug. 2019, Art. no. 108320.
- [24] H. Nie, S. Tu, and L. Xu, "RecSleepNet: An automatic sleep staging model based on feature reconstruction," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2021, pp. 1458–1461.
- [25] Z. Yubo, L. Yingying, Z. Bing, Z. Lin, and L. Lei, "MMASleepNet: A multimodal attention network based on electrophysiological signals for automatic sleep staging," *Frontiers Neurosci.*, vol. 16, Aug. 2022, Art. no. 973761.
- [26] W. Wenjian, X. Qian, X. Jun, and H. Zhikun, "DynamicSleepNet: A multi-exit neural network with adaptive inference time for sleep stage classification," *Frontiers Physiol.*, vol. 14, p. 762, May 2023.
- [27] Y. Li, J. Chen, W. Ma, G. Zhao, and X. Fan, "MVF-SleepNet: Multi-view fusion network for sleep stage classification," *IEEE J. Biomed. Health Informat.*, early access, Sep. 21, 2022, doi: [10.1109/JBHI.2022.3208314](https://doi.org/10.1109/JBHI.2022.3208314).
- [28] P. An, J. Zhao, B. Du, W. Zhao, T. Zhang, and Z. Yuan, "Amplitude-time dual-view fused EEG temporal feature learning for automatic sleep staging," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 10, 2022, doi: [10.1109/TNNLS.2022.3210384](https://doi.org/10.1109/TNNLS.2022.3210384).
- [29] X. Chen, Y. Xiao, Y. Tang, J. Fernandez-Mendoza, and G. Cao, "ApneaDetector: Detecting sleep apnea with smartwatches," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 5, no. 2, pp. 59:1–59:22, 2021.
- [30] S. I. Dimitriadis, C. I. Salis, and D. Liparas, "An automatic sleep disorder detection based on EEG cross-frequency coupling and random forest model," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 046064.
- [31] M. Sharma, J. Darji, M. Thakrar, and U. R. Acharya, "Automated identification of sleep disorders using wavelet-based features extracted from electrooculogram and electromyogram signals," *Comput. Biol. Med.*, vol. 143, Apr. 2022, Art. no. 105224.
- [32] M. M. Siddiqui, G. Srivastava, and S. H. Saeed, "Diagnosis of narcolepsy sleep disorder for different stages of sleep using short time frequency analysis of PSD approach applied on EEG signal," in *Proc. Int. Conf. Comput. Techn. Inf. Commun. Technol. (ICCTICT)*, Mar. 2016, pp. 500–508.
- [33] D. Y. Son et al., "Changes in physiological network connectivity of body system in narcolepsy during REM sleep," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104762.
- [34] Y. Zhang et al., "Comparative polysomnography parameters between narcolepsy type 1/type 2 and idiopathic hypersomnia: A systematic review and meta-analysis," *Sleep Med. Rev.*, vol. 63, Jun. 2022, Art. no. 101610.
- [35] Y. H. Um et al., "Differential characteristics of repeated polysomnography and multiple sleep latency test parameters in narcolepsy type 1 and type 2 patients: A longitudinal retrospective study," *Sleep Breathing*, vol. 26, pp. 1939–1946, Dec. 2022.
- [36] J. B. Stephansen et al., "Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy," *Nature Commun.*, vol. 9, no. 1, p. 5229, Dec. 2018.
- [37] M. Crawshaw, "Multi-task learning with deep neural networks: A survey," 2020, *arXiv:2009.09796*.
- [38] Y. Dauvilliers, S. Rompré, J.-F. Gagnon, M. Vendette, D. Petit, and J. Montplaisir, "REM sleep characteristics in narcolepsy and REM sleep behavior disorder," *Sleep*, vol. 30, no. 7, pp. 844–849, Jul. 2007.
- [39] L. Liebel and M. Körner, "Auxiliary tasks in multi-task learning," 2018, *arXiv:1805.06334*.
- [40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [41] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.
- [42] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.
- [43] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," 2019, *arXiv:1912.01703*.
- [44] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [45] L. Prechelt, "Early stopping-but when?" in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 1996, pp. 55–69.
- [46] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Jun. 1998.
- [47] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.