A Developed LSTM-Ladder-Network-Based Model for Sleep Stage Classification

Ruichen Li, Bei Wang, *Member, IEEE*, Tao Zhang, *Senior Member, IEEE*, and Takenao Sugi, *Member, IEEE*

***Abstract*— Sleep staging is crucial for diagnosing sleep-related disorders. The heavy and time-consuming task of manual staging can be released by automatic techniques. However, the automatic staging model would have a relatively poor performance when working on unseen new data due to individual differences. In this research, a developed LSTM-Ladder-Network (LLN) model is proposed for automatic sleep stage classification. Several features are extracted for each epoch and combined with the following epochs to form a cross-epoch vector. The long short-term memory (LSTM) network is added into the basic ladder network (LN) to learn the sequential information of adjacent epochs. The developed model is implemented based on a transductive learning scheme to avoid the issue of accuracy loss caused by individual differences. In this process, the labeled data pre-trains the encoder, and the unlabeled data re-fine the model parameters by minimizing the reconstruction loss. The proposed model is evaluated on the data from public database and hospital. Comparison experiments were conducted where the developed LLN model achieved rather satisfied performance while dealing with the unseen new data. The obtained results demonstrate the effectiveness of the proposed approach in addressing individual differences. This can improve the quality of automatic sleep staging when assessed on different individuals and has strong application potential as a computer aided approach for sleep staging.**

***Index Terms*— Sleep stage, EEG, long short-term memory network, ladder network, transductive learning.**

I. Introduction

SLEEP is the most crucial part of humans' daily physiological activities. Good sleep with suitable circadian is beneficial to maintaining our physical and mental health [1], [2]. However, sleep deprivation increases the risk of potential diseases and the probability of accidents [3], [4]. According to research, around 47% of accidents are caused by people falling asleep while driving. Those who slept for less than 6 hours had about 25% more accidents than those who got enough sleep [5], [6].

Generally, sleep is defined by several sleep stages that are commonly used to interpret the overnight sleep process. The golden standard for sleep staging is the Rechtschaffen and Kales criteria (R&K) [7]. It is further developed by the American Academy of Sleep Medicine (AASM) [8]. Both standards suggest dividing sleep into three primary stages: Wake, Non-Rapid Eye Movement (NREM), and Rapid Eye Movement (REM). NREM is further divided into Stage I (S1), Stage II (S2), Stage III (S3), and Stage IV (S4). S1 and S2 tend to show periods of light sleep, while S3 and S4 represent periods of deep sleep. Additionally, S3 and S4 are considered as one stage of slow wave sleep (SS). A person's overnight sleep process contains regular sleep cycles consisting of those sleep stages, while the sleep structure of patients having diseases may have abnormal characteristics. The sleep of patients with depression often include changes of sleep continuity, and impaired non-REM sleep [9]. The deprivation of REM sleep is linked to behavioral deviation, insomnia, brain shrinkage, and abnormally high rates of neural cell death [10].

Polysomnography (PSG) is primarily used to monitor sleep stages and sleep cycles. The measurement of PSG includes electroencephalograph (EEG), electrocardiograph (ECG), electromyograph (EMG), and electro-oculograph (EOG), etc. Generally, sleep stages are inspected by clinicians with clinical experience and qualified skill. Each 30-second epoch of PSG is demonstrated by a suitable sleep stage. Due to the complexity of sleep signals, it takes several hours for a clinician to mark the overnight PSG recording [11]. Thus, visual inspection is a quite heavy and laborious task. Furthermore, it is rather objective that may be varied among clinicians [12].

Automatic sleep staging has an advantage on processing efficiency compared with visual inspection. The rule-based approaches [13], [14] had been used to realize automatic sleep stage identification, where the features were primarily extracted from the time and frequency domains of neurophysiological signals [15] [16], or by newly developed time-frequency feature representations [17], [18]. Machine learning approaches, such as HMM [19], SVM [20], [21],

Manuscript received 1 April 2022; revised 19 September 2022 and 16 January 2023; accepted 14 February 2023. Date of publication 20 February 2023; date of current version 24 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61773164 and in part by the National Key Research and Development Program of China under Grant 2017YFB13003002. *(Corresponding author: Bei Wang.)*

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

Ruichen Li and Bei Wang are with the Department of Automation, School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China (e-mail: ruichen_li@mail.ecust.edu.cn; beiwang@ecust.edu.cn).

Tao Zhang is with the Department of Automation, Tsinghua University, Beijing 100086, China (e-mail: taozhang@tsinghua.edu.cn).

Takenao Sugi is with the Department of Electrical and Electronic Engineering, Faculty of Science and Engineering, Saga University, Saga 840-8502, Japan (e-mail: sugi@cc.saga-u.ac.jp).

Digital Object Identifier 10.1109/TNSRE.2023.3246478

and Decision Tree [22], [23] were also used to realize the sleep stage classification by combining the extracted features. However, machine learning requires data to be highly separable [24], making it more demanding on the effectiveness of features. Recently, the deep learning model had been widely used for sleep staging. Deep learning approaches extract features automatically for classifiers. As the depth of the layers increases, more detailed features can be learned [25]. With the rapid development of different deep learning models, the effectiveness has been validated compared with the traditional classification approaches [26], [27], [28]. However, the limitation is also obvious. The deep learning models require many training samples and qualified labels, which are almost impractical, especially in clinics. The reasons are not only due to the complex and unknown characteristics of brain waves of EEG, but also the hard work of visual labeling by clinicians. The differences from individual data or recordings may cause the model to perform poorly on the unknown new data. Furthermore, the sleep recordings at the hospital are primarily of patients with sleep disorders. There are more complicated and varied cases compared with the typical definition for normal sleep in the criteria, which would affect the performance of a trained model.

Due to the above limitations, several research focused on the transfer learning techniques for sleep staging to enhance the adaptability of the deep learning model to different data sets. In [29], one database consisting of 200 subjects was used as the source domain and three target domain databases to reduce the data variability issue caused by equipment differences. In [30], eight data sets were used for transfer learning to enhance the model's generalization ability on data from individuals with different sleep disorders. The transfer learning technique employs source domain data to pre-train the model and obtains the prediction results on the unseen target data set. Although the existing public data can better pre-train the model and transfer it to the other data sets, the recent studies still demonstrated the requirement for large-scale training data sets. It is still necessary to develop the deep learning models to meet the actual requirements for clinical application.

In this research, a LSTM-Ladder-Network based on a transductive learning scheme is developed for automatic sleep stage classification dealing with fewer training samples and unknown new data. The aim is to enhance the application performance that is more convenient for clinicians to employ and more adaptive to different individuals. The primary contributions are: (1) Several features are extracted from sleep EEG and EOG of subsequent epochs and combined to form a cross-epoch vector as inputs. The structure of the ladder network is developed, where the sequential information of the changing regularity of sleep stages is learned using LSTM with cross-epoch vectors. (2) A transductive learning scheme is implemented, where the labeled data pre-trains the encoder and the unlabeled new data re-fine the model parameters by minimizing the reconstruction loss. By fully using the distribution structure of the new data, the model parameters are more consistent with the data distribution of the predicted individuals. Finally, the developed model is validated on the sleep data of subjects with different ages from a public database and of patients from the hospital. The classification performance is assessed on the unknown new data that would be flexible for real clinical application.

TABLE I
DESCRIPTION OF SLEEP RECORDINGS

Dataset	Subject information	PSG measurement
I	Healthy subjects Configuration: Bipolar Gender: 10M, 10F Age: 25-97 Amount: 20 recordings	Total channel: 6 EEG: Fpz-Cz, Pz-Oz EOG: EOG Others: EMG, Resp oro-nasal, Temp rectal.
II	Patients with sleep disorders Configuration: Monopolar Gender : unknown Age: unknown Amount: 3 recordings	Total channel: 22 EEG: C3-M2, C4-M1, O1-M2, O2-M1, F3-M2, F4-M1 EOG: E1-M2, E2-M2. Others: Chin EMG, ECG, Snore, Therm, etc.

In section II, the description of two data sets to be analyzed is introduced, the extraction of features and the composition of cross-epoch vectors are described, and the basic classification model and its improvements are explained. In section III, the preparations of experiments are given, several comparisons are conducted to examine the performance of the proposed approach, and the obtained hyponograms of normal subjects and patients are analyzed. The originality and effectiveness are discussed in section IV, and the summary of our work is given in section V.

II. MATERIAL AND METHOD

A. Data Sets

The sleep recordings from two data sets were employed for analysis and evaluation. Table I summarizes the description of sleep recording. (1) Dataset I is from the PhysioNets Sleep-EDF Expanded (Sleep-EDFx) database [31]. The data of each subject were recorded for two consecutive nights. The PSG measurement included two channels of EEG, one channel of EOG, and others. (2) Dataset II contains the overnight sleep recordings from two man-machine versus sleep stage classification competitions that were held in 2021 in Beijing and Shenzhen, China. Each subject had one night of recording. The data are of standard PSG measurement, including 22 channels. EEGs were recorded in the frontal, central, and occipital areas. EOGs were of two channels corresponding to horizontal and vertical eye movements. The other channels were chin-EMG, ECG, Snore, Therm, etc.

Totally 20 subjects from Dataset I (10 males and 10 females, with different ages from 25 to 97 years old, as shown in Table II) are selected and three subjects from Dataset II are used for testing and evaluation. For Dataset I, the two EEGs and one EOG are examined for sleep staging. For Dataset II, two channels are selected from the frontal and occipital EEGs and one channel based on the substraction of two EOGs is used. The sampling rate of EEG and EOG in the Dataset I was 100 Hz, and 256 Hz in Dataset II. Furthermore, the recording of Dataset I contained nearly 20 hours of PSG data. To focus on the overnight sleep staging, the data corresponding to the night sleep period were extracted (ranging from 8 to

TABLE II
THE INFORMATION OF THE 20 SUBJECTS FROM DATASET I

No.	Gender	Age	Epochs	No.	Gender	Age	Epochs
1	Female	25	914	11	Male	28	1177
2	Female	26	1098	12	Male	31	1065
3	Female	26	745	13	Male	31	1120
4	Female	28	1205	14	Male	54	927
5	Female	50	1041	15	Male	54	1094
6	Female	51	1142	16	Male	56	1130
7	Female	60	957	17	Male	67	1176
8	Female	66	1056	18	Male	69	1127
9	Female	85	1136	19	Male	88	1314
10	Female	97	1366	20	Male	95	1109

11 hours among subjects). Both data sets had been inspected by clinicians. The labels of sleep stage scoring include Wake, REM, S1, S2 and SS, where S3 and S4 are treated as a single stage of SS according to AASM standard.

B. Feature Extraction

EEG characteristics primarily describe the definition of sleep stages in the staging criteria, while the other signals are also supplementarily considered. Wake is characterized by dominant α activity (8-13 Hz) and low voltage fast wave of EEG. S1 and S2 are identified as light sleep with low voltage θ activity (2-7 Hz), where the transient characteristic waveforms of the sleep spindle and K-complex appeared in EEG. Deep sleep SS is based on high amplitude slow wave activity δ (0.5-2 Hz). The EEG in REM indicates relatively low amplitude mixed frequency, primarily θ activity and intermittent α activity. With the depth of sleep, the dominant frequency component becomes slower whereas the amplitude becomes higher. In REM, EEG becomes slightly active which may be closer to S1. Furthermore, the eye movements are also decreased from Wake to light sleep and deep sleep. Episodic rapid eye movement in EOG is the primary characteristic for distinguishing REM from the other sleep stages.

According to the above description, different sleep stages are characterized by different dominant rhythms in brain waves measured by EEG. Additionally, the eye movements recorded by EOG are useful to identify REM from NREMs. Here, the continuous recording of EEG and EOG is divided into consecutive 30-second epochs. Features are calculated for each epoch. There are two types of characteristic parameters calculated from the time domain and frequency domain for sleep staging.

1) Time Domain: Three Hjorth parameters of activity, mobility and complexity are computed to obtain the statistical characteristics of the data from EEG channels.

$$\text{Activity}(s) = var(s) \tag{1}$$

$$\text{Mobility}(s) = \sqrt{\frac{var(s')}{var(s)}} \tag{2}$$

$$\text{Complexity}(s) = \frac{\text{Mobility}(s')}{\text{Mobility}(s)} \tag{3}$$

where s represents the input time series, $(\bullet)'$ denotes first order derivative, and $var(\bullet)$ represents the variance calculation. The time domain features are employed to distinguish the intensity of human brain activity by the change in signal frequency, which is gradually decreased from waking to light sleep and deep sleep states.

2) Frequency Domain: The proportion of the energy of a certain frequency activity is examined for EEG and EOG. FFT (Fast Fourier transform) is applied on the time series to obtain the periodogram. Based on the peridogram, the energy of EEG characteristic waveforms are analyzed to extract features in the frequency domain. There are six features of EEG including the frequency activity of δ: 0.5-2 Hz, θ: 2-7 Hz, low amplitude mixing frequency waves (LAMF): 4-7 Hz, α: 8-13 Hz, σ: 12-14 Hz, β: 15-30 Hz. The proportion of the energy of each frequency activity is obtained by comparing it with the total frequency activity of 0.5-30 Hz. Furthermore, one feature of EOG is the energy of frequency activity within 2-10 Hz corresponding to the eye movements. The frequency domain features are fitting to the description of sleep stages in the sleep staging criteria.

C. Cross-Epoch Vector

One's overnight sleep consists of several sleep cycles with the changes of sleep stage from light to deep sleep and back to light sleep. In the sleep staging criteria, for the recorded epoch cannot be directly interpreted, the scoring results of the neighboring epochs are necessary to be considered together. For the same stage, it would have continuity before the appearance of other stages. For the change from current stage to other stages, it would be accord with reasonable transition principles. It is obvious that the inspection of sleep stage requires context information among the recorded epochs.

Generally, several deep learning approaches used the original signal as input. Unlike the other studies, features are used instead of the raw data for sleep stage classification. Each recorded epoch is represented by several characteristic features. In order to include the context information for sleep staging, the features of several epochs are combined to form a cross-epoch vector to integrate the sequential information into the model training.

Fig. 1 illustrates the framework of the developed automatic sleep stage classification approach. Fig. 1 (a) shows the detailed procedures for obtaining the cross-epoch vector. The synchronously and continuously recorded raw EEG and EOG time series are segmented into 30-second epochs without overlapping before feature extraction. Several characteristic features are computed from the time and frequency domains of EEG and EOG, as indicated by x_n. The obtained features of the current epoch combined with the subsequent epochs to form a cross-epoch vector $\widetilde{x}_n$. $\widetilde{x}_n$ is treated as the input for the classification model.

D. Classification Process

The automatic sleep stage classification is implemented using a transductive learning scheme. Unlike the other training schemes, the unseen new data are added to the training process

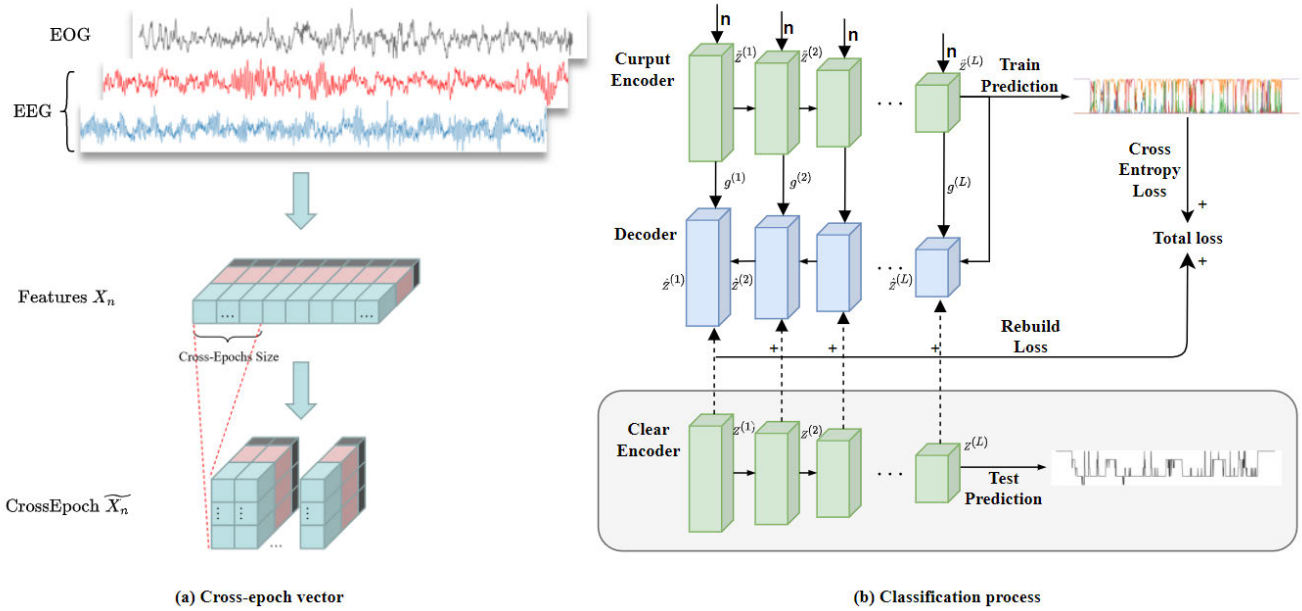

Fig. 1. The framework of the developed automatic sleep stage classification approach. (a) describes the composition of cross-epoch vectors. The features are extracted from EEG and EOG time series for each 30-second epoch. The obtained features of the current epoch are combined with the subsequent epochs to form a cross-epoch vector. (b) illustrates the main structure of the developed model where the inputs are in form of cross-epoch vectors. The classification process is implemented on sleep stage classification based on a transductive semi-supervised learning scheme.

as unlabeled data to reduce the distribution difference between new data and labeled data. Fig. 1 (b) shows the structure of the developed classification model. The ladder network is the basis where the LSTM network is employed to enhance the sleep staging performance by dealing with the sequential characteristics.

1) Ladder Network: The ladder network consists of encoders and decoders. During the training process, the cross-epoch feature vectors of the labeled data $\widetilde{x}_l(n)$ and the unlabeled data $\widetilde{x}_u(n)$ are input to the encoder after adding the Gaussian noise. The corrupted latent variables $\widetilde{z}_l^{(l)}$ of labeled data and $\widetilde{z}_u^{(l)}$ of unlabeled data are obtained from the corrupted encoder.

After the softmax activation layer, the noisy prediction result $\widetilde{y}$ is obtained based on the labeled latent variable $\widetilde{z}_l^{(l)}$. $\widetilde{y}$ is used to compute the loss between the real labels $r(n)$ with the labeled data. The supervised loss function is,

$$\text{Cost}_l = -\frac{1}{N}\sum_{n=1}^{N}\log P\left(\widetilde{y} = r(n)|\widetilde{x}(n)\right) \tag{4}$$

Additionally, the labeled latent variable $\widetilde{z}_l^{(l)}$ is used to train the decoder with the unlabeled latent variable $\widetilde{z}_u^{(l)}$.

The decoder and the encoder correspond one-to-one from high to low, and decode layer by layer to obtain the decoding latent variable $\hat{z}^{(l)}$. The reconstruction loss is calculated by

$$\text{Cost}_u^{(l)} = \frac{\lambda^l}{N}\sum_{n=1}^{N}||\hat{z}^{(l)}(n) - z^{(l)}(n)||^2 \tag{5}$$

where $z^{(l)}(n)$ represents the clean latent variable obtained from the clean encoder, N represents the total number of samples, and λ^l represents the weight of the l-th layer. The total cost is,

$$\text{Cost} = \text{Cost}_l + \sum_{l=1}^{L}\text{Cost}_u^{(l)} \tag{6}$$

The cross-epoch feature vectors of the new data to be identified are input into the clean encoder during the testing process. The prediction results are obtained using softmax. According to the transductive learning scheme, the new data are the unlabeled data to train the model. The model is fine-tuned according to the reconstructing error.

2) LSTM-Ladder-Network: Based on the ladder network, the LSTM network is employed to improve the learning ability to extract sequential information from the input cross-epoch vectors. The encoders and decoders are constructed using the LSTM network to match the input cross-epoch vectors containing inter-stage information.

Each encoder and decoder consists of a LSTM network layer, a batch normalized layer, a parameter layer, and an activation layer. As shown in Fig. 2, the structure of an encoder is illustrated in (a) and a decoder is in (b). To ensure that the input of each LSTM network layer can maintain the same distribution, a batch normalized layer is added after the LSTM network layer. In Fig. 2 (a), $\gamma^{(l)}$ and β^l represent the scale parameter and the bias parameter respectively, which are selected depending on the activation function. The two parameters are both used for tanh and softmax activation functions. In Fig. 2 (b), $g(\bullet, \bullet)$ represents the denoising function under the condition of Gaussian distribution.

The output of the first two encoders is a three-dimensional vector, including the number of samples, the duration of cross-epoch, and the number of features. Tanh is used in the activation layer to make the model converge faster. The

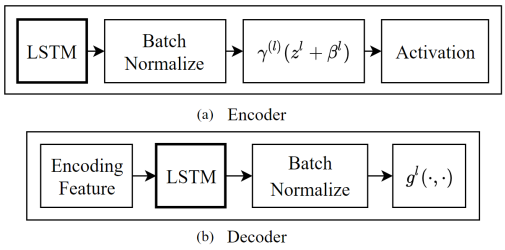

Fig. 2. The structures of one encoder unit and one decoder unit. The developed classification model consists of three encoders and decoders. The first two encoders' activation layers use tanh. The last encoder extracts the coding features and takes the first duration dimension which activation layer employs softmax for classification. The encoded features obtained by the last encoder are input into the last decoder, and decoded in reverse order.

last encoder further extracts the first epoch data from the input three-dimensional vector, i.e. a two-dimensional vector of current epoch for sleep staging. The final activation layer employs the softmax layer to obtain the classification results with the two-dimensional vector.

III. Results

Experimental works are conducted by applying the proposed approach to the two data sets. The obtained results are compared with the visual inspection. First, the effectiveness of extracted features for sleep staging is validated comparing with the original raw time series. Second, the improvements of the LSTM-Ladder-Network are assessed by comparing it with the basic ladder network, and the effectiveness of the transductive learning scheme is investigated by comparing the cases with and without unlabeled data. Third, the classification performance of proposed model is compared with several models in recent literatures. Finally, the hyponogram is obtained to interpret the overnight sleep process of normal subjects and patients with sleep disorders.

A. Experimental Description

For Dataset I, there are 20 sleep recordings. The sleep recordings are divided into consecutive 30-second epochs for sleep staging. The number of epochs of one's overnight sleep recording ranges from 745 to 1,366. The total number of epochs of 20 sleep recordings is 21,849. The computed features for sleep staging include six frequency domain features and three time domain features extracted from two EEG channels, respectively, and one frequency domain feature from one EOG channel, with a total of 19 features. The leave-one-out cross-validation is conducted to evaluate the classification performance on individual data. Each time, the sleep recordings of 19 subjects are employed as labeled data while the remaining one is used as unlabeled data for the test. It is repeated 20 times. For Dataset II, similar processing is conducted. The models in the following comparison experiments are build using Tensorflow 2.2 and trained on a AMD Ryzen 7 5800H 16 cores CPU (with no GPU).

B. Performance Indicator

The classification performance on the sleep staging task is evaluated based on several indicators, including overall accuracy (ACC), precision (PR), recall (RE), and F1-score (F1).

TABLE III
Comparison Between Raw Data and Extracted Features for Sleep Stage Classification

LSMT inputs	Training	Total time	Accuracy
Raw data with CNN	120 round	60s*120=2h	80%
Extracted features	50 round	2s*50=1.6min	81%

The confusion matrix is employed to show the comparison between the automatic sleep staging and the visual inspection. The indicators are computed based on the confusion matrix,

$$\mathrm{ACC} = \frac{1}{N}\sum_{m=1}^{M} \mathrm{TP}_m \tag{7}$$

$$\mathrm{PR}_m = \frac{\mathrm{TP}_m}{\mathrm{TP}_m + \mathrm{FP}_m} \tag{8}$$

$$\mathrm{RE}_m = \frac{\mathrm{TP}_m}{N_m} \tag{9}$$

$$\mathrm{F1}_m = \frac{2 \times \mathrm{PR}_m \times \mathrm{RE}_m}{\mathrm{PR}_m + \mathrm{RE}_m} \tag{10}$$

where TP_m and FP_m are the true positive and false positive of one sleep stage, M is the total number of sleep stages for classification, N_m is the number of epochs of each sleep stage, and N is the total amount of epochs for sleep staging.

C. Features for Sleep Staging

A comparison experiment between the features and raw data is conducted in order to evaluate the feature extraction for sleep staging. The original sleep EEG and EOG time series are processed by a 4-layers CNN (Convolutional Neural Network) [32], and input into a LSTM network for sleep staging. The number of units of LSTM is 128. On the other hand, the extracted features by section II-B are directly input into the representation layer of the same LSTM network.

The RMSProp optimization algorithm is used to adjust the network weight. The learning rate is 0.0001, and the decay rate is 0.9. The model training batch is 128. The number of training rounds is 300. The training is stopped early when the accuracy rate does not increase within recent 5 rounds. The leave-one-out cross-validation is utilized. Each time, the 19 individuals is used for training and 1 individual for testing. The comparison results are given in Table III. By using the original data with CNN units, it tended to be stable after about 120 rounds of training, the total time was 2 hours, and the averaged classification accuracy was 80%. By using the extracted features, it tended to be stable after about 50 rounds of training, the total time was 1.6 minutes, and the averaged classification accuracy was 81%.

The utilized 4-layer CNN unit can extract the features from the original time series automatically. The raw data with CNN obtained rather good classification accuracy. However, the whole sleep staging model is more complex where large amounts of parameters need to be trained, and more data is required to have a better training. It increased the cost for model training. By using our method, the extracted feature from time domain and frequency domain achieved better

classification accuracy and the training cost can be reduced significantly. The comparison experiment demonstrated the effectiveness of extracted features for sleep staging.

D. Sleep Stage Classification

The main developments of proposed model are evaluated based on the following conditions. The input data of the LSTM-Ladder-Network is composed of labeled data and unlabeled data. The data size is (sample size, 6, 19), where the first dimension is the sample number, the second is of cross-epoch size and the third is of feature number. The model is trained in the batch training approach with a batch size of 32. The Gaussian noise variance is 0.01, and the learning rate is fixed at 0.05. The unit sizes of the three-layer LSTM are 100, 50, and 10 respectively. 300 training times are conducted, and the total loss is employed to evaluate the training model. An early stop is used to terminate the training process when the loss is not further decreased within recent 20 rounds.

1) Improvements of Model Structure: To validate the improvements of the LSTM-Ladder-Network (LLN), the classification performance is compared with the basic ladder network (LN). Here, LN is the traditional ladder network where the inputs are single-epoch feature vectors while LLN corresponds to the developed network where the inputs are cross-epoch feature vectors. Both approaches are applied on the sleep recordings of Dataset I.

The confusion matrix is obtained by comparing the automatic classification result with the visual inspection. The performance indicators are computed based on the obtained confusion matrices by using two models, as shown in Table IV and V respectively. The numbers on the diagonal of the confusion matrix represent the number of epochs where the predicted results are consistent with the visual inspection. Among the performance indicators, the F1-score is a comprehensive indicator based on precision and recall.

By employing the basic LN with a single-epoch feature vector, the F1-score of S2 and Wake were better than the other sleep stages, REM was about 70% while SS was lower, and S1 was only 23.4%. The evaluation results in Table V showed that the classification performance of each sleep stage was enhanced significantly compared with Table IV. The F1-score of almost all of the stages were near and higher than 80%. Although S1 was still lower than the other sleep stages, it increased by about 16%. The total classification accuracy using the proposed LLN reached 81.5% while the basic LN was 74.5%. The comparison results showed that the improvements in the model structure obtained more reasonable classification results for sleep staging.

2) Improvements of Model Training: The developed LLN is implemented on Dataset I with and without unlabeled data, respectively, to demonstrate the effect of unlabeled data on the model training. Table V corresponds to the training by only using the labeled data, while Table VI is by using the labeled and unlabeled data.

Comparing the confusion matrices in Table V and VI, the classification performance of every sleep stage has been further improved. The indicators in Table VI revealed that the classification performance of sleep stages was better than that in Table V. The F1-score of Wake and S2 were around 90%, SS and REM were over 80%, and S1 was further improved by 1.6%. The total classification accuracy was achieved at 84.3%. The improvements of using the unlabeled data on model training is effective to implement the proposed model for the unseen new data.

The detail classification result of individual data is investigated in Fig. 3. The accuracies of using the developed approach by the two cases of with and without unlabeled data are compared for each subject. It can be observed that the classification accuracy was remarkably improved using the transductive learning scheme in most of the subjects, while slightly lower in the five subjects. The overall classification performance of the developed LLN under a transduction learning scheme was rather satisfied on sleep staging.

TABLE IV
CONFUSION MATRIX FOR RESULTS OF LN SLEEP STAGING MODEL

LN	Predicted					Performance		
	Wake	REM	S1	S2	SS	RE/%	PR/%	F1/%
Wake	**5098**	226	356	180	65	90.2	86.0	88.1
REM	101	**2387**	175	640	2	65.7	72.2	68.8
S1	304	429	**328**	706	5	31.8	18.5	23.4
S2	142	587	170	**7635**	444	75.3	85.0	79.9
SS	10	2	1	982	**627**	54.9	38.7	45.4

*The total classification accuracy is 74.5%.

TABLE V
CONFUSION MATRIX FOR RESULTS OF LLN SLEEP STAGING MODEL WITHOUT UNLABELED DATA

LLN	Predicted					Performance		
	Wake	REM	S1	S2	SS	RE/%	PR/%	F1/%
Wake	**5413**	143	461	106	72	93.2	87.4	90.2
REM	55	**2745**	222	398	1	80.4	80.2	80.3
S1	266	224	**706**	631	1	40.0	38.6	39.3
S2	72	302	372	**7824**	211	82.0	89.1	85.4
SS	4	2	5	585	**989**	77.6	62.4	69.2

*The total classification accuracy is 81.5%.

TABLE VI
CONFUSION MATRIX FOR RESULTS OF LLN SLEEP STAGING MODEL WITH UNLABELED DATA

LLN	Predicted					Performance		
	Wake	REM	S1	S2	SS	RE/%	PR/%	F1/%
Wake	**5541**	108	334	73	5	92.6	91.4	92.0
REM	40	**2820**	221	282	1	80.8	83.8	82.3
S1	354	254	**722**	581	6	44.8	37.7	40.9
S2	43	309	332	**8218**	170	86.5	90.6	88.5
SS	6	1	3	348	**1127**	86.1	75.9	80.7

*The total classification accuracy is 84.3%.

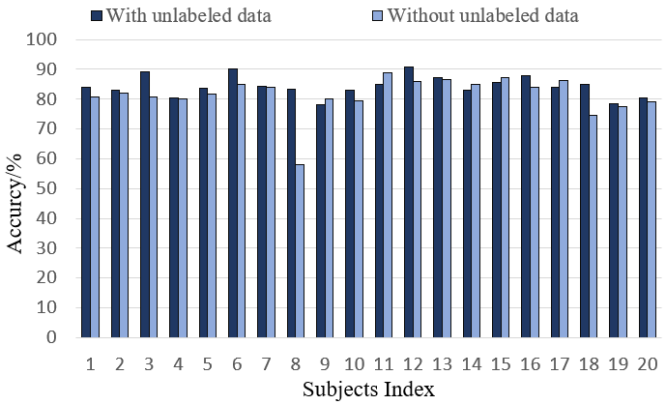

Fig. 3. Accuracy of sleep stage classification for each object when the model is trained with or without the unlabeled data.

3) Comparison With Other Methods: Several models developed in recent literatures are compared with our proposed model. Table VII summarizes the comparison results, where the other approaches are categorized into types of Supervised Learning (SL), Transfer Learning (TL) and Semi-supervised Learning (SSL). The data set is the same as the public sleep-EDF database. The classification performance of three indexes of ACC, Macro-F1-score (MF1) and Kappa coefficient are compared, and the cross-validation way is also described (k-fold or leave-one-out).

The SL models were commonly employed for sleep staging. The obtained sleep staging results in references [26], [27], [28], and [33], were over an accuracy of 82%, MF1 of 74% and kappa coefficient of 0.7. In [28], the presented model achieved rather high classification performance using k-fold cross-validation. However, the performance is decreased obviously when facing the new data using the leave-one-out cross-validation. The individual differences affected the classification performance under SL.

The application of the TL scheme can be found in the recent literature of references [29] and [30]. The classification performance is better than the SL scheme under leave-one-out cross-validation. However, the model still needs to be pre-trained by many labeled training data. The SSL scheme is another way to deal with such problem. It seems that the classification performance of references [34] and [35] were lower than SL and TL even though the k-fold cross-validation was employed.

Our proposed model is developed and implemented based on a transductive learning scheme to avoid the problem of accuracy loss caused by individual differences. The classification performance is examined on the new data by leave-one-out cross-validation. Furthermore, the subjects were of different ages from 25 to 97 years old both males and females. According to the comparison results in Table VII, the overall classification accuracy and MF1 were satisfied and the kappa coefficient was maintained at a good level. The developed model was more effective in meeting the requirements for clinical application.

Additionally, for the issue of computational complexity, the parameters of proposed model to be trained are mainly

TABLE VII
PERFORMANCE COMPARISON WITH OTHER SLEEP STAGING METHODS

Methods	Subjects	ACC(%)	MF1(%)	Kappa	Cross-validation
SL: Ref. [26]	20	83.2	74.7	0.76	leave-one-out
SL: Ref. [27]	20	82.0	76.9	0.76	leave-one-out
SL: Ref. [28]	8	91.2	91.6	0.86	k-fold
		82.4	75.1	0.72	leave-one-out
SL: Ref. [33]	20	82.3	74.7	0.75	leave-one-out
TL: Ref. [29]	20	84.6	79.0	0.78	leave-one-out
TL: Ref. [30]	20	-	81.7	-	leave-one-out
SSL: Ref. [34]	78	80.0	67.6	0.72	k-fold
SSL: Ref. [35]	20	80.1	-	-	k-fold
Proposed model	20	84.3	77.0	0.78	leave-one-out

SL: Supervised Learning; TL: Transfer Learning;
SSL: Semi-supervised Learning.

in the LSTM layer in the encoder and decoder and the fully connected layer of the final output. Here, a 3-layer encoder and decoder structure is adopted for sleep staging. The total number of parameters for training is 167,090, which is extremely fewer than the several million parameters of the models in literatures with multi-layer CNN. When conducting the comparison experiments, each time 19 individuals (about 20,000 epochs) were trained, and 1 individual (about 1,000 epochs) was used for testing. Under the hardware configuration given in section III-A, the time for training and prediction process was of an average 12 minutes. The computational complexity of proposed model is reasonable, and the training and prediction process are efficient.

E. Hypnograms

1) Normal Subject From Dataset I: According to the sleep staging results, the overnight sleep process can be interpreted by using the hyponogram. Fig. 4 shows the hyponograms of one subject from Dataset I. Fig. 4 (a) is by using the basic LN while Fig. 4 (b) is by the developed LLN. In these figures, the black solid line is the visual inspection (VI) and the red dotted line is the automatic sleep staging results. The sleep staging results by using the proposed method are closer to the visual inspection comparing the obtained two hyponograms. The overall accuracy was 83.2% by using the basic LN while improved to 90.7% by using the developed LLN. The enhancements in model structure and training scheme are effective.

2) Patient From Dataset II: Furthermore, the sleep recordings of patients from Dataset II are examined. Fig. 5 shows the obtained hyponograms of one patient. The identification of sleep stages is analyzed into three types of categories. Fig. 5 (a) shows the interpretation of wake from the other sleep stages. The classification accuracy of the two states reached 98.2%. Fig. 5 (b) illustrates the identification result of Wake, REM, and NREM states, where the classification accuracy achieved 90.3%. As shown in Fig. 5 (c), the changes in sleep states, including Wake, REM, S1, S2 and SS are

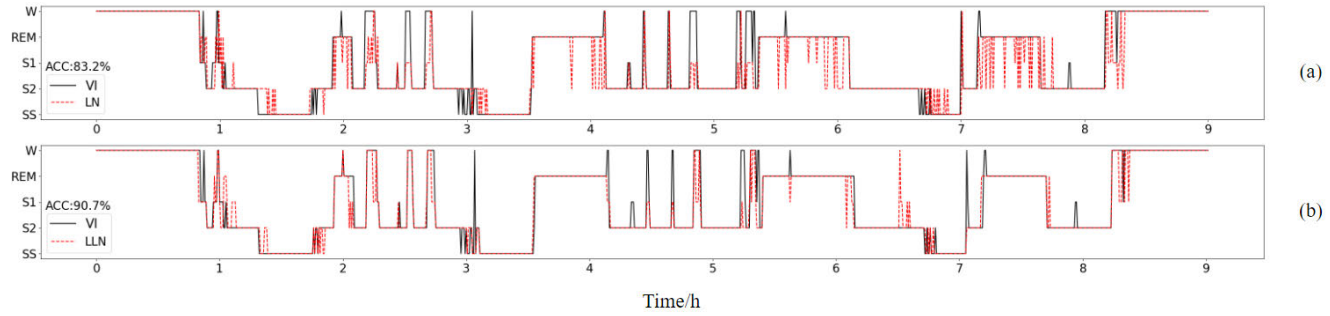

Fig. 4. Hyponograms of one normal subject from Dataset I. (a) the comparison result of sleep staging between the basic LN model with single-epoch feature vector and the visual inspection, (b) the comparison result of sleep staging between the developed LLN model with cross-epoch feature vector and the visual inspection.

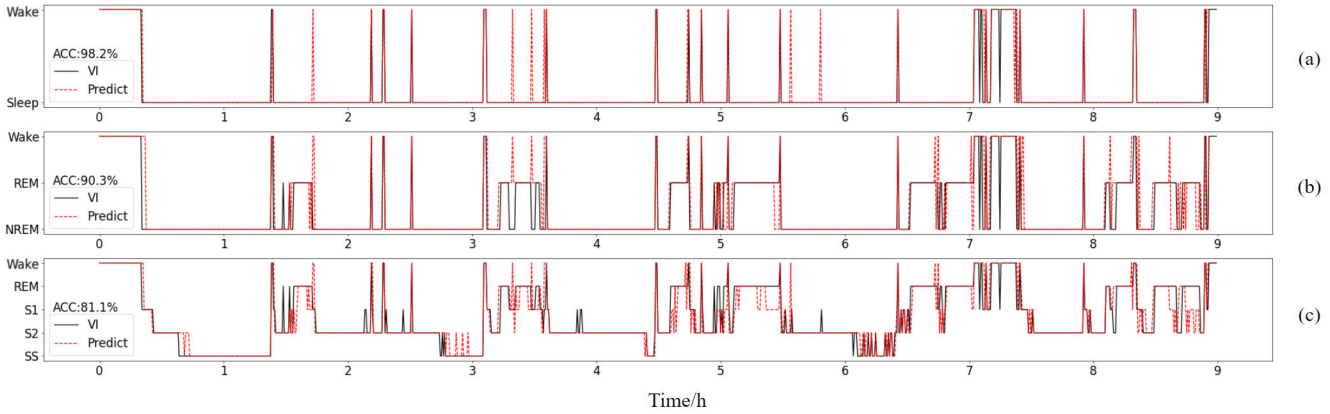

Fig. 5. Hyponograms of one patient from Dataset II. The automatic sleep stage prediction results are compared with the visual inspection by three types of categories: (a) the interpretation results of wake and sleep, (b) the identification results of Wake, REM and NREM, and (c) the classification results of Wake, REM, S1, S2 and SS.

illustrated in the hyponogram. The automatic sleep staging result was quite consistent with the visual inspection, where the accuracy was 81.1%. The experiments revealed that the developed LLN under a transduction learning scheme presented a feasible assistant tool for clinical application.

IV. Discussion

A. Individual Differences

Sleep stage scoring is useful and crucial for describing the change in sleep states during one's overnight sleep. In clinical practice, the data to be interpreted often come from new subjects. We investigated on the sleep data of different individuals. Fig. 6 shows the distributions of the density of four main EEG components including δ, θ, LAMF and α. Fig. 6 (a) illustrates the distribution of each parameter computed from the two sleep recordings of the same subject. The consistency of feature distributions from the same individual was relatively high. However, Fig. 6 (b) shows the feature distributions of the two different subjects. There were quite obvious differences between the two individuals compared with Fig. 6 (a). The influence of individual factors such as age, gender, and physical fitness of different subjects will result in differences in the feature distribution.

There is primary two kinds of method commonly utilized for neurophysiological signal processing. One is subject-dependent method, i.e. to establish a model for each subject to avoid the individual difference effect. The typical application fields include brain computer interface and emotion recognition [36], [37]. In those applications, the EEG signals were examined under the designed experimental environment. It is possible to obtain more recordings from the same subject to train the classification model for individuals by repeating the experiment. Different from those applications, it would be difficult to obtain enough sleep recording for individual training. The subject-dependent method may not be applicable to deal with the sleep staging task. Another is subject-independent method focusing on training a classification model to address the variations in EEG signals among different subjects. It is commonly utilized for sleep stage classification. Recently, studies using transfer learning scheme and un/semi-supervised learning schemes were applied for sleep staging trying to avoid the individual difference effect on automatic sleep staging models. However, large number of training samples with manual labeling is required and classification performances on the new data still need to be developed.

By considering the effect of individual difference, the aim of our research is to develop the automatic sleep staging model dealing with fewer training samples and unknown new data for clinical application. During the training process, the labeled training sample and the unlabeled test samples are used to overcome the individual difference effect. Although the proposed approach is subject-independent, it needs not large number of labeled training samples. Furthermore, the

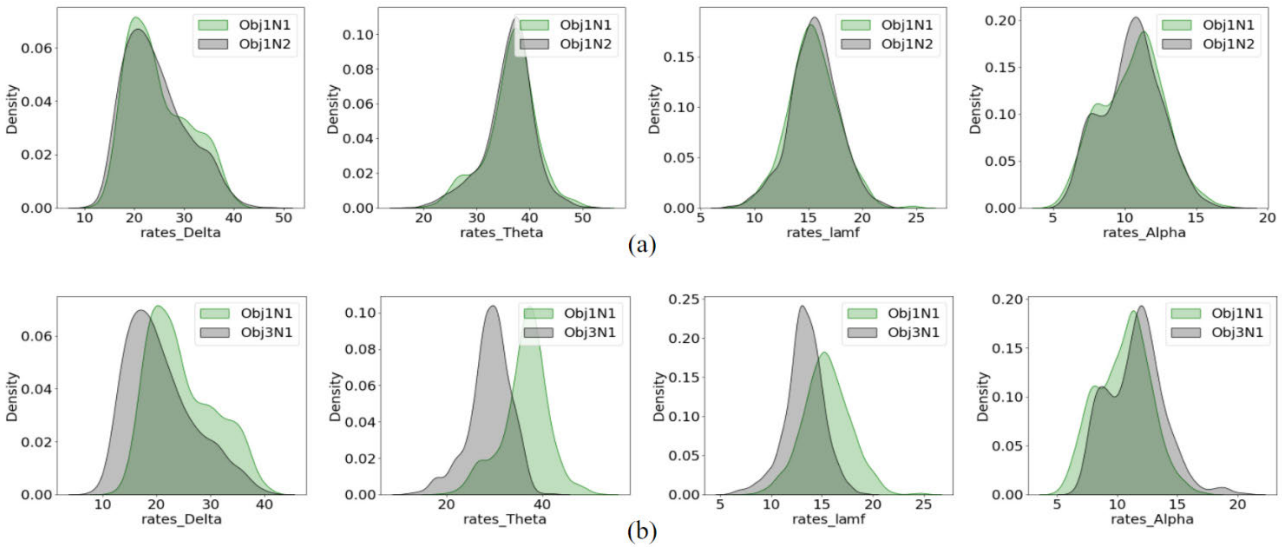

Fig. 6. The distributions of features. (a) is from the two sleep recordings of the same subject, and (b) is from the sleep recordings of two different subjects.

training process has the similar role as the subject-dependent method even there is few individual data for each subject. These demonstrate that our method would be more feasible to deal with the issue of individual differences for sleep staging.

B. Developed LSTM-Ladder-Network

Automatic sleep stage classification had been investigated by various models. In most of the studies, the supervised model is adopted and validated by k-fold cross-validation. Generally, the k-fold cross-validation is performed under the assumption of consistent data distribution. During the validation, part of the sleep data from all individuals is employed to train the models. No matter which individual's data are predicted, some of his/her data have been trained to meet the requirement of distribution consistency. Although better results can be obtained, the trained models may have poor performance when applied to the unseen new data. In real applications, consistent data distribution is almost impossible. Poor data consistency would be difficult to train a model and achieve good results when there are fewer labeled data related to various actual cases. Recently, studies using TL scheme were applied for sleep staging which achieved better classification performance than SL and SSL. The TL models employs source domain data to pre-train the model and obtains the prediction results on the unseen target data set. However, the cost for the acquisition of the data with expert labeling is rather high and sometimes difficult to solve.

Due to the above limitations, we investigated a classification model under a transductive learning scheme. The transductive learning is a kind of SSL scheme. In our experimental works, one individual sleep recording is treated as unlabeled data while the others are labeled data. Features are extracted from the EEG and EOG. The developed model is realized by a transductive learning scheme through the encoder-decoder structure, where LSTM is employed to build the encoder and decoder dealing with the input cross-epoch feature vectors. By training the encoder with labeled and unlabeled data, the reconstruction error is minimized after decoding to avoid large distribution differences between the unlabeled new data and labeled data.

The classification performance is examined on the new data by using leave-one-out cross-validation. Furthermore, the subjects are of different ages from 25 to 97 years old both males and females (as shown in Table II). The results shown in Table VI and Fig. 3 proved the effectiveness of our approach for sleep stage classification based on fewer training samples for unseen new data. The comparison with other models is also given in Table VII. Comparing with SL and TL, our model achieved a good classification level which needs not require large number of training data with expert labeling. Comparing with the other SSL models, the classification performance is significantly improved. The developed model is more effective in meeting the requirements for clinical application.

C. Cross-Epoch Vector

Sleep recording has sequential characteristics where the sleep states are changed based on certain regularities. Most of the studies used LSTM trying to improve the learning ability of sequential information within one epoch [26], [38]. Some studies used several epochs to enlarge the duration of time series for learning [27]. However, the recognition of sleep stages is for epochs. To some extent, the change in features is more understandable than the original time series for learning the sequential information.

We considered the sequential information of the continuity and transfer of sleep stage among the epochs rather than the time series within the epochs. Firstly, several typical features are computed instead of the common convolutional neural network. According to the comparison results in Table III, the model's structure can be simplified and the computational costs can be reduced by using the defined features rather than the raw data. Moreover, the extracted features also achieved

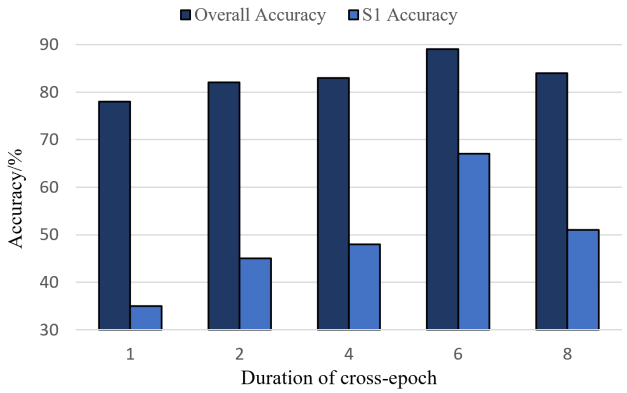

Fig. 7. The sleep staging accuracy using the developed model under a different duration of the cross-epoch feature vector.

fairly well classification accuracy. Secondly, the features of the current epoch are combined with subsequent epochs to form a cross-epoch vector as the input. The representation of cross-epoch makes better use of LSTM to learn the sequential information throughout the recorded epochs as described in the sleep staging criteria. Accordingly, the learning of the transfer rule in consecutive epochs can further improve the classification performance for automatic sleep staging as shown in Table VI. Even without using unlabeled data to fine-tune the model, the LSTM encoding and decoding structure with cross-epoch vectors still achieved better classification performance comparing the evaluation results in Table V with Table IV.

The number of epochs used to form the cross-epoch vector is also investigated. The same experimental test is repeated where the duration of the cross-epoch is increased by 1, 2, 4, 6, and 8. Fig. 7 shows the obtained results. The duration equals 1 implies that the model does not use the subsequent epochs. As shown in Fig. 7, the proposed model achieved the best classification accuracy when the duration of the cross-epoch was 6. Moreover, the identification of light sleep S1 is also analyzed. S1 is a transient state from wake to sleep. The characteristics of S1 are quite similar to the adjacent states of Wake and S2. It is rather difficult to achieve high accuracy on S1. The discrimination of S1 must refer to the characteristics of the consecutive epochs. The accuracy of S1 was increased when considering more context, as illustrated in Fig. 7. The cross-epoch is beneficial for enhancing the recognition of the transient state as S1. According to our research, the number of epochs to form the cross-epoch vectors can be adjusted within a certain range. We used the six consecutive epochs. The constructed model was well-behaved on the unseen new individual data as evaluated on the normal subjects and patients of the two data sets.

D. Clinical Data Applications

Besides evaluating the sleep recordings of normal subjects from Dataset I, the patients with sleep disorders from Dataset II are tested and examined, as shown in Fig. 5. We separated the predicted results by the proposed approach into three situations. First, the identification of wake from sleep was shown to recognize the wake stage during the sleep process. Second, the results of Wake, REM and NREM were given to illustrate the rough information related to sleep cycles. Third, the detailed classification of five sleep stages was all given for the interpretation of one's overnight sleep process. According to Fig. 5, the identification results by our proposed approach were rather satisfied closer to the visual inspection. The descriptions of "Wake and Sleep", "Wake, REM and NREM", and "all sleep stages" may offer useful identification information for the diagnosis and treatment of sleep-related disorders such as apnea syndrome, and insomnia. Additionally, the TL approach has a similar consideration of unseen new data that is concentrated on the differences between the source domain and target domain such as different hospitals or institutions. In the case of the application in the hospital, the individual differences would be more crucial to be considered when developing the automatic classification approach. Unlike the TL technique, our proposed approach needs not rely on large-scale training data. The transductive learning scheme revealed effectiveness to make sleep stage classification on the unseen new data. Furthermore, the classification effectiveness was validated on the sleep recordings of subjects whose age was distributed with large differences. This would be more feasible and practical to be implemented for clinical application.

V. Conclusion

An automatic sleep stage classification approach was developed. The influence of individual differences is considered and investigated. Several characteristic features are extracted from the time and frequency domains of sleep EEG and EOG for each 30-second epoch. The current epoch and several subsequent epochs are connected to form a cross-epoch vector as input. The potential features of new data are learned by encoding and decoding the structure of ladder network. The LSTM network is employed to learn the time sequence relationship of sleep features between epochs. The classification model is rematched to the new data by minimizing the reconstruction loss. The proposed approach's effectiveness is assessed by comparison experiments and applied to the sleep recordings of normal subjects and patients with sleep disorders from two data sets. The obtained results revealed that the developed LLN based on fewer training samples under a transductive learning scheme had a rather satisfactory classification performance on the unseen new data. Moreover, it is also possible to adopt more advanced time-frequency feature extraction techniques for feature representation in current model as future works. This can be a feasible assistant tool for clinical application.

References

[1] C. M. Shapiro, R. Bortz, D. Mitchell, P. Bartel, and P. Jooste, "Slow-wave sleep: A recovery period after exercise," *Science*, vol. 214, no. 4526, pp. 1253–1254, Dec. 1981.

[2] C. M. Swanson et al., "The importance of the circadian system & sleep for bone health," *Metabolism*, vol. 84, pp. 28–43, Jul. 2018.

[3] F. S. Luyster, P. J. Strollo, P. C. Zee, and J. K. Walsh, "Sleep: A health imperative," *Sleep*, vol. 35, no. 6, pp. 727–734, Jun. 2012.

[4] J. F. S. Junior, A. L. Eckeli, C. C. C. Ribeiro, R. F. L. Batista, A. A. M. da Silva, and C. M. C. Alves, "Influence of excessive daily sleeping and sleep quality on BDNF and NGF serum levels in adolescents," *Sleep Med.*, vol. 84, pp. 415–423, Aug. 2021.

[5] J. Kalsi, T. Tervo, A. Bachour, and M. Partinen, "Sleep versus non-sleep-related fatal road accidents," *Sleep Med.*, vol. 51, pp. 148–152, Nov. 2018.

[6] T. Ohida et al., "The influence of lifestyle and health status factors on sleep loss among the Japanese general population," *Sleep*, vol. 24, no. 3, pp. 333–338, May 2001.

[7] T. Hori et al., "Proposed supplements and amendments to 'a manual of standardized terminology, techniques and scoring system for sleep stages of human subjects', the rechtschaffen & Kales (1968) standard," *Psychiatry Clin. Neurosciences*, vol. 55, no. 3, pp. 305–310, Jun. 2001.

[8] R. B. Berry et al., "Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events: Deliberations of the sleep apnea definitions task force of the American academy of sleep medicine," *J. Clin. Sleep Med.*, vol. 8, no. 5, pp. 597–619, Oct. 2012.

[9] C. Gottesmann and I. Gottesman, "The neurobiological characteristics of rapid eye movement (REM) sleep are candidate endophenotypes of depression, schizophrenia, mental retardation and dementia," *Prog. Neurobiol.*, vol. 81, no. 4, pp. 237–250, Mar. 2007.

[10] L. Yang, Z. Huang, J. Zhong, S. Qiu, and D. Liu, "The characteristics and clinical significance of REM< 10% in children with sleep-disordered breathing," *Int. J. Pediatric Otorhinolaryngology*, vol. 110, pp. 53–56, Jul. 2018.

[11] S. Özşen, "Classification of sleep stages using class-dependent sequential feature selection and artificial neural network," *Neural Comput. Appl.*, vol. 23, no. 5, pp. 1239–1250, Oct. 2013.

[12] R. S. Rosenberg and S. Van Hout, "The American academy of sleep medicine inter-scorer reliability program: Respiratory events," *J. Clin. Sleep Med.*, vol. 10, no. 4, pp. 447–454, Apr. 2014.

[13] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, and Y.-S. Cheng, "A rule-based automatic sleep staging method," *J. Neurosci. Methods*, vol. 205, no. 1, pp. 169–176, 2012.

[14] H.-J. Park, J.-S. Oh, D.-U. Jeong, and K.-S. Park, "Automated sleep stage scoring using hybrid rule- and case-based reasoning," *Comput. Biomed. Res.*, vol. 33, no. 5, pp. 330–349, Oct. 2000.

[15] S. Khalighi, T. Sousa, G. Pires, and U. Nunes, "Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels," *Exp. Syst. Appl.*, vol. 40, no. 17, pp. 7046–7059, Dec. 2013.

[16] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2959–2978, Oct. 2017.

[17] V. Gupta and R. B. Pachori, "FBDM based time-frequency representation for sleep stages classification using EEG signals," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102265.

[18] V. Bajaj and R. B. Pachori, "Automatic classification of sleep stages based on the time-frequency image of EEG signals," *Comput. Methods Programs Biomed.*, vol. 112, no. 3, pp. 320–328, 2013.

[19] D. Jiang, Y.-N. Lu, Y. Ma, and Y. Wang, "Robust sleep stage classification with single-channel EEG signals using multimodal decomposition and HMM-based refinement," *Exp. Syst. Appl.*, vol. 121, pp. 188–203, May 2019.

[20] E. Alickovic and A. Subasi, "Ensemble SVM method for automatic sleep stage classification," *IEEE Instrum. Meas.*, vol. 67, no. 6, pp. 1258–1265, Jun. 2018.

[21] A. Jameer Basha, B. Saravana Balaji, S. Poornima, M. Prathilothamai, and K. Venkatachalam, "Support vector machine and simple recurrent network based automatic sleep stage classification of fuzzy kernel," *J. Ambient Intell. Humanized Comput.*, pp. 6189–6197, Jul. 2022.

[22] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," *Knowl.-Based Syst.*, vol. 128, pp. 115–124, Jul. 2017.

[23] T. Lajnef et al., "Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines," *J. Neurosci. Methods*, vol. 250, pp. 94–105, Nov. 2015.

[24] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, pp. 1–16, 2016.

[25] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[26] B. Yang, X. Zhu, Y. Liu, and H. Liu, "A single-channel EEG based automatic sleep stage classification method leveraging deep one-dimensional convolutional neural network and hidden Markov model," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102581.

[27] 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.

[28] J. Zhou et al., "Automatic sleep stage classification with single channel EEG signal based on two-layer stacked ensemble model," *IEEE Access*, vol. 8, pp. 57283–57297, 2020.

[29] H. Phan et al., "Towards more accurate automatic sleep staging via deep transfer learning," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 6, pp. 1787–1798, Jun. 2021.

[30] A. Guillot and V. Thorey, "RobustSleepNet: Transfer learning for automated sleep staging at scale," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1441–1451, 2021.

[31] A. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[32] Q. Li, B. Wang, J. Jin, and X. Wang, "Comparison of CNN-uni-LSTM and CNN-Bi-LSTM based on single-channel EEG for sleep staging," in *Proc. 5th Int. Conf. Intell. Informat. Biomed. Sci. (ICIIBMS)*, Nov. 2020, pp. 76–80.

[33] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "Joint classification and prediction CNN framework for automatic sleep stage classification," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 5, pp. 1285–1296, May 2019.

[34] B. Haoran and L. Guanze, "Semi-supervised end-to-end automatic sleep stage classification based on pseudo-label," in *Proc. IEEE Int. Conf. Power Electron., Comput. Appl. (ICPECA)*, Jan. 2021, pp. 83–87.

[35] X. Wuzheng, S. Zuo, L. Yao, and X. Zhao, "Semi-supervised sparse representation classification for sleep EEG recongition with imbalanced sample sets," *J. Mech. Med. Biol.*, vol. 21, no. 5, Jun. 2021, Art. no. 2140006.

[36] Y. Liu and O. Sourina, "Real-time subject-dependent EEG-based emotion recognition algorithm," in *Transactions on Computational Science XXIII*. Berlin, Germany: Springer, 2014, pp. 199–223.

[37] M.-H. Lee, S. Fazli, J. Mehnert, and S.-W. Lee, "Subject-dependent classification for robust idle state detection using multi-modal neuroimaging and data-fusion techniques in BCI," *Pattern Recognit.*, vol. 48, pp. 2725–2737, Aug. 2015.

[38] H. Phan, F. Andreotti, N. Cooray, O. Y. Chén, and M. D. Vos, "Automatic sleep stage classification using single-channel EEG: Learning sequential features with attention-based recurrent neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 1452–1455.