A Novel Sleep Stage Classification via Combination of Fast Representation Learning and Semantic-to-Signal Learning

Hongxin Xiang*, Ting Zeng*, Yun Yang*†
*National Pilot School of Software, Yunnan University, Kunming 650091, China
†Kunming Key Laboratory of Data Science and Intelligent Computing, Kunming 650500, China
Email: xianghx@mail.ynu.edu.cn, zengting@mail.ynu.edu.cn, yangyun@ynu.edu.cn

***Abstract*—Electroencephalogram (EEG) signal is often used to assess sleep quality and treat sleep disorders. Many existing methods usually obtain high accuracy through a large number of feature preprocessing and feature extraction of EEG signals, which need a lot of prior knowledge as the basis. In this paper, a novel sleep stage classification framework, named FRL&S2SL, is proposed. The framework combines fast representation learning (FRL) and semantic-to-signal learning (S2SL) and uses single-channel EEG without any preprocessing of EEG signals. In the proposed framework, we utilize convolutional neural networks (CNN) to extract time-invariant features and bidirectional long short-term memory (BiLSTM) models to extract temporal features. Furthermore, auxiliary classifier generative adversarial network (ACGAN) is used to embed semantic features into signal features and to extract knowledge domain features of EEG signals for the first time. According to the American Academy of Sleep Medicine (AASM), sleep is divided into five stages: awake, rapid eye movement (REM) and three non-rapid eye movement (N1/N2/N3). We evaluated our framework using single-channel EEG (Fpz-Oz) from Sleep-EDF dataset, which is subject to the standards specified by AASM. The results show that our framework has achieved state-of-the-art in many evaluation metrics.**

***Index Terms*—electroencephalogram (EEG) signal, sleep stage classification, ACGAN, fast representation learning, semantic-to-signal learning.**

I. Introduction

Sleep is one of the most important physiological activities of the human body. Most people spend a third of their time sleeping, which is closely related to their physical and mental health. Modern medical research shows that more than 80 kinds of human diseases are closely related to long-term sleep disorders [1]. The effective diagnosis and treatments of sleep-related diseases have become an urgent and in-depth research topic in the current medical field. Many doctors and researchers have long debated how to understand it best. In recent years, sleep has become a branch of medicine and has been found to play a role in seemingly unrelated clinical problems [2].

This work was supported in part by the Natural Science Foundation of China (NSFC) under Grant 61876166 and Grant 61663046, in part by the Yunnan Applied Fundamental Research Project under Grant 2016FB104, in part by the Yunnan Provincial Young Academic and Technical Leaders Reserve Talents under Grant 2017HB005, in part by the Program for Yunnan High Level Overseas Talent Recruitment, and in part by the Program for Excellent Young Talents of Yunnan University.

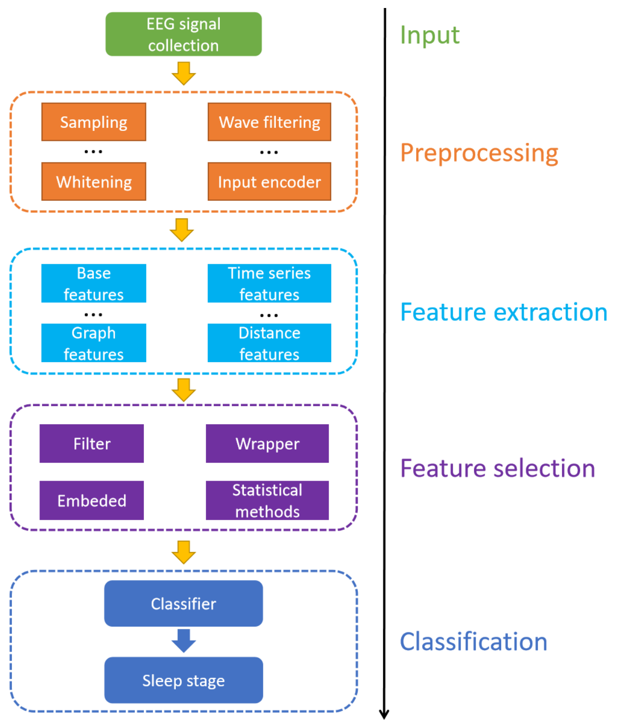


Fig. 1. Flow chart of common sleep stage methods.

Sleep stage classification is the first step in the diagnosis of sleep-related diseases. Physiologically, the sleep stage is divided into two main parts: the rapid eye movement period (REM) and the non-rapid eye movement (NREM) [3]. The sleep process usually circulates in REM and NREM. NREM was divided into the shallow sleep period (stage N1 and N2) and the deep sleep period (stage N3). Polysomnography (PSG) is mainly used as the main tool of sleep assessment, which includes electroencephalography (EEG), electromyography (EMG), and electrooculography (EOG), etc. Among them, EEG signal is most commonly used in sleep stage scoring because clinically acceptable sleep stage assessment mainly reads EEG signal according to R&K standard, which was standardized by Rechtschaffen and Kales [4] in 1968 and further developed by American Society of sleep medicine (AASM) in 2007 [5].

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Traditional visual sleep stages scoring is very tedious, time-consuming and subjective, which involves analyzing the signals in PSG records to score about 8 hours of sleep [6]. Therefore, many automatic sleep assessment methods [7]–[10] have been proposed. The process of these methods is summarized in Fig. 1. These researches mainly extract various sleep-related features from EEG signals, such as time domain features, frequency features, correlation features, entropy features, and so on. Potential sleep data information does not play a key role in automatic sleep stage classification.

In recent years, more and more deep learning methods have been applied to sleep classification [11]–[15]. CNN [16]–[23] is used to extract time-invariant local features. RNN [18], [24], [25] is used to learn temporal information such as mining time association between different sequences in the same 30-s epoch EEG. However, due to the complexity of the model, the speed of training and prediction of these deep learning methods is very slow. Considering the real-time nature of EEG data collection, these models are difficult to be used for online learning and real-time prediction. Therefore, we propose a new fast representation learning model (FRL), which uses a shallower network to improve the training and prediction speed without losing accuracy.

In addition, the method of improving prediction performance based on semantic knowledge [26], [27] has been widely used in image classification. For the processing of EEG signals, doctors observe the sleep stages of EEG signals through prior knowledge. For example, in the labeling REM period, doctors need to know that the performance of the REM period is low amplitude fast wave, which is a classification process with semantic knowledge. Similarly, it is inspired by the use of semantic information in image classification to improve the performance of the model. We can use semantic knowledge to assist our model in sleep stage evaluation. In this paper, semantic-to-signal learning(S2SL), which employed ACGAN [28] as the basic structure, is proposed to mine semantic features related to sleep stages.

This paper is organized as follows. In Section II, we summarize the theoretical background of the proposed method. Then, the experiment related content is described in Section III. Finally, the paper ends with conclusion and future works in Section IV. The main contributions of this paper are as follows:

1) A novel simple and efficient deep learning model is proposed, which uses CNNs with different filter sizes and bidirectional-LSTMs [29] to extract time-invariant features and time-variant features respectively. This model has fewer parameters than other deep learning models, such as [18], [20]–[22], and it has faster training and testing speed without losing accuracy.
2) The problem of data imbalance is solved by adding weighted softmax loss to the fast representation learning, i.e., only learning to classify the majority of sleep stages.
3) Semantic knowledge is applied to sleep classification for the first time. The performance of the model is further improved by fusing results of representation learning and semantic knowledge.
4) An experimentally good performance is demonstrated on publicly available Sleep-EDF dataset [30], [31]. It represented our framework can automatically learn features for sleep stage scoring without utilizing any hand-engineered features.

II. Methodology

In recent years, more and more automatic sleep stage evaluation methods tend to use deep learning to extract EEG features [18], [20]–[23]. Similar works [18], [21] show that the deep learning methods can achieve very optimistic results in the automatic sleep stage evaluation. However, most deep learning frameworks are complex in structure and slow in training time, so they are difficult to be used in the scene of streaming EEG data collection. Our method uses the shallower representation learning framework to extract the distinguishing features. This structure greatly speeds up the training and testing time of the model and obtains promising performance. In addition, the introduction of semantic knowledge further improves the performance of our model. In the end, our proposed framework achieves state-of-the-art results on several evaluation metrics. Fig. 2 shows the overall architecture of our method. The three main modules in the Fig. 2 will be described below.

A. Fast Representation Learning Module

The single-channel 30-s EEG epochs are divided into n sub-segments on average and these sub-segments are jointed vertically. Next, in the first layer of the model, m CNNs with the same size of $kernel_x$ and the different size of $kernel_y$ are both used to extract time-invariant features, and temporal features are extracted by bidirectional LSTM (BiLSTM). The design of this structure is mainly to simulate experts' sleep stage of EEG data: they need to understand the whole EEG data and need to observe the changes of some sub-segments of the EEG data. Small filters can better capture the EEG signal features of a specific mode, while large filters can capture the frequency information of the global EEG signal better. After each convolutional layer, three series operations are performed: batch normalization (BN) is used to speed up the training and to optimize the performance of the convolutional layer; Rectified linear unit (ReLU), which is formalized as $relu(x) = max(0, x)$, is used as the activation function; The maximum pooling layer is used for downsampling. Finally, the features, which extracted by multiple parallel convolution layers and two serial BiLSTM, are concatenated and transmitted forward to the softmax layer. Because the number of samples of different classes is very imbalanced in the original EEG data, each output of the softmax layer is given different weight. The weight is calculated by Formula.1, where s_c represents the number of all samples labeled c and $c = 1, 2, ..., C$ is a set of classes. The final loss function is defined as Formula.2, where x_i represents a 30-s epoch EEG sample, the corresponding label is y_i and $P_1(c|x_i)$ is contributed by posterior probability

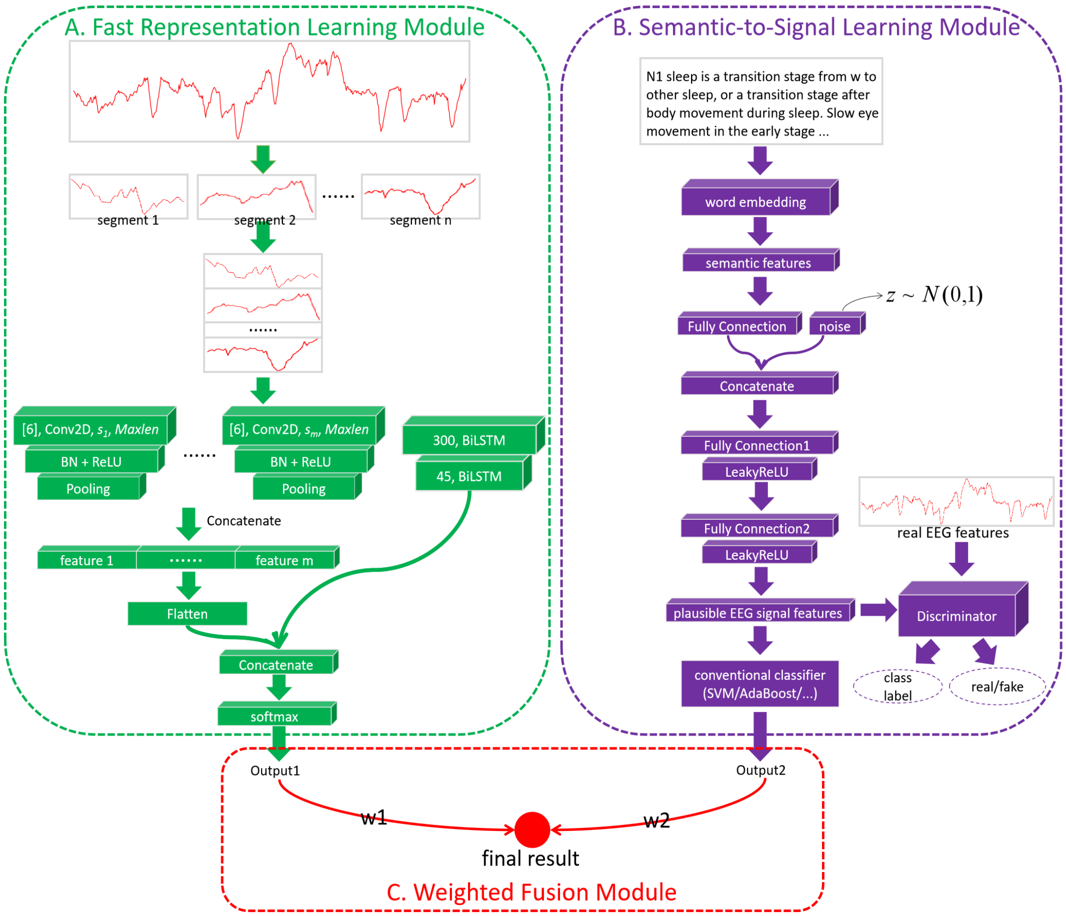


Fig. 2. Overall architecture: the fast representation learning module (green part) takes the segmented EEG data as input, and uses convolutional layer and BiLSTM layer to extract the signal features of EEG data quickly. This model will obtain the prediction probability $output_1$ for each class. For each class of text of EEG data, they are input into the semantic-to-signal learning module (purple part). According to these texts, the model can generate pseudo EEG signal features with semantics. The traditional classifier is trained to use these pseudo signal features to predict the real EEG signal features and get the prediction probability of each class $output_2$. Finally, the weighted fusion module (red part) is used to fuse $output_1$ and $output_2$ to get the final prediction probability.

of each class. The detailed parameter settings of FRL are described in Section III-D.

$$w_c = 1 - \frac{s_c}{\sum^{c=1...C} s_c} \tag{1}$$

$$L_{FRL} = \sum_{c=1}^{C} w_c * [\mathbb{I}(c \neq y_i)(1 - P_1(c|x_i)) + \mathbb{I}(c = y_i)(P_1(c|x_i))] \tag{2}$$

B. Semantic-to-Signal Learning Module

Semantic Features Extraction: For each class c of description text T_c, we collect them from the AASM manual published in 2007 [5] and Wikipedia articles that describe these sleep stages. Then, we only keep the text about the characteristics of sleep stages. And text preprocessing needs to be performed by removing stop words and tokenizing texts. Finally, semantic features can be extracted by text encoder ϕ, which can embed the texts into semantic spaces. In this paper, Term Frequency-Inverse Document Frequency (TF-IDF) [32] is used as the method of text embedding.

Semantic-to-Signal Embedding: Generator (G_θ for short, θ is the parameters of the generator) is used to transform semantic features into EEG signal features. Firstly, the semantic feature $\phi(T_c)$ of class c is passed two fully connected(FC) layers to extract features to reduce text noise. The extracted features are concatenated to a random noise $z \in \Re^Z$, which is sampled from Gaussian distribution $\aleph(0,1)$. Next, two FC layers with LeakyReLU activators are used to complete the inference process. The plausible EEG signal feature $\widetilde{x}$ can be generated through $G_\theta(T_c, z)$. The loss of the generator is defined in Formula.3, where the first term is Wasserstein loss [33] and the second term is cross-entropy loss of classes. It is worth noting that we can generate a large number of different signal features by using the same semantic features, because the noise z can be sampled repeatedly from the Gaussian distribution. These signal features generated by $G_\theta(T_c, z)$ are used to train a traditional classifier. In this paper, the AdaBoost model [34], [35] is trained to predict the original EEG signal features. The posterior probability of each class $P_2(c|x)$ in the original EEG data x will be obtained from the AdaBoost model.

TABLE I
THE NUMBER OF 30-S EPOCHS IN EACH SLEEP PHASE OF SLEEP-EDF DATASET

Dataset	W	N1	N2	N3(N4)	REM	Total
Sleep-EDF	8285	2804	17799	5703	7717	42308

Discriminator: The discriminator (D) receives the pseudo signal features generated by the generator and the original EEG signal features, and propagates them forward to the full connection layer with LeakyReLU. Then two subnetworks are designed: one is used to judge whether the input sample is real or fake, and the other one is used to judge the correct class label of the input sample. The loss function of the discriminator is defined in Formula.4, where the first two terms approximate Wasserstein distance of the distribution of real features and synthesized features, the last two terms are cross-entropy loss class of synthesized and real features, respectively.

$$L_G = E_{z\sim p_z(z)}[1 - D_\omega(G_\theta(T_c, z))] + L_{cls}(G_\theta(T_c, z)) \quad (3)$$

$$\begin{aligned} L_D = {} & E_{x\sim p_{data}}[1 - D_\omega(x)] + E_{z\sim p_z(z)}[D_\omega(G_\theta(T_c, z))] \\ & + L_{cls}(G_\theta(T_c, z)) + L_{cls}(D_\omega(x)) \end{aligned} \quad (4)$$

C. Weighted Fusion Module

Model fusion can improve the results in different machine learning tasks [36]–[42]. Weighted fusion method, which is an effective and the most simple method in model fusion, is used to fuse the results of FRL and S2SL (FRL&S2SL) in the paper. Its formal representation is defined as follows:

$$P(c|x_i) = \alpha_1 * P_1(c|x_i) + \alpha_2 * P_2(c|x_i) \quad (5)$$

where, $c \in \{1, 2, ..., C\}$ is a set of classes, $P(c|x_i)$ is the posterior probability of single 30-s epoch EEG sample x_i.

Finally, the prediction label $\widetilde{y}$ of FRL&S2SL will be obtained by the following formula:

$$\widetilde{y} = \arg\max_c P(c|x_i) \quad (6)$$

III. RESULTS

A. Data

The Sleep-EDF benchmark dataset [30], [31], which is widely applied in [11,12,13,14], is used to make a comparison between our approach and the state-of-the-art approaches. The sleep data which is publicly available online at https://www.physionet.org/content/sleep-edf/1.0.0/ has been used in this study. The recordings were obtained from Caucasian males and females (21-35 years old) without any sleep-related medication. The recordings contain horizontal EOG, Fpz-Cz and Pz-Oz EEG, and each sampled at 100 Hz. In this dataset, these records are artificially classified into eight periods according to R&K criteria, namely W, N1, N2, N3, N4, REM, MOVEMENT, UNKNOWN. To comply with AASM standards, we merged N3 and N4 periods into a single N3 period and removed data labeled MOVEMENT and UNKNOWN. Because the dataset has a long W phase before and after sleep, we only keep the data for 30 minutes.

In this paper, we evaluated the performance of our model using the single Fpz-Cz channel without any further preprocessing. Table I shows the number of sleep stages in sleep-EDF dataset, which are classified as a time window of 30s.

B. Experimental Design

The k-fold cross-validation strategy is used to evaluate our model. The each of subjects in the dataset are independent when the training set and the test set are divided. In our experiments, k was set to 20 for the Sleep-EDF dataset. Specifically, we used $N - (N/k)$ subject records for training and the remaining N/k subject records for testing, where N is the number of subjects in the dataset. In the training process, two subjects in the training set are randomly selected as development set, which will be used as the criteria for selecting the appropriate epoch and early stopping the training. Our model requires 20 training and tests on the sleep-EDF dataset, and then the sleep stage prediction results of each fold test data were combined to calculate the evaluation metrics. The evaluation metrics used in this paper will be discussed in Section III-C.

C. Performance Metrics

We use six metrics to evaluate the performance of our model, namely, precision of each class (PR_c), recall rate of each class (RE_c), F1-score of each class ($F1_c$), macro-averaging F1-score ($MF1$), the overall accuracy (ACC) and the Cohen's Kappa coefficient (κ) [43], [44]. The formulas for calculating these evaluation metrics are as follows:

$$PR_c = \frac{TP_c}{TP_c + FP_c} \quad (7)$$

$$RE_c = \frac{TP_c}{TP_c + FN_c} \quad (8)$$

$$F1_c = \frac{2PR_c * RE_c}{PR_c + RE_c} \quad (9)$$

$$MF1 = \frac{\sum_{c=1}^{C} F1_c}{N} \quad (10)$$

$$ACC = \frac{\sum_{c=1}^{C} TP_c}{N} \quad (11)$$

$$kappa = \frac{PR - p_e}{1 - p_e} \quad (12)$$

where TP_c is the number of true positive instances of class c, FP_c is the number of false positive instances of class c, FN_c is the number of false negative instances of class c, N is the number of unique classes, PR is overall precision, p_e is the chance agreement probability.

D. Training Parameters

In the fast representation learning (FRL), the 3000-dimensional EEG signal is first divided into n bi-section, and the n bisection EEG data is concatenated into the shape of (n, $3000/n$). Four parallel 2D convolutions layers and one RNN layer are used to extract EEG signal features. The number of filters per convolution layer is $filters = 6$, step size is $strides = 1$, and the sizes of these filters are (5, $3000/n$), (15, $3000/n$), (45, $3000/n$), (80, $3000/n$). Two bidirectional LSTM(BiLSTM) models are used in RNN layer to extract EEG features, and the number of neurons in hidden layer of BiLSTMs is units = 300, units = 45, respectively.

In the semantic-to-signal learning (S2SL), the TF-IDF method is used to transform knowledge text into semantic features with dimension of 157. These semantic features are propagated forward to the 2000 dimensional full connection layer, and the output of this full connection layer will be concatenated with the 50 dimensional random Gaussian noise. The two full connection layers with LeakyReLU activation function, which have dimensions of 3500 and 3000 respectively, are used to further extract features. The signal features generated by S2SL and the original EEG signal are trained as the input of the discriminator. Finally, the AdaBoost model with $n_{estimators} = 100$ is trained by synthesized signal features to predict the real EEG data.

Our method is trained with Adam, using the default parameters $\beta_1 = 0.9, \beta_2 = 0.999$. Other hyper-parameters $batchsize = 512, epoch = 30$ are set. After each epoch, the model will be used to predict the development set. If the $MF1$ measure on the development set for 3 epochs does not improve compared with the current maximum $MF1$ measure, the model will stop training and use the model with the highest $MF1$ measure as the final model. Additionally, we set $\alpha_1 = 0.8$, and $\alpha_2 = 0.2$ as the result weight of the two modules because the parameters can steadily improve the performance on the development set. A reasonable explanation that the weight of FRL module is greater than that of S2SL module is that the effect of S2SL module is limited due to many noises in the text. The final prediction result will be calculated using Formula.5 and Formula.6.

E. Implementation

Our model is mainly divided into two modules: fast representation learning (FRL) and semantic-to-signal learning (S2SL). The two modules are implemented using Keras with TensorFlow backend, which is a deep learning library and can be used for the design, debugging and evaluation of the deep learning model. In the semantic-to-signal learning, the conventional classifier is implemented by scikit-learn, which is a widely used python-based machine learning library. As shown in Table II, the server configuration we used for training and evaluation is shown. The training time required for cross-validation of each fold is about 25 minutes. Furthermore, it takes only 26 milliseconds to predict a 30-s epoch EEG data in test stage.

TABLE II
SERVER SPECIFICATIONS

parameters	specifications
RAM	32G
CPU	Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz
GPU	GeForce RTX 2060
OS	windows 10

TABLE III
CONFUSION MATRIX USING FAST REPRESENTATION LEARNING(FRL) METHOD. IT IS OBTAINED FROM 20-FOLD CROSS-VALIDATION ON FPZ-CZ CHANNEL FROM THE SLEEP-EDF DATASET

	Predicted					Per-class Metrics		
	W	N1	N2	N3	REM	PR	RE	F1
W	**7300**	522	144	26	293	86.71	88.11	87.40
N1	492	**1032**	491	6	783	42.54	36.81	39.47
N2	328	367	**15283**	624	1197	87.78	85.86	86.81
N3	51	4	610	**5025**	13	88.36	88.11	88.24
REM	248	501	882	6	**6080**	72.68	78.79	75.61

F. Visualization and Sleep Stage Scoring Performance

In order to reveal the differences of features, which come from the former layer of softmax in fast representation learning (FRL) module, among different sleep stages, our visualization attempts to use randomly selected 30-s epoch EEG samples from different sleep stages. In Fig. 3, the hidden layer outputs of the same object in different sleep stages are shown in each row, and those of different objects in the same sleep stage are shown in each column. A phenomenon can be observed: the features of same sleep stages seem to have some similar trend, but have obviously different amplitudes between different sleep stages. Although the difference between different sleep stages can explain that the FRL module can achieve a good performance in the benchmark dataset, the difference is still not intuitive for individual.

Table III and IV show confusion matrix obtained from the 20-fold cross-validation on the Fpz-Cz channel from Sleep-EDF dataset. Each row and column represent the number of 30 second EEG cycles for each sleep stage classified by sleep experts and our methods, respectively. The bold numbers indicates the number of samples with correct classification.

TABLE IV
CONFUSION MATRIX USING FAST REPRESENTATION LEARNING AND SEMANTIC-TO-SIGNAL LEARNING(FRL&S2SL) METHOD. IT IS OBTAINED FROM 20-FOLD CROSS-VALIDATION ON FPZ-CZ CHANNEL FROM THE SLEEP-EDF DATASET

	Predicted					Per-class Metrics		
	W	N1	N2	N3	REM	PR	RE	F1
W	**7301**	494	154	26	310	87.68	88.12	87.90
N1	469	**1072**	536	6	721	45.66	38.18	41.58
N2	273	308	**15637**	582	999	87.11	87.85	87.48
N3	41	2	654	**5001**	5	88.95	87.69	88.32
REM	243	472	970	7	**6025**	74.72	78.07	76.36

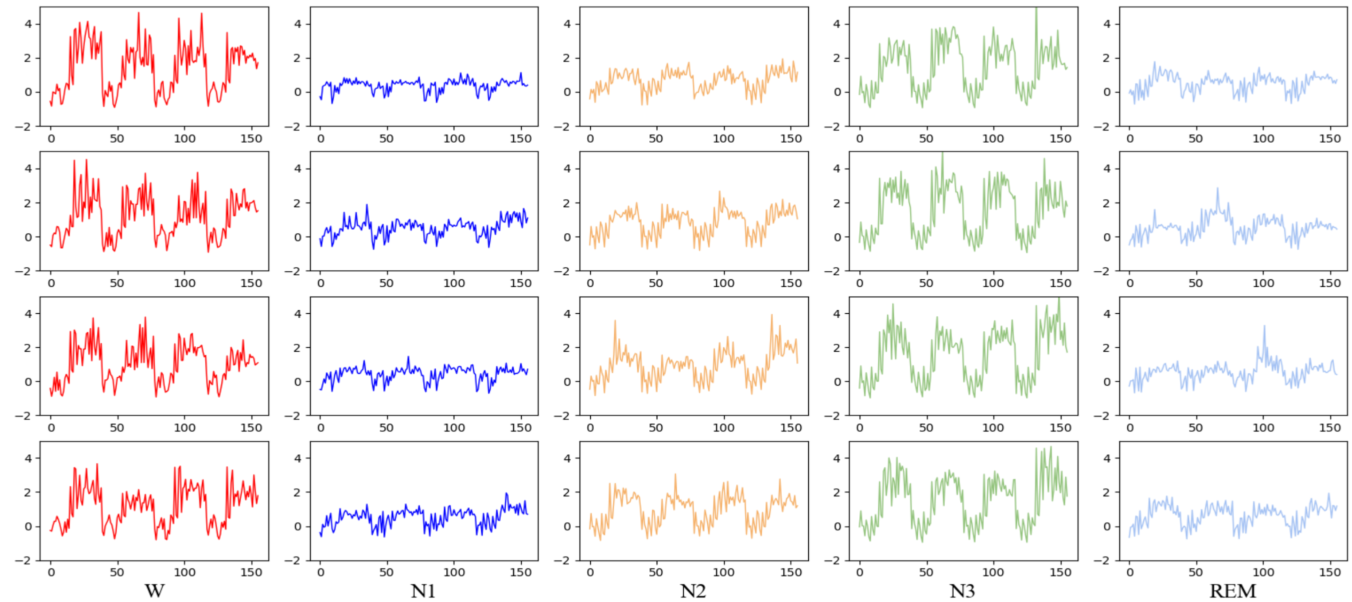


Fig. 3. Sample hidden layer outputs of fast representation learning module (FRL) for the five sleep stages. The hidden layer outputs are obtained by concatenating all convolution layer results and BiLSTM results of the last layer.

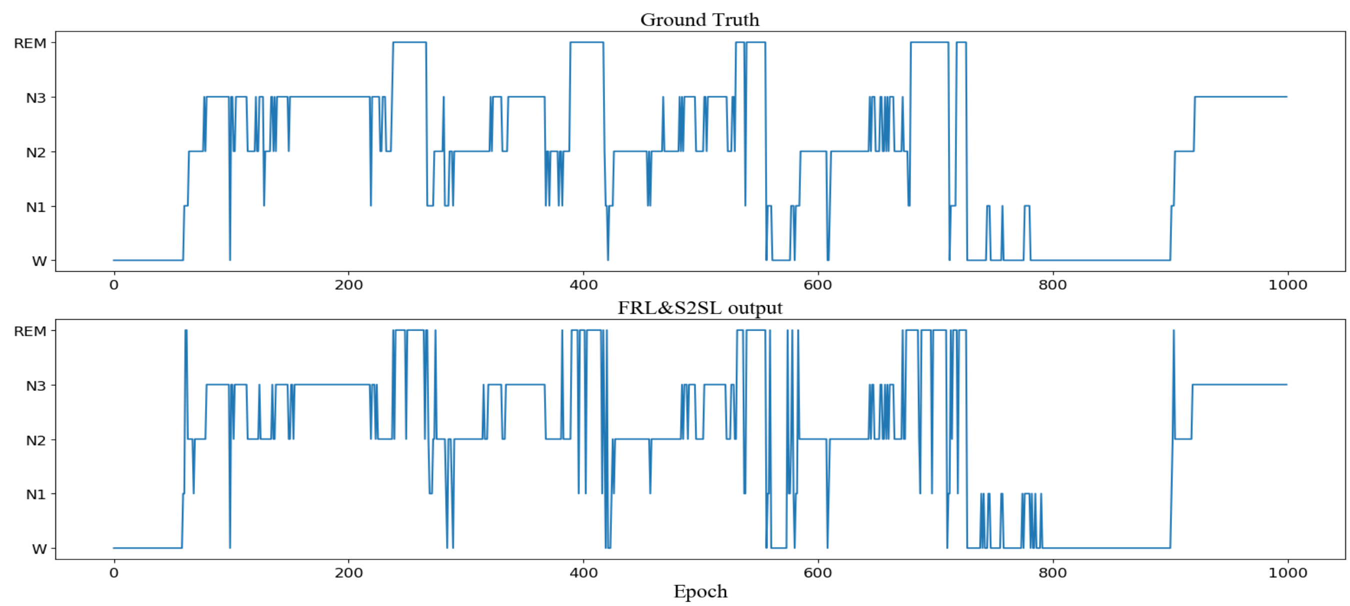


Fig. 4. The hypnogram classified by sleep experts(top) and the hypnogram classified by FRL&S2SL(bottom) are used for the subject "sc4001e0" in Sleep-EDF dataset.

The last three columns in each row indicate per-class performance metrics computed from the confusion matrix. Table III and IV represent the use of FRL method and FRL&S2SL method to evaluate sleep staging performance, respectively. The tables both show that our methods performed the worst in N1, followed by REM. In other sleep stages, F1 values were significantly better, ranging from 86.81 to 88.24. Meanwhile, it can be seen that the confusion matrix is almost diagonally symmetric and the prediction results of our methods do not tend to majority class N2. This indicates that the misclassifications are unlikely to be caused by imbalanced-class problem [18]. It is worth noting that most of the misclassification results in N1 are predicted as REM. A rational reason is that the physiological characteristics of stage N1 and stage REM are very similar, and it is difficult for the model to learn the discriminative features. Currently, prediction performance in stage N1 is still a challenge.

In addition, we compared the hypnogram scored by expert and hypnogram scored by FRL&S2SL method in Fig. 4. Obviously, the hypnogram obtained by our method is highly

TABLE V

PERFORMANCES OF THE PROPOSED METHOD COMPARED TO PREVIOUS METHODS ON THE SLEEP-EDF DATASET USING INDEPENDENT TRAINING AND TESTING. OUR FRL REPRESENTS THE RESULT OF FAST REPRESENTATION LEARNING METHOD, WHILE OUR FRL&S2SL REPRESENTS THE RESULT OF WEIGHTED FUSION OF FAST REPRESENTATION LEARNING AND SEMANTIC-TO-SIGNAL LEARNING METHOD. BOLD NUMBERS, ITALICS, AND RED FONTS INDICATE THE HIGHEST ACCURACY, THE SECOND HIGHEST ACCURACY, AND THE PERFORMANCE DIFFERENCE OF FRL AND FRL&S2SL METHODS, RESPECTIVELY.

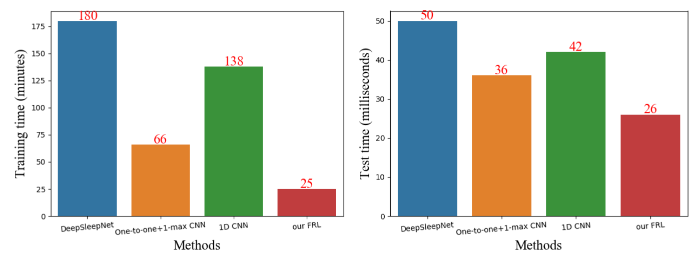
Methods	Dataset	EEG Channel	Overall Metrics			Per-class F1-score (F1)				
			ACC	MF1	κ	W	N1	N2	N3	REM
Ref. [19]	Sleep-EDF	Fpz-Cz	78.9	73.7	0.71	71.6	**47**	84.6	84	81.4
Ref. [16]			74.8	69.8	0.65	65.4	43.7	80.6	84.9	74.5
Ref. [18]			82	**76.9**	**0.76**	84.7	*46.6*	85.9	84.8	*82.4*
Ref. [20]			81.6	72	0.73	56	**47**	*87*	87	**83**
Ref. [23]			*82.3*	74.7	*0.75*	77.3	40.5	**87.4**	86	82.3
Our FRL			82.1	75.5	*0.75*	*87.4*	39.5	86.8	*88.2*	75.6
Our FRL&S2SL			**82.5**(+0.4)	*76.3*(+0.8)	**0.76**(+0.1)	**87.9**(+0.5)	41.6(+2.1)	**87.5**(+0.7)	**88.3**(+0.1)	76.3(+0.7)

consistent with that obtained by experts, which shows that our method is effective in the evaluation of sleep stage.

G. Comparison With State-of-the-Art Approaches

Our methods are used to compare with the five latest methods which are from [19], [16], [18], [20], [23] on the Fpz-Cz channel from the Sleep-EDF dataset. Table V shows three evaluation metrics, namely $ACC, F1, \kappa$, and the results of comparative experiments on independently divided subjects. On the overall evaluation metrics, FRL&S2SL obtained the highest accuracy ACC and the highest Cohen's kappa coefficient κ, which were 82.5% and 0.76%, respectively. In addition, FRL&S2SL achieved the second best result in $MF1$ metric, which is 76.3%, only 0.6% lower than the best [18]. For each class of $F1$ metric, FRL&S2SL method achieved the highest $F1$ results in stage W, stage N2 and stage N3, which were 87.9%, 87.5% and 88.3%, respectively. Compared with the best methods, stage N1 and stage REM show some decline. At the same time, FRL method is slightly inferior to FRL&S2SL method, but it exceeds the average performance in almost all evaluation metrics. In addition, compared with the results of only using FRL method, FRL&S2SL method shows a consistent improvement in all evaluation metrics. This proves the effectiveness of increasing semantic information in sleep stage scoring.

Furthermore, the time efficiency of our FRL method is compared with the latest three methods, which are DeepSleepNet [18], one-to-one+1-max CNN [23], 1D CNN method [22]. The time performances is calculated using a unified computer configuration as shown in Table II. The Fig. 5(a) and Fig. 5(b) show the training time comparison and test time comparison between FRL method and other methods, respectively. Our FRL method is obviously faster than other methods in training time and testing time.



(a) Training time (minutes) comparison (b) Test time (milliseconds) comparison

Fig. 5. Training or test time comparison with the latest methods. The proposed method, fast representation learning(FRL), is compared with the three latest methods. DeepSleepNet and 1D CNN methods are respectively selected in the paper [18], [22]. The one-to-one+1-max CNN model is used as time performance comparison because it is the simplest of many models proposed in paper [23].

IV. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we proposed a novel framework (FRL&S2SL) that combined fast representation learning with semantic-to-signal learning for automatic sleep stage scoring based on single-channel EEG without using any hand-engineering features. The fast representation learning(FRL) uses multiple parallel convolutional neural networks and multiple serial BiLSTM models to extract frequency-domain and time-domain features respectively. In order to alleviate class-imbalance problems, the weighted softmax loss is applied to the FRL. For the first time, the semantic features of EEG signals are extracted through the semantic-to-signal learning(S2SL). Our results show that the FRL has faster training and prediction speed than other algorithms [18], [22], [23]. Furthermore, the results of the benchmark dataset, Sleep-EDF, show that the FRL can extract effective EEG features, and achieve the performance of state-of-the-art on some evaluation metrics. In addition, the fusion of FRL and S2SL has achieved consistent improvement in many evaluation metrics compared with FRL, which shows that S2SL can learn effective EEG semantic information. The final results show that our framework, FRL&S2SL, achieves the performance of state-of-the-art on most of the evaluation metrics.

However, our model can be improved in many ways. In the future, the structure of FRL module can be further improved to learn more distinguishing features of N1 and REM. Secondly, the optimal S2SL module structure or the introduction of knowledge graph can be explored to extract more semantic features related to EEG signal. Finally, more fusion strategies of FRL module and S2SL module will be studied to find the optimal fusion results.

REFERENCES

[1] N. Kettner, H. Voicu, M. Finegold, C. Coarfa, A. Sreekumar, N. Putluri, C. Katchy, C. Lee, D. Moore, and L. Fu, "Circadian homeostasis of liver metabolism suppresses hepatocarcinogenesis," *Cancer Cell*, p. S1535610816304949, 2016.

[2] H. T. Wu, R. Talmon, and Y.-L. Lo, "Assess sleep stage by modern signal processing techniques," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1159–1168, 2015.

[3] S. A. Gharib, S. A. Gharib, and S. A. Gharib, "Sleep medicine: Essentials and review.," 2009.

[4] E. A. Wolpert, "A manual of standardized terminology and scoring system for sleep stages of human subjects," *Electroencephalography & Clinical Neurophysiology*, vol. 20, no. 2, p. 246, 1969.

[5] C. Iber and C. Iber, *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*, vol. 1. American Academy of Sleep Medicine Westchester, IL, 2007.

[6] S. zen, "Classification of sleep stages using class-dependent sequential feature selection and artificial neural network," *Neural Computing and Applications*, vol. 23, 10 2012.

[7] A. Stochholm, K. Mikkelsen, and P. Kidmose, "Automatic sleep stage classification using ear-eeg," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4751–4754, IEEE, 2016.

[8] A. R. Hassan and M. I. H. Bhuiyan, "An automated method for sleep staging from eeg signals using normal inverse gaussian parameters and adaptive boosting," *Neurocomputing*, vol. 219, pp. 76–87, 2017.

[9] Y. Ma, W. Shi, C.-K. Peng, and A. C. Yang, "Nonlinear dynamical analysis of sleep electroencephalography using fractal and entropy approaches," *Sleep medicine reviews*, vol. 37, pp. 85–93, 2018.

[10] H. Kassiri, A. Chemparathy, M. T. Salam, R. Boyce, A. Adamantidis, and R. Genov, "Electronic sleep stage classifiers: A survey and vlsi design methodology," *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 1, pp. 177–188, 2016.

[11] K. Mikkelsen and M. De Vos, "Personalizing deep learning models for automatic sleep staging," 2018.

[12] J. Zhang and Y. Wu, "A new method for automatic sleep stage classification," *IEEE Transactions on Biomedical Circuits & Systems*, pp. 1–14, 2018.

[13] H. Phan, F. Andreotti, N. Cooray, O. Chn, and M. de Vos, "Dnn filter bank improves 1-max pooling cnn for single-channel eeg automatic sleep stage classification," vol. 2018, 07 2018.

[14] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, pp. 758–769, 2017.

[15] A. Vetek, K. Müller, and H. Lindholm, "A compact deep learning network for temporal sleep stage classification," *2018 IEEE Life Sciences Conference (LSC)*, pp. 114–117, 2018.

[16] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic Sleep Stage Scoring with Single-Channel EEG Using Convolutional Neural Networks," *arXiv e-prints*, p. arXiv:1610.01683, Oct 2016.

[17] S. Stober, D. J. Cameron, and J. A. Grahn, "Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings," in *Advances in neural information processing systems*, pp. 1449–1457, 2014.

[18] A. Supratak, H. Dong, C. Wu, and Y. Guo, "Deepsleepnet: a model for automatic sleep stage scoring based on raw single-channel eeg," *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, pp. 1–1, 2017.

[19] O. Tsinalis, P. M. Matthews, and Y. Guo, "Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders," *Annals of Biomedical Engineering*, vol. 44, no. 5, pp. 1587–1597, 2016.

[20] M. Sokolovsky, F. Guerrero, S. Paisarnsrisomsuk, C. Ruiz, and S. A. Alvarez, "Deep learning for automated feature discovery and classification of sleep stages," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1–1, 2019.

[21] H. Dong, A. Supratak, W. Pan, C. Wu, and Y. Guo, "Mixed neural network approach for temporal sleep stage classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, no. 99, pp. 1–1, 2016.

[22] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, "A convolutional neural network for sleep stage scoring from raw single-channel eeg," *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018.

[23] 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 Transactions on Biomedical Engineering*, pp. 1–1, 2018.

[24] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded lstm recurrent neural network for automated sleep stage classification using single-channel eeg signals," *Computers in biology and medicine*, vol. 106, pp. 71–81, 2019.

[25] S. Mousavi, F. Afghah, and U. R. Acharya, "Sleepeegnet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PloS one*, vol. 14, no. 5, p. e0216456, 2019.

[26] C. Menglong, J. Detao, Z. Ting, Z. Dehai, X. Cheng, C. Zhibo, and X. Xiaoqiang, "Image classification based on image knowledge graph and semantics," in *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 81–86, May 2019.

[27] M. Yeh and Y. Li, "Multilabel deep visual-semantic embedding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.

[28] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," 2016.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[30] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[31] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberye, "Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 9, pp. 1185–1194, 2000.

[32] "Term frequency by inverse document frequency," 2009.

[33] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv e-prints*, p. arXiv:1701.07875, Jan 2017.

[34] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proceedings of the Second European Conference on Computational Learning Theory*, 1995.

[35] J. Zhu, A. Arbor, and T. Hastie, "Multi-class adaboost," vol. 2, no. 3, pp. 349–360, 2006.

[36] Y. Yang and J. Jiang, "Hybrid sampling-based clustering ensemble with global and local constitutions," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 27, no. 5, pp. 952–965, 2017.

[37] M. R. Hassan, B. Nath, and M. Kirley, "A fusion model of hmm, ann and ga for stock market forecasting," *Expert Systems with Applications*, vol. 33, no. 1, pp. 171–180, 2007.

[38] S. Iqbal, M. U. G. Khan, T. Saba, Z. Mehmood, and R. Abbasi, "Deep learning model integrating features and novel classifiers fusion for brain tumor segmentation," *Microscopy Research and Technique*, vol. 82, no. 3, 2019.

[39] Y. Yang and J. Jiang, "Adaptive bi-weighting toward automatic initialization and model selection for hmm-based hybrid meta-clustering ensembles," *IEEE Transactions on Cybernetics*, pp. 1–12, 2019.

[40] Y. Xu and Y. Lu, "Adaptive weighted fusion: A novel fusion approach for image classification," *Neurocomputing*, vol. 168, pp. 566–574, 2015.

[41] J. Liu, W. Shang, and W. Lin, "Improved stacking model fusion based on weak classifier and word2vec," in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, 2018.

[42] Y. Yang and J. Jiang, "Bi-weighted ensemble via hmm-based approaches for temporal data clustering," *Pattern Recognition*, vol. 76, pp. 391–403, 2018.

[43] K. Jin and W. Wang, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 74, no. 1, pp. 116–138, 2014.

[44] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.