IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 23, 2020, accepted May 10, 2020, date of publication May 14, 2020, date of current version June 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2994593

Self-Supervised Learning From Multi-Sensor Data for Sleep Recognition

AITE ZHAO[1], JUNYU DONG[1], (Member, IEEE), AND HUIYU ZHOU[2]
[1]Department of Information Science and Engineering, Ocean University of China, Qingdao 266100, China
[2]Department of Informatics, University of Leicester, Leicester LE1 7RH, U.K.

Corresponding author: Junyu Dong (dongjunyu@ouc.edu.cn)

This work was supported in part by the Natural Science Foundation of Shandong under Grant ZR2014FQ023, and in part by the National Natural Science Foundation of China (NSFC) under Grant 41576011 and Grant 41927805.

ABSTRACT Sleep recognition refers to detection or identification of sleep posture, state or stage, which can provide critical information for the diagnosis of sleep diseases. Most of sleep recognition methods are limited to single-task recognition, which only involves single-modal sleep data, and there is no generalized model for multi-task recognition on multi-sensor sleep data. Moreover, the shortage and imbalance of sleep samples also limits the expansion of the existing machine learning methods like support vector machine, decision tree and convolutional neural network, which lead to the decline of the learning ability and over-fitting. Self-supervised learning technologies have shown their capabilities to learn significant feature representations. In this paper, a novel self-supervised learning model is proposed for sleep recognition, which is composed of an upstream self-supervised pre-training task and a downstream recognition task. The upstream task is conducted to increase the data capacity, and the information of frequency domain and the rotation view are used to learn the multi-dimensional sleep feature representations. The downstream task is undertaken to fuse bidirectional long-short term memory and conditional random field as the sequential data recognizer to produce the sleep labels. Our experiments shows that our proposed algorithm provide promising results in sleep identification and can further be applied in clinical and smart home environments as a diagnostic tool. The source code is provided at: ''https://github.com/zhaoaite/SSRM ''.

INDEX TERMS Sleep recognition, sleep diseases, multi-sensor, self-supervised learning, bidirectional LSTM, CRF, feature representations, temporal information.

I. INTRODUCTION

Sleep quality is a proven biometric that plays an eminent role in health status evaluation of patients with mental or physical disorders. According to the survey of the World Health Organization (WHO), about 1/3 of people in the world have sleep problems, and the global sleep disorder rate is 27%, which seriously affect people's health and quality of life [1]. In order to better evaluate and monitor sleep quality and state, we are committed to establishing a reliable and safe sleep model for analysis and diagnosis of sleep state.

Sleep recognition problem is divided into three levels: sleep posture recognition, sleep stage recognition and insomnia detection. Sleep quality is directly related to sleep posture. Supine sleeping positions will not suppress organs such as viscera and organs, and can effectively relieve symptoms of pain in the neck and back. However, this kind of sleeping posture is easy to lead to the fall of the tongue root and block breathing, which is not suitable for people who often snore or have respiratory diseases. Lying on the right side is conducive to the normal operation of the gastrointestinal tract and will not compress the heart, but it can affect the movement of the right lung [2]. On the other hand, monitoring the sleep stage is also an important way to evaluate sleep state. In the period of 90-100 minutes, there are two different stages: non-rapid eye movement sleep (NREMs), and rapid eye movement sleep (REMs) [3]. Insomnia is the most prominent and frequent manifestation of sleep disorders. Our paper provides an effective method for the diagnosis of insomnia, as well as the monitoring and classification of sleep posture and stage.

For sleep recognition, multiple sensors have been applied in gathering data. Some visual sensors or pressure sensors can be used to capture sleep posture. In this paper, two different types of pressure sensing mats are used to collect in-bed pose

The associate editor coordinating the review of this manuscript and approving it for publication was Yasar Amin.

pressure data for sleep posture recognition. In sleep stage recognition, polysomnography (PSG) [4] is a sleep diagnostic tool which uses electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), and other physiological sensors to collect data and diagnose sleep disorders. However, it is not convenient to use many sensors during sleep. Sleep experiments in special sleep facilities due to the psychological pressure may not accurately reflect the real sleep problems. Therefore, we use wrist-worn wearable devices to obtain heart rate and motion information to monitor sleep state. For insomnia detection, bioradar is installed for non-contact sleep monitoring and bioradiolocation signals acquisition.

Based on the multi-sensor sleep data, a large quantity of feature representation and classification methods have been applied in the field of sleep recognition. Histogram of oriented gradients (HoG) and local binary patterns (LBP) from data, restricted Boltzmann machines (RBM), convolutional neural network (CNN), k-nearest neighbour (KNN), support vector machines (SVM), have been involved to perform the in-bed posture classification [5]–[7]. Besides, there are also a number of studies that make efforts to identify sleep stages, such as hand-crafted feature classification, k-means and random forests [8]–[10]. These methods are also suitable to distinguish insomnia patients.

Although these methods have shown promising performance in classifying sleep data, most of them are only applied to one type of sleep dataset, and do not involve multi-modal sensor data like images and signals. They are also used for single task recognition, with end-to-end recognition characteristics without a pre-training process. In the case of insufficient samples or uneven sample distributions, these methods are even less able to learn effective feature representations, and deep learning frameworks may also experience over-fitting.

To solve these shortcomings, we put forward a self-supervised sleep recognition model (SSRM), which utilizes frequency information and spatio-temporal information of sleep signals as the evidence of recognition. Additionally, it is noticed that the self-supervised sleep model can increase the amount of data, which enables a deep learning model to be successfully used with relatively little or unbalanced data. It can also enhance the capacity of each class of data and reaches the optimal state of the training model.

The upstream task is the pre-training stage and feature representation stage of sleep recognition. After having collected sleep data, we extract the frequency-domain and rotation features of the data to generate new labels together with the original data. We build a four-layer CNN which can learn nonlinear spacial information to generate sleep feature representations through pre-training based on multi-modal sleep data. The features in various classes generated by this self-supervised model will be integrated with the original data to participate in the downstream classification task.

In the downstream task, we describe a dynamic bidirectional long-short term memory (BiLSTM) approach to model the temporal sleep data. The fused sleep data is the input of BiLSTM, and the time step is also part of the input with varied length sequences. BiLSTM can not only learn the temporal features, but also consider the context. Moreover, the conditional random field (CRF) is the suffix of BiLSTM, which improves the efficiency of the model. The prediction probability of BiLSTM is regarded as the input of CRF to further learn parameters.

The main contributions of this paper are summarized as follows:

- We study the problem of sleep recognition aiming at three levels: sleep position recognition, sleep stage recognition and insomnia detection, including the analysis and understanding of multi-sensor sleep data.
- A self-supervised sleep recognition model (SSRM) is proposed to solve the problem of multi-sensor sleep recognition, including upstream pre-training, feature interpretation and downstream recognition tasks. Data capacity and feature representation can be realized through pre-training, and label prediction and recognition can be realized through BiLSTM-CRF. To a certain extent, this method successfully solves the problem of multi-level sleep recognition.
- The proposed method achieves superior performance on three challenging datasets for multi-sensor sleep recognition.

After introducing the related work in Section II, we present the proposed self-supervised sleep recognition model (SSRM) in Section III, and experiments that include various similarity criteria, and several advanced approaches on three sleep datasets in Section IV. Finally, Section VI shows limits and advantages of our method, and Section 6 concludes the paper.

II. RELATED WORK

After motivating our choice of a self-supervised model, we will discuss related works exploring the latent space of sleep recognition methods.

A. SLEEP POSTURE RECOGNITION

Xu *et al.* proposed a new distance measurement method for sleep posture differencing. By projecting the pressure distribution to the horizontal and vertical directions, distribution differences can be identified and it achieved 90.78% high accuracy through a KNN classifier [11]. They improved the model and proposed a novel matching-based approach in the next year achieving 91.21% accuracy [12]. To estimate body postures reliably and comfortably, the ballistocardiogram (BCG) signals were collected for sleep posture recognition, and a Bayesian classifier with piecewisesmoothing correction was used for classification [13].

A multi-stream CNN was also applied in this field, which was based on depth images for identifying ten sleep postures with high accuracy [14]. A sleep monitoring system by embedding radio frequency identification (RFID) tags

was proposed for sleeping posture recognition and body movement detection, which used a convolutional neural network (CNN) to identify the sleeping postures. Meanwhile, the movements and their durations can also be detected by using k-means [15].

B. SLEEP STAGE RECOGNITION

Walsh *et al.* presented the evaluation of an under-mattress sleep monitoring system for non-contact sleep/wake discrimination, which compared different classifiers (SVM, KNN, ANN and LDA) based on the extracted temporal, spatial, and statistical features [16]. By using electroencephalogram (EEG) signals, the structural graph similarity and the k-means (SGSKM) are combined to identify six sleep stages, and four existing methods and the support vector machine (SVM) classifier were compared with the proposed method [17]. An importance weighted kernel logistic regression (IWKLR) was applied for classification of the EEG, EOG and EMG signals [18].

The multi-taper spectral analysis was presented to create visually interpretable images of sleep patterns from EEG signals as inputs to a deep convolutional network trained to solve visual recognition tasks and the classification of sleep stages [19]. Furthermore, ANN and HMM also contributed to the macro-sleep stages (MSS) detection by using full night audio time series, which took advantages of differences in sound properties within each MSS due to breathing efforts (or snores) and body movements in bed [20].

C. INSOMNIA DETECTION

In paper [21], a support vector machine (SVM) classifier was employed to distinguish the control group and insomnia patients. A k-means classifier was presented using Hjorth parameters extracted from the central electroencephalogram (EEG) signals to accurately detect insomnia [22].

Abdullah *et al.* used an artificial neural network (ANN) to extract linear and nonlinear features from the denoised signals such as sleep EEG and ECG signals, and classify healthy people and insomnia patients [23]. Moreover, deep neural networks (DNNs) for insomnia detection was trained, which were fed by a set of temporal and spectral features derived from 2 EEG channels [24].

It can be seen that the combination of hand-crafted features and statistical methods with a traditional machine learning classifier has become the mainstream means of sleep recognition, and the application of deep learning model in this field is still lacking.

III. THE SELF-SUPERVISED SLEEP RECOGNITION MODEL

Self-supervised learning aims to input a group of unsupervised data and construct new labels artificially through the structure or characteristics of the data itself. With new labels, we can train the input data using supervised learning. Self-supervised learning has been firstly adopted within the computer vision community to learn representations by solving various auxiliary tasks, such as colorize gray scale images or solving puzzles from image patches. Self-supervised learning has also been applied successfully in language modeling, leading to models like BERT [25]. In this problem, the self-supervised learning model can increase the data capacity very well, so that the model can be trained using unlabeled data.

In this section, we design a self-supervised model and present the insights gained from doing so. We first conduct a pre-training task for feature representation, adopting the feature of rotation and frequency domain, which can enlarge the data capacity to 2-5 times of the original data. The frequency spectrum describes the frequency structure of the signals and the relationship between the frequency and the amplitude of frequency signals, so we use the Fourier transform [26] to calculate the frequency characteristics of the signal. The features in the frequency domain and rotation are fed into a four-layer CNN for supervised training by using new added labels. After training, we concatenate the features extracted by CNN as the input of the downstream tasks.

The downstream task is BiLSTM-CRF [27]. The sleep data is collected continuously according to the chronological order, which fully fits the function of the expandable nodes of the BiLSTM. RNN and LSTM can only predict the output at the next moment in terms of the temporal information of the previous moment. However, in some problems, the output of the current time is related to the previous state, as well as the future state. For example, to calculate the pressure signal output at a certain time point, it is necessary to analyse the previous information and consider the later content. The BiLSTM consists of two RNNs stacked up and down, and the output is determined by the status of the two RNNs.

First of all, we show the framework and formulation of the proposed self-supervised model. The structure is illustrated in FIGURE 1. Sleep data from the sensors are the input of the upstream pre-training model, including the raw data and four workers. Workers are used to extract features $X_0 - X_4$ of different dimensions of the raw data, create new labels and feed them into the four-layer CNN together with features in order. The results are sent to the BiLSTM-CRF model for feature acquisition of time series.

Assuming the input data is $X_0 \in \mathbb{R}^N$, N is the sample number of the raw data. After the pre-processing of the input, we obtain multiple features X_1^m, and the final output $I^m = \{I^m \in \mathbb{R}^N, m = 0, 1, 2, 3, 4\}$ is calculated by a four-layer CNN model. By the fusion of all the features, in the downstream process, $I^n = \{I^n \in \mathbb{R}^K, n = 0, 1, 2, \ldots, N\}$, K is the sequence length of feature in one sample, which is fed into BiLSTM model with corresponding label sequences $L^n = \{L^n \in \mathbb{R}^N, n = 0, 1, 2, \ldots, N\}$.

A. PRE-TRAINING AND FEATURE REPRESENTATIONS

In the process of pre-training and feature representations, we consider two types of transformations for self-supervised data augmentation, rotation (3 transformations) and frequency domain feature, as illustrated in FIGURE 2. We take the sleep posture data as the input of the pre-training model.

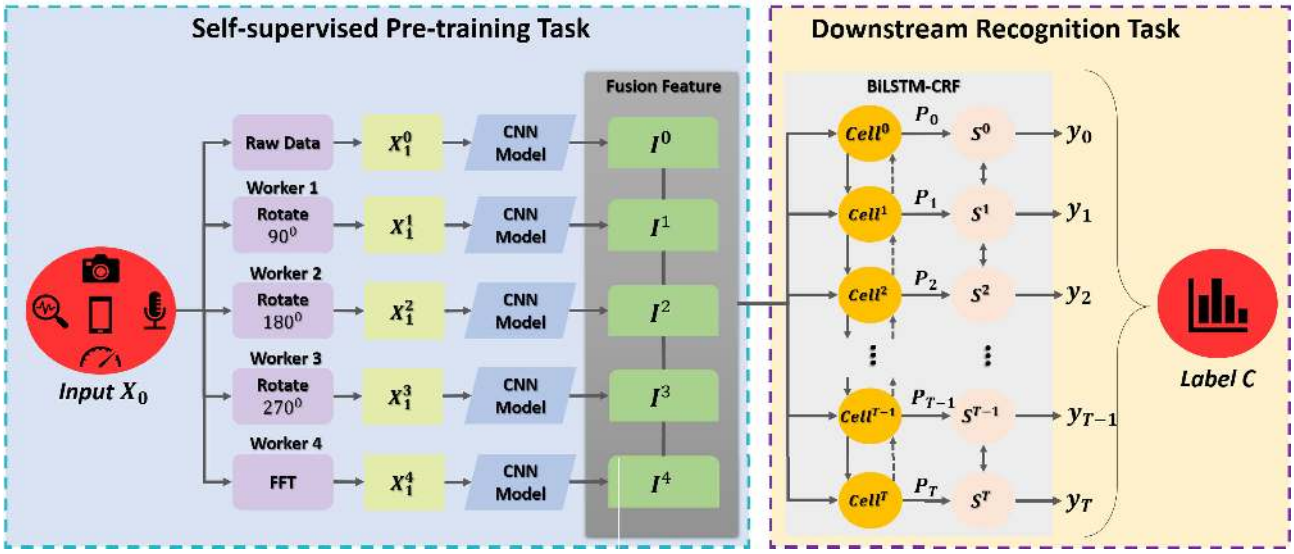


FIGURE 1. The structure of the self-supervised sleep recognition model (SSRM). The whole model is divided into self-supervised pre-training model and downstream recognition model. The pre-training process includes the input of the original sleep data X_0 in the first layer, the rotation and frequency-domain feature extraction of the original data using four workers in the second layer to get X_1^m, and the generation of corresponding labels. The third layer uses a standard CNN model to pre-train and expand the data to get I^m. The first layer of the downstream task sends the features extracted from the pre-training to BiLSTM, and the second layer uses CRF to analyze the prediction probability of the BiLSTM, so as to get the correct identification label C.

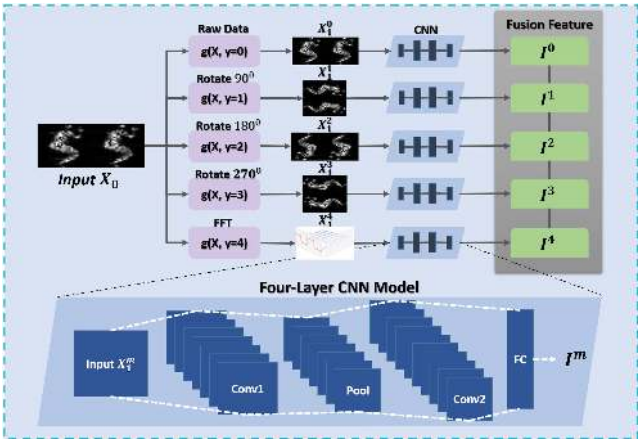


FIGURE 2. The self-supervised pre-training process. We take the pose image in the pressure map dataset as an example to introduce the self-supervised pre-training process. First, the original input X_0 is rotated at different angles to get $X_1^1 - X_1^3$ and get X_1^4 through FFT, generating labels 0-4. A four-layer CNN with 2 hidden layers and one input-output layer, which uses 5 * 5 convolution kernel for convolution operation, we take the output of the second hidden layer as the final result I^m of pre-training.

First, the input data is rotated to 90 degrees, 180 degrees and 270 degrees respectively. Fourier transform is used to extract frequency-domain features, and after that, 5 × labels and samples are generated for self-supervision. Then the four-layer CNN, which is composed of two hidden layers, an input layer and an output layer, is constructed for supervised training.

1) ROTATION SELF-SUPERVISION

For data augmentation, e.g., rotation or cropping, which systematically enlarge the training dataset by explicitly generating more training samples, have been popularly used to improve the generalization performance of deep neural networks [28].

A novel method for self-supervised feature learning is presented. By training the CNN model, we can recognize four rotation degrees (0°, 90°, 180°, 270°) applied to its input image [29]. It successfully forces the CNN model trained on it to learn semantic features useful for various visual perception tasks, such as object recognition, object detection and object segmentation. Let X_0 be an input, the output matrix of rotation is X_1^m, $m = 0, 1, 2, 3$.

2) FREQUENCY DOMAIN SELF-SUPERVISION

In the frequency domain, frequency is the independent variable, which has the horizontal axis of the frequency and the vertical axis of the amplitude. The frequency spectrum describes the frequency structure of the signal and the relationship between frequency and amplitude of the frequency signal.

When analyzing signals in the time domain that the signals are exactly the same. Because the signal changes with time, as well as frequency, phase and other information. It is necessary to further analyze the frequency structure of the signal. The dynamic signal is from the time domain to the frequency domain mainly by Fourier series and Fourier transform. Therefore, we utilize fast Fourier transform to extract the frequency domain features of the input data. The fast discrete Fourier transform (FFT) is the first choice because the input of each frame is discrete points. The calculation of FFT is shown in Eq.(1).

$$X_1^m = X(k) = \sum_{n=0}^{N-1} X_0(n) W_n^{kn}$$
$$m = 4, \quad k = 0, 1 \ldots, N-1, \; W_N = e^{-j\frac{2\pi}{N}} \qquad (1)$$

FFT decomposes $X(k)$ into the sum of even and odd sequences, the length of $x_1(n)$ and $x_2(n)$ is $N/2$, $x_1(n)$ is an even sequence, and $x_2(n)$ is an odd sequence. The output $X_1^m, m = 4$ of FFT is shown as follows:

$$\begin{aligned} X(k) &= X_1(k) + W_N^k X_2(k), k = 0, 1 \ldots, N-1 \\ &= \sum_{n=0}^{\frac{N}{2}-1} x_1(n) W_{N/2}^{kn} + W_N^k \sum_{n=0}^{\frac{N}{2}-1} x_2(n) W_{N/2}^{kn} \end{aligned} \tag{2}$$

3) CNN MODEL

In order to extract the spatial features of the input data, we use CNN with two hidden layers and a full connection layer for training. The objective function is as follows:

$$\begin{cases} l_1 = \sigma(W_1 X_1^m + b_1) \\ l_2 = \sigma(W_2 l_1 + b_2) \\ \ldots \\ I^m = l_d = \sigma(W_d l_{d-1} + b_d) \end{cases} \tag{3}$$

where $l_1, l_2, \ldots, l_d$ are the output of each layer in CNN, W denotes the shared weight of neuron, σ is activation function. The I^m is obtained after the convolutional operation.

4) LOSS FUNCTION

This section introduces the loss function of the pre-training process, which is determined by KL divergence and cross entropy function, and is gradually reduced by Adam optimizer with a learning rate of 0.001. The specific calculation is illustrated in Ep.(4).

KL divergence works for representing the similarity between $I^1 - I^4$ and corresponding raw data I^0, and cross entropy loss is calculated to represent the similarity between the generated real and the predicted labels. These two types of losses are minimized simultaneously to ensure the separability of the self-supervised data and the aggregation of similar data. Algorithm 1 shows the process of the self-supervised pre-processing model.

$$\begin{aligned} \mathbf{L} &= \alpha * l_{kl} + (1-\alpha) * l_{cross}, \alpha \in [0, 1] \\ &= \alpha * \left(\sum_x p(x) \log \frac{p(x)}{q'(x)} \right) \\ &\quad + (1-\alpha) * \left(- \sum_x p(x) \log q(x) \right) \end{aligned} \tag{4}$$

where l_{cross} represents the classification loss of the data while l_{kl} represents the similarity of the self-supervised features and the raw data. $p(x)$ is the the true label sequence y, $q(x)$ is the probability of prediction label y' through softmax, $q'(x)$ is the prediction probability of the self-supervised feature label after softmax. α means the loss weight coefficient from 0 to 1. In the optimization process, Adam optimizer is applied to process the first and second moments of the gradients in order to reduce the loss quickly.

Algorithm 1 Self-Supervised Pre-Processing Model

Require: The sleep data, X_0;
Output: Fused features I^n extracted from the raw data;

1: Rotating the raw sleep data X_0 to $X_1^m, m = 0, 1, 2, 3$ according to different angles 0°, 90°, 180° and 270°.
2: Using FFT to describe the frequency information of the sleep data X_1^4;
3: Generating label $L_m, m = 0, 1, 2, 3, 4$ and extract feature I^m using a four-layer CNN;
4: Concatenating the feature vectors I^m to I^n;
5: **return** I^n;

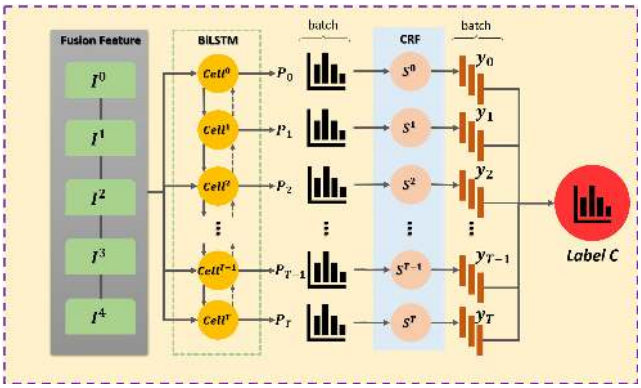


FIGURE 3. The downstream recognition task of the model. After concatenating features, we use the extended BiLSTM to model these time series, and the score probability of labels is used as the unnormalized emission probability in the CRF model for parameter learning. Moreover, the output of all BiLSTM will be the input of the CRF layer, and the final prediction result will be obtained by learning the order dependence information between labels.

B. DOWNSTREAM SLEEP RECOGNITION

1) TEMPORAL FEATURE MODELING

After obtaining the output feature $I^n = \{I^n \in \mathbb{R}^K, n = 0, 1, 2, \ldots, N\}$ of the upstream task, the downstream task uses BiLSTM to model it and input the integrated temporal features into the network under the expandable memory unit.

The standard recurrent neural network often ignores the future context information in the processing sequence. The bidirectional recurrent neural network proposes that each training sequence is two recurrent neural networks, one is forward and the other is backward, and both of them are connected with an output layer. This structure provides complete past and future context information for each point in the input sequence of the output layer. It is worth noting that there is no information flow between the forward and backward hidden layers, which ensures that the expansion graph is acyclic.

For the hidden layer of the bi-directional recurrent neural network, forward prediction is the same as that of an unidirectional recurrent neural network (RNN), except that the input sequence is in the opposite direction for the two hidden layers, and the output layer does not update until all the input sequences have been processed by the two hidden layers. The backward prediction of the bi-directional recurrent neural

network is similar to that of the standard RNN through time propagation. All the output layer activation functions are estimated first, and then return to two different directions of the hidden layer.

In this section, we improve the structural path of the internal expanded node of the bidirectional long-shot term memory (BiLSTM). The structure of the GRU node is simpler than that of the basic LSTM, but it ignores the consideration of front and rear time states. With contextual information from two directions, we can get more relevant spatio-temporal content at the current node.

The calculation process of the internal nodes is illustrated as follows:

$$\begin{cases} z_t = \sigma(W_z \cdot [O_{t-1}, I_t^n]) \;\; r_t = \sigma(W_r \cdot [O_{t-1}, I_t^n]) \\ \tilde{h}_t = \tanh(W \cdot [r_t \odot O_{t-1}, I_t^n]) \\ ctemp = \tanh(W_{ctemp} \cdot [O_{t-1}, I_t^n]) \\ c_t = (1 - z_t) \odot \tilde{h}_t + z_t \odot O_{t-1} \\ O_t = c_t \odot \sigma(ctemp) \\ P_t = g(W[\overrightarrow{O_t}, \overleftarrow{O_t}] + b) \end{cases} \tag{5}$$

where I_t^n and O_{t-1} are the current input and previous output in the LSTM node. σ and tanh are the activation functions. z_t denotes the output of the update gate at time step $t \in \{1, 2, 3 \ldots, T\}$, which determines whether or not the hidden state will be updated with hidden state $\tilde{h}$. r_t denotes the reset gate. $ctemp$ is a temporary state in order to determine I_t^n and O_{t-1} while c_t denotes the final state of the original node. BiLSTM is composed of two LSTMs stacked up and down calculating $P_t \in \mathbb{R}^{N \times C}$ (C is the number of classes and N is the sample number) as the final output of BiLSTM. Finally, the output P_t is obtained as the input of CRF.

2) SLEEP RECOGNITION

The conditional random field (CRF) [30] layer takes the output of the BiLSTM layer as the input, which can modify the output of the BiLSTM by learning the transfer probability between different labels in the dataset. BiLSTM extracts the features and inputs them to the conditional random field, calculates the likelihood of the label sequence as the loss, and then uses Viterbi method to predict the labels of the current batch [31].

The output dimension of the BiLSTM layer is $N \times C$, which is equivalent to the emission probability value of feature i mapped to class j. The output matrix of BiLSTM is P_t, where $P_{i,j}$ represents the non-normalized probability of the feature i mapped to class j. For CRF, we assume that there is a transition matrix A, then $A_{i,j}$ represents the transition probability of class i to class j.

For the output label sequence y corresponding to the input sequence X, the score is defined as:

$$s(X, y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=0}^{n} P_{i, y_i} \tag{6}$$

Given the feature X, the probability of getting the real label y is $p(y|X)$, and Y_X represents all possible label sequences

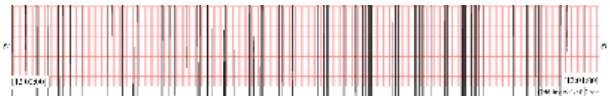


FIGURE 4. The input bioradiolocation data of one subject.

corresponding to X. In order to maximize the probability of X corresponding to the real tag sequence, the loss function $\log(p(y|X))$ needs to be minimized. The computing process is demonstrated in Eq.(7).

$$\begin{cases} p(y|X) = \dfrac{e^{s(X,y)}}{\sum_{\tilde{y} \in Y_X} e^{s(X,y)}} \\ \log(p(y|X)) = s(X, y) - \log(\sum_{\tilde{y} \in Y_X} e^{s(X,\tilde{y})}) \end{cases} \tag{7}$$

Algorithm 2 shows the process of the supervised temporal sleep recognition model.

Algorithm 2 Supervised Temporal Sleep Recognition Model

Require: The sleep time series data, I^n; Labels of sleep time series data, L_t;

Output: The classification result is the corresponding label C of the data;

1: Reshaping the data to (input*time step), and feeding them based on the expanded nodes in BiLSTM.
2: The predictions P_t is fed into CRF to calculate the log likelihood loss.
3: Using viterbi algorithm to predict label sequence L_p.
4: Computing the max score of BiLSTM, which is the output of the softmax function. $C_{label} = \arg\max(scores)$; **return** C_{label};

To sum up, the pre-training process of our proposed sleep recognition method can extract representative features, as well as expand the dataset, and successfully apply self-supervised learning in the field of sleep recognition. The downstream recognition task highlights the identification process based on time series features. The two tasks closely cooperate with each other to achieve better results than the existing methods.

IV. EXPERIMENT

The experiment is divided into three parts: data introduction, experimental settings and the results. The data used in the experiment is publicly accessible on the PhysioNet website. The existing methods for comparisons are based on the latest work of these sleep data.

A. DATASETS

Three sleep datasets were used in the experiment, including sleep bioradiolocation dataset (bioradar) [32], pressure map dataset [33] and polysomnography (PSG) dataset [34].

Sleep Bioradiolocation Dataset contained 32 records of non-contact sleep monitoring by a bioradar which was developed at the Remote Sensing Laboratory of Bauman Moscow State Technical University. The records were accompanied

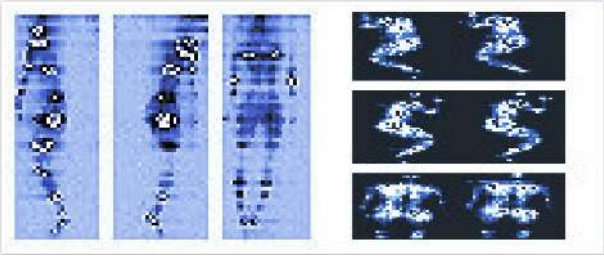


FIGURE 5. The in-bed posture pressure data in the dataset.

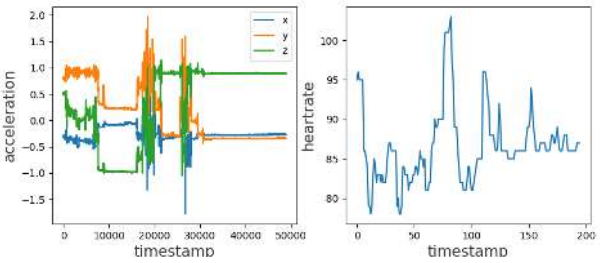


FIGURE 6. The acceleration data and heart rate of one subject in the dataset. The fluctuation value of x,y,z in the left image changes greatly twice, and the corresponding heart rate also shows in the center of the right image.

by results of sleep scoring and all subjects were free from sleep-disordered breathing and sleep related movement disorders. There were 4 subjects diagnosed as Insomnia. The type of bioradar is continuous waves by a quadrature receiver. The radar adopts the modulation mode of stepped frequencies from 3.6GHz to 4.0 GHz with the maximum emitted power of 3mW and the sampling rate of 50Hz.

Pressure Map Dataset were collected from adult participants using two types of pressure sensing mats, including various sleeping postures and pressure data from two separate experiments. And we also deal with the corresponding pressure data according to these two groups of experiments.

Experiment I: The pressure data was obtained from 13 participants in 8 standard postures and 9 additional states. Data was collected by a pressure mat - Vista Medical FSA SoftFlex 2048 with the size of 32*64. Its output figure was in the range of [0,1000] with a sampling rate of 1Hz. Each output file contained approximately 2 minutes (120 frames) of image frames. As shown in FIGURE 5, the three images on the right demonstrate the supine, right and left postures in this experiment, and there are two subjects in one sample.

Experiment II: The pressure data was obtained from 8 participants in 29 different states of 3 standard postures. Data was collected by two pressure mats (both sponge and air mattresses). The type of pressure mat was Vista Medical BodiTrak BT3510 with the size of 27*64. Its output figure was in the range of [0,500] for each sensor with a sampling rate of 1Hz. Each output file contained the average of around 20 frames. As illustrated in FIGURE 5, the three images on the left demonstrate the supine, right and left postures in this experiment.

PSG Dataset contained acceleration (in units of g) and heart rate (*bpm*, measured from photoplethysmography) recorded from the Apple Watch, as well as the labeled sleep data scored from gold-standard polysomnography. There were 31 subjects wearing Apple Watch to collect their ambulatory activity patterns for a week before spending one night in a sleep lab. Each type of data recorded from the Apple Watch and the labeled sleep from polysomnography was saved in a separate file, tagged with a random subject identifier.

In addition, there were 5 sleep stages in this dataset: Wake, NREM 1 (N1), NREM 2 (N2), NREM 3 (N3), REM (wake = 0, N1 = 1, N2 = 2, N3 = 3, REM = 5). We did different classifications based on the specific condition.

The features, acceleration and heart rate, of the sleep data are shown in FIGURE 6. We can see that the peaks of the two sets of data are similar.

B. EXPERIMENTAL SET-UP

This section introduced the implementation details and settings of the experiment. The experiments were conducted with Tensorflow [35] and Python [36] libraries. In addition, mean testing time for the best experimental results of the three datasets were 0.95s (bioradar), 0.76s (pressure map I), 0.12 (pressure map II), 0.66s (PSG) which ran on the system of GTX1050Ti GPU, i5-7300HQ CPU and 8G RAM. In order to reduce noise and differences presented in the sleep data, the data was normalized to [0, 1].

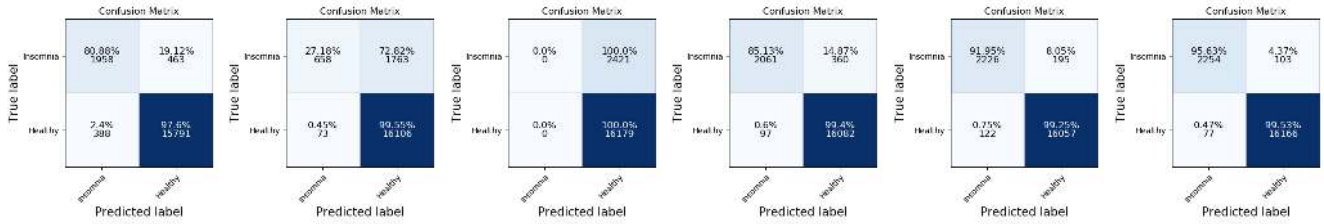
1) SLEEP BIORADIOLOCATION DATASET

For the sleep bioradiolocation dataset, we first undertook data preprocessing, combined all sleep data and created corresponding labels, which were divided into two classes: healthy subjects and insomnia.

In the pre-training process, we only fed 80% of the data into the upstream model for training, after frequency-domain feature extraction, temporary labels 0,1 (the number of workers was 2) were generated. The four-layer CNN was trained with the hidden layers of 32 and 64, kernel size (1,1), input-output layer of 4 and 2, dropout of 0.5, batch size of 128 and loss weight of 0.5. The downstream recognition model was trained by the BiLSTM-CRF with true labels, and the time step of 20, input dimension of 4 and learning rate of 0.001.

2) PRESSURE MAP DATASET

For the experiment I of the pressure map dataset, we reshaped the input dimension of each sample into 2048 due to the size of pressure mat was 32*64, and merge all files in the order of three sleep postures (supine, right side and left side). Because the posture dataset was in the form of gray images, we enabled rotation and frequency-domain feature extraction to generate temporary labels 0-4 (the number of workers was 5). The input and output layers of the four-layer CNN were 2048d and 5d, kernel size (5,5), the hidden layers were 128d and 64d, dropout was 0.5, batch size was 128, and loss weight



(a) DT (b) GBDT (c) LR (d) RF (e) KNN (f) SSRM

FIGURE 7. The confusion matrix of different classifiers compared with SSRM on sleep bioradiolocation dataset.

TABLE 1. Performance of classification of insomnia patients and healthy subjects compared with state-of-the-art models on sleep bioradiolocation dataset.

Classifier	DT	GBDT	LR	RF	KNN
Healthy	97.60%	99.55%	100.00%	99.40%	99.25%
Insomnia	80.88%	27.18%	0.00%	85.13%	91.95%
Deep Model	CNN	GRU	LSTM	BiLSTM	SSRM
Healthy	99.07%	98.52%	97.36%	98.46%	*99.53%*
Insomnia	89.01%	90.03%	89.95%	90.22%	*91.34%*

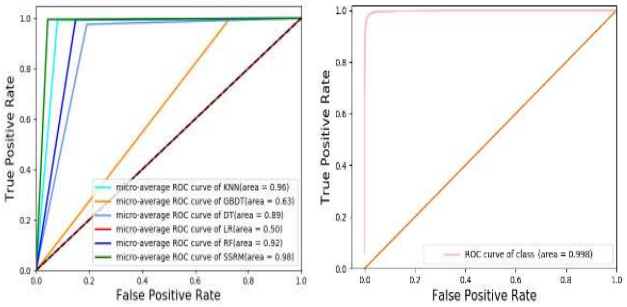
was 0.5. BiLSTM-CRF was trained with a learning rate of 0.001 with time step of 32 and the input dimension of 64.

In experiment II, we reshaped the input dimension of each sample into 2048 due to the size of the pressure mat was 27*64. Five workers were employed before training the CNN with 1728d (1728-dimension) of the input dimension, kernel size (5,5), 0.5 of the dropout, 128d of the hidden output and 5 of the final output dimension. Some parameters in BiLSTM-CRF were the time step of 27, input dimension of 64 and learning rate of 0.001.

3) PSG DATASET

In the PSG Dataset, there were three classifications. In order to remove the noisy data, we utilized the source code of [37] provided with the dataset to preprocess and extract four-dimensional features: cosine feature, heart rate, time feature and count feature. In the light of the settings of literature [34], we choose 70% data for training and 30% for testing.

To conduct the two-classification wake-sleep task, we fed the data into the self-supervised model by frequency-domain feature extraction. The four-layer CNN was trained with hidden layers of 64 and 32, kernel size (1,1). The dimensions of the input and output layers were 4 and 2, dropout was 0.8, batch size was 128 and loss weight was 0.5. The time step and input dimension of BiLSTM-CRF were 20 and 4. For the classification of wake, NREM and REM, the settings of the self-supervised model and BiLSTM-CRF remained unchanged, except that the final classification output dimension was 3. The third experiment was classification of wake, N1/N2, N3 and REM, with the final output dimension of 4.



(a) The ROC curve of models. (b) The ROC curve of classes.

FIGURE 8. The ROC curve of different models and classes on sleep bioradiolocation dataset.

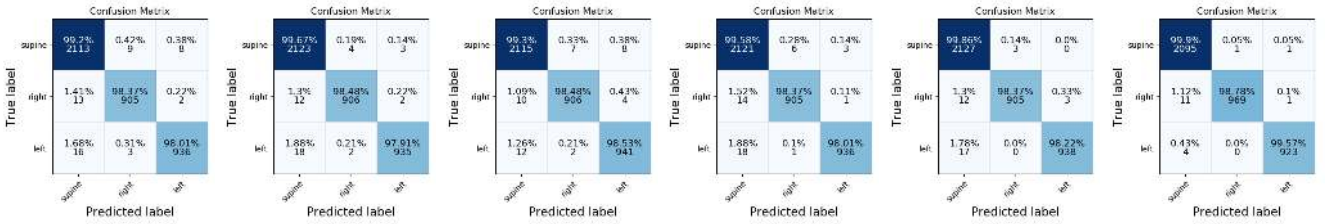
C. EXPERIMENTAL RESULTS

The experimental results of the three datasets will be illustrated in this section. We have compared the classification outcomes of different classifiers on these three datasets, random forest (RF), decision tree (DT), logical regression (LR), k-nearest neighbor (KNN), gradient boosting decision tree (GBDT), and the recognition rates of some popular deep learning frameworks, convolutional neural network (CNN), gate recurrent unit (GRU), long-shot term memory (LSTM), and BiLSTM.

1) RESULTS ON SLEEP BIORADIOLOCATION DATASET

The purpose of this experiment is to distinguish insomnia patients and healthy subjects. According to the experimental settings, the experimental results on this dataset are represented in FIGURES 7 and 8 by confusion matrix and ROC curve respectively. In FIGURE 7, we can see that the sample number of the healthy subjects is significantly larger than that of insomnia patients, which leads to the low classification accuracy and over-fitting of some classifiers. KNN and RF show superiority than other classifiers like LR. As shown in FIGURE 8, we calculate the micro-average value of the multiple classifiers, SSRM outperforms the others and AUC of ROC curve of two classification is 0.998, which proves that the model can reach the best classification performance.

The experimental results of the deep model are shown in TABLE 1, which demonstrates the classification accuracy of



(a) DT (b) GBDT (c) LR (d) RF (e) KNN (f) SSRM

FIGURE 9. The confusion matrix of different classifiers compared with SSRM on pressure map dataset I.

different models for insomnia patients and healthy subjects. Compared with other classifiers, KNN achieves the best result which is a little lower than SSRM. However, the performance of deep model like CNN and LSTM are not satisfactory, do not exceed KNN and other traditional classifiers due to the small dimension of the input features and the extracted features have poor separability. Over-fitting of LR and GBDT results in a low detection rate of insomnia. To conclude, apart from SSRM, the most accurate model for the classification of healthy subjects is RF, and for the detection of insomnia patients is KNN, which reflects the applicability of KNN and RF.

2) RESULTS ON PRESSURE MAP DATASET

The pressure map dataset includes two sub datasets. Experiments I and II are implemented according to these two datasets respectively. Dataset I contains more than 2×10^4 samples, and dataset II contains more than 400 samples.

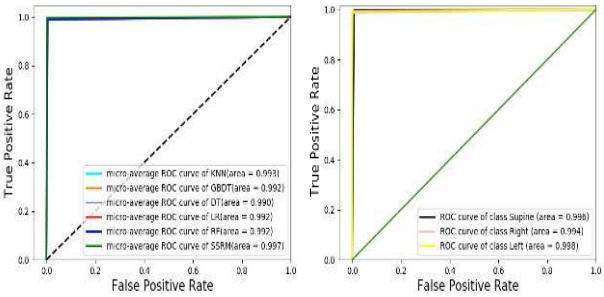
a: RESULTS OF THE CLASSIFICATION OF SLEEP POSTURE IN EXPERIMENT I

The task of this section is to distinguish three different sleep positions: supine, right side and left side. The confusion matrix and ROC curve of pose recognition are shown in FIGURES 9 and 10. In terms of the sample distribution of the three postures, supine posture accounts for the largest proportion, and its correct identification samples will also be more relative, leading to the accuracy higher than the other two postures. The classification results of all the classifiers and the deep models for the three postures are all over 95%, and the SSRM is 99.9%. The separability of the data is confirmed. FIGURE 10 illustrates that the ROC of the three postures are all over 99.9%, which verifies the stability of our proposed model. The AUC of SSRM is higher than that of other models.

In TABLE 2, for the classification of supine posture, KNN performs the best among the traditional classifiers and SSRM outperforms other deep models. LR, GBDT and SSRM are the best for right and left postures.

b: RESULTS OF THE CLASSIFICATION OF SLEEP POSTURE IN EXPERIMENT II

This section describes the recognition of sleep postures captured by different mattresses. The three postures are the same



(a) The ROC curve of models. (b) The ROC curve of classes.

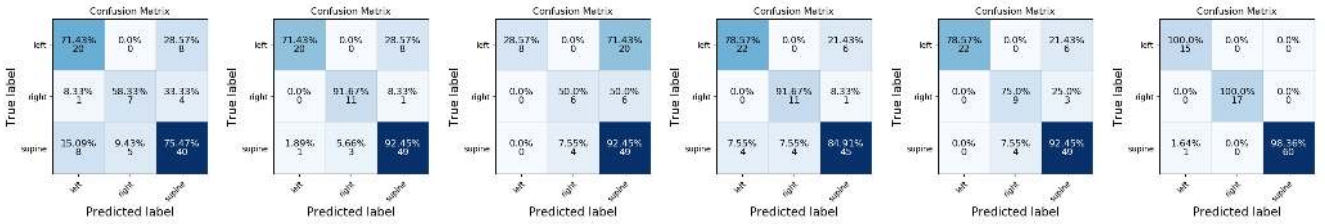
FIGURE 10. The ROC curve of different models and classes on pressure map dataset I.

TABLE 2. Performance of classification compared with state-of-the-art models on pressure map dataset I.

Classifier	DT	GBDT	LR	RF	KNN
Supine	99.20%	99.67%	99.30%	99.58%	99.86%
Right	98.37%	98.48%	98.48%	98.37%	98.37%
Left	98.01%	97.91%	98.53%	98.01%	98.22%
Deep Model	CNN	GRU	LSTM	BiLSTM	SSRM
Supine	99.72%	99.58%	99.51%	99.40%	*99.90%*
Right	98.38%	98.01%	98.07%	98.42%	*98.48%*
Left	97.93%	98.05%	97.89%	98.77%	*99.57%*

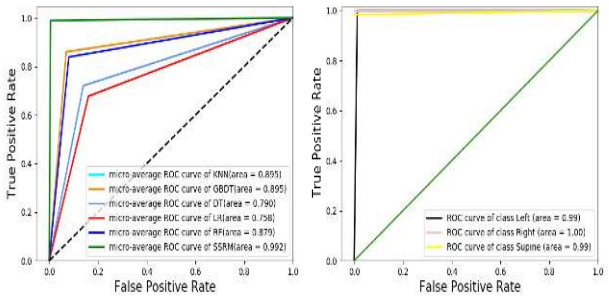
as those presented in the previous sections. As is illustrated in FIGURE 11, supine data accounts for the largest proportion. Because of the small number of samples, some classifiers can not determine a certain sleep posture very well. For example, the probability of LR recognizing left posture is only 28.57%, and the recognition rate of DT on right posture is 58.33%. SSRM uses self-supervised expansion method to surpass all the classification results, and it can recognize left and right posture samples better, which is preferable than the existing advanced methods. FIGURE 12 represents the performance of our proposed model, the AUC area of SSRM is larger than that of other classifiers reaching 99.2%.

TABLE 3 shows the recognition accuracy of all the learning models. For the classification of the supine position, SSRM and CNN are superior to other models because CNN can rely on convolution to check the pressure data collected by 27 * 64 mattresses for spatial feature analysis. Several different types of LSTM can extract the temporal features of sleep data without any fitting phenomenon. It is clear that



(a) DT (b) GBDT (c) LR (d) RF (e) KNN (f) SSRM

FIGURE 11. The confusion matrix of different classifiers compared with SSRM on pressure map dataset II.



(a) The ROC curve of models. (b) The ROC curve of classes.

FIGURE 12. The ROC curve of different models and classes on pressure map dataset II.

TABLE 3. Performance of classification compared with state-of-the-art models on pressure map dataset II.

Classifier	DT	GBDT	LR	RF	KNN
Supine	75.47%	92.45%	92.45%	84.91%	92.45%
Right	58.33%	91.67%	50.00%	91.67%	75.00%
Left	71.43%	71.43%	28.57%	78.57%	78.57%
Deep Model	CNN	GRU	LSTM	BiLSTM	SSRM
Supine	98.72%	90.25%	89.37%	95.49%	*98.36%*
Right	93.94%	96.85%	95.44%	96.69%	*100.00%*
Left	92.86%	97.79%	96.98%	97.60%	*100.00%*

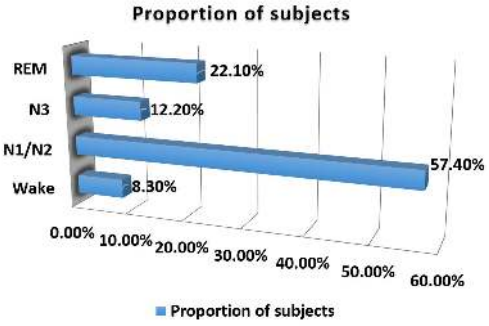
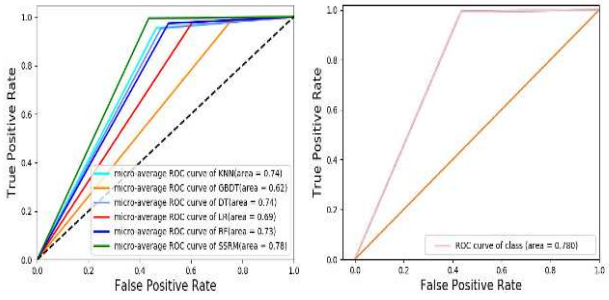


FIGURE 13. The proportion of the subject.

the input data features are rich and the separability is strong, resulting in better experimental results. Furthermore, SSRM has also expanded the data capacity and achieved outstanding results.



(a) The ROC curve of models. (b) The ROC curve of classes.

FIGURE 14. The ROC curve of different models and classes on PSG dataaset (Wake/Sleep).

TABLE 4. Performance of classification compared with advanced models on PSG dataaset (Wake/Sleep).

Classifier	DT	GBDT	LR	RF	KNN
Wake	51.76%	23.51%	39.54%	48.85%	53.59%
Sleep	95.25%	99.66%	97.70%	97.30%	95.35%
Deep Model	CNN	GRU	LSTM	BiLSTM	SSRM
Wake	8.8%	32.84%	36.25%	35.05%	*56.52%*
Sleep	98.27%	99.01%	98.94%	99.07%	*99.39%*

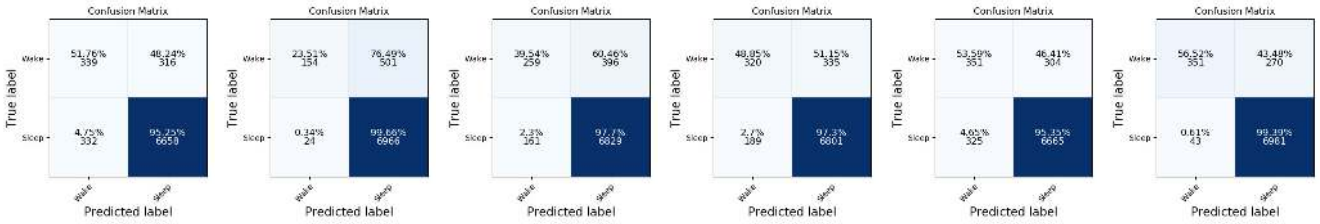
3) RESULTS ON PSG DATASET

In this experiment, one dataset is divided into different classes for sleep stage recognition. The classified objects of two-classification are wake and sleep, for three-classification are wake, NREM and REM, for four-classification are wake, N1/N2, N3 and REM. Non rapid eye movement (NREM) can be further divided into three sub stages: N1, N2 and N3 according to different brain waves. Generally speaking, N1 and N2 are light sleep, and N3 is deep sleep.

It should be noted that there is data imbalance in this dataset, which leads to the decline of recognition accuracy, as well as the inapplicability of some models. As is shown in FIGURE 13, N1 and N2 account for 57% of the samples, more than half of the whole dataset, while the proportion of wake is only 8%, which will directly lead to the over-fitting of the experimental result. SSRM method extends the experimental data and solves this problem to a certain extent.

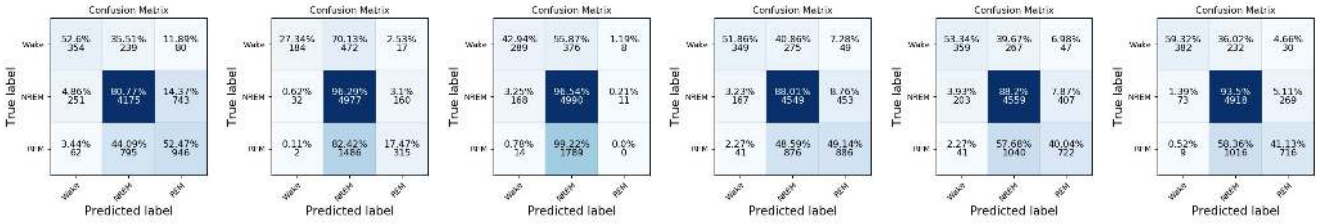
a: RESULTS OF THE CLASSIFICATION OF WAKE/SLEEP

For the classification of sleep and wake, we also use the traditional classifier and deep model to learn the sample features.



(a) DT (b) GBDT (c) LR (d) RF (e) KNN (f) SSRM

FIGURE 15. The confusion matrix of different classifiers compared with SSRM on PSG dataaset (Wake/Sleep).



(a) DT (b) GBDT (c) LR (d) RF (e) KNN (f) SSRM

FIGURE 16. The confusion matrix of different classifiers compared with SSRM on PSG dataaset (Wake/NREM/REM).

TABLE 5. Performance of classification compared with advanced models on PSG dataaset (Wake/NREM/REM).

Classifier	DT	GBDT	LR	RF	KNN
Wake	52.60%	27.34%	42.94%	51.86%	53.34%
NREM	80.77%	96.29%	96.54%	88.01%	88.20%
REM	52.47%	17.47%	0.00%	49.14%	40.04%
Deep Model	CNN	GRU	LSTM	BiLSTM	SSRM
Wake	19.12%	41.75%	30.21%	47.71%	*59.32%*
NREM	98.27%	90.05%	93.67%	91.04%	*93.50%*
REM	3.81%	27.83%	16.79%	25.06%	*41.13%*

The partial confusion matrix is shown in FIGURE 15, and the ROC curve FIGURE 14 also explains the outstanding effect of SSRM. The imbalance of data can be observed clearly. The recognition rate of SSRM to the wake data is higher than that of all the classifiers, and KNN to the wake state is superior to other classifiers. It can also be seen from the calculation of the micro-average AUC value that SSRM and KNN rank first and second.

In order to compare our method with other deep learning approaches, we illustrate the results in TABLE 4. For sleep classification, the accuracy of all the models are over 90%, while for wake classification, most of them are less than 50%, and the depth model CNN's accuracy is only 8.8%. There is severe over-fitting, and the average recognition rate of the time series model is also lower than the traditional model, which shows that the learning efficiency of the deep model is limited under the condition of uneven samples. Although SSRM can recognize the wake state better, there is a big gap between SSRM and sleep recognition rate, and there is a large space for improvement.

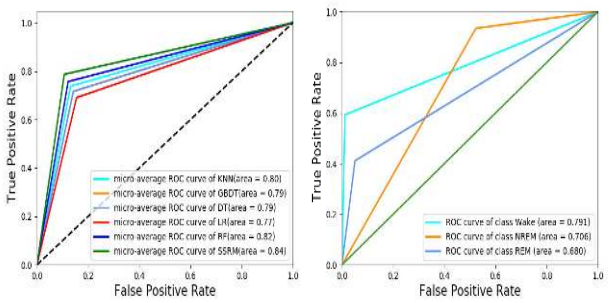
b: RESULTS OF THE CLASSIFICATION OF WAKE/NREM/REM

In terms of the three-classification of wake, NREM and REM, the largest number of the samples of NREM leads to the best classification of each classifier here. The experimental results are shown in FIGURES 17 and 16. Some classifiers have limitations, for example, the accuracy of LR and GBDT for classification of REM state is too low, and the samples with accurate classification are all focused on the NREM class, which is obviously over fitting. In FIGURE 17, the performance of SSRM is better than that of the other classifiers based on the value of micro-average AUC. Looking at the ROC curve of three states, the trend of NREM is different from that of other states due to the largest number of samples. The lowest AUC of REM state also indicates the lowest recognition accuracy.

TABLE 5 demonstrates the accuracy of the three stages of different model identification. For the deep models, the most stable performance is achieved by the temporal model BiLSTM, CNN appears the same situation as LR, because the recognition rate of REM is 3.81%. RF and KNN outperform all the deep learning models, which proves the advantages of the traditional classifier. Moreover, excluding the over-fitting model, the most of the excellent models for Wake, NREM and REM classification are SSRM, GBDT and RF respectively.

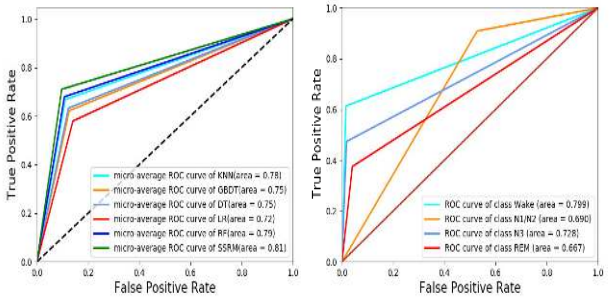
c: RESULTS OF THE CLASSIFICATION OF WAKE/N1,N2/N3/REM

The classification results are shown in FIGURES 18 and 19. Similar to the previous two experiments, each machine learning model has the highest classification accuracy for N1



(a) The ROC curve of models. (b) The ROC curve of classes.

FIGURE 17. The ROC curve of different models and classes on PSG dataaset (Wake/NREM/REM).



(a) The ROC curve of models. (b) The ROC curve of classes.

FIGURE 18. The ROC curve of different models and classes on PSG dataaset (Wake/N1N2/N3/REM).

TABLE 6. Performance of classification compared with advanced models on PSG dataaset (Wake/N1N2/N3/REM).

Classifier	DT	GBDT	LR	RF	KNN
Wake	50.39%	35.01%	47.88%	55.10%	57.14%
N1/N2	69.65%	93.46%	93.43%	79.78%	81.10%
N3	53.10%	13.46%	2.78%	49.25%	46.15%
REM	57.06%	18.15%	0.00%	59.14%	44.42%
Deep Model	CNN	GRU	LSTM	BiLSTM	SSRM
Wake	20.73%	42.33%	41.63%	42.72%	*61.34%*
N1/N2	99.04%	85.17%	86.30%	86.18%	*90.92%*
N3	0.00%	29.59%	29.78%	29.77%	*47.40%*
REM	0.00%	24.53%	23.76%	22.81%	*37.52%*

and N2, and the recognition rate for the other three categories is relatively average. But there are also models like LR experienced over-fitting problem that have no discrimination ability to identify REM stage. RF and KNN still perform well and steadily surpassing most models. In terms of AUC values, SSRM, RF and KNN sit on the top ranking places. For the trend of the ROC curve, N1/N2 class differes from the other three classes (Wake/N3/REM) due to the large quantity, the AUC value of REM ranks the last and the result is the worst due to the small amount of sleep samples.

The recognition rates of ten machine learning models for four sleep stages are shown in TABLE 6. In addition to the severely over fitted LR and CNN models, the models with the highest discrimination performance for wake, N1/N2, N3 and REM stages are SSRM, GBDT, DT and RF, respectively. The number of accurate samples identified by SSRM is the largest, and the overall recognition performance of SSRM will be explained in the next section.

4) PERFORMANCE OF THE SPLIT COMPONENTS IN SSRM

In this section, we describe the performance of SSRM by self-supervised pre-training and recognition process.

a: THE PRE-TRAINING PROCESS

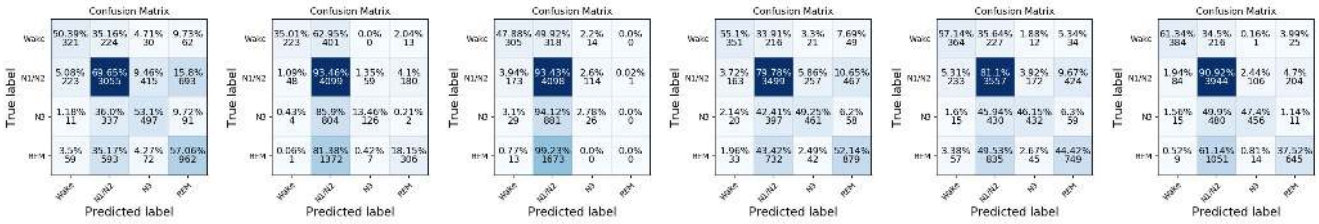
The self-supervised model is used to increase the data capacity, and the four-layer CNN is used as the training model to extract the original data features. We use the data visualization tool t-SNE (t-stochastic neighbour embedding) to reduce the high-dimensional data to 2 dimensions (dimension reduction), to compare the original features and the features extracted by self-supervised pre-training. Experiments show that the features extracted by self-supervised model are more separable, and the specific results are shown in FIGURE 20. Different colors represent different classes.

FIGURES 20(a) and 20(b) represent the sample distribution after dimensionality reduction of the original data and the features obtained after using the self-supervised model respectively on the bioradar dataset. It can be seen from the color distribution that FIGURE 20(a) has many scattered points of insomnia samples. FIGURE 20(b) aggregates these insomnia samples, indicating that the similarity between the same classes and the distance between different classes are increased. Because of the large number of samples (9×10^5 +), t-SNE costs much time. Insomniac samples are much smaller than healthy samples, so the sample distribution is not obvious.

FIGURES 20(c) and 20(d) show the comparison of data dimension reduction in experiment I of the pressure map dataset. Obviously, the correct boundaries of different sleep postures are clearer after using the self-supervised model. In experiment II, due to the small amount of data, we can see a significant aggregation effect. It is difficult to identify the three postures in FIGURE 20(e) when they are fused together, and it is easier to identify the class type of sample points in FIGURE 20(f) when they are close to the same class.

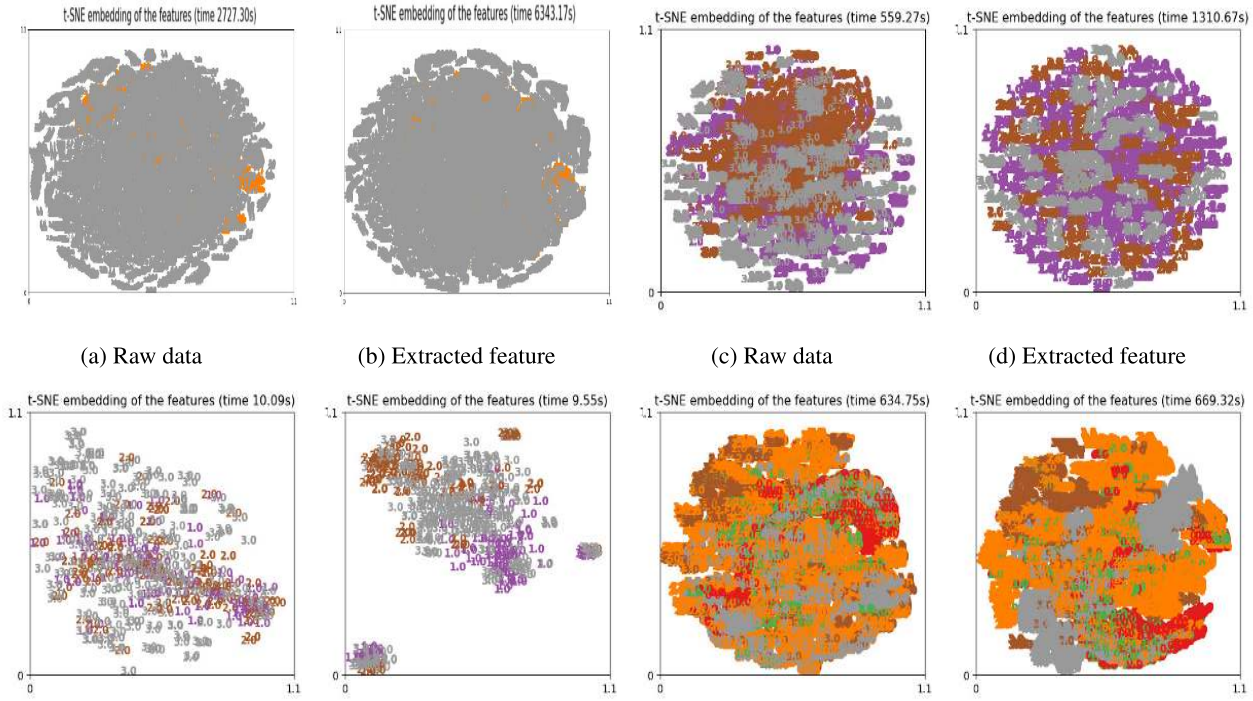
For the PSG dataset, we can see that the dimension reduction representation of the original data is chaotic, which makes all four colors merge together, indicating that its separability is poor. After the processing of the self-supervised model, the similar classes around have been generally aggregated, and the boundaries of different classes have become discriminative (FIGURES 20(g) and 20(f)).

These four experiments are enough to prove the effectiveness of the upstream self-supervised model. Although there is no true labels used for classification, the rotation and frequency-domain features can make the original data more separable and promote the implementation of the downstream tasks.



(a) DT (b) GBDT (c) LR (d) RF (e) KNN (f) SSRM

FIGURE 19. The confusion matrix of different classifiers compared with SSRM on PSG dataaset (Wake/N1N2/N3/REM).



(a) Raw data (b) Extracted feature (c) Raw data (d) Extracted feature

(e) Raw data (f) Extracted feature (g) Raw data (h) Extracted feature

FIGURE 20. Dimension reduction visualization of features extracted from the pre-training process. (a) and (b) represent the original data and extracted features on the bio-radar dataset. (c), (d) and (e), (f) show the results of experiment I and II on the pressure map dataset. (g) and (h) show the results of the four-class (wake/N1,N2/N3/REM) data on the PSG dataset.

b: DOWNSTREAM SLEEP RECOGNITION

Referring to the description in the above section, the downstream task employs BiLSTM-CRF to train and extract the supervised temporal series features, and regards the prediction probability of CRF as the final accuracy. The classification results of ten learning models on all the datasets are described in TABLE 7. SSRM stays ahead of other learning approaches on all the datasets, so we will discuss other models later.

For insomnia detection on the bio-radar dataset, KNN and GRU outperform the other models, which can represent the best results of traditional classifiers and deep learning models. GRU and BiLSTM among the deep models are slightly better than CNN, which shows that they are more suitable for continuous temporal sleep data.

For sleep posture recognition on the pressure map dataset, experiment I shows the notable data separability, which makes all the classifiers distinguish different sleep postures successfully. In experiment II, although the amount of the data is small, the average result of the deep model surpasses that of traditional model. Because each sample contains rich spatio-temporal features, deep model can be more sensitive to analyze non-linear spatio-temporal information.

For sleep stage recognition on the PSG dataset, the results of the four-classification are far inferior to those of the two-classification because of the data imbalance. Furthermore, the average classification outcome of KNN and RF is superior to that of all the deep models, which shows the advantages of the traditional classifier in dealing with the unbalanced data. However, the excellent results obtained by

TABLE 7. Performance of classification compared with advanced models.

Bio-radar		Pressure-e1		Pressure-e2		PSG-2class		PSG-3class		PSG-4class	
DT	95.42%	DT	98.73%	DT	72.04%	DT	91.52%	DT	71.62%	DT	63.24%
GBDT	90.12%	GBDT	98.98%	GBDT	86.02%	GBDT	93.13%	GBDT	71.63%	GBDT	62.18%
LR	86.98%	LR	98.93%	LR	67.74%	LR	92.71%	LR	69.05%	LR	57.93%
RF	97.54%	RF	98.93%	RF	83.87%	RF	93.15%	RF	75.66%	RF	67.89%
KNN	98.29%	KNN	99.12%	KNN	86.02%	KNN	91.77%	KNN	73.77%	KNN	66.74%
CNN	97.77%	CNN	98.98%	CNN	96.40%	CNN	69.80%	CNN	69.61%	CNN	56.98%
GRU	98.95%	GRU	98.20%	GRU	96.47%	GRU	90.36%	GRU	70.63%	GRU	62.35%
LSTM	97.12%	LSTM	98.65%	LSTM	95.17%	LSTM	93.60%	LSTM	70.77%	LSTM	61.44%
BiLSTM	98.44%	BiLSTM	99.03%	BiLSTM	96.62%	BiLSTM	93.72%	BiLSTM	71.72%	BiLSTM	60.72%
-		[38]	99.20%	[5]	97.90%	[34]	92.00%	[34]	73.00%	[34]	59.00%
SSRM	*99.03%*	SSRM	*99.55%*	SSRM	*98.92%*	SSRM	*95.91%*	SSRM	*78.69%*	SSRM	*71.01%*

SSRM verify the suitability of the deep learning algorithm on this dataset.

In addition, we compared the results of other studies on these three datasets, as shown in TABLE 7. Davoodnia and Etemad *et al.* [38] utilizes a convolutional neural networks (CNN) to classify sleeping posture, which consists of four main blocks is designed to convert the pressure map data manifold into a feature space. Matar *et al.* [5] extracted HoG and LBP features of sleep postures and trained a feed-forward artificial neural network for classification. Olivia *et al.* [34] adopted multiple classifiers for sleep stage recognition. Inspired by the combination of the traditional classifier and the deep network, the SSRM algorithm proposed in this paper has achieved satisfying results.

V. DISCUSSION

As mentioned above, SSRM has been proposed for sleep recognition and achieved remarkable results. In the three tasks, the results of insomnia detection and sleep posture recognition are much better than that of sleep stage detection. It can be seen that it is difficult to judge different sleep stages, because the boundaries of different sleep stages are not clear and there are certain gaps between different stages. In particular, the recognition rate of wake/N1,N2/N3/REM is only 71.01%, and there is still a lot of room for improvement.

The investigation of different sleep stages is helpful for researchers and medical experts to treat some insomnia disorders. By locating different sleep stages and studying the changes of brain waves in these stages, doctors can stimulate the neurons in the cerebral cortex and increase the duration of N3 (deep sleep stage) through non-drug and physical therapy, such as transcranial magnetic stimulation (TMS), low frequency impulse electrical therapy, etc.

VI. CONCLUSION

In this work, we have proposed a novel sleep recognition method enabling self-supervised approach for medical sleep data obtained from different sensors. To the best of our knowledge, the use of a self-supervised model for sleep recognition is a novel solution. We have considered sleep data from multiple sensors which is input to a BiLSTM-CRF classifier after performing self-supervision. To evaluate the quality of the learned features, a dimension reduction method t-SNE is adopted to regulate the feature representation of the self-supervised model and the original data, and show the recognition results of several state-of-the-art deep models and traditional classifiers, which verifies the effectiveness of our proposed model.

A number of open explorations still remain, for instance, the choice of using an image generation technology such as GAN to make artificial dreams to solve insomnia is an interesting future work. Monitoring different sleep stages and calculating a state transition probability to evaluate sleep quality is also an interesting future direction. Considering the deep learning algorithm based on multi-sensor prediction of sleep heart rate, blood pressure and other biological parameters, improving the prediction accuracy is also an interesting work in the future.

REFERENCES

[1] A. Crivello, P. Barsocchi, M. Girolami, and F. Palumbo, "The meaning of sleep quality: A survey of available technologies," *IEEE Access*, vol. 7, pp. 167374–167390, 2019.

[2] Y.-Y. Li, Y.-J. Lei, L. C.-L. Chen, and Y.-P. Hung, "Sleep posture classification with multi-stream CNN using vertical distance map," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, Jan. 2018, pp. 1–4.

[3] C. Sun, J. Fan, C. Chen, W. Li, and W. Chen, "A two-stage neural network for sleep stage classification based on feature learning, sequence learning, and data augmentation," *IEEE Access*, vol. 7, pp. 109386–109397, 2019.

[4] C.-T. Lin, M. Prasad, C.-H. Chung, D. Puthal, H. El-Sayed, S. Sankar, Y.-K. Wang, J. Singh, and A. K. Sangaiah, "IoT-based wireless polysomnography intelligent system for sleep monitoring," *IEEE Access*, vol. 6, pp. 405–414, 2018.

[5] G. Matar, J.-M. Lina, and G. Kaddoum, "Artificial neural network for in-bed posture classification using bed-sheet pressure sensors," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 101–110, Jan. 2020.

[6] M. B. Pouyan, J. Birjandtalab, M. Heydarzadeh, M. Nourani, and S. Ostadabbas, "A pressure map dataset for posture and subject analytics," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, 2017, p. 1.

[7] M. Enayati, M. Skubic, J. M. Keller, M. Popescu, and N. Z. Farahani, "Sleep posture classification using bed sensor data and neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, p. 1.

[8] S. Güne , K. Polat, and . Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7922–7928, Dec. 2010.

[9] J. Yang, J. M. Keller, M. Popescu, and M. Skubic, "Sleep stage recognition using respiration signal," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, p. 1.

[10] M. Schwaibold, R. Harms, B. Schöller, I. Pinnow, W. Cassel, T. Penzel, H. F. Becker, and A. Bolz, "Knowledge-based automatic sleep stage recognition-reduction in the interpretation variability," *Somnologie-Schlafforschung und Schlafmedizin*, vol. 7, no. 2, pp. 59–65, 2010.

[11] X. Xu, F. Lin, A. Wang, C. Song, Y. Hu, and W. Xu, "On-bed sleep posture recognition based on body-earth mover's distance," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2015, p. 1.

[12] X. Xu, F. Lin, A. Wang, Y. Hu, M.-C. Huang, and W. Xu, "Body-earth mover's distance: A matching-based approach for sleep posture recognition," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 5, pp. 1023–1035, Oct. 2016.

[13] M. Liu and S. Ye, "A novel body posture recognition system on bed," in *Proc. IEEE 3rd Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2018, pp. 38–42.

[14] Y.-Y. Li, Y.-J. Lei, L. C.-L. Chen, and Y.-P. Hung, "Sleep posture classification with multi-stream CNN using vertical distance map," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, Jan. 2018, p. 1.

[15] X. Hu, K. Naya, P. Li, T. Miyazaki, K. Wang, and Y. Sun, "Non-invasive sleeping posture recognition and body movement detection based on RFID," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul. 2018, pp. 1817–1820.

[16] L. Walsh, S. McLoone, J. Ronda, J. F. Duffy, and C. A. Czeisler, "Noncontact pressure-based Sleep/Wake discrimination," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1750–1760, Aug. 2017.

[17] M. Diykh, Y. Li, and P. Wen, "EEG sleep stages classification based on time domain features and structural graph similarity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 11, pp. 1159–1168, Nov. 2016.

[18] S. Khalighi, T. Sousa, and U. Nunes, "Adaptive automatic sleep stage classification under covariate shift," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 2259–2262.

[19] A. Vilamala, K. H. Madsen, and L. K. Hansen, "Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring," in *Proc. IEEE 27th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2017, pp. 1–6.

[20] E. Dafna, M. Halevi, D. Ben Or, A. Tarasiuk, and Y. Zigel, "Estimation of macro sleep stages from whole night audio analysis," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, vol. 1, no. 1, pp. 2847–2850.

[21] L. Mulaffer, M. Shahin, M. Glos, T. Penzel, and B. Ahmed, "Comparing two insomnia detection models of clinical diagnosis techniques," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 3749–3752.

[22] S. T.-B. Hamida, B. Ahmed, and T. Penzel, "A novel insomnia identification method based on hjorth parameters," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2015, pp. 548–552.

[23] H. Abdullah, T. Penzel, and D. Cvetkovic, "Detection of insomnia from EEG and ECG," in *Proc. 15th Int. Conf. Biomed. Eng.*, 2014, pp. 687–690.

[24] M. Shahin, L. Mulaffer, T. Penzel, and B. Ahmed, "A two stage approach for the automatic detection of insomnia," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 466–469.

[25] M. Ravanelli, J. Zhong, S. Pascual, P. Swietojanski, J. Monteiro, J. Trmal, and Y. Bengio, "Multi-task self-supervised learning for robust speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1–5.

[26] A. Valiūnienė, T. Sabirovas, J. Petronienė, and A. Ramanavi ius, "Towards the application of fast Fourier transform-scanning electrochemical impedance microscopy (FFT-SEIM)," *J. Electroanal. Chem.*, vol. 864, May 2020, Art. no. 114067.

[27] H. Poostchi and M. Piccardi, "BiLSTM-SSVM: Training the BiLSTM with a structured hinge loss for named-entity recognition," *IEEE Trans. Big Data*, early access, Aug. 28, 2019, doi: 10.1109/TBDATA.2019.2938163.

[28] P. Goyal, D. Mahajan, A. Gupta, and I. Misra, "Scaling and benchmarking self-supervised visual representation learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6391–6400.

[29] H. Lee, S. J. Hwang, and J. Shin, "Rethinking data augmentation: Self-supervision and self-distillation," 2019, *arXiv:1910.05872*. [Online]. Available: https://arxiv.org/abs/1910.05872

[30] H. Shen and J. Zhang, "Fully connected CRF with data-driven prior for multi-class brain tumor segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1727–1731.

[31] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, *arXiv:1508.01991*. [Online]. Available: http://arxiv.org/abs/1508.01991

[32] A. Tataraidze, L. Korostovtseva, L. Anishchenko, M. Bochkarev, Y. Sviryaev, and S. Ivashov, "Bioradiolocation-based sleep stage classification," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 2839–2842.

[33] M. B. Pouyan, J. Birjandtalab, M. Heydarzadeh, M. Nourani, and S. Ostadabbas, "A pressure map dataset for posture and subject analytics," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Informat. (BHI)*, 2017, pp. 65–68.

[34] O. Walch, Y. Huang, D. Forger, and C. Goldstein, "Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device," *Sleep*, vol. 42, no. 12, p. 12, Dec. 2019.

[35] Google. (2020). *Tensorflow Library*. [Online]. Available: https://tensorflow.google.cn

[36] Python. (2020). *Python*. [Online]. Available: https://tensorflow.google.cn

[37] O. Walch, "Motion and heart rate from a wrist-worn wearable and labeled sleep from polysomnography," PhysioNet, 2019, doi: 10.13026/hmhs-py35.

[38] V. Davoodnia and A. Etemad, "Identity and posture recognition in smart beds with deep multitask learning," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 3054–3059.

AITE ZHAO received the bachelor's degree in software engineering from the Qingdao University of Technology, in 2013. She is currently pursuing the Ph.D. degree with the College of Information Science and Engineering, Ocean University of China. She is a Visiting Ph.D. Researcher with the Department of Informatics, University of Leicester, Leicester, U.K. Her research interests include computer vision, pattern recognition, machine learning, data analysis, and robotics.

JUNYU DONG (Member, IEEE) received the B.Sc. and M.Sc. degrees from the Department of Applied Mathematics, Ocean University of China, in 1993 and 1999, respectively, and the Ph.D. degree from Heriot–Watt University, U.K., in November 2003. He is currently a Professor and the Vice-Dean of the College of Information Science and Engineering, Ocean University of China. He has published more than 100 journals and conference papers. His research interests include computer vision, underwater image processing, and machine learning, with more than ten research projects supported by the NSFC, MOST, and other funding agencies.

HUIYU ZHOU received the B.Eng. degree in radio technology from the Huazhong University of Science and Technology, China, the M.Sc. degree in biomedical engineering from the University of Dundee, U.K., and the Dr.Phil. degree in computer vision from Heriot–Watt University, Edinburgh, U.K. He is currently a Professor with the Department of Informatics, University of Leicester, U.K. He has authored over 250 peer-reviewed articles in his research field. His research work has been or is being supported by the U.K. EPSRC, MRC, EU, Royal Society, Leverhulme Trust, Puffin Trust, Invest NI, and the industry