Extracting Multi-Scale and Salient Features by MSE Based U-Structure and CBAM for Sleep Staging

Zhi Liu, Sixin Luo, Yunhua Lu, Yihao Zhang, Linfeng Jiang, and Hanguang Xiao

***Abstract*—According to the World Health Organization, more and more people in the world are suffering from somnipathy. Automatic sleep staging is critical for assessing sleep quality and assisting in the diagnosis of psychiatric and neurological disorders caused by somnipathy. Many researchers employ deep learning methods for sleep stage classification and have achieved high performance. However, there are still no effective methods to modeling intrinsic characteristics of salient wave in different sleep stages from physiological signals. And transition rules hidden in signals from one to another sleep stage cannot be identified and captured. In addition, class imbalance problem in dataset is not conducive to building a robust classification model. To solve these problems, we construct a deep neural network combining MSE(Multi-Scale Extraction) based U-structure and CBAM (Convolutional Block Attention Module) to extract the multi-scale salient waves from single-channel EEG signals. The U-structured convolutional network with MSE is utilized to extract multi-scale features from raw EEG signals. After that, the CBAM is used to focus more on salient variation and then learn transition rules between successive sleep stages. Further, a class adaptive weight cross entropy loss function is proposed to solve the class imbalance problem. Experiments in three public datasets show that our model greatly outperform the state-of-the-art results compared with existing methods. The overall accuracy and macro F1-score (Sleep-EDF-39: 90.3%-86.2, Sleep-EDF-153: 89.7%-85.2, SHHS: 86.8%-83.5) on three public datasets suggest that the proposed model is very promising to completely take place of human experts for sleep staging.**

***Index Terms*—Sleep stage classification, multi-scale extraction, convolutional block attention module, deep learning.**

I. Introduction

Good sleep can supplement the energy of the human body, increase body's resistance to disease, and enhance the quality of life. However, poor sleep will cause serious and long-term disease. In recent years, sleep problems have received much public attention. According to the research [1], 35.7% of people in the world are suffering from sleep disorders. More worrying, however, is that this number is increasing with the acceleration of life pace and social pressures. In addition, the function of 711 genes in the human body will be changed with less than 6 hours of sleep per night for a week, including metabolism, inflammation, immunity and stress resistance and so on. It can be seen that lack of sleep will not only lead to the physical diseases, such as cardiovascular disease, metabolic diseases, cancer, but also mental diseases such as depression and other psychiatric diseases.

In the study of sleep physiology, the classification of sleep stages is extremely important. Sleep researchers often use polysomnography (PSG) to study human brain activity during different stages of sleep. PSG includes electroencephalogram (EEG), electrocardiogram (ECG), electrooculography (EOG), myocardium Electrograms (EMG) and other biomedical records. Sleep experts use visual observation to label characteristic waves to classify sleep stages. At present, there are mainly two standards for sleep stages. One is proposed by Rechtschaffen and Kales (R&K) [2] in 1968, they divided Non-Rapid Eye Movement (NREM) into four steps (S1, S2, S3, S4) based on the changes in EEG and EOG. The other one is established by the American Academy of Sleep Medicine (AASM) [3] in 2004. In AASM standard, the S3 and S4 of NREM are combined based on the R&K sleep staging standard. Moreover, the effects of arousal, respiratory, cardiac and motor events on sleep quality were also supplemented. Although these rules can help sleep experts classify sleep stages, manual labeling is time-consuming and susceptible to subjective perception. Therefore, automatic sleep stage classification will be more efficient than that of manual one, and it also can exert important clinical value.

In earlier research, conventional machine learning methods such as Decision Trees [4], Random Forests [5], [6], [7] and Support Vector Machines [8], [9] are often used to classify sleep stages. And they extract features mainly from time domain signals [5], [8], frequency domain signals [6], [7] or nonlinear parameters [4], [9]. The performance of these models, however, is heavily reliant on the extracted features, and building feature extractors typically necessitates researchers with relevant domain knowledge. And unfortunately the feature extractors only can apply to some specific

Manuscript received 31 May 2022; revised 30 September 2022; accepted 13 October 2022. Date of publication 19 October 2022; date of current version 30 January 2023. This work was supported in part by the Natural Science Foundation of Chongqing, China, under Grant cstc2021jcyj-msxmX0605; in part by the National Natural Science Foundation of China under Grant 61971078 and Grant 61501070; and in part by the Science and Technology Foundation of Chongqing Education Commission under Grant KJQN202001137, Grant CQUT20181124, and Grant 2022CJZ042. *(Corresponding author: Hanguang Xiao.)*

The authors are with the School of Artificial Intelligence, Chongqing University of Technology, Chongqing 400050, China (e-mail: hgxiao@cqut.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2022.3216111

data. On the other hand, the non-linearity of EEG data, the differences in acquisition equipment, and the diversity of individuals, make model construction time-consuming and unsuitable for widespread use. With the breakthrough progress of deep learning in various fields, its learning ability without manual feature extraction has attracted much attention of researchers. For example, Yang et al. [10] used Convolutional Neural Networks (CNNs) to extract features from raw EEG, and then used Hidden Markov Model (HMM) to correct unreasonable classification of sleep stages. Results showed that the classification accuracy reaches 83.98%. In [11], the authors convert each raw signal into a time-frequency image using signal processing techniques, and then used a multi-task CNN to perform classification on the current stage epoch and prediction tasks on neighbouring epochs. In general, the above models achieved relatively good results in classifying sleep stages. However, most of them are unable to learn sleep transition rules effectively and cannot focus on salient waves in the raw EEG signal.

In order to learn sleep transition rules, some researchers began to use Recurrent Neural Network (RNN). Supratak et al. [12] proposed DeepSleepNet, which utilizes two CNNs for time-invariant feature extraction, and then bi-LSTM is utilized to learn the transition rules based on the extracted features. The classification accuracy of their models achieved 76.94%. In order to reduce the amount of model parameters, the author subsequently employed the hybrid model of CNN and RNN in TinySleepNet [13], and then the classification accuracy reached 85.4%. SleepEEGNet [14] adopted the CNN feature extraction framework in DeepSleepNet, and then two Bidirectional Recurrent Neural Networks(BiRNN) and attention mechanism were used as encoder-decoder and for classification, respectively, and finally the classification accuracy reached 84.3%. In SeqSleepNet [15], the authors firstly used Short-Time Fourier Transform (STFT) to process raw EEG signals into time-frequency images, and the sleep stages were then classified using a parallel convolutional network and a bidirectional RNN encoding the sleep sequence information. The results showed that the accuracy reached 86%. ResnetLSTM [16] used ResNet to extract features, and then LSTM was used for classification with the accuracy reaching 82.5%. Hogeon Seo et al. proposed the IITNet [17] model, which extracted representative features of sub-segments through Fourier transform and then classified the time-series data by analyzing their time correlations. IITNet firstly decomposed each half-minute EEG segment into an overlapping sub-segment and then encoded each sub-epoch to its corresponding representative feature. After that, modified ResNet-50 was used to extract features and then biLSTM was applied to classify sleep stages. However, due to the recurrent characteristic of RNN, the models based on RNN usually have great complexity, resulting in much training time and difficulty to adjust and optimize. Therefore, much efforts have been made to find new classifiers to replace RNN.

In order to overcome the shortcomings of RNN, researchers have developed some new methods. For example, AttnSleep [18] used convolutional neural networks to extract different features in EEG signals and then recalibrated them via an Adaptive Feature Recalibration module. Then, causal convolution and multi-head attention mechanism were used to extract transition rules from the captured features. In addition, the authors addressed the data imbalance problem by improving the loss function,but it only took effect for the N1 level. This model also could not fully capture the characteristic waves of physiological signals, and the classification accuracy was 84.4%. Zhu et al. [19] used CNN to extract local signal features and learn sleep transition rules using an attention mechanism, and then classified sleep stages. U-Time [20] took use of a fully convolutional encoder-decoder network to classify sleep stages from an input raw EEG signal of arbitrary length. The authors further optimized the model in U-sleep [21] and improved the performance. However, sleep transition rules are still not fully utilized. SalientSleepNet [22] adopted a dual-stream structure trained on EEG signals and EOG signals, which attempted to improve the accuracy of sleep stage classification by extracting features of different salient waves from EEG signals and EOG signals. Results showed that the classification accuracy reached 87.5%. However, the use of multi-channel physiological data would increase the model complexity and training time, and certain specific requirements were also required for data set collection. This model also dose not solve the data imbalance problem.

Although the above methods have been able to classify the sleep stages well, the challenges and difficulties are still not well resolved. According to AASM sleep standard, in physiological signals, there are different waveforms in different sleep stages. For instance, the waveform feature of N2 is spindle wave and K-complex wave, while N3 stage is δ wave. Therefore, it is important, also difficult to automatically capture the waveform multi-scale characteristics of different sleep stages to improve classification efficiency. Except observing the wave features of sleep stages, experts also can classify the sleep stage through analyzing its adjacent stages. In the process of sleep, this change between different sleep stages is considered as transition rules of sleep standards. However, the saliency of sleep transition rules is not fully extracted and exploited in many methods. Furthermore, the duration of different sleep stages is different for different people, and N2 stage generally occupies most of the sleep time, which resulting a serious imbalance in sleep time of different stages. Oversampling is often applied to balance the data, but it will lead to the increase of the training time.

In this paper, we propose a deep neural network based on the U-structure of MSE combined with CBAM, which can effectively capture multi-scale features of different sleep stages, learn sleep transition rules and solve the data set imbalance problem. The following are the paper's main contributions:

1) An end-to-end U-structured network is proposed to classify sleep stages. In the proposed network, U-structure with MSE is used to extract multi-scale features of salient waves in different sleep stages and CBAM is utilized to capture the saliency in transition rules between successive sleep stages from raw EEG signals.

2) A class adaptive weight cross entropy loss function is proposed to solve the problem of data imbalance without adding extra computation.

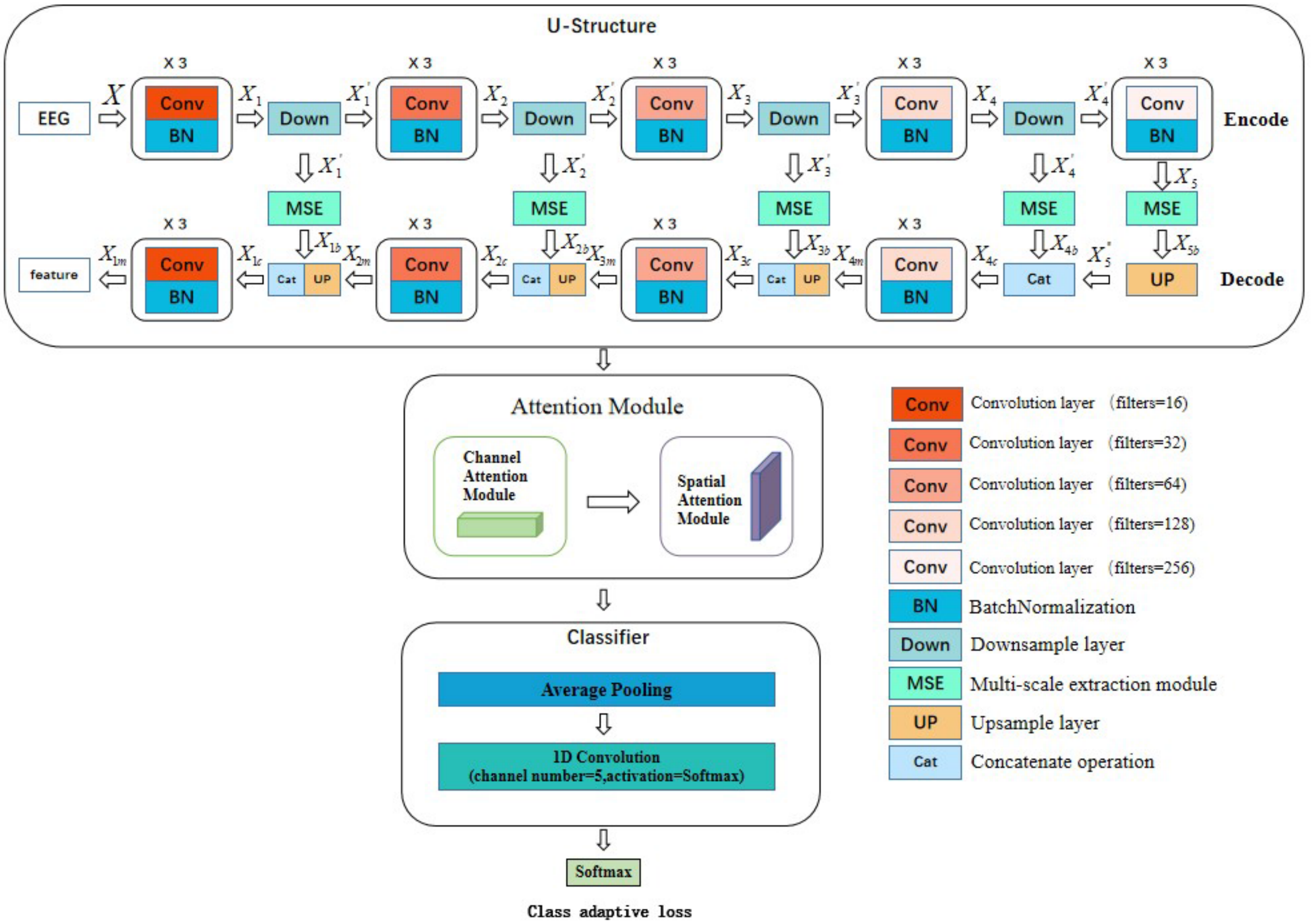


Fig. 1. The architecture Overview of the proposed model. It composed of U-structure with MSE, Attention Module, and the Segment-wise classifier. The U-structure with MSE is utilized to extract salient and multi-scale features from raw EEG signals. The depth of U-structure is 5(l = 5). The Attention Module is used to adaptively pay more attention to salient variation of wave pattern and then learn transition rules between successive sleep stages. The EEG signal is input to the U-Structure as X to get the output X_{1m}. Then X_{1m} will be input to the Attention Module, and the feature map obtained by Attention Module is input to the Classifier Module. Finally the result is output by Softmax.

3) The proposed model is extensively experimented on three public datasets; the results demonstrate that our model has a significant improvement in sleep stage classification and outperforms the state-of-the-arts.

II. The Proposed Approach

In this section, the components of our proposed sleep stage classification model based on single channel EEG data will be introduced, namely U-Structure, MSE, Attention Module, Segment classifier, and class adaptive weight cross entropy loss function.

A. Model Overview

As shown in Fig. 1, the proposed model includes three components, the Encode-Decode U-structure with MSE, attention module and classifier. Multi-scale features of salient waves are learned by U-structure and MSE which consists of dilated convolution with different-scale receptive fields and the bottleneck layer to reduce model parameters and lower computational costs. Then CBAM for channel attention and spatial attention is used to pay more attention to saliency of transition rules to improve classification accuracy. Finally, a segment-wise classifier is used to map the feature map to a sequence of predicted labels.

According to AASM and R&K, the EEG signal was taken as a segment of 30 seconds. Each sleep epoch is defined as $x \in R^n$, where n is the number of sampling points for 30 seconds. The proposed model inputs is X,where $X = \{x_1, x_1 \ldots x_L\}$ is a sequence of consecutive epochs, x_i $(i \in \{1, 2, \ldots, L\})$ is the target epoch and L is the number of input epochs. The proposed model maps a sequence of sleep segments X to a corresponding sequence of sleep stages Y, where $Y = \{y_1, y_1 \ldots y_L\}$ and y_i is the classification result of x_i. According to the AASM standard, each $y_i \in \{0, 1, 2, 3, 4\}$ matches each of the five sleep stages W, N1, N2, N3, and REM, respectively. Compared with other methods, our model is more flexible because any length of sleep segments can be input, which is similar to the process of labeling sleep stages by humans.

B. U-Structure

Sleep specialists often classify sleep stages based on salient waves in the EEG signal. In order to capture EEG salient

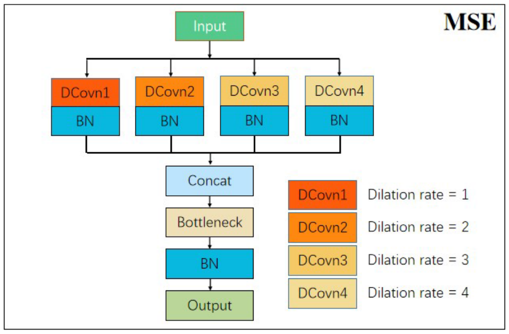


Fig. 2. The structure of Multi-Scale Extraction module (MSE). Dconvr is the dilation convolution; Concat: Concatenate operation; Bottleneck: Bottleneck is the Bottleneck layer; BN: BN is the Batch normalization.

waves and extract multi-scale features in the EEG signal, an encoder-decoder based U-shaped structure is designed, in which MSE module is combined.

1) Encoder: The encoder consists of five convolutional blocks, each of which has three convolutional layers, and batch normalization is performed after each convolutional layer to avoid gradients vanishing and speed up convergence. Inputting the given1D feature map X into a convolution block:

$$X_l = Conv_l(X'_{l-1}), l \in \{1, 2, 3, 4, 5\} \tag{1}$$

where X_l is the feature map output by the convolution block, l is the depth of the encoder($l = 5$ in our model), when $l = 0, X'_0 = X$.

Four of the layers are downsampling, and the pooling sizes are 10, 8, 6, and 4:

$$X'_l = Down(X_l), l \in \{1, 2, 3, 4\} \tag{2}$$

where X'_l is the output of the downsampling, and $Down$ is the downsampling operation.

2) Multi-Scale Feature Extraction: Inspired by the idea of Feature Pyramid Networks [23], shortcut in ResNet [24] and SalientSleepNet [22], MSE modules are introduced to directly connect corresponding layers of encoder and decoder in U-structure. Shortchut is effective for training deep convolutional models. And MSE can better boost U-structure to extract multi-scale features of salient waves with its dilated convolution. The bottleneck layer makes the shortcuts introduce few extra parameters and computation complexity. The MSE is shown in Fig. 2.

To obtain a multi-scale feature map, the multi-scale feature extraction module is designed to consist of four dilated convolutions with dilation rates ranging from 1 to 4. The feature maps learned from the different scales are concatenated and defined as:

$$X_d^r = Dconv_r(X'_l), r \in \{1, 2, 3, 4, 5\} \tag{3}$$

$$X_{ls} = Concat(X_d^1, X_d^2, X_d^3, X_d^4) \tag{4}$$

where X'_l is the input feature map, $Dconv_r$ is the dilated convolution of the dilation rate r. The output of $Dconv_r$ is X_d^r. Finally, the output feature map is X_{ls}, when $l = 5, X'_5 = X$. Then, bottleneck layer is added to MSE. It enables the channels of the input feature map to be reduced. Thus a large number of parameters in the model can be reduced, reducing computational costs. The bottleneck layer is defined as:

$$X_{lb} = Bottleneck(X_{ls}) \tag{5}$$

where X_{lb} is the multi-scale feature map obtained after the bottleneck layer operation. $Bottleneck$ is the operation of bottleneck layers.

3) Decoder: The decoder consists of four convolutional blocks and upsampling. An upsampling operation is performed before inputting in each convolution block, and three convolution operations are applied on each convolution block, then followed by batch normalization after each convolution. Then the multi-scale feature map output by MSE and the up-sampled output are concatenated as the input of the current layer:

$$X_{lm} = Cnov(X_{lc}), l \in \{1, 2, 3, 4\} \tag{6}$$

$$X''_l = Up(X_{lm}), l \in \{2, 3, 4, 5\} \tag{7}$$

$$X_{lv} = concat(X''_{l+1}, X_{lb}), l \in \{1, 2, 3, 4\} \tag{8}$$

where X''_l is the upsampling output and Up is the upsampling operation, when $l = 5, X_{5m} = X_{5b}$. X_{lc} is the feature map after connection, X_{lm} is the output of the convolution block in the decoder, and the final output feature map of the U-structure is X_{1m}.

C. Attention Module

In order to learn sleep transition rules, the model needs to pay attention to the overall trend of multiple sleep stages of the input, which is often easily ignored. For example, when the sleep stages cannot be classified, sleep specialists usually classify the current sleep stage according to the previous and subsequent sleep stages. In deep learning, the attention module solves this problem by enabling the model to focus more on what it needs to focus on. In the proposed model, we use the attention module to learn transition rules between sleep stages by adaptively paying more attention to salient variations in wave pattern hidden in EEG. Previous researchers [14], [18] [19], [22] often only utilized the channel attention and ignored the spatial features of the feature map. Therefore, a lightweight CBAM [25] combining channel and spatial attention module is used in the proposed model to improve its performance.

1) Channel Attention Module: The channel attention module can pay attention to the valuable information of the input and calculate the internal relationship between each channel. Performing max-pooling and average pooling operations on the input feature map can simultaneously compress the $C \times H \times W$ feature map to a size of $C \times 1 \times 1$, which is conducive to integrating information of each spatial channel and obtain finer features in the feature map. Then after a Shared MLP, the number of channels is compressed to C/r (reduction=16), and then expanded back to C. To generate the final channel attention feature, the shared MLP output features are summarized and then sigmoid activated, as follows:

$$F_c = \sigma(MLP(AvgPool(X_{1m})) + MPL(MaxPool(X_{1m}))) \tag{9}$$

TABLE I
NUMBER AND PROPORTION OF SLEEP STAGES FOR THE THREE DATASETS USED IN THE EXPERIMENT

Datasets	EEG Channel	W	N1	N2	N3	REM	Total Samples
Sleep-EDF-39	Fpz-Cz	8285 19.6%	2804 6.6%	17799 42.1%	5703 13.5%	7717 18.2%	42308
Sleep-EDF-153	Fpz-Cz	65951 33.7%	21522 11.0%	69132 35.4%	13039 6.7%	25835 13.2%	195479
SHHS	C4-A1	46319 14.3%	10304 3.2%	142125 43.7%	60153 18.5%	65953 20.3%	324854

where σ is the activation function and F_c is the final channel attention feature.

2) Spatial Attention Module: The spatial attention module is designed to focus on the location information of the target. The input of the spatial attention module is an element-wise summation of the output feature map of the channel attention module with the input feature map of the channel attention module. Then, two $H \times W \times 1$ feature maps are obtained using maximum pooling and global average pooling. After that, a channel-based concat operation is performed on both feature maps, and then the convolution operation is implemented to reduce the number of channel dimensions to one. After sigmoid activation, spatial attention feature map is generated, and finally this spatial attention feature map is multiplied by the feature map of input from the spatial attention module to obtain the final generated features, as follows:

$$F_s = \sigma(Conv(AvgPool(F_c); MaxPool(F_c)) \tag{10}$$

where σ is the sigmoid function and F_s is the final spatial attention feature.

D. Segment Classifier

Unlike models in other areas of computer vision, models for segment-wise classification of physiological signals have a continuous sequence of EEG signals as input. There is a need to capture local structural information between adjacent points.

In this study, a segmentation classifier in SalientSleepNet [22] is used to map pixel-level feature maps to segment-level prediction tag sequences. Firstly, an average pooling operation is performed on the 1D feature map, and $F_s \in R^{L \times n}$ is reshape into $F_{pool} \in R^L$, where L is the number of sleep stages, and n is the sampling point in a sleep stage. Then F_{pool} is subjected to the dimension reduction via a convolutional layer, and then scaled from 0 to 1 using the softmax function. Finally, it will be mapped to the predicted label sequence Y.

E. Class Adaptive Weight Cross Entropy Loss Function

As can be seen from Table I, the number of each category in the sleep staging dataset varies greatly from each other. An adaptive weight cross-entropy loss function is designed, which is an improvement on the cross-entropy loss function, to solve the data imbalance problem. The weights of the cross-entropy loss function can be adjusted adaptively based on the number of categories in the dataset. The following is the definition of the adaptive weight cross-entropy loss function:

$$Loss = -\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N}\eta_k yi^k \log(\overline{y}_i^k) \tag{11}$$

$$\eta_k = \frac{\prod_{p=1}^{k}(N_p/N_k)}{\sum_{q=1}^{k}\prod_{p=1}^{k}(N_p/N_q)} \tag{12}$$

where y_i^k is the probability of the ground truth of the i-th sample, $\overline{y}_i^k$ is the predicted probability of the i-th sample, N is the total number of samples, K is the number of categories, η_k is the weight of the k-th category, and N_k is the data amount of the k-th category.

Through the above formula, the following relationship can be obtained between each class:

$$\eta_1 \times N_1 = \eta_2 \times N_2 = \cdots = \eta_k \times N_k \tag{13}$$

$$\eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5 = 1 \tag{14}$$

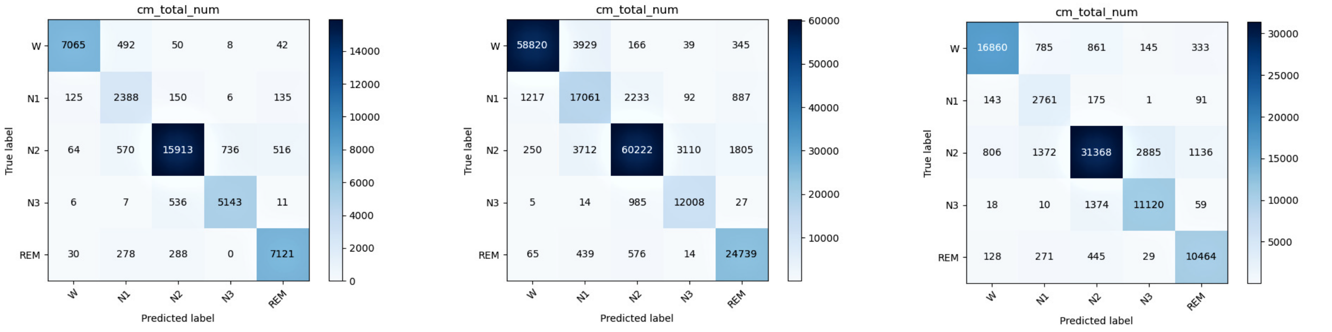
From formulas (13) and (14), it can be concluded that the weight of each category depends on its corresponding number of samples. In this way, it will not introduce complexity of the model, while achieves better training and testing performance for each category.

III. EXPERIMENT AND RESULTS

In this section, three public available datasets used in the experiments are introduced. For each dataset, only one EEG channel is used for experiments. Extensive experiments are conducted on all these three datasets to demonstrate the validity of the model.

A. EEG Datasets and Preprocessing

Sleep-EDF [26] is obtained from a 1987–1991 study on the effect of sleep in healthy whites aged 25-101 years without any sleep-related drugs. In this study, PSG recordings from 20 healthy subjects (aged 25-34 years) of corresponding to the age effect is used. SleepEDF-153 is an extended version consisting of PSG recordings from 78 healthy subjects aged 25-101 years. In both datasets, each subject has two diurnal PSG recordings. Each PSG recording contains two scalp EEG signals from the Fpz-Cz and Pz-Cz channels, one EOG channel (horizontal), one chin EMG channel and one oronasal respiration signal. All EEG and EOG are sampled at 100 Hz. For each document, sleep experts have manually scored these records according to R&K standard. It is noteworthy that the



Sleep-edf-39 Sleep-edf-153 SHHS

Fig. 3. Visualization of the experimental confusion matrix.

TABLE II
RESULTS OF THE SLEEP-EDF-39 AND SLEEP-EDF-153 DATASETS COMPARED WITH PREVIOUS METHODS

Method	Sleep-EDF-39							Sleep-EDF-153						
	overall results		F1-score for each class					overall results		F1-score for each class				
	MF1	Accuracy	W	N1	N2	N3	REM	MF1	Accuracy	W	N1	N2	N3	REM
DeepSleepNet	76.9	82.0	85.0	47.0	86.0	85.0	82.0	75.3	78.5	91.0	47.0	81.0	69.0	79.0
SeqSleepNet	79.7	86.0	91.9	47.8	87.2	85.7	86.2	78.2	83.8	92.8	48.9	85.4	78.6	85.1
SleepEEGNet	79.7	84.3	89.2	52.2	86.8	85.1	85.0	77.0	82.8	90.3	44.6	85.7	81.6	82.9
ResnetLSTM	73.7	82.5	86.5	28.4	87.7	89.8	76.2	71.4	78.9	90.7	34.7	83.6	80.9	67.0
MultitaskCNN	75.0	83.1	87.9	33.5	87.5	85.8	80.3	72.8	79.6	90.9	39.7	83.2	76.6	73.5
SleepUtime	79.0	–	87.0	52.0	86.0	85.0	82.0	76.0	–	92.0	51.0	84.0	75.0	80.0
TinySleepNet	80.5	85.4	90.1	51.4	88.5	88.3	84.3	78.1	83.1	92.8	51.0	85.3	81.1	80.3
AttnSleep	77.7	84.4	89.7	42.6	88.8	90.2	79.0	75.1	81.3	92.0	42.0	85.0	82.1	74.2
SalientSleepNet	83.0	87.5	92.3	56.2	89.9	87.2	89.2	79.5	84.1	93.3	54.2	85.8	78.3	85.8
Proposed Method	**86.2**	**90.3**	**94.5**	**73.0**	**91.6**	**88.7**	**91.6**	**85.2**	**89.7**	**95.1**	**73.1**	**90.4**	**84.9**	**92.2**

EEG signals in this dataset are all based on the Fpz-cz channel and thus do not need any further preprocessing.

SHHS [27], [28] is a multicentre cohort study in which subjects with various diseases were selected to study the effects of sleep breathing disorders on cardiovascular disease and other diseases. In previous research [18], [29], the author selected 329 subjects who had regular sleep in order to reduce the effects of the disease. We have followed these studies and we also used EEG data from these subjects for our experiments. In our experiments, we chose the C4-A1 channel with a sampling rate of 125 Hz.

To compare with other methods, we treated the three datasets in the same way as the previous study, merging the N3 and N4 phases into one N3 phase and removing the motion and unknown phases. Each PSG file contains a large number of wake-up periods, while we only focus on the sleep periods, and thus only the 30 minutes records before and after sleep are kept.

B. Experiment Settings

Our model is implemented based on TensorFlow 2.3 and trained it on an NVIDIA GeForce RTX3080Ti. Adam optimizer trained our model with a learning rate of $\eta = 10^{-3}$. The batch size is 8 and the training epoch is 100. The length of the input sleep epoch sequence is 20 ($L = 20$). In addition, the downsampling rate of the bottleneck layer is 4. The subjects in each dataset are divided into 20 groups and 20-fold cross-validation is used to evaluate our model. The average value of the predicted sleep stages of all 20 test samples is calculated. Fianlly, various performance metrics are obtained.

C. Experiment Results

As shown in Tables II and III, the proposed model with state-of-the-art methods on three public datasets are compared. Table II shows the comparable results of the proposed model with other methods on Sleep-EDF-39 and Sleep-EDF-153. Table III illustrates the results of the proposed model with other methods on SHHS. In comparison to other methods, our models have achieved huge improvements on accuracy and F1 scores. Because the N1 class is frequently misclassified as the W, REM, and N2 classes, stage N1 has the lowest performance, and the F1 scores of N1 in the other methods are all below 50%. In the proposed model, the F1 score of N1 is greater than 70% in Sleep-EDF-39 and Sleep-EDF-153, and greater than 60% in SHHS. It shows that the designed class adaptive loss function is useful for dealing with data imbalance problem. After using the 20-fold method, the confusion matrices of the experimental results for each dataset are visualized in Fig. 3. Each square represents the number of that stage, with the darker the color indicating the higher the number. On these three datasets, it can be seen that the proposed model outperforms existing SOTA methods significantly.

Although models such as DeepSleepNet, SeqSleepNet, ResnetLSTM, and TinySleepNet can utilize CNN and RNN to capture the salient waves patterns and transition rules between sleep stages, they contain extensive parameters, resulting in difficulty to adjust and optimize. Furthermore, ResnetLSTM, SleepEEGNet, IITNet and MultitaskCNN require time-frequency images as input, however using signal processing techniques on physiological signals can result in some information loss. In terms of detecting salient waves patterns

TABLE III
RESULTS OF THE SHHS DATASETS COMPARED WITH PREVIOUS METHODS

Method	SHHS						
	overall results		F1-score for each class				
	MF1	Accuracy	W	N1	N2	N3	REM
DeepSleepNet	73.9	81.0	85.4	40.5	82.5	79.3	81.9
SleepEEGNet	68.4	73.9	81.3	34.4	73.4	75.9	77.0
ResnetLSTM	69.4	83.3	85.1	9.4	86.3	87.0	79.1
MultitaskCNN	71.2	81.4	82.2	25.7	83.9	83.3	81.1
AttnSleep	75.3	84.2	86.7	33.2	87.1	87.1	82.1
IITNet	78.8	86.3	88.7	21.3	86.1	84.9	78.1
Proposed Method	**83.5**	**86.8**	**91.3**	**66.0**	**87.4**	**83.1**	**89.4**

of physiological signals and resolving the data imbalance problem, our model is superior to the SalientSleepNet, which uses EEG and EOG as dual-channel U^2 structure flow. Our model uses U-Structure based on encoder-decoder and dilated convolution in the MSE module to capture multi-scale features in the EEG signal. More importantly, the dual attention modules can pay spatial and channel attention simutaneously to salient variation of wave pattern and then learn transition rules between successive sleep stages. In addition, the weights of loss function will be dynamically balanced according the number of samples in each class (sleep stage) by introducing the proposed class adaptative weight cross-entropy loss function. So that the data in each sleep stage can be fully valued and the model can be trained well. The large improvement in F1-score for each class also demonstrates that our proposed class adaptative weight cross-entropy loss function can enhance the training of the model for each class very well. In summary, the overall performance of our model is clearly superior to that of other methods.

D. Visualisation of the Results of the Classification

In this study, the t-SNE (t-distributed Stochastic Neighbor Embedding) method [30] is employed to visualize the results of the model classification to demonstrate the validity of the proposed model. In Fig. 4, it is evident that the points representing each type of sleep stage are clearly differentiated, demonstrating the strong ability of our model to extract significant waveform features. Also, it can be seen that the number of misclassification of sleep stage is small, and most of the them are related to the N2 class, as it is the majority class.

E. Ablation Experiments

Several different models are designed to evaluate the effectiveness of each module in the model, as follows:

U-structure: U-structure module only.

U-structure + MSE: Based on the U-structure module, a multi-scale feature extraction module is added.

U-structure+MSE+CBAM: The CBAM module added to U-structure and MSE.

Proposed Method: This model combine the U-structure, the multi-scale feature extraction module, the CBAM module and class adaptive weight cross entropy loss function for training.

Table IV shows the results of the ablation experiments. When MSE is added to the U-structure, the accuracy is greatly

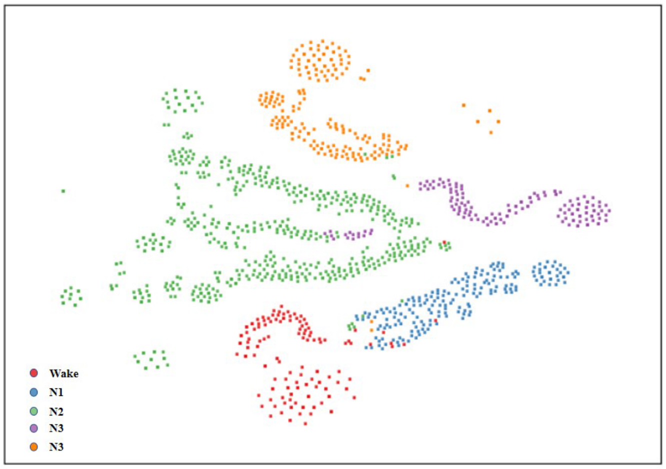


Fig. 4. Results of probabilistic post-processing for each class extracted from proposed model, with different colors and labels denoting different stages of sleep (W, N1, N2, N3, R).

TABLE IV
THE RESULTS OF ABLATION EXPERIMENT

Method	Accuracy	MF1
SalientSleepNet(SOTA)	87.5	83.0
U-structure	85.5	80.1
U-structure+MSE	87.8	84.0
U-structure+MSE+CBAM	88.0	84.1
Proposed Method	**90.3**	**86.2**

improved, indicating that MSE can effectively learns the multi-scales features. After adding the CBAM module, the accuracy is further enhanced because the CBAM can adaptively pay more attention to salient variation of wave pattern. Finally, the use of class adaptive weight cross entropy loss function solves the problem of data imbalance well, making the model achieve a further higher performance.

In order to demonstrate the effectiveness of the class adaptive weight cross entropy loss function, experiments are executed on the Sleep-EDF-39 dataset. The performance using various weighted cross entropy loss functions are compared with our proposed loss function. They are conventional weighting, the uniform weight and the Class-Aware Loss Function proposed in AttnSleep [18]. As shown in Table V, the F1-score when using class adaptive weight cross entropy loss function exceeded that of all other weighted cross-entropy loss functions in each class. Experimental results demonstrate that proposed adaptative loss function can effectively solve the problem of data imbalance, which is mainly because the class adaptive weight cross entropy loss function is capable

TABLE V
RESULTS OF DIFFERENT WEIGHTING METHODS

Weighting Method	Accuracy	F1-score for each class				
		W	N1	N2	N3	REM
conventional weight	87.6	92.2	58.7	90.3	87.5	88.0
uniform weight	88.4	91.3	63.2	90.4	86.7	91.2
Class-Aware weight	88.5	93.2	63.9	90.4	86.7	90.5
Class adaptive weight	**90.3**	**94.5**	**73.0**	**91.6**	**88.7**	**91.6**

of making the model well trained for each class. Moreover, the accuracy of the proposed loss function reaches 90.3, suggesting its superiority than other weighted cross entropy loss function in resolving data imbalance problem.

IV. CONCLUSION

In this paper, an end-to-end deep neural network is proposed to classify sleep stages only from raw EEG signal. The MSE based U-structure is used to extract multi-scale features for salient waves in sleep stages. To learn transition rules between successive stages, CBAM module is utilized to pay more attention to channel and spatial feature simultaneously. In order to reduce the influence of data imbalance on model training, a class adaptative loss function is designed to balance the contribution to loss according to the ratio of different sleep stages. The results show that our model achieves state-of-the-art performance. Specifically, the classification accuracy on Sleep-edf-39, Sleep-edf-153 and SHHS data set reaches 90.3%, 89.7%, and 86.8% respectivly. Ablation Experiments on Sleep-edf-39 data set show that MSE, CBAM and adaptative loss function can all contribute to improve the performance of sleep stage classification. The proposed model is expected to free sleep specialists from heavy sleep staging, which is important in the diagnosis, of psychiatric and neurological disorders, such as depression, anxiety disorder, insomnia, etc.. However, in this study, the proposed model failed to consider noise reduction of physiological signals to obtain higher performance, which will be a focus of our future research.

REFERENCES

[1] H. Jahrami, A. S. Bahammam, N. L. Bragazzi, Z. Saif, and M. V. Vitiello, "Sleep problems during the COVID-19 pandemic by population: A systematic review and meta-analysis," *J. Clin. Sleep Med.*, vol. 17, no. 2, pp. 299–313, 2020.

[2] E. A. Wolpert, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Arch. General Psychiatry*, vol. 20, no. 2, pp. 246–247, 1969.

[3] R. B. Berry et al., "The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications," Am. Acad. Sleep Med., Darien, IL, USA, 2012, p. 2012, vol. 176.

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

[5] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel EEG signal," *Comput. Biol. Med.*, vol. 42, no. 12, pp. 1186–1195, 2012.

[6] L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," *Comput. Methods Programs Biomed.*, vol. 108, no. 1, pp. 10–19, 2012.

[7] A. R. Hassan and M. I. H. Bhuiyan, "Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating," *Biomed. Signal Process. Control*, vol. 24, pp. 1–10, Feb. 2016.

[8] A. R. Hassan and M. I. H. Bhuiyan, "A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features," *J. Neurosci. Methods*, vol. 271, pp. 107–118, Sep. 2016.

[9] G. Zhu, Y. Li, and P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 6, pp. 1813–1821, Nov. 2014.

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

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

[12] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.

[13] A. Supratak and Y. Guo, "TinySleepNet: An efficient deep learning model for sleep stage scoring based on raw single-channel EEG," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 641–644.

[14] 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, May 2019, Art. no. e0216456.

[15] H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 400–410, Mar. 2019.

[16] Y. Sun, B. Wang, J. Jin, and X. Wang, "Deep convolutional network method for automatic sleep stage classification based on neurophysiological signals," in *Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2018, pp. 1–5.

[17] H. Seo, S. Back, S. Lee, D. Park, T. Kim, and K. Lee, "Intra- and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG," *Biomed. Signal Process. Control*, vol. 61, Aug. 2020, Art. no. 102037.

[18] E. Eldele et al., "An attention-based deep learning approach for sleep stage classification with single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 809–818, 2021.

[19] T. Zhu, W. Luo, and F. Yu, "Convolution- and attention-based neural network for automated sleep stage classification," *Int. J. Environ. Res. Public Health*, vol. 17, no. 11, p. 4152, Jun. 2020.

[20] M. Perslev, M. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–12.

[21] M. Perslev, S. Darkner, L. Kempfner, M. Nikolic, P. J. Jennum, and C. Igel, "U-sleep: Resilient high-frequency sleep staging," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–12, Dec. 2021.

[22] Z. Jia, Y. Lin, J. Wang, X. Wang, P. Xie, and Y. Zhang, "SalientSleepNet: Multimodal salient wave detection network for sleep staging," 2021, *arXiv:2105.13864*.

[23] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

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

[27] G.-Q. Zhang et al., "The national sleep research resource: Towards a sleep data commons," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1351–1358, 2018.

[28] S. F. Quan et al., "The sleep heart health study: Design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, 1997.

[29] P. Fonseca, N. den Teuling, X. Long, and R. M. Aarts, "Cardiorespiratory sleep stage detection using conditional random fields," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 4, pp. 956–966, Jul. 2016.

[30] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.