An Attention-based Deep Learning Approach for Sleep Stage Classification with Single-Channel EEG - Supplementary material

S.I. Activation functions

We provide the results of experiments using different activation functions in the MRCNN module other than GELU. We notice degradation in the performance of ReLU due to the suppression of the negative weights in the MRCNN module, which denies the AFR of making use of them.

Additionally, we also notice a degradation in the performance with the use of Leaky-ReLU, which allows strong negative activations to generate undesirable impact on the sum of activations feeding the next layers.

TABLE S.1: Comparison of the performance of AttnSleep with using different activation functions in MRCNN module applied on Sleep-EDF-20 dataset.

	Per-Class F1-score					Overall Metrics			
Activation	W	N1	N2	N3	REM	ACC	MF1	κ	MGm
ReLU	**89.9**	42.0	88.2	89.0	78.9	83.8	77.7	0.78	85.4
Leaky-ReLU	88.7	42.3	88.4	89.5	78.7	83.7	77.5	0.78	85.4
GELU	89.7	**42.6**	**88.8**	**90.2**	79.0	**84.4**	**78.1**	**0.79**	**85.5**

S.II. Loss function parameters

In this section, we provide the results of several experiments performed to show the effect of the different variants of a, b, and c values, as shown in Table S.1. We conclude that the choice of the values itself does not have a considerable effect on the performance as long as they follow the above recommendation, and their values are not equal to each other or greater than K. Note that we choose $a = 1, b = 1.5$, and $c = 2$ in this paper as it achieved the best performance.

TABLE S.2: Different variants of (a,b,c) values used in our loss applied on Sleep-EDF-20 dataset.

	a	b	c	ACC	MF1
Different	0.5	1	1.5	84.3	78.0
	1	2	3	84.2	78.0
	1	1.5	2	**84.4**	**78.1**
	2	2.5	3	84.0	77.9
Similar	1	1	1	83.8	77.7
	2	2	2	83.9	77.6
	6	6	6	83.6	77.4

S.III. Datasets

A. Sleep-EDF-20

It contains two different studies on healthy subjects: Sleep Cassette (SC), which studies the effect of age on sleep, and Sleep Telemetry (ST), which studies the effect of temazepam on sleep. We adopt SC study in our experiments. The data include whole-night PSGs sleep recordings with a sampling rate of 100 Hz for about 20 hours recordings during two subsequent day-night periods except for few subjects with only one night recording. Each PSG record contains two EEG channels (Fpz-Cz and Pz-Oz), one EOG channel, one chin EMG channel, and event markers. The sleep scoring was performed on 30-second epochs and their corresponding hypnograms were manually scored according to Rechtschaffen & Kales manual. Each epoch was labeled into one stage of W, N1, N2, N3, N4, REM, Movement (M) and UNKNOWN.

B. Sleep-EDF-78

An expanded version of Sleep-EDF-20 dataset, and it contains data files for 78 subjects. It has the exact similar characteristics of Sleep-EDF-20

Download links:

- Sleep-EDF-78
  - The full expanded version,
  - Can be downloaded from Physionet website using the following link:
  - *https://physionet.org/static/published-projects/sleep-edfx/sleep-edf-database-expanded-1.0.0.zip*
- Sleep-EDF-20
  - A subset from Sleep-EDF-78 dataset,
  - Consists of the files named: "SC4xx1E0-PSG.edf" with xx ranging from (00 to 19),
  - Download links can be found in the following link:
  - *https://github.com/emadeldeen24/AttnSleep/blob/main/prepare_datasets/download_edf20.sh*

C. SHHS

Sleep Heart Health Study (SHHS) dataset contains recordings of patients in two visits, but we only use the data from the first visit as its recordings have fixed sampling rate of 125 Hz, while the second has multiple sampling rates. This visit contains recordings of 6,441 men and women aged 40 years and older, including the records of six-class sleep stage classification for each subject manually labelled by specialists.

The subjects in this dataset suffer from sleep-related diseases such as: lung diseases, sleep apnea syndromes, cardiovascular diseases and coronary disease. These diseases would result in a high bias to the training model, so to minimize their impact, we filter the data based on the following criteria:

1) We use the metric Apnea Hypopnea Index (AHI), which represents the number of apneas or hypopneas recorded during the study per hour of sleep. Based on the AHI, the severity of the obstructive sleep apnea can be classified as follows: *None* if AHI $<$ 5 per hour, *Mild* if 5 $\leq$ AHI $<$ 15 per hour, *Moderate* if 15 $\leq$ AHI $<$ 30 per hour, *Severe* if 30 $\leq$ AHI $<$ 50 per hour, and *Very Sever* if AHI $\geq$ 50. We chose the subjects with AHI $<$ 5 per hour, as the severity of the obstructive sleep apnea is the least in this case.
2) Among the selected subjects from step 1, we then select the subjects with at least 5% of deep sleep (N3, N4 stages), 15% REM stage during their sleep periods, a sleep efficiency of at least 75% (sleep efficiency is the percentage of time spent asleep while in bed), and a minimum of 7 hours in bed.

Subjects selected based on these criteria are close be having a regular sleep. Following these two steps, we end up with 329 subjects which are our dataset from SHHS database. A list of these subjects is provided in our github repository.

A common simple preprocessing steps we apply for the three datasets are:

1) We exclude both M and UNKNOWN stages as they don't belong to any of the sleep stages.
2) We merge stages N3 and N4 into one stage N3 according AASM standard.
3) We include only 30 minutes of wake periods before and after the sleep periods to add more focus on the sleep stages.

S.IV. Parameters details

TABLE S.3: Parameters values for our model. b1, b2 represent small and wide kernel branches respectively in MRCNN. For SHHS dataset, the value of d becomes 100 because of its higher sampling rate, and hence its longer signal length.

	Layer	Kernel	Stride	Dropout
Convolutions	Conv1 (b1)	50	6	-
	Conv2,3 (b1)	8	1	-
	Conv1 (b2)	400	50	-
	Conv2,3 (b2)	7	1	-
	Conv1,2, AFR	1	1	-
	Causal Conv	7	1	-
MaxPooling	MP1 (b1)	8	2	0.5
	MP2 (b1)	4	4	0.5
	MP1 (b2)	4	2	0.5
	MP2 (b2)	2	2	0.5
		In	Out	Dropout
Fully Connected	FC1, AFR	30	1	-
	FC2, AFR	1	30	-
	FC1, TCE	80	120	0.1
	FC2, TCE	120	80	-
		Value		
TCE parameters	Num. of heads (h)	5		
	Input dimension (d)	80 (Sleep-EDF) 100 (SHHS)		
	Num. of layers	2		

S.V. Baselines Comparison

We show a summarized comparison between the baselines and our proposed AttnSleep model.

TABLE S.4: A summarized comparison with other research work in sleep stage classification.

Method	Architecture	Temporal Dependency	Handling data Imbalance	Remarks
DeepSleepNet [20]	Low- and high- frequency information using CNNs + two biLSTM layers	LSTM	Oversampling with replication	Increased training burden with two-steps training
SleepEEGNet [24]	Low- and high- frequency information using CNNs + LSTM encoder-decoder	LSTM with attention	Oversampling with SMOTE	Slower training with oversampled data
ResNetLSTM [43]	ResNet CNN + LSTM	LSTM	N/A	Didn't consider class imbalance, Overfits for small datasets
MultitaskCNN [17]	CNN with varying sizes kernels	CNN with MaxPooling	Equal number of samples for each class in the mini-batch	CNN is not the best to capture temporal dependencies
SeqSleepNet [36]	End-to-end hierarchical RNN	LSTM with attention	N/A	Requires 3-epoch input
AttnSleep (ours)	MRCNN + AFR + Multi-Head self-attention	Self-Attention	Class-aware loss	Fast and efficient

S.VI. Ablation Study

In this section, we show the ablation study of the different components of AttnSleep for the other two adopted datasets: Sleep-EDF-78 and SHHS.

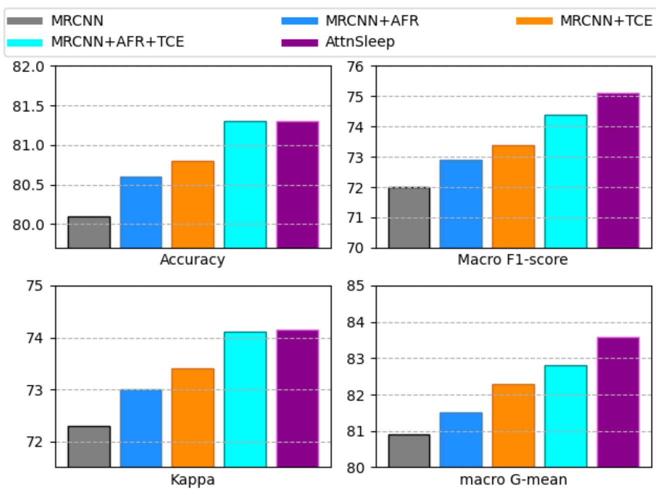

Fig. S.3: Ablation study conducted on Sleep-EDF-78 dataset.

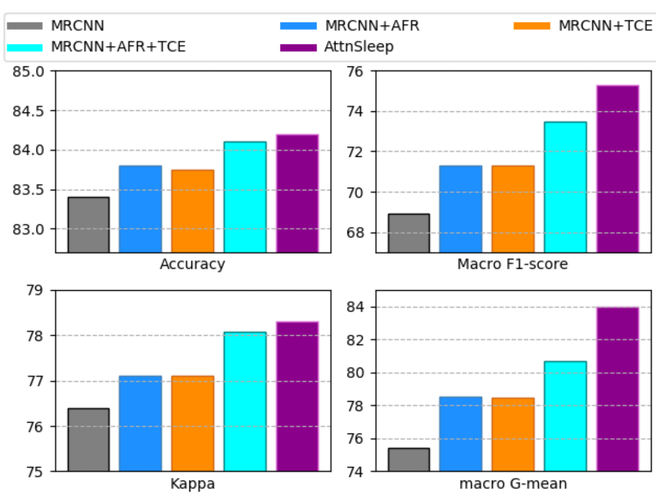

Fig. S.4: Ablation study conducted on SHHS dataset.

S.VII. Sensitivity analysis

We show the results of our sensitivity analysis for the other two adopted datasets: Sleep-EDF-78 and SHHS.

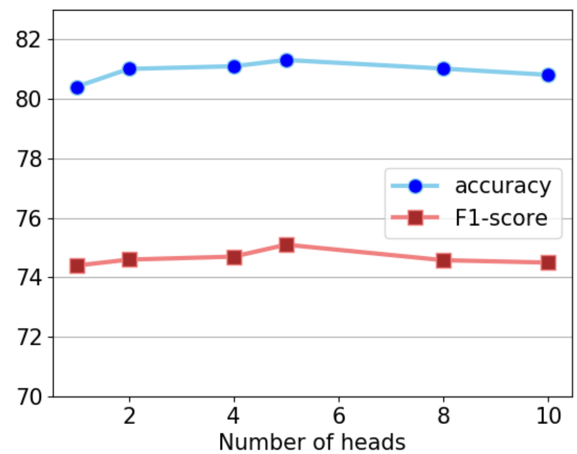

Fig. S.5: Sensitivity analysis of the number of heads in Sleep-EDF-78.

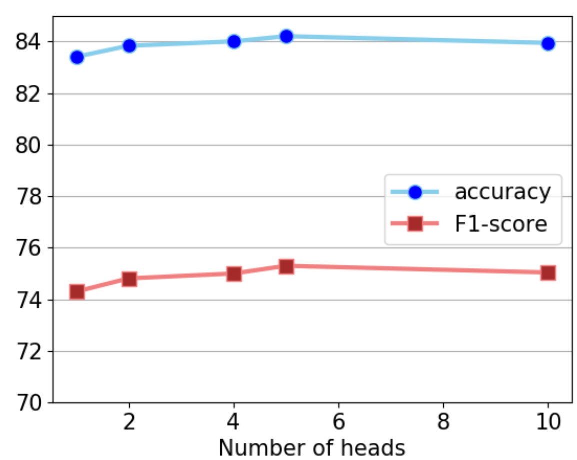

Fig. S.6: Sensitivity analysis of the number of heads in SHHS.