2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-1179-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/EMBC46164.2021.9630808

Sex Difference in Emotion Recognition under Sleep Deprivation: Evidence from EEG and Eye-tracking

Rui-Xiao Ma*, Xu Yan*, Yu-Zhong Liu, Hua-Liang Li and Bao-Liang Lu† *Fellow, IEEE*

***Abstract*— Many psychiatric disorders are accompanied with sleep abnormalities, having significant influence on emotions which might worsen the disorder conditions. Previous studies discovered that the emotion recognition task with objective physiological signals, such as electroencephalography (EEG) and eye movements, provides a reliable way to figure out the complicated relationship between emotion and sleep. However, both of the emotion and EEG signals are affected by sex. This study aims to investigate how sex differences influence emotion recognition under three different sleep conditions. We firstly developed a four-class emotion recognition task based on various sleep conditions to augment the existing dataset. Then we improved the current state-of-the-art deep-learning model with the attention mechanism. It outperforms the best model with higher accuracy about 91.3% and more stabilization. After that, we compared the results of the male and the female group given by this model. The classification accuracy of happy emotion obviously decreases under sleep deprivation for both males and females, which indicates that sleep deprivation impairs the stimulation of happy emotion. Sleep deprivation also notably weakens the discrimination ability of sad emotion for males while females maintain the same as under common sleep. Our study is instructively beneficial to the real application of emotion recognition in disorder diagnosis.**

I. Introduction

Sleep abnormalities and psychiatric disorders are inherently linked, each being a cause and consequence of the other. Although researchers are still engaged in exploring the mechanisms, it has been discovered that sleep deprivation wreaks havoc in the brain, including emotional regulation impairment. Zohar *et al.* [1] found that sleep loss amplifies the negative emotive effects while reducing the positive effects by analyzing self-reported questionnaires about emotional ratings. However, the subjective approaches cannot model the real estimation of emotional arousal and are unable to deal with individual differences on evaluation scaling across subjects. These drawbacks urge scientists to figure out objective parameters to measure emotion recognition.

[1]R. Ma and B. L. Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, the Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, the Brain Science and Technology Research Center, and Qing Yuan Research Institute, Shanghai Jiao Tong University, 800 Dongchuan Rd., Shanghai 200240, China; the Center for Brain-Machine Interface and Neuromodulation and RuiJin-Mihoyo Laboratory, RuiJin Hospital, Shanghai Jiao Tong University School of Medicine, 197 Ruijin 2nd Rd., Shanghai 200020, China

[2]X. Yan is with the Department of Linguistics, University of Washington, Seattle, WA, 98195, USA.

[3]Y. Liu and H. Liang are with Key Laboratory of Occupational Health and Safety of Guangdong Power Grid Co., Ltd, Electric Power Research Institute of Guangdong Power Grid Co.,Ltd., Guangzhou, Guangdong, China

*R. Ma and X. Yan contribute equally to this work.

†Corresponding author: Bao-Liang Lu, bllu@sjtu.edu.cn

The past decade has seen a renewed importance in exploring the more general objective criteria in emotion processing. Some studies go from the physiological level. Among all kinds of signals, EEG signals and eye movements, regarded as the representatives for the internal emotional states and external subconscious behaviors, respectively, receive much more attention than fMRI, etc. due to lower cost and higher accessibility. Zheng *et al.* [2] first attempted to do emotion recognition with EEG data using deep learning models and got decent results. With the hope of modeling this psycho-physiological process comprehensively, some innovative studies [3] started to combine multimodal signals like EEG and eye movements. And the better performances indicate that using both types of information works. In Tao and Lu's experiment [4], the model accuracy can even reach 89.16% on the four-class emotion recognition task.

Nevertheless, sex-based factors should be taken into consideration in emotion recognition [5]. This vulnerability definitely hinders developments and applications of using EEG as the key parameter for emotion recognition on a large scale of people in the diagnosis of psychiatric disorders in the future. Therefore, further investigation on the sex we discuss in this paper, is indispensable and vital. The differences between the male and the female on EEG-based emotion recognition task has been demonstrated by several studies. Yan *et al.* [6] revealed the sex-based lateralization existing in the brain and found that there are differences between males and females in the discrimination ability of the same emotion, especially for happy and sad emotions.

The previous research of the current study adopted a multimodal residual LSTM network combining both EEG and eye movements, which had a state-of-the-art performance on the four-class emotion recognition task. In this paper, we first enlarged the dataset to be roughly three times as large as before to cover 40 subjects. Then we meliorated the model with the attention mechanism to better extract the key temporal characters from the physiological signals. After that, we used this updated system to dive into the sex differences under three sleep conditions: sleep deprivation, sleep recovery, and the baseline normal sleep. The results are reported from multiple aspects including confusion matrices and topographic maps to provide a well-rounded analysis on the effect of sex.

II. Experiment Setup

A. Experiment Design

Since we hoped to enlarge the previous dataset, all conditions were kept as same as the original [4]. We took the criterion from the National Sleep Foundation to guide our design that eight hours is the threshold for a normal sleep baseline [7]. For each subject, the experiment comprises a sequence of sessions under three different sleep conditions which are sleep deprivation, sleep recovery, and baseline conditions. In the first session, a sleep deprivation experiment is conducted after 30-hour sleep deprivation. In the second session, subjects are tested after an 8-hour sleep recovery the day after the first session. The third session, also the baseline experiment, is operated at least 14 days after the second session. In this period, subjects are asked to maintain a regular sleep schedule of 8-hours per day. The sleep conditions are monitored by the smart bracelets during the whole experiment period. The 62-channel ESI NeuroScan System and SMI eye-tracking glasses are utilized to record the EEG and the eye movement signals simultaneously during the experiment.

B. Subjects

For augmentation, 24 healthy subjects (12 females with mean age: 24.03, std: 2.67, and 12 males with mean: 23.87, std: 2.96) participated in the experiments. All selected subjects kept a regular daily routine with the habit of sleeping for 7-8 hours each day and were prohibited to take psychotropic drugs during the whole experiment. The study was approved by the local ethics committee and all subjects signed an informed agreement before the experiments. Considering that the previous dataset has data from 16 subjects in which 12 subjects' data are available after preprocessing, our analysis later is conducted based on all 36 subjects.

C. Stimuli Materials

To elicit emotions efficiently, we borrowed the video materials from the popular public emotion EEG dataset SEED-IV[1] as our stimuli. The precisely selected video clips in SEED-IV have been proved on their effectiveness of arousing four emotions: happy, fear, sad and neutral [8]. The duration of the stimuli clips ranges from 2-4 min. Regarding each clip as a trial, there are 24 trials without repetition in one session, and they are divided equally into the four emotion categories.

III. Methods

A. Preprocessing and Feature Extraction

Curry 7 System was used to preprocess the EEG signals with a constant baseline correction. A band-pass filter with the frequency range between 1 to 50 Hz and a notch filter with 50 Hz were applied to each channel. Finally, eye movement artifacts were detected and removed in virtue of signals from EOG and FPZ channels.

The Short-Time Fourier Transform with a time window of 1 s and no overlapping Hanning window was applied to extract the EEG features of EEG signals from 62 channels into five frequency bands: delta (1-3 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (14-30 Hz), and gamma (31-50 Hz)

[1]http://bcmi.sjtu.edu.cn/seed/seed-iv.html

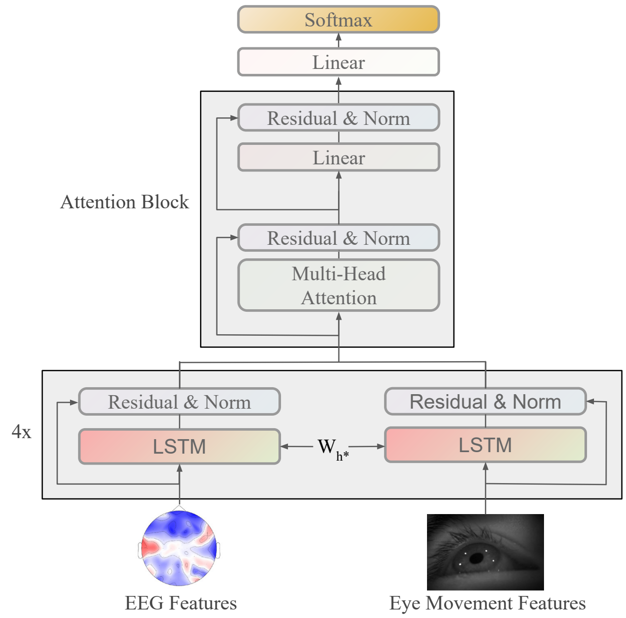

Fig. 1. The architecture of multimodal residual LSTM with attention mechanism model.

[9]. Additionally, the linear dynamic system algorithm was employed for feature smoothing, which subsequently added up to a feature vector of length 310. Eye movements were characterized by 50-dimensional features obtained with SMI BeGaze, aiming to represent information about pupil diameter (X and Y), dispersion (X and Y), fixation duration, blink duration, and saccade [8].

B. Multimodal Residual LSTM with Attention Mechanism

To take full advantage of intra-modality and inter-modality correlations, we adopted a novel multimodal residual LSTM with the attention mechanism. The input signals obtained from each subject (24 periods with 63 s duration each) were divided into 1 s non-overlapping intervals, shuffled, and then fed into the LSTM networks. As illustrated in Fig. 1, we firstly reduced the dimensions of EEG features and eye movement features separately by four connected multimodal residual LSTM blocks. Then we extracted the concatenated high-level features with the attention mechanism. Emotion labels are eventually predicted by linear layers with softmax activation. Mathematical details are explained below.

LSTM, replacing the summation units of standard Recurrent Neural Networks in the hidden layer with memory cells, has shown its effectiveness for extracting temporal information from long biosignals [10]. Recently, multimodal residual LSTM networks have received much attention due to their efficiency in learning the correlation between the information from EEG and other physiological signals [11]. As shown in Fig. 1, the parallel LSTM blocks in the model share the weights W_{h*}. The formulas of multimodal residual LSTM networks, excluding the bias terms, are as follows:

$$\tilde{c}_t^s = tanh(W_{hg} * h_{t-1}^s + W_{xg}^s * x_t^s) \quad (1)$$

$$f_t^s = \sigma(W_{hf} * h_{t-1}^s + W_{xf}^s * x_t^s) \quad (2)$$

$$i_t^s = \sigma(W_{hi} * h_{t-1}^s + W_{xi}^s * x_t^s) \tag{3}$$

$$o_t^s = \sigma(W_{ho} * h_{t-1}^s + W_{xo}^s * x_t^s) \tag{4}$$

$$c_t^s = f_t^s \odot c_{t-1}^s + i_t^s \odot \tilde{c}_t^s \tag{5}$$

$$h_t^s = o_t^s \odot tanh(c_t^s) \tag{6}$$

where the superscripts s and t indicate the type of modality and the time step in the input sequence, respectively. f_t^s, i_t^s, o_t^s, c_t^s, and h_t^s denote the forget gate, input gate, output gate, cell states, and hidden states, respectively.

The shared weights W_{h*} across the three parallel LSTM structures including W_{hg}, W_{hf}, W_{hi}, and W_{ho} are the weight matrices of the hidden states at previous time step, while W_{xg}^s, W_{xf}^s, W_{xi}^s, and W_{xo}^s are the weight matrices of the input at current time step t. σ represents the sigmoid function. The operators $*$ and $\odot$ indicate the matrix multiplication and the Hadamard product, respectively. The LSTM networks are capable of learning separate temporal features from different modalities because of its corresponding weights W_{x*}^s, hidden states h_t^s and cell states c_t^s. Meanwhile, the correlation between the EEG signals and the eye movement is captured by the shared weights W_{h*}.

Residual architecture was adopted to solve the problem of vanishing gradients [12]. Moreover, residual blocks can eliminate the complexity and expressiveness of the networks when increasing the number of layers. The formula of residual architecture is the linear combination of the input and a learnable residual. It therefore provides a shortcut across the layers to train more effectively.

Layer normalization [13], effective at stabilizing the hidden state dynamics in recurrent networks, was conducted following the residual architecture. Dropout is applied in each layer to avoid overfitting. High-level representations of two modalities' features are concatenated before entering the attention block. The attention mechanism [14] has shown its potentials in a wide range of tasks concerning time sequence processing. The temporal information of EEG features and eye movement features can be extracted by the attention function. The high-level features of two modalities are firstly mapped to queries, keys, and values with the same dimension $\sqrt{d_k}$ by learnable linear projections respectively. The attention function is defined as follows:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{8}$$

where Q, K, V indicate the matrices packing together the set of queries, keys and values, respectively.

Inspired by the attention mechanism, we adopted the multi-head attention mechanism to make the model automatically explore the critical channels and bands for EEG features and critical local information for eye movement features. It not only reduces the model complexity while maintaining a good structure of the networks but also allows the model to learn relevant information in different representation subspaces and extract richer feature information. We performed the attention function in parallel with two heads so that the model is able to pay attention to EEG features and eye movement features in separate subspaces.

TABLE I
RESULTS (%) OF EMOTION CLASSIFICATION TASKS

Session	Model	Mean ± Std. (M)	Mean ± Std. (F)
Deprivation	SVM	55.64 ± 15.49	56.47 ± 13.72
	LSTM	85.60 ± 4.19	87.36 ± 5.05
	LSTM*	**86.96 ± 3.63**	**88.81 ± 4.03**
Recovery	SVM	57.57 ± 6.26	59.10 ± 13.96
	LSTM	86.46 ± 5.07	88.75 ± 4.76
	LSTM*	**88.54 ± 4.32**	**90.15 ± 4.45**
Baseline	SVM	68.92 ± 10.63	68.22 ± 14.32
	LSTM	89.64 ± 6.31	90.32 ± 5.13
	LSTM*	**90.83 ± 6.16**	**91.13 ± 4.70**

Notes: In the Model column, LSTM represents the multimodal residual LSTM Networks, while LSTM* stands for multimodal residual LSTM Networks with attention mechanism proposed in this paper.

IV. Results and Discussion

A. Model Validation

To validate the efficiency of our improvement with the attention mechanism, we compared the performance with the original state-of-the-art model and the popular used Support Vector Machine (SVM) method on the four-class subject-dependant emotion recognition task with the data we collected. The results represented by mean accuracies and standard deviations are reported in Table I.

Our model outperforms the other two methods with higher accuracy and more stable performance as shown above. Remarkably, these differences exist in all three conditions, especially under sleep deprivation. Moreover, sleep deprivation impairs the stimulation of emotions more for males. Our model provides additional support for digging out the influence of sex in several sleep conditions.

B. Classification Performance

To better understand the discrepancies between the male and the female on different emotions, we made the confusion

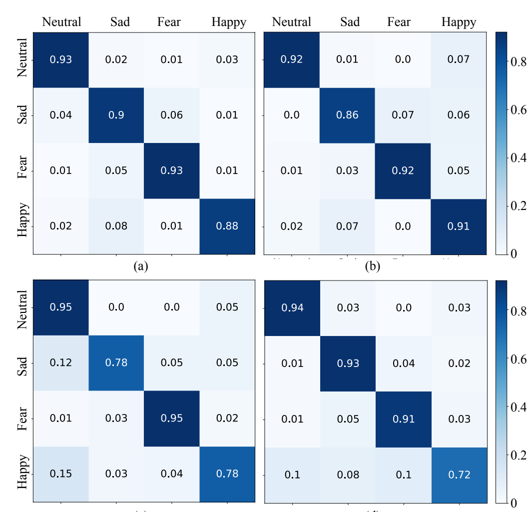

Fig. 2. The confusion matrices under baseline (upper) and sleep deprivation (lower) condition for males (left) and females (right).

matrices of baseline and those of sleep deprivation as shown in Fig. 2. For males, the recognition accuracies of sad and happy drop significantly by 12% and 10% under 30-hour sleep deprivation while results of the neutral and the fear emotions remain almost the same as the baseline. In contrast, the deficiency in sleep only drastically impacts the happy emotion of the female, pushing the accuracy down to 72% from 91%. These numbers reveal that sleep conditions play an important role in people's happy emotion for both males and females, and make males become sensitive to the sad emotion. Comparatively, the neutral and the fear emotions are much more stable and insensitive to the sleep conditions.

C. Neural Patterns from Topographs

The brain topographic maps give a direct insight into the distinctions about underlying neural patterns. As depicted in Fig. 3, we selected the maps of the average energy distributions of EEG features in the gamma band as a representation since they are the most distinguishable compared to the others, which is consistent with the previous studies [4].

For the only positive emotion, happy, it is clear that males need more energy to arouse happy feelings when sleep is deprived, whereas females' demands do not change much with the sleep status. Besides, the active regions of males' brain become centralized under sleep deprivation, but females' go from the inverse way in that more areas get involved. The two groups are similar in the pattern of the neutral emotion. More energy concentrates on the lateral temporal lobe compared with the baseline situation. As for the negative emotions, sleep deprivation makes females more vulnerable to fear and sadness while the overall activation level of males varies little. Interestingly, less prefrontal lobe area is activated for both groups under sleep deprivation.

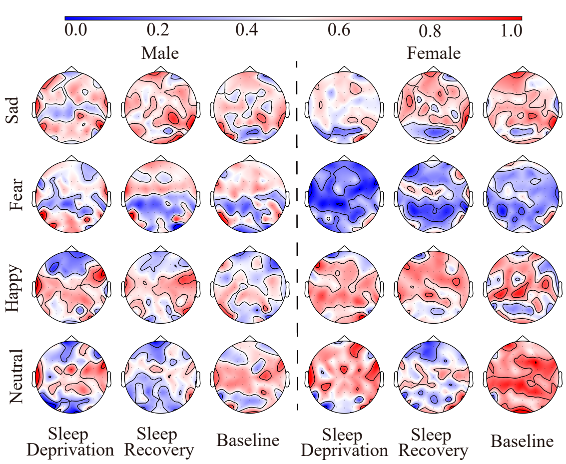

Fig. 3. Topographic maps of the four emotions in Gamma band for males and females.

V. CONCLUSIONS

In this paper, we have investigated the sex differences after enlarging the previous dataset utilizing the multimodal residual LSTM networks with the attention mechanism which captures well the temporal intra-modality and inter-modality correlations based on EEG and eye movement signals. With this advanced model, experimental results indicated that males become more susceptible to the sad emotion under sleep deprivation and the elicitation of the happy emotion is heavily impacted for both by sleep conditions. Relatively speaking, there is not much difference on the neutral and the fear emotions between these two groups. In our view, these results constitute an excellent initial step toward the fine-grained emotion recognition system which has great potential to benefit the diagnosis of psychiatric disorders. We hope that our research will serve as a base for future studies on other EEG or emotion-sensitive factors.

ACKNOWLEDGMENT

This work was supported in part by grants from the National Natural Science Foundation of China (Grant No. 61976135), SJTU Trans-Med Awards Research (WF540162605), the Fundamental Research Funds for the Central Universities, the 111 Project, and the China Southern Power Grid (Grant No. GDKJXM20185761).

REFERENCES

[1] D. Zohar, O. Tzischinsky, R. Epstein, and P. Lavie, "The effects of sleep loss on medical residents' emotional reactions to work events: a cognitive-energy model." *Sleep*, 28(1), pp. 47-54, 2005.

[2] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "EEG-based emotion classification using deep belief networks," in *IEEE International Conference on Multimedia and Expo*, pp. 1-6, 2014.

[3] W.-L. Zheng, B.-N. Dong, and B.-L. Lu, "Multimodal emotion recognition using EEG and eye tracking data," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5040-5043, 2014.

[4] L.-Y. Tao, and B.-L. Lu, "Emotion recognition under sleep deprivation using a multimodal residual LSTM network," in *International Joint Conference on Neural Networks*, pp. 1-8, 2020.

[5] L.-Q. Bao, J.-L. Qiu, H. Tang, W.-L. Zheng, and B.-L. Lu, "Investigating sex differences in classification of five emotions from EEG and eye movement signals," in *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2019, pp. 6746-6749

[6] J. G. Kreifeldt, "An analysis of surface-detected EMG as an amplitude-modulated noise," in *1989 International Conference of Medicine and Biological Engineering*, Chicago, IL.

[7] M. Hirshkowitz, K. Whiton, S. M. Albert, E. S. Katz *et al.*, "National sleep foundation's sleep time duration recommendations: Methodology and results summary," *Sleep Health*, vol. 1, no. 1, pp. 40–43, 2015.

[8] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE Transactions on Cybernetics*, vol. 49, pp. 1110–1122, 2019.

[9] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6627–6630, 2013.

[10] H. Tang, W. Liu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using deep neural networks," in *International Conference on Neural Information Processing*. Springer, pp. 811–819, 2017.

[11] J.-X Ma, H. Tang, W.-L Zheng, and B.-L. Lu, "Emotion recognition using multimodal residual LSTM network," in *Proceedings of ACM International Conference on Multimedia*, pp. 176-183, 2019.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[13] J.L. Ba, J.R. Kiros, and G.E. Hinton, "Layer normalization," arXiv preprint arXiv:1607.06450, 2016.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, A. N. Gomez *et al.*, "Attention is all you need". in *NeurIPS*, 2017.