2023 IEEE/SICE International Symposium on System Integration (SII) | 979-8-3503-9868-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/SII55687.2023.10039222

A Real-Time and Two-Dimensional Emotion Recognition System Based on EEG and HRV using Machine Learning*

Yongxin Wei[1], Yunfan Li[1], Mingyang Xu[1], Yifan Hua[1], Yukai Gong[1], Keisuke Osawa[1], *Member, IEEE*, and Eiichiro Tanaka[1*], *Member, IEEE*

***Abstract*— With the research on mental health, rehabilitation training and other fields, obtaining people's real emotion feelings is frequently required in many fields. Emotion recognition method based on physiological signals can directly obtain people's emotion states and avoid pretending expression and emotional expression disorder. In physiological signals, Electroencephalogram (EEG) signal is commonly used in the emotion evaluation, and Heart Rate Variability (HRV) signal is related to people's excited feeling. This paper proposed an emotion recognition method based on EEG and HRV to do the emotion recognition work. This method aims to solve the accuracy problem of instant emotion recognition, and achieve a higher accuracy. According to Russell's model of emotion, the system in this paper use two dimensions, "valence" and "arousal", to describe people's emotion. The emotion recognition system we proposed combines more advanced neural network models and eigenvalues closely related to emotional states. This system uses DenseNet as the neural network model for machine learning process, which is more accurate than the general deep neural network. Using differential entropy as the main eigenvalue makes the system's ability to analyze emotions based on EEG more efficient.**

*This work was supported by JSPS KAKENHI Grant Numbers 19H04505 and 22H03997.

[1*]Correspondence author: E. Tanaka is with the Graduate School of Information, Production and Systems, Waseda University, 2-7 Hibikino, Wakamatsu-ku, Kita-Kyushu, Fukuoka 808-0135, Japan. E-mail: tanakae@waseda.jp.

[1]Y. Wei, Y. Li, M. Xu, Y. Hua, Y. Gong, and K. Osawa are with the Graduate School of Information, Production and Systems, Waseda University, Fukuoka, Japan.

I. INTRODUCTION

In countries and regions with more developed infrastructure and medical conditions, telemedicine is widely used as a new medical method. This method breaks through the space limitations of traditional medical care and can integrate medical resources in different regions [1]. As the global aging phenomenon worsens, geriatric diseases account for an increasing proportion of medical problems, and in some countries, this phenomenon will be particularly serious. Many diseases of old age can be relieved or prevented by increasing exercise, walking assistance combined with emotion recognition can help stroke patients or stroke risk groups to improve their exercise experience, prolong exercise time, and reduce the risk of disease. Symptoms of onset patients and help patients regain normal mobility. At the same time, emotion recognition can be better used in various fields if it can give instant evaluation results.

Several kind of emotion recognition method have been developed so far. There are mainly two classes, discrete model based and dimensional model based. With discrete models, K. Gouizi et al. developed a method using Support Vector Machine (SVM) technique [2]. The first class of methods uses a finite number of discrete basic emotions to represent people's emotional states. The second class of methods are dimensional model based methods. J. Kim et al. developed a method using Linear and Quadratic Discriminant Analysis [3] and A.R. Aguinaga et al. developed a method using SVM technique [4]. The second type of method is easier to quantify the description of emotions, which is more

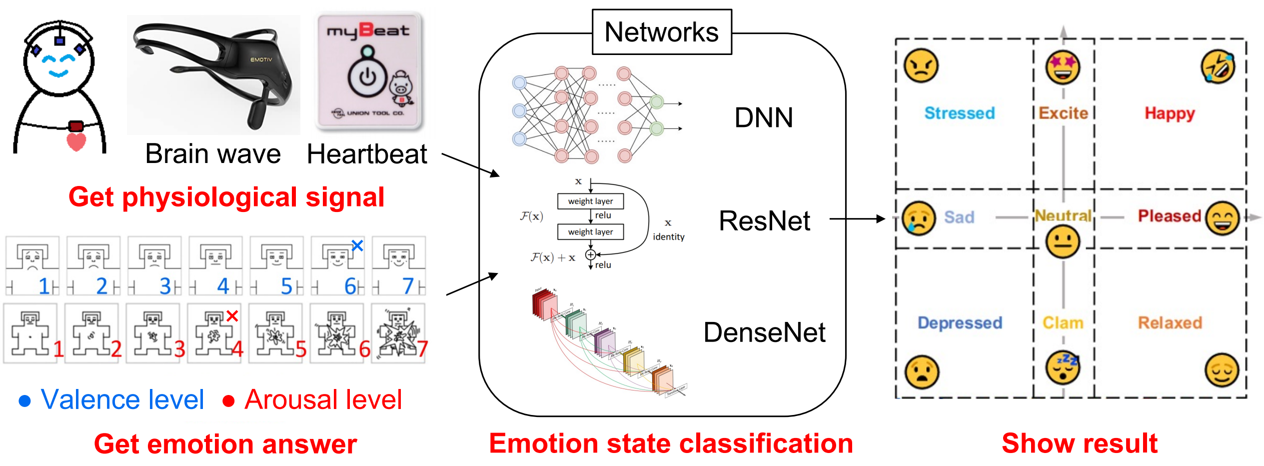

Fig. 1. Overview of emotion recognition system

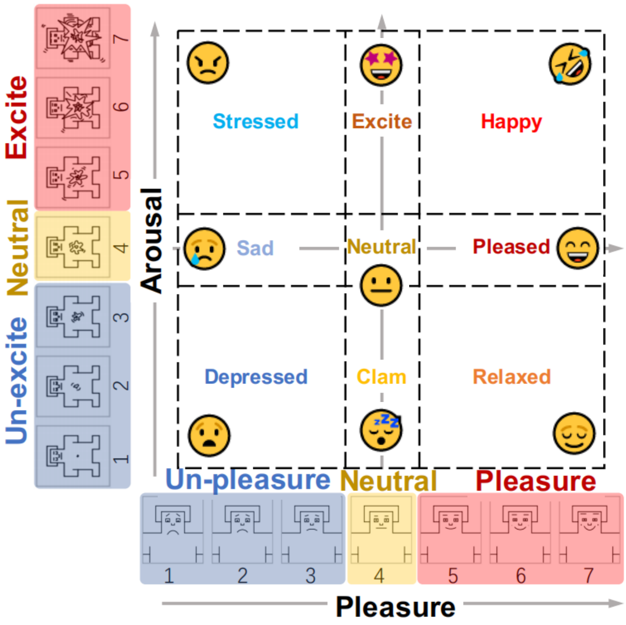

Fig. 2. 2-D emotion map

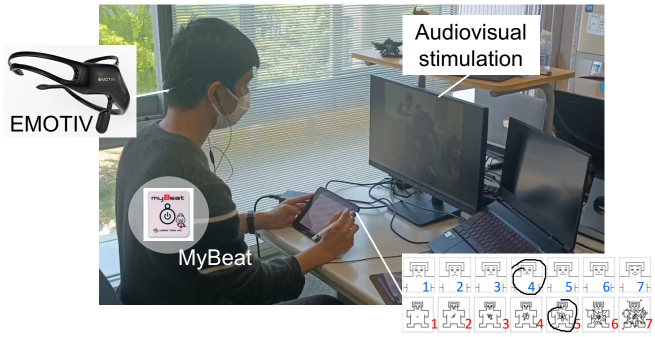

Fig. 3. Experimental setup

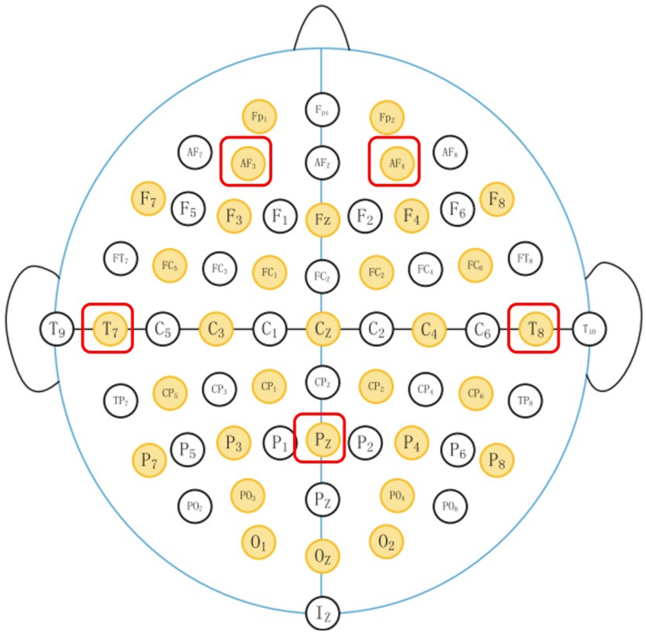

Fig. 4. 5-channels EEG area distribution

beneficial when it is necessary to accurately discriminate emotions. At the same time, this type of emotion model requires higher accuracy of emotion recognition methods.

Based on the application requirements of emotion recognition technology in exercise assistance and other fields, our laboratory have been researched for providing comfortable walking by combining a walking assistive device [5], [6], [7] and an emotion recognition system [8], [9], [10]. Using the variation of the physiological signal as emotional recognition way is better than other methods, we recognized the human emotion based on using the variation of the physiological signal [11]. We also proposed a method which could be used for mapping people's emotion state on the two-dimensional arousal-valence model of effect [12]. Moreover, we implement various algorithms (k-means, T method of MTS (Mahalanobis Taguchi System) and DNN (Deep Neural Network)) for determining the emotional state from physiological data. Finally, the findings indicate that deep neural network method can precisely recognize the human emotional state [13], [14]. On the other hand, both further improvement in accuracy and reduction in time are required.

This study aims to propose a fast-response emotion recognition method to provide more accurate and real emotion recognition services, as shown in Fig. 1. In this paper, we report that the construction of emotion recognition system used both DenseNet as the neural network model and differential entropy as the main eigenvalue, and its evaluation results.

II. MATERIALS AND METHODS

A. Emotion Model

This study uses a Valence-Arousal model [15] with three levels of positive, neutral, and negative, as shown in Fig. 2. Valence dimension describes people's happiness level, while the Arousal dimension describes people's activation level. The proposed emotion recognition method uses EEG and HRV data as data sources for emotion recognition, both of which are physiological signals generated by human body during daily activities. Among them, the Valence dimension used EEG signal and its related characteristics, and the Arousal dimension used EEG signal and HRV signal and its related characteristics.

B. Emotional Data Collection

A total of 20 subjects participated in the emotional stimulation experiment, including 18 males and 2 females, with an age range of 22-27 years. All participants were in a healthy state when they participated in the experiment, with no history of cardiovascular and cerebrovascular diseases, audio-

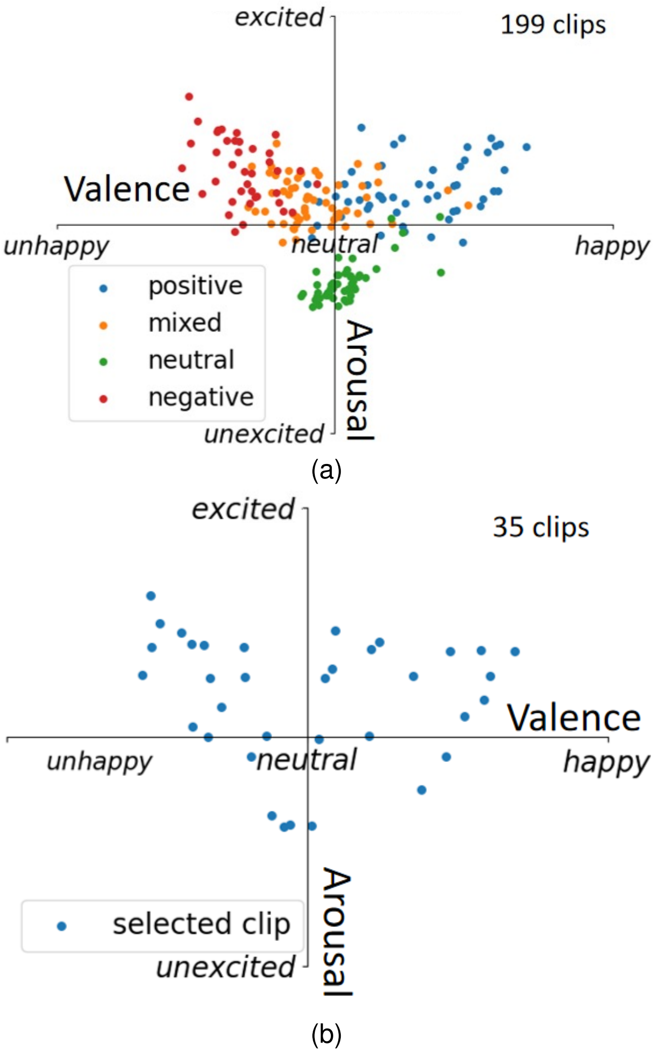

Fig. 5. Videos label distribution. (a) All videos, (b) Selected videos.

visual disorders, and mental diseases that affect emotional perception and expression. In this experiment, physiological signal data under specific emotions were obtained by allowing subjects to watch emotion-guided videos (Fig. 3).

To obtain the dataset needed to train the neural network, we used 5 Channel Mobile Brainwear (Insight 2.0, EMOTIV Inc., USA) to collect the data. The device can monitor the state of human brain waves in real time, and read the data into local files in real time. This feature fits nicely into our target system and can continue to be used throughout the system in the future. At the same time, compared with the 14-channel EEG acquisition device, the device has fewer sensors, lighter weight, and better portability. The 5-channels scheme uses AF3, T7 , AF4, T8 and Pz total 5 area channels, as shown in Fig. 4. Additionally, we used a wearable heart rate sensor myBeat (WHS-1, UNION TOOL Co., Japan) to monitor many metrics including heartbeat waveforms, including and simultaneously output in the form of waveform graphs or numbers. The device transmits the collected heartbeat information to the computer GPD Pocket3 (Intel Core i7 1195 G7, 16GB, 1TB SSD) by wireless connection, has good portability, and is an ideal device for collecting physiological signals in sports assistance.

TABLE I

FREQUENCY TABLE

Band Name	Range (Hz)	Corresponding emotion
Delta (δ)	0.5-4	Subconscious
Theta (θ)	4-8	Drowsy Mediative
Alpha (α)	8-12	Relaxed Reflective
Low Beta (β)	12-20	Alert Working
High Beta (β)	20-30	Alert Working
Gamma (γ)	30-100	Active thought

This resource library is a collection of videos based on a wide range of people as samples. 411 subjects participated in the statistics, and the video clips were evaluated, and quantitative descriptions were made based on multiple dimensions such as pleasure level, activation level, and so on. The data volume of the database is 199 videos [16], and the videos are classified according to the comprehensive evaluation received by each video. Our main target is video objects in the emotional distribution that are close to the edge of the distribution range, and these videos elicit relatively stronger emotional responses. According to the needs of the experiment, we selected 35 clips in the video collection mentioned above as media resources for emotional stimulation, as shown in Fig. 5. The video resources are 35 groups of video, and each group is composed of multiple parts, including a scene video with soothing music to make the subjects relax, a black screen and mute picture, an emotionally stimulating video, and another black screen and mute picture. The landscape video in the first part lasts about 30 seconds, the black screen in the second part lasts for 10 seconds, the emotional stimulation video in the third part varies from 20 to 30 seconds depending on the resource itself, the black screen in the fourth part and the second Part of the same is ten seconds. This process looped until subjects watched all sets of videos.

C. Data Processing

The collected raw data of physiological signals include brain wave signals and heartbeat signals. The sampling frequency of the EEG signal acquisition device is theoretically 128Hz. EEG signals are generally distributed in the range of 0-30 Hz. From low frequency to high frequency, it can be divided into Delta band, Theta band, Alpha band, Beta band and Gamma band. We used Delta, Theta, Alpha and Beta four bands, Beta wave band is divided into low band and high band here, as Table 1. The heartbeat signals include Heart Rate (HR) and LF/HF. Heart rate is the number of heartbeat cycles per minute; LF/HF is the ratio of neural activity, which has no units. LF is the low frequency band of 0.04-0.15Hz, which represents the activity of sympathetic and parasympathetic nerves, and HF is the high frequency band of 0.15-0.4Hz. The variability of data in this range reflects the activity of parasympathetic nerves. This paper retains the signal data of the subjects at the end of each emotionally

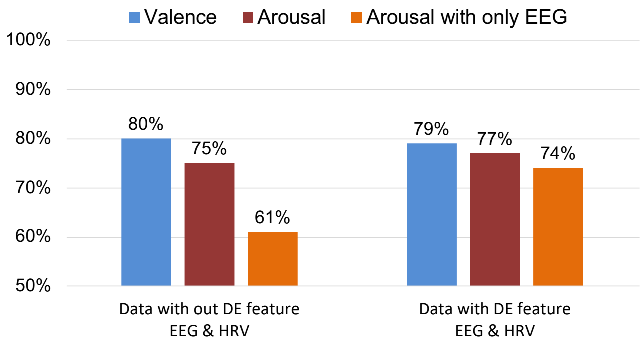

Fig. 6. DNN classification accuracy

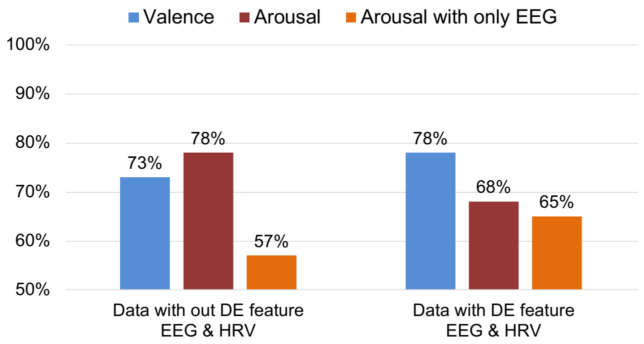

Fig. 7. ResNet classification accuracy

stimulating video. During this time period, the subjects have just watched the entire video and just reacted to the overall content of the video. It is relatively easy to capture emotional fluctuations moment. The physiological signal dataset used in this paper is mainly obtained in this way, including two acquisitions, the length of time collected in experiment is 1 second. In the past research of our laboratory, due to various constraints, in order to ensure that the collected data can reflect the feelings of the subjects after watching the entire video clip, the length of time collected in the data set is about 0.3 seconds. As the sampling time increases, the data set can store more information, and the newly obtained results have better stability and are closer to the real emotional state of people. This paper also extracts the eigenvalues of differential entropy (DE). DE is a continuous form of traditional information entropy. In practice, the calculation method of this feature can be simplified as follows.

$$
\begin{aligned}
H[x] &= -\int p(x)\log p(x)dx \\
&= -\int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx \\
&= -\frac{1}{2}\log(2\pi e\sigma^2)
\end{aligned} \tag{1}
$$

x is the input of function, $p(x)$ is the probability of x, and σ^2 is the variance of x.

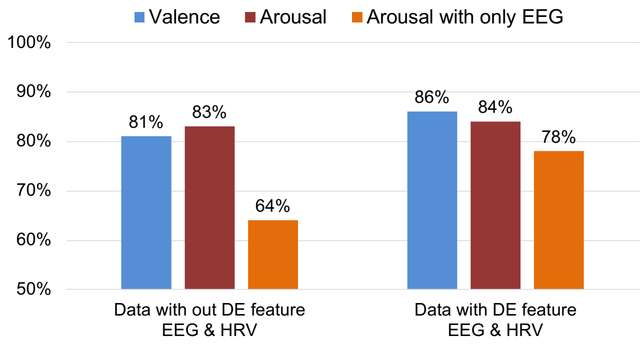

Fig. 8. DenseNet classification accuracy

D. Emotion Classification Method

The way to deal with the emotion classification problem is the neural network model. The three models of DNN, ResNet and DenseNet are trained separately from the two datasets mentioned in Section II.B. The neural network model used in this paper is mainly the ReLU function [17], and the Sigmoid function [18] is used in the output layer. The loss function used in the neural network is mainly the cross-entropy-softmax function [19]. The performance of DNN, ResNet [20] and DenseNet [21] in this dataset is shown in Fig. 6-8. Among the three neural network models, on average, DenseNet has the highest accuracy.

The dataset has a longer sampling time than previous research, and its accuracy is compared with the original dataset. While using this dataset, this paper attempts to use only brain wave signals for emotion recognition in the Arousal dimension, and compares it with the results using both brain wave signals and heartbeat signals. The preprocessing of the new dataset includes the extraction of DE features, and the results of whether or not the feature values are used are also compared according to the type of neural network.

The classification accuracy of DNN is shown in Fig. 6. After using the data set for extracting differential entropy features, the accuracy of the Valence dimension decreases slightly, and the accuracy of the Arousal dimension increases slightly. If only the brain wave signal data is used to After training the Arousal dimension Neural network, the accuracy increased by about 13%.

The classification accuracy of ResNet is shown in Fig. 7. After using the data set for extracting differential entropy features, the accuracy of the Valence dimension increases slightly, and the accuracy of the Arousal dimension decreases significantly, reaching 20%. If only the brain wave signal data is used When training a neural network with Arousal dimension, the accuracy increased by about 8%.

The classification accuracy of DenseNet is shown in Fig. 8. After using the data set for extracting differential entropy features, the accuracy of the Valence dimension increases by about 5%, and the accuracy of the Arousal dimension is basically unchanged. If only the brain wave signal data is used for training Arousal dimension neural network, the

TABLE II
TIME COST EACH MODELS (UNIT : SECOND)

Average time cost	Without DE feature	With DE feature
DNN	0.21 s	0.45 s
ResNet	0.98 s	1.26 s
DenseNet	0.38 s	0.49 s

accuracy rate increased by about 14%.

Among the neural network models trained on the new dataset, the best accuracy on the test set is DenseNet using the differential entropy feature. On the other hand, the time consumption of these models to complete the classification evaluation on newly input data is also different, in the case of using brain wave signals and heartbeat signals in the Arousal dimension, as shown in Table 2.

The time spent by the neural network model in the experimental environment is about 1 second, which can be applied in the immediate response system. At the same time, we found that the accuracy of DenseNet is relatively high and the time consumption is acceptable, so DenseNet is selected as the neural network model for building an emotion recognition system.

III. VALIDATIONS AND RESULTS

A. System Implementation

The emotion recognition system consists of multiple modules, including a data input module, an emotion evaluation calculation module, and a result output module. The physiological signals obtained by the acquisition equipment are sent to the program in real time, and the program calls the trained neural network model for calculation and outputs the results. Finally, the output module, that is, the result display module, displays the analysis results on the two-dimensional emotional map.

B. Validation Test

The validation dataset is obtained in the same way as the dataset in Section II. A total of 4 subjects participated in the validation experiment, and after obtaining the questionnaire results and physiological signal data, the accuracy of the neural network model was calculated. For neural network models, these data are more unfamiliar than the data in the training dataset, and are closer to the accuracy rate in working environment.

C. Results

In previous studies, we tested the trained network. Although the accuracy is measured using independently divided test sets, the accuracy still decreases when another independently collected data set is used. The degree of accuracy drop is 6-14%, the reason for this phenomenon may be that the number of existing subject groups is limited and cannot include the emotional data of all groups in all states, so the actual measurement accuracy has declined relative to the ideal experimental environment. In order to test the actual use accuracy of the emotion recognition system, we tested the accuracy of the DenseNet (DE) model by using the dataset in Section III.B. For the possible drop in accuracy in this test, we tested the system, and the results are shown in Fig. 9.

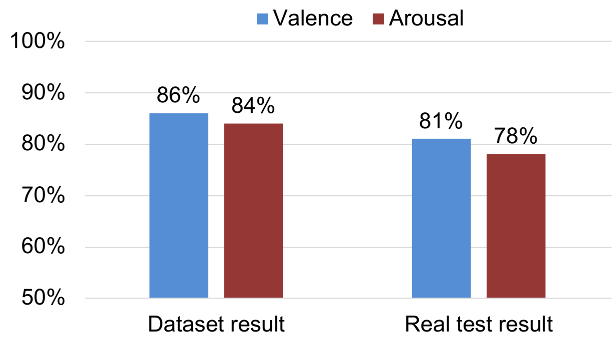

Fig. 9. Accuracy comparison (using DenseNet with DE)

Compared with the evaluation on the test set, in the experiment, the new model also has the problem of decreasing the accuracy. The degree of decrease in Fig. 9 is about 6%. This shows that the model may have better resistance to interference factors in actual use, such as individual differences between people, in the face of the accuracy limit caused by the size of the data set. In practical use, the system has better performance than previous emotion recognition systems. In the preliminary tests, we have already confirmed that noise caused by external factors such as vibration of head is filtered by an original algorithm installed in EMOTIV Insight, and that it has almost no effect on the accuracy of emotion recognition in this study. On the other hand, it is necessary to further investigate how disturbance of impulsive biosignals due to blinking and differences in frequency characteristics due to age affect the accuracy of emotion recognition.

IV. CONCLUSIONS AND DISCUSSIONS

In this study, we developed an emotion recognition system based on EEG and HRV physiological signals. This system uses DenseNet as the neural network model for machine learning process, which is more accurate than the general deep neural network. In addition, using differential entropy as the main eigenvalue makes the system's ability to analyze emotions based on EEG more efficient. These contributions have been successfully showed using a dataset obtained from subjects in the experiment. Emotion recognition system based on physiological signals can provide a relatively stable and effective solution for research in this field, providing services such as healthcare, sports assistance, and psychological health examinations. On the other hand, it is necessary to discuss for further improvement due to some limitations in this study.

First, this experiment adopts a multimodal physiological signal acquisition method. Generally speaking, the more signal sources, the higher the accuracy of emotion recognition and the better stability. However, not all use environments are suitable for the operation of a large number of heavy equipment. In the direction of lightweight equipment, if you can use light acquisition equipment with good signal quality, the scope of application can be expanded.

Second, in the case of using only EEG signals, the accuracy of part of the results is reduced. If suitable eigenvalues are found or some optimization measures are taken, this reduction in accuracy can be partially compensated. EEG equipment needs to apply physiological saline on the skin surface to increase the conduction efficiency of electrical signals. If this operation can be simplified, the portability of the equipment can be greatly improved.

Third, a device like myBeat has a small weight and collects a lot of information. Lightweight devices that can collect indicators such as LF/HF, including smart bracelets, are also useful in emotion recognition. Although these devices cannot give classification results according to complex emotional models when running alone, they can give answers to questions with fewer (no more than 3) types of emotions, especially the degree of excitement or tension of people and Heart activity is closely related to emotions, and heartbeat data is a more sensitive source of information.

Finally, this paper compares the classification performance of 3 neural networks, which are collected based on 5-channel brain wave data. If there is no requirement for portability, you can consider using more channels of brain wave data and increasing the size of the convolution kernel of the neural network, which can greatly improve the classification accuracy. In the case of collecting long-term data conditionally, using a neural network model similar to LSTM to analyze data in a longer period of time and improve the classification accuracy should be considered. In the future, we will address these issues and improve to make it easier to implement.

REFERENCES

[1] K.M. Rump, J.L. Giovannelli, N.J. Minshew, and M.S. Strauss. “The Development of Emotion Recognition in Individuals with Autism,” *Child Development*, vol. 80, no. 5, pp. 1434-1447. 2009.

[2] K. Gouizi, F.B. Reguig, and C. Maaoui. “Analysis physiological signals for emotion recognition,” *Proceedings of International Workshop on Systems, Signal Processing and their Applications, WOSSPA2011*, pp. 147-150, 2011.

[3] J. Kim, and E. Andre. “Emotion recognition based on physiological changes in music listening,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067-2083, 2008.

[4] A.R. Aguinaga, M.A.L. Ramirez, and M.R.B. Flores. “Classification model of arousal and valence mental states by EEG signals analysis and Brodmann correlations,” *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 6, pp. 230-238, 2015.

[5] E. Tanaka, R. Niwa, K. Osawa, K. Nakajima, K. Muramatsu, K. Watanuki, S. Saegusa, L. Yuge. “Motion assistance apparatus enabled for neuro-rehabilitation of patients and for the promotion of exercise for the elderly,” *Proceedings of IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM2015*, pp. 937-942, 2015.

[6] E. Tanaka, K. Muramatsu, K. Watanuki, S. Saegusa, L. Yuge. “Walking Assistance Apparatus Enabled for Neuro-Rehabilitation of Patients and its Effectiveness,” *Mechanical Engineering Letters*, vol. 1, no. 15-00530, pp. 1-15, 2015.

[7] J. Zhuang, H. Nagayoshi, H. Kondo, K. Muramatsu, K. Watanuki, E. Tanaka. “Development of a torque limiter for the gear of an assistive walking device,” *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 11, no. 6, 17-00376, 2017.

[8] J. Zhuang, G. Wu, H. Lee, E. Tanaka. “Evaluation of the Walking-Emotion Relationship for Applying to an Assistive Walking Device,” *Proceedings of 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2019*, pp. 3147-3150, 2019.

[9] E. Tanaka, J. Zhuang, B. Yang, G. Wu, H. Lee, L. Yuge. “A Walking Assistive Device of Ankle Joint Motion and Control Method Taken into Account the Emotion Condition,” *Proceedings of the 2020 IEEE/SICE International Symposium on System Integration, SII 2020*, pp. 623-627, 2020.

[10] J. Yang, J. Zhuang, G. Wu, E. Tanaka. “Study of Current Emotion and Muscle Fatigue Evaluation Method for a Walking Assistive Device,” *Proceedings of 2020 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM 2020*, pp. 1884-1889, 2020.

[11] E. Tanaka, K. Muramatsu, Y. Osawa, S. Saegusa, L. Yuge, and K. Watanuki. “A Walking Promotion Method Using the Tuning of a Beat Sound Based on a Two-Dimensional Emotion Map,” *Advances in Intelligent Systems and Computing*, vol. 483, pp. 519-525, 2017.

[12] Z. Zhang, and E. Tanaka. “Affective computing using clustering method for mapping human’s emotion,” *Proceedings of IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM2017*, pp. 235-240, 2017.

[13] J.R. Zhuang, Y.J. Guan, H. Nagayoshi, L. Yuge, H.H. Lee, and E. Tanaka. “Two-Dimensional Emotion Evaluation with Multiple Physiological Signals,” *Advances in Intelligent Systems and Computing*, vol. 774, pp. 158-168, 2019.

[14] J.R. Zhuang, Y.J. Guan, H. Nagayoshi, K. Muramatsu, K. Watanuki, and E. Tanaka. “Real-time emotion recognition system with multiple physiological signals,” *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 13 no. 4, pp. JAMDSM0075, 2019.

[15] M.M. Bradley, and P.J. Lang. “Measuring emotion: The self-assessment Manikin and the Semantic Differential,” *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, iss. 1, pp. 49-59, 1994.

[16] A.C. Samson, S.D. Kreibig, B. Soderstrom, A.A. Wade, and J.J. Gross. “Eliciting positive, negative and mixed emotional states: A film library for affective scientists,” *Cognition and Emotion*, vol. 30, no. 5, pp. 827-856, 2016.

[17] V. Nair, and G.E. Hinton. “Rectified Linear Units Improve Restricted Boltzmann Machines,” *Proceedings of the 27th International Conference on Machine Learning, ICML2010*, pp. 807-814, 2010.

[18] J. Han, and C. Moraga, Claud. “The Influence of the Sigmoid Function Parameters on the Speed of Backpropagation Learning,” *Proceedings of the International Workshop on Artificial Neural Networks: From Natural to Artificial Neural Computation, IWANN1995* pp. 195-201, 1995.

[19] S. Raschka, and V. Mirjalili. *Python machine learning, 2nd Ed*, Packt Publishing, Birmingham, UK, 2017.

[20] K. He, X. Zhang, S. Ren, and J. Sun. “Deep Residual Learning for Image Recognition,” *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR2016*, pp. 770-778, 2016.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger. “Densely Connected Convolutional Networks,” *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR2017*, pp. 2261-2269, 2017.