2019 IEEE 15th International Colloquium on Signal Processing & its Applications (CSPA 2019), 8 -9 March 2019, Penang, Malaysia

# Subject-Independent Emotion Recognition During Music Listening Based on EEG Using Deep Convolutional Neural Networks

Panayu Keelawat
Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand
panayu.k@student.chula.ac.th

Nattapong Thammasan
Human Media Interaction
University of Twente
Enschede, Netherlands
n.thammasan@utwente.nl

Boonserm Kijsirikul
Department of Computer Engineering
Chulalongkorn University
Bangkok, Thailand
boonserm.k@chula.ac.th

Masayuki Numao
The Institute of Scientific and Industrial Research
Osaka University
Osaka, Japan
numao@sanken.osaka-u.ac.jp

***Abstract*—Emotion recognition during music listening using electroencephalogram (EEG) has gained more attention from researchers, recently. Many studies focused on accuracy on one subject while subject-independent performance evaluation was still unclear. In this paper, the objective is to create an emotion recognition model that can be applied to multiple subjects. By adopting convolutional neural networks (CNNs), advantage could be gained from utilizing information from electrodes and time steps. Using CNNs also does not need feature extraction which might leave out other related but unobserved features. CNNs with three to seven convolutional layers were deployed in this research. We measured their performance with a binary classification task for compositions of emotions including arousal and valence. The results showed that our method captured EEG signal patterns from numerous subjects by 10-fold cross validation with 81.54% and 86.87% accuracy from arousal and valence respectively. The method also showed a higher capability of generalization to unseen subjects than the previous method as can be observed from the results of leave-one-subject-out validation.**

***Keywords*—emotion, electroencephalogram, convolutional neural networks, music**

## I. Introduction

Nowadays, many researchers are interested in emotion recognition [1]. An electroencephalogram (EEG), which is adopted to track and record brainwave patterns, is one of the tools popularly used in this task. Numerous studies mainly focus on accuracy on one subject. However, another aspect that should be emphasized is the capability of generalization to different subjects. It will be more useful to have a model which is independent of any subjects. Creating a subject-independent emotion recognition model is a challenging topic. There is still no standard approach. Yet, the performance was limited due to numerous reasons such as high disparity in EEG settings for each subject and fluctuation of brainwaves.

There could be plentiful types of stimuli in order to evoke emotions such as videos [2], images [3], or even HCI games [4]. Nevertheless, music is one of the most frequently used tools in this research field because it is a powerful method that can arouse a wide diversity of emotions [5]. Moreover, incorporating music in EEG-based emotion recognition can enable several useful applications such as music therapy [6], music recommendation system [7], and multimedia tagging [8].

To classify emotions, many studies have demonstrated the use of multimodal approach which can significantly enhance accuracy of this quest. For instance López-Gil et al. incorporated eye-moving factor in the study [9]. Additionally, during measuring stress response while using a wheelchair [10], four sensors were used which included EEG, heart inter-beat interval (IBI), galvanic skin response (GSR) and stressor level lever. In addition, according to the study of Verma and Tiwary [11], assorted elements were accounted for identifying depression, such as EEG, GSR, blood volume pressure, respiration pattern, skin temperature, electromyogram (EMG) and electrooculogram (EOG). Musical features and EEG signals were combined in several studies as well [12, 13].

Approaching emotion recognition task using only EEG signals may gain lower accuracy, especially in subject-independent fashion due to signals discrepancy across subjects. Nonetheless, the test cannot be intervened by other modalities. Previously, there have been several single-modality ways based on feature extraction [14]. On the other hand, this step could be cumbersome for implementation. Features that are taken out also have to certainly be embedded in every individual's brainwave. Additionally, it could not be guaranteed that all essential features have been addressed. All of these issues are possibly obstacles that can back performance. Finding alternative ways to improve classification performance is always a captivating topic.

Recently, convolutional neural networks (CNNs) were introduced to EEG-based emotion recognition to obtain higher performance in subject-dependent task [15]. CNN utilizes information directly from different electrodes and time steps to predict emotions unlike other traditional methods where feature engineering is necessary due to the non-stationary and complexity of raw EEG signals, e.g., support vector machine (SVM). As stated previously, this according step cannot warrant that all relevant informative features will be extracted. Applying CNN with deep layers can learn more complex features with advantage of being one of the end-to-end models. Furthermore, the spatiotemporal patterns tend to be simultaneously learned. From these differences, CNN may have ability to capture patterns to gain higher performance.

In this paper, we constructed a hypothesis that CNN may have tendency to create a subject-independent emotion recognition model based solely on EEG signals. Moreover, multiple network architectures were deployed to investigate their performance in this task.

978-1-5386-7563-2/19/$31.00 ©2019 IEEE

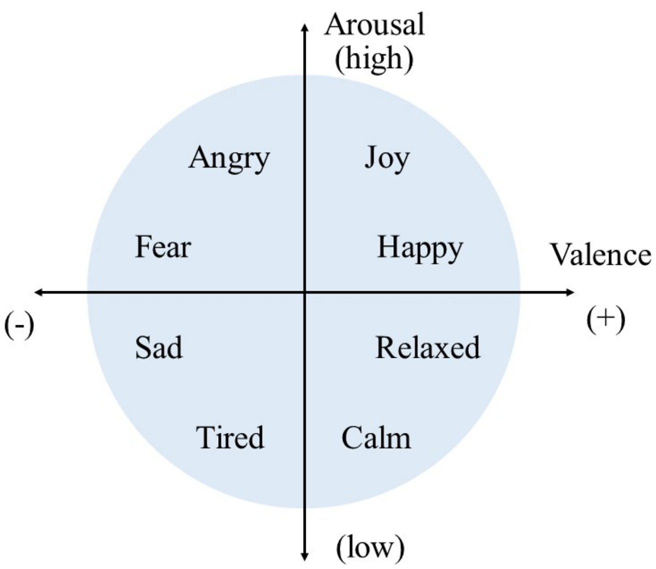

Fig. 1. The two-dimensional emotion model.

## II. Background

A systematic emotion representation is crucial for a computational emotion recognition. One of the most outstanding models is the emotion model proposed by Russell [16]. Its principle is that human emotions could be represented as points in a two-dimensional space of arousal and valence. The illustration can be seen from Fig. 1. Arousal is the vertical axis designating activation levels of emotions while valence is the horizontal axis indicating positivity or negativity of emotions. In this study, the emotion model was adopted to represent emotions.

From our review, music has been used for inducing emotions. According to Lin et al. [17], multilayer perceptron (MLP) and SVM were employed for recognizing four pre-labeled classes of emotions from EEG signals captured from subjects undergoing a music-listening task. Their results were evaluated by 10-fold cross validation achieving accuracy of 81.52 ± 3.71 for MLP and 82.29 ± 3.06 for SVM. As noted in the study of Khalili and Moradi [18], pictures from IAPS were used as stimuli and emotions were categorized into three classes. The outcomes from five subjects using only EEG signals were 66.66% and 76.66% when correlation dimension was combined. One study [19] used deep learning network with covariate shift adaptation whereas another study employed music videos from the DEAP dataset [20] as stimuli. They measured with a leave-one-out cross validation scheme on three levels of valence states and arousal states obtaining 53.42% and 52.05% accuracy. In the study of Lin et al. [13], SVM was used on leave-trial-out evaluation depending on individual's dataset. Furthermore, in the research conducted by Thammasan et al. [12], time-varying characteristics of emotion during music listening were addressed resulting in both subject-dependent and subject-independent. In this study, the same dataset was also employed. Comparison with this work can be found in the section beyond.

CNN has been used to recognize emotions obtained from the DEAP dataset [20] but it was evaluated separately on each subject [15] obtaining 77.98% and 72.98% accuracy on valence and arousal respectively. Additionally, performance of Long Short-Term Memory (LSTM) [21] and combination of CNN and Recurrent Neural Network (RNN) [22] were also investigated on the same dataset. They achieved average subject-dependent accuracy of 85.65% and 85.45% respectively from LSTM model and 5-fold accuracy of 74.12% and 72.06% from CNN and RNN for arousal and valence respectively.

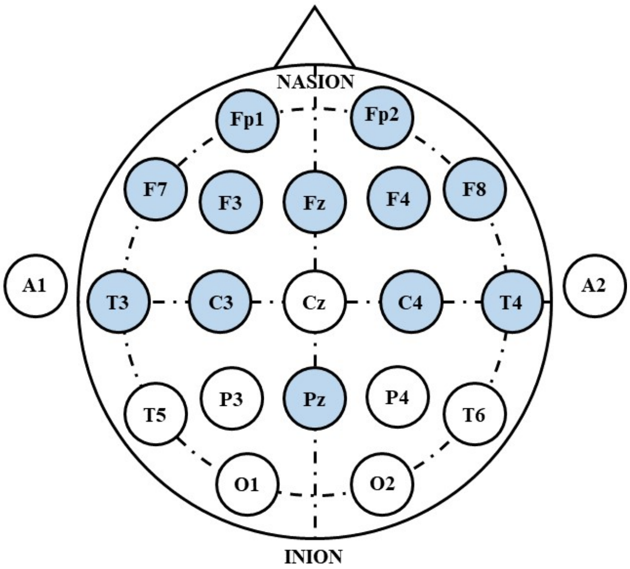

Fig. 2. The 10-20 system of electrode placement showing the selected electrodes.

## III. Research Methodology

### A. Experimental Protocol

The experimental data was collected from twelve healthy male students of Osaka University. The average age was 25.59 years and SD was 1.69 years. None of them had formal music education. Music collection was used as a source of emotion stimulation. Our collection comprised 40 MIDI files with different instrument and tempo. Therefore, emotions induced by lyrics were abolished. Each subject was instructed to select 16 MIDI songs from 40 MIDI songs. Next, they were assigned to listen to the selected songs which were synthesized by the Java Sound API's MIDI package[1]. Each song ended with a 16-second silent resting period in order to reduce any effect from the previous song. EEG signals were recorded at sampling rate of 250 Hz from twelve electrodes on a Waveguard EEG cap[2] placed in accordance with the 10-20 international system using Cz as a reference electrode. Twelve electrodes located near frontal lobe which is a significant part in emotion regulation [6] were picked out of total 21 electrodes, *i.e.*, Fp1, Fp2, F3, F4, C3, C4, Fz, Pz, F7, F8, T3 and T4. The placement illustration is shown as Fig. 2. The impedance of all electrodes was less than 20 kΩ. EEG signals were passed to Polymate AP1532 amplifier and visualized using APMonitor. Both tools were developed by TEAC Corporation[3]. The amplifier was set to include a 60-Hz notch filter. Thus, power line artifact was removed. During each session, all subjects were asked to close their eyes and stay still to avoid unrelated artifacts. After listening to all songs, each subject had to remove the EEG cap and start

[1] https://docs.oracle.com/javase/7/docs/technotes/guides/sound/

[2] https://www.ant-neuro.com/products/waveguard_caps

[3] https://www.teac.co.jp/int/

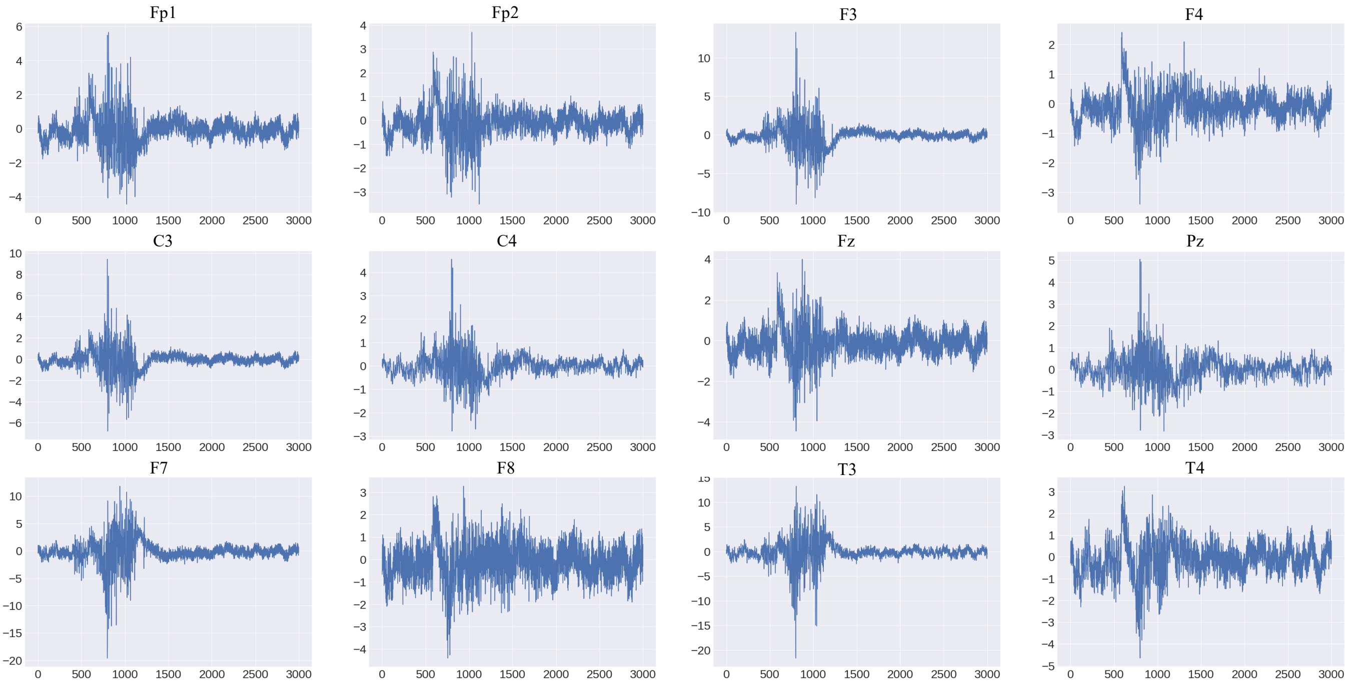

Fig. 3. Sample plots of scaled signals from all 12 electrodes in a 12-second duration.

annotating. The subject listened to the same songs and annotated by continuously clicking on a corresponding point in arousal-valence displayed on a screen. Arousal and valence were recorded independently. After all these processes, the collected data were passed to preprocessing step.

### B. *EEG Preprocessing*

A bandpass filter was applied to filter signals between 0.5-60 Hz. EEGLAB [23], an open source environment for EEG processing under MATLAB, was utilized to remove distinct artifacts contaminating data based on independent component analysis (ICA) such as eye movement, muscle activity and noise. Then, we associated signals with emotion annotation via timestamps.

Next, mean and standard deviation from the signals of every electrode at every time step were calculated to perform feature scaling using standardization. For the sake of simplicity, emotion recognition was measured with a binary classification task for arousal and valence. Arousal values were separated into high and low classes while valence values were divided into positive and negative classes. Following previous preprocessing steps, EEG signals were recorded in a form of columns and rows which represent electrodes and time steps respectively. Consecutive time steps of EEG signals from twelve electrodes with the same valence and arousal classes were considered as a single block. Sample plots could be seen in Fig. 3. Blocks may have different sizes, so they were divided into the size of 55 × 12. If necessary, zero padding was added equally to particular sub blocks.

### C. *Emotion Classification*

As mentioned earlier, our task was to measure emotion recognition with binary classification of arousal and valence classes. Deep CNNs were applied to recognize these values. CNNs with different number of layers ranging from three to seven layers were trained allowing the comparison of performance from each network architecture. Every model was designed to keep information of all electrodes although there were max pooling layers. For regularization, dropouts [24] were employed in every model. After detecting high level features from numerous convolutional layers, fully connected layers were attached at the end of each model. At this point, networks were separated into two parts which may independently predict arousal and valence classes. Detailed architecture models are shown in Table 1.

During training, Adam optimizer [25] was also used to speed up the training. Cross-entropy loss was employed in order to adjust weights in the networks. We used the 10-fold cross validation method (10-fold CV) to investigate the networks potential on capturing wave patterns depending on different time steps and electrodes. Moreover, we also used leave-one-subject-out cross validation method (LOSO CV) to evaluate performance focusing on generalizing to unseen subject. Validation set was randomly selected from training sets of both evaluations. Importantly, the selected validation set was taken out of the training set. Once validation loss increased, training was terminated to avoid overfitting.

Regarding a performance evaluation, emotion recognition accuracy was calculated by finding the percentage of the number of test instances that were classified correctly in the total number of test instances. This can be calculated by

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\,, \tag{1}$$

while *TP* is the number of true positives, *TN* is the number of true negatives, *FP* is the number of false positives and *FN* is the number of false negatives.

TABLE I. NETWORK ARCHITECTURES

<table>
<tr><th>Index</th><th colspan="2">Conv3</th><th colspan="2">Conv4</th><th colspan="2">Conv5</th><th colspan="2">Conv6</th><th colspan="2">Conv7</th></tr>
<tr><td><b>1</b></td><td colspan="2">Conv2D 12x12x32</td><td colspan="2">Conv2D 12x12x16</td><td colspan="2">Conv2D 12x12x16</td><td colspan="2">Conv2D 12x12x16</td><td colspan="2">Conv2D 12x12x16</td></tr>
<tr><td><b>2</b></td><td colspan="2">Conv2D 2x1x32</td><td colspan="2">Conv2D 2x1x32</td><td colspan="2">Conv2D 2x1x32</td><td colspan="2">Conv2D 2x1x32</td><td colspan="2">Conv2D 2x1x32</td></tr>
<tr><td><b>3</b></td><td colspan="2">MaxPooling 4x1</td><td colspan="2">MaxPooling 2x1</td><td colspan="2">Conv2D 2x1x64</td><td colspan="2">Conv2D 2x1x64</td><td colspan="2">Conv2D 2x1x64</td></tr>
<tr><td><b>4</b></td><td colspan="2">Conv2D 2x1x64</td><td colspan="2">Conv2D 2x1x64</td><td colspan="2">MaxPooling 2x1</td><td colspan="2">MaxPooling 2x1</td><td colspan="2">Conv2D 2x1x128</td></tr>
<tr><td><b>5</b></td><td colspan="2">Dropout 0.5</td><td colspan="2">MaxPooling 2x1</td><td colspan="2">Conv2D 2x1x128</td><td colspan="2">Conv2D 2x1x128</td><td colspan="2">Conv2D 2x1x256</td></tr>
<tr><td><b>6</b></td><td>FC 128x1</td><td>FC 128x1</td><td colspan="2">Conv2D 2x1x128</td><td colspan="2">Conv2D 2x1x256</td><td colspan="2">Conv2D 2x1x256</td><td colspan="2">Conv 2x1x512</td></tr>
<tr><td><b>7</b></td><td>Dropout 0.5</td><td>Dropout 0.5</td><td colspan="2">Dropout 0.25</td><td colspan="2">Dropout 0.25</td><td colspan="2">Conv2D 2x1x512</td><td colspan="2">Conv2D 2x1x1024</td></tr>
<tr><td><b>8</b></td><td>FC 2x1</td><td>FC 2x1</td><td>FC 128x1</td><td>FC 128x1</td><td>FC 128x1</td><td>FC 128x1</td><td colspan="2">Dropout 0.25</td><td colspan="2">MaxPooling 2x1</td></tr>
<tr><td><b>9</b></td><td></td><td></td><td>Dropout 0.5</td><td>Dropout 0.5</td><td>Dropout 0.5</td><td>Dropout 0.5</td><td>FC 128x1</td><td>FC 128x1</td><td colspan="2">Dropout 0.25</td></tr>
<tr><td><b>10</b></td><td></td><td></td><td>FC 2x1</td><td>FC 2x1</td><td>FC 2x1</td><td>FC 2x1</td><td>Dropout 0.5</td><td>Dropout 0.5</td><td>FC 128x1</td><td>FC 128x1</td></tr>
<tr><td><b>11</b></td><td></td><td></td><td></td><td></td><td></td><td></td><td>FC 2x1</td><td>FC 2x1</td><td>Dropout 0.5</td><td>Dropout 0.5</td></tr>
<tr><td><b>12</b></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>FC 2x1</td><td>FC 2x1</td></tr>
</table>

Moreover, our self-reporting emotion annotation method might lead to data imbalance. Matthews Correlation Coefficient (MCC) [26] was also adopted. It reflects the performance including class imbalance factor. MCC can be calculated by

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (2)$$

Equation (2) shown above uses the same abbreviations as (1). MCC ranges from -1 to 1. The maximal coefficient could imply 100% accuracy while the minimal coefficient could represent 0% accuracy.

This value could help accuracy report become more reliable since it reduces effects from class imbalance during the experiment.

## IV. RESULTS AND DISCUSSION

From 10-fold CV, the networks showed their potential in recognizing EEG signals directly from different electrodes and time steps to predict emotions with high accuracy. Among them, the model with six convolutional layers had the greatest performance. The model predicted arousal with the accuracy of 81.54% and MCC value of 0.630892; at the same time, it achieved valence accuracy of 86.87% and MCC value of 0.737470. Results are shown as Table. 2 and Fig. 4.

TABLE II. RESULTS FROM 10-FOLD CROSS VALIDATION

| **#Conv Layers** | **Accuracy (MCC)** | |
|---|---|---|
| | ***Arousal*** | ***Valence*** |
| 3 | 77.97% (0.559440) | 84.14% (0.682812) |
| 4 | 79.41% (0.588295) | 85.24% (0.704820) |
| 5 | 80.91% (0.618231) | 85.97% (0.719308) |
| 6 | 81.54% (0.630892) | 86.87% (0.737470) |
| 7 | 80.59% (0.611720) | 86.35% (0.727043) |

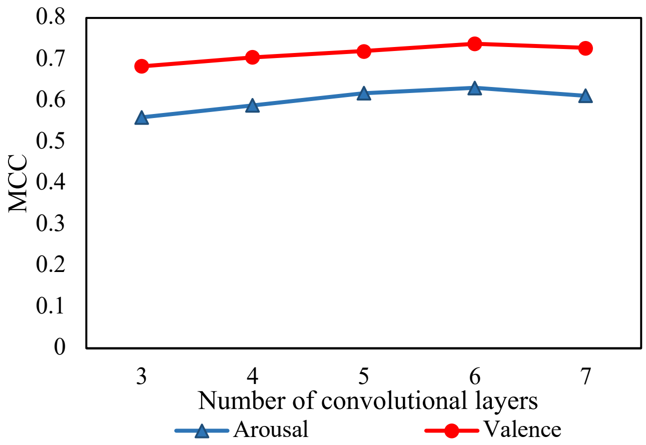

Fig. 4. MCC values of arousal and valence over number of convolutional layers on 10-fold cross validation method.

As reported by LOSO CV, their performance dropped since they were tested with an unseen subject. Regarding arousal accuracy, the model with six convolutional layers gained the highest performance with 56.22% accuracy and MCC value of 0.124365. For valence, the model with four convolutional layers had the highest accuracy of 68.75% and MCC value of 0.374912. Results of our LOSO CV are depicted in Table. 3 and Fig. 5.

Compared to the supplemental file of the previous study using SVM on the same dataset [12], all of our models demonstrated superior performance. As reported by this research, LOSO CV using only EEG signals had highest arousal accuracy of 45.17% and valence accuracy of 55.47%. Although we counted instances differently, the trends could be obviously observed. For clarity, our models with four and six convolutional layers were selected as representatives of the best valence and arousal results, respectively, to be illustrated alongside SVM results in Table 4.

According to our obtained results, valence recognition gained higher accuracy in both validation methods. This could be implied that valence is more observable than arousal from EEG signals during music listening.

Our proposed methodology employed CNN that analyzed data from different electrodes and from different time steps to see signal dynamics through time, and thus successfully improved the generalization of the classification model to unseen subjects. Additionally, adding more convolutional layers could possibly be helpful to 10-fold CV since the networks might be more capable of detecting features in higher dimensions. On the other hand, this was not likely to benefit the LOSO CV since the test set was less similar to the training set. Under the nonstationary characteristic of EEG signals, obtaining elevated performance has shown progress in this field. Even though computation power has increased compared to our previous work [12], accuracy of classifiers from this study have clearly shown improvement. However, there are still opportunities for further enhancements since evaluation from 10-fold CV had significantly higher performance. This suggests that emotion classification model can achieve accuracy as high as results from 10-fold CV if the model is completely independent to subjects. Therefore, future work should focus on increasing the number and diversity of subjects, e.g., including female subjects, collecting more samples for training and testing models. In addition, thoroughly examining relations between subject's characteristics and his brainwave patterns is also encouraged for further studies. Besides, furtherance of EEG recording device could reasonably increase this quest's performance and decrease inter-subject variation.

TABLE III. Results from Leave-one-subject-out Validation Method

| #Conv Layers | Accuracy (MCC) | |
|---|---|---|
| | *Arousal* | *Valence* |
| 3 | 50.15% (0.002999) | 64.86% (0.297071) |
| 4 | 51.82% (0.036507) | 68.75% (0.374912) |
| 5 | 52.00% (0.036507) | 61.59% (0.374912) |
| 6 | 56.22% (0.124365) | 65.15% (0.303020) |
| 7 | 50.81% (0.016220) | 68.54% (0.370895) |

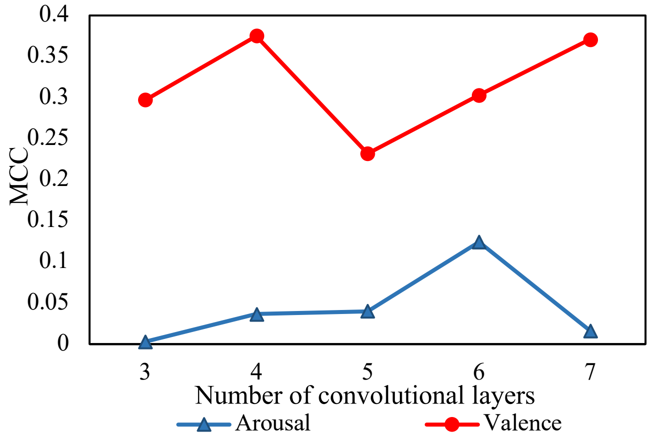

Fig. 5. MCC values of arousal and valence over number of convolutional layers on leave-one-subject-out validation method.

TABLE IV. CNNs and SVM Leave-one-subject-out Results Comparison

| Architecture | Accuracy (MCC) | |
|---|---|---|
| | *Arousal* | *Valence* |
| SVM | 45.17% (-) | 55.47% (-) |
| CNN 4 Conv Layers | 51.82% (0.036507) | 68.75% (0.374912) |
| CNN 6 Conv Layers | 56.22% (0.124365) | 65.15% (0.303020) |

## V. Conclusion

In this work, we have presented a study of subject-independent emotion recognition during music listening based on EEG using deep CNNs according to the hypothesis that CNNs could recognize patterns from EEG by having advantage from using data from electrodes and time steps with automatic feature extraction. According to the experiment, EEG signals and emotion annotation on arousal-valence space were collected. CNNs with distinct architectures were employed to classify arousal and valence independently using 10-fold CV and LOSO CV. The results showed that all of our models were superior to the old methodology which used SVM classifier under fluctuation constraint from EEG signals. Using CNN with information from different electrodes and time steps could gain higher performance. Nevertheless, there are still rooms for improvement since results from 10-fold CV had significantly higher performance as all subjects were seen. Obtaining more diverse and bigger datasets along with deliberately examining relation between subject's characteristics and brainwave will probably achieve higher accuracy.

## Acknowledgment

We would like to thank Panyawut Sri-iesaranusorn for his helpful technical advices.

## References

[1] R. A. Calvo and S. D. Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," *IEEE Transactions on Affective Computing,* vol. 1, no. 1, pp. 18-37, 2010.

[2] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection," *IEEE Transactions on Affective Computing,* vol. 7, no. 1, pp. 17-28, 2016.

[3] C.-T. Wu, G. D. Dillon, H.-C. Hsu, S. Huang, E. Barrick, and Y.-H. Liu, "Depression Detection Using Relative EEG Power Induced by Emotionally Positive Images and a Conformal Kernel Support Vector Machine," *Applied Sciences,* vol. 8, no. 8, 2018.

[4] H. Kandemir and H. Kose, "Effects of Physical Activity Based HCI Games on the Attention, Emotion and Sensory-Motor Coordination," in *Advances in Service and Industrial Robotics*, Cham, 2019, pp. 718-727: Springer International Publishing.

[5] S. Koelsch, "Brain and music," in *Brain and music.*, ed: Wiley-Blackwell, 2012, pp. xiii, 308-xiii, 308.

[6] S. Koelsch, "Brain correlates of music-evoked emotions," *Nature Reviews Neuroscience,* Review Article vol. 15, p. 170, 2014.

[7] H.-Y. Chang, S.-C. Huang, and J.-H. Wu, "A personalized music recommendation system based on electroencephalography feedback," *Multimedia Tools and Applications,* vol. 76, no. 19, pp. 19523-19542, 2017.

[8] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," *IEEE Transactions on Affective Computing,* vol. 3, no. 1, pp. 42-55, 2012.

[9] J.-M. López-Gil, J. Virgili-Gomá, R. Gil, and R. García, "Method for Improving EEG Based Emotion Recognition by Combining It with Synchronized Biometric and Eye Tracking Technologies in a Non-invasive and Low Cost Way," *Frontiers in computational neuroscience,* vol. 10, pp. 85-85, 2016.

[10] J. Abdur-Rahim *et al.*, "Multi-Sensor Based State Prediction for Personal Mobility Vehicles," *PLoS ONE 11(10): e0162593,* 2016.

[11] G. K. Verma and U. S. Tiwary, "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals," *NeuroImage,* vol. 102, pp. 162-172, 2014.

[12] N. Thammasan, K.-i. Fukui, and M. Numao, "Multimodal Fusion of EEG and Musical Features in Music-Emotion Recognition," in *2017 AAAI Conference on Artificial Intelligence*, 2017.

[13] Y.-P. Lin, Y.-H. Yang, and T.-P. Jung, "Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening," *Frontiers in Neuroscience,* vol. 8, no. 94, 2014.

[14] M.-K. Kim, M. Kim, E. Oh, and S.-P. Kim, "A Review on the Computational Methods for Emotional State Estimation from the Human EEG," *Computational and Mathematical Methods in Medicine,* vol. 2013, p. 13, 2013.

[15] Z. Wen, R. Xu, and J. Du, "A novel convolutional neural networks for emotion recognition based on EEG signal," in *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, 2017, pp. 672-677.

[16] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology,* vol. 39, no. 6, pp. 1161-1178, 1980.

[17] Y. Lin *et al.*, "EEG-Based Emotion Recognition in Music Listening," *IEEE Transactions on Biomedical Engineering,* vol. 57, no. 7, pp. 1798-1806, 2010.

[18] Z. Khalili and M. H. Moradi, "Emotion recognition system using brain and peripheral signals: Using correlation dimension to improve the results of EEG," in *2009 International Joint Conference on Neural Networks*, 2009, pp. 1571-1575.

[19] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation," *The Scientific World Journal,* vol. 2014, p. 10, 2014.

[20] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," *IEEE Trans. Affect. Comput.,* vol. 3, no. 1, pp. 18-31, 2012.

[21] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion Recognition based on EEG using LSTM Recurrent Neural Network," *International Journal of Advanced Computer Science and Applications(IJACSA),* vol. 8, no. 10, 2017.

[22] L. Xiang, S. Dawei, Z. Peng, Y. Guangliang, H. Yuexian, and H. Bin, "Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 352-359.

[23] A. Delorme *et al.*, "EEGLAB, SIFT, NFT, BCILAB, and ERICA: New Tools for Advanced EEG Processing," *Computational Intelligence and Neuroscience,* vol. 2011, p. 12, 2011.

[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research,* vol. 15, no. 1, pp. 1929-1958, 2014.

[25] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[26] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochimica et Biophysica Acta (BBA) - Protein Structure,* vol. 405, no. 2, pp. 442-451, 1975.