

An approach of re-organizing input dataset to enhance the quality of emotion recognition using the bio-signals dataset of MIT

Van-Dung Pham, Thanh-Long Cung

School of Electrical Engineering, Hanoi University of Science and Technology, Vietnam

Article Info

Article history:

Received Jul 27, 2021

Revised Sep 7, 2021

Accepted Oct 12, 2021

Keywords:

Emotion recognition

Emotion state

Machine learning

Overlapped segments

Physiological signal

ABSTRACT

The purpose of this paper is to propose an approach of re-organizing input data to recognize emotion based on short signal segments and increase the quality of emotional recognition using physiological signals. MIT's long physiological signal set was divided into two new datasets, with shorter and overlapped segments. Three different classification methods (support vector machine, random forest, and multilayer perceptron) were implemented to identify eight emotional states based on statistical features of each segment in these two datasets. By re-organizing the input dataset, the quality of recognition results was enhanced. The random forest shows the best classification result among three implemented classification methods, with an accuracy of 97.72% for eight emotional states, on the overlapped dataset. This approach shows that, by re-organizing the input dataset, the high accuracy of recognition results can be achieved without the use of EEG and ECG signals.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Thanh-Long Cung

School of Electrical Engineering

Hanoi University of Science and Technology

Dai-Co-Viet street, Hai-Ba-Trung district, Hanoi, Vietnam

Email: long.cungthanh@hust.edu.vn

1. INTRODUCTION

Emotions are natural responses of the human brain to circumstances or external impacts. Positive emotions can be highly effective at work or daily life, whereas prolonged negative emotions can cause human health problems. Emotion recognition is a research topic of interest because it has a myriad of potential applications in life. Emotion recognition models can be used to estimate the level of stress, measure the emotional changes in autism patients, build distance learning programs that can self-adjust to learners' emotions, or develop robots that can interact emotionally with people. Generally, emotion recognition can be divided into three main groups, which are using visual signals (facial expressions, facial gestures), audio signals (human's speech), and physiological signals (galvanic skin response - GSR, electrocardiogram-ECG, electroencephalogram-EEG) to recognize human emotional states.

Previous researches using images and speech have shown remarkable results in accuracy [1]-[6]. However, the reliability of systems using these methods may not be guaranteed in exceptional cases, such as when subjects intentionally change facial expressions and speech, or subjects have defects on face or speech. Meanwhile, physiological signals are the human body's natural responses, they change according to emotional states and cannot be easily controlled. The results of emotion recognition using physiological signals are therefore highly reliable.

Various types of physiological signals have been used in previous studies on emotion recognition. The author of [7] used fast fourier transform (FFT) and wavelet transform features of 2 channels Fp1 and Fp2 of EEG signals to recognize two classes of (high and low) valence and arousal and achieved the accuracy of 76.34% and 75.18%, respectively. The combination of features extracted from respiration (RESP) and blood volume pulse (BVP) signals and convolutional neural network (CNN) showed an accuracy of 94.02% in identifying six emotional states [8]. The distinguish between negative emotions and neutral state were achieved in [9]. In this research, three physiological signals GSR, heart rate (HR), and RESP were used, and the accuracy of the obtained recognition results is 94%.

Various types of classification techniques have been applied to emotion recognition. Traditional methods, such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), k-nearest neighbor (KNN), random forest (RF), support vector machine (SVM),... were adopted in many studies and showed good results. The work in [10] used the KNN classifier, and the result of recognizing five affective states “disgust”, “happy”, “fear”, “neutral”, and “sad” are 90.83%, 100%, 94.17%, 90.28%, and 43.89% respectively. The authors used the SVM technique to identify two classes (high and low) of valence and arousal on EEG signals [11]. The obtained results are 91.3% and 91.1% of accuracy. Other techniques using deep learning, such as deep neural network (DNN), CNN, long-short term memory (LSTM), have been studied and applied to physiological emotion recognition when working with a large amount of data [8], [12]-[15]. Alhagry *et al.* [12] proposed an LSTM network for classifying two classes (high and low) of valence, arousal, and liking on EEG signal and achieved an accuracy of 85.65%, 85.45%, and 87.99%, respectively. A method of using pre-trained CNN model was used in [14], which achieved 72.81% and 81.8% accuracy in recognizing two emotion classes on DEAP and the Loughborough University Multimodal Emotion Dataset (LUMED) dataset, respectively.

Many studies used EEG signal [7], [11], [12], [16]-[18] or ECG signal [19]-[23] showed favorable results. However, the use of these signals requires very complicated data collection systems. Meanwhile, the data segments used to identify emotions in these studies are often large (longer than 1 minute), which is unsuitable for identifying transient emotional states and online recognition applications.

The MIT physiological dataset has been used in a lot of previous researches. Picard *et al.* applied sequential forward floating selection (SFFS) and LDA simultaneously for dimensional reduction and used maximum a posteriori estimation (MAP) for classification [24]. The obtained results showed an accuracy of 81.25% over eight emotional states. Sequential forward selection (SFS) and linear discriminant function (LDF) were applied [25]. The recognition results showed an accuracy of 85% for high/low valence and 87.5% for high/low arousal.

In this work, we proposed an approach to enhance the recognition quality by re-organizing the input data and using statistical features of BVP, EMG, GSR, RESP signals to classify eight emotional states of the MIT dataset. All raw segments of signals in the dataset are divided into shorter segments for data augmentation purposes. It also allows us to speed up the procedure from data preprocessing to classification. The paper is organized into the following sections: section 2 presents data preprocessing and feature extraction; section 3 explains the classification techniques applied in our work, section 4 is the recognition results and discussions, and the conclusions are presented in section 5.

2. DATA PREPROCESSING AND FEATURE EXTRACTION

The MIT dataset consists of 4 different physiological signals: EMG, BVP, GSR, and RESP; which represent eight emotional states: No emotion, Anger, Hate, Grief, P-love, R-love, Joy, and Reverence. All data were recorded throughout many days and divided into two subsets (set A and set B-Figure 1). In particular, set A consists of data recording in 20 days without the faults of sensors. This subset includes 20 files containing signal segments of 4 types of signals (EMG, BVP, GSR, and RESP). Set B includes all the collected data. Each signal segment of set-B is longer than that of set A and contains a corresponding signal segment of the set A within in Figure 1.

2.1. Data preprocessing

Data from set A are sampled at 20Hz, including 20 files corresponding to 20 days of data collection. Each file contains 32 signal segments of 4 physiological signals (BVP, EMG, GSR, and RESP), corresponding to 8 different emotional states. Each segment lasts 100 seconds (including 2000 discrete values). We grouped four signal segments of 4 physiological signals of the same emotional state into a signal set from each file, yielding eight signal sets per file. In total, 160 signal sets corresponding to 8 emotional states (20 signal sets for a state) have resulted from the grouping process.

We implemented a method for segmenting and re-organizing the dataset to speed up the recognition process and augment the amount of data for recognition. For every 100 second long segment of a certain affective state, we cut it into shorter segments that last only 10 seconds. There are two different ways of data

re-organizing. For the first way, the original segment is divided into ten separate segments with the length of 10 seconds for each (non-overlapped data). For the second way, each original segment is divided into overlapped segments (overlapped data) with an overlap rate of 70%, and the length of each new data segment is still 10 seconds. It means the next segment contains 7 seconds of the previous segment. Each short segment is then labeled corresponding to the label of its original segment. By applying these ways of data re-organizing, we created two new datasets: the first one called "non-overlapped dataset" containing 1600 sets of signals (200 sets per one emotional state, each set has four different types of physiological signal), and the second one called "overlapped dataset" consisting of 4960 signals sets (620 sets per one emotional state). These two new datasets were then divided into the ratio of 80% and 20% for training and testing, respectively.

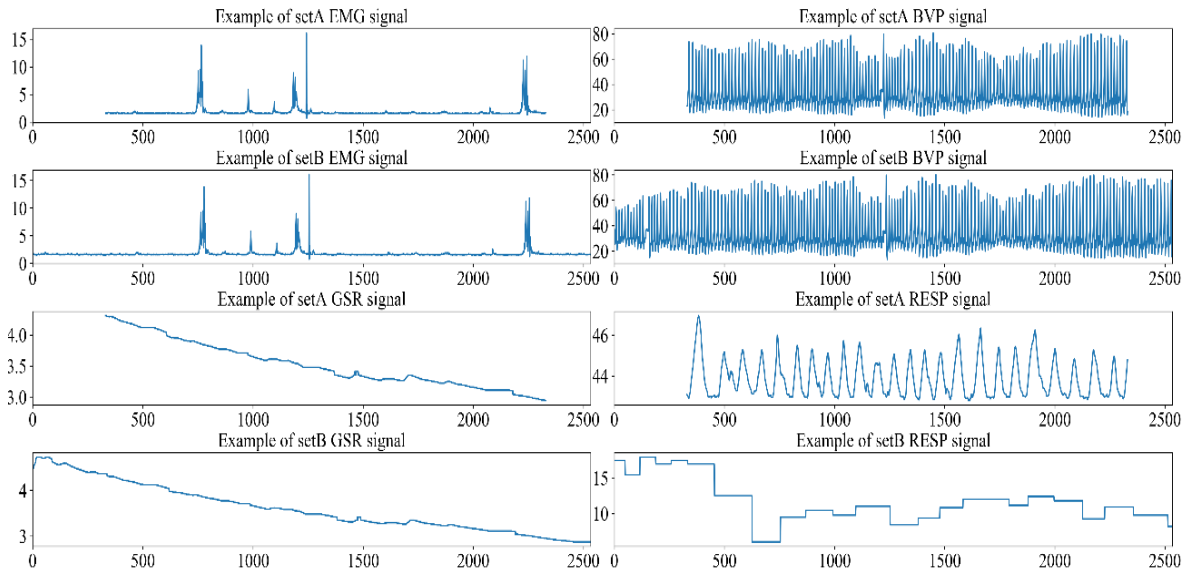


Figure 1. Physiological signals from two subsets of MIT dataset

2.2. Feature extraction

We used statistical features like mean, median, maximum, minimum, standard deviation, variance, range, skewness, and kurtosis to simplify the calculation for short signal segments. Besides, two features: first absolute mean difference (first-degree difference) and second absolute mean difference (second-degree difference), which were used in [24] are also calculated for raw and Z-score normalized signal segments. All features are calculated following the formulas given in Table 1.

Table 1. The formula of statistical features

Feature's name	Formula	Feature's name	Formula
minimum	$\min(X)$	range	$\max(X) - \min(X)$
maximum	$\max(X)$	skewness	$\frac{\sqrt{N(N-1)} \sum_{n=1}^N (x_n - \mu_x)^3 / N}{s^3}$
mean	$\frac{1}{N} \sum_{n=1}^N x_n$	kurtosis	$\frac{\sum_{n=1}^N (x_n - \mu_x)^4 / N}{s^4}$
Standard deviation	$\sqrt{\frac{1}{N} \sum_{n=1}^N (x_n - \mu_x)^2}$	first-degree difference	$\frac{1}{N-1} \sum_{n=1}^{N-1} x_{n+1} - x_n $
variance	$\frac{1}{N} \sum_{n=1}^N (x_n - \mu_x)^2$	second-degree difference	$\frac{1}{N-2} \sum_{n=1}^{N-2} x_{n+2} - x_n $

X denotes a signal segment; x_n denotes the n th value of the raw signal segment ($n = 1, 2, \dots, N - N$ is the length of the segment); s denotes the standard deviation

In this work, we choose SFFS and sequential backward floating selection (SBFS) as techniques for feature selection. These are two algorithms of the sequential feature algorithms (SFAs) family, which performed well in previous works [7], [24], [25]. The purpose of these algorithms is to reduce the dimension of the initial feature space to a lower-dimensional subspace by choosing only a subset of features, which is considered the most relevant to the problem. SFFS starts with an empty set and then each feature from the initial features set is added in. The best subset of a given number of features is selected according to an evaluation criterion. The SBFS algorithm is similar to the SFFS, but it starts with all initial features, then each feature is consecutively eliminated.

The best feature set was selected from the results of two algorithms SFFS, and SBFS. To do so, the accuracy of the KNN classification technique is used as the evaluation criterion. Using the tool of "grid search with cross-validation" to evaluate the accuracy of KNN, we obtained the cross-validation (CV) scores of SBFS and SFFS which are shown in Figure 2. For the non-overlapped dataset, a subset of 13 features was selected (with the highest CV of 62%). For the overlapped dataset, the subset of 19 features was selected (with the highest CV of 87%). These features will be used as input for the classification algorithm.

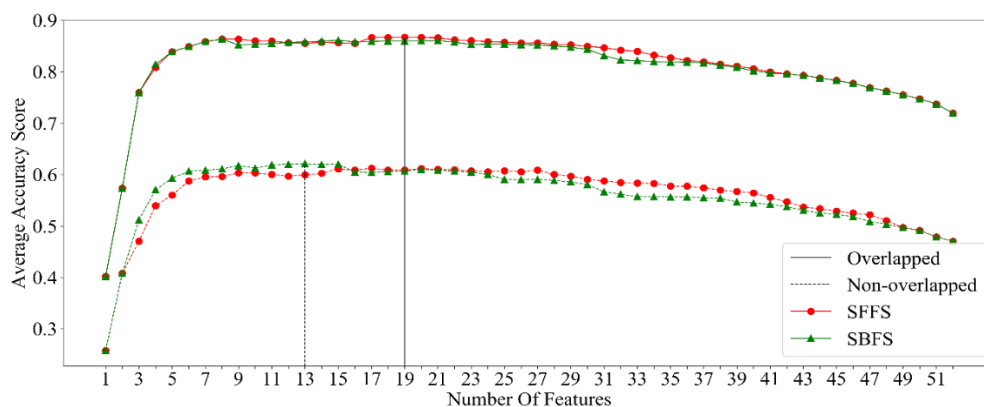


Figure 2. Cross-validation score of SFFS and SBFS algorithms

3. CLASSIFICATION

In this paper, SVM, RF, and multi-layer perceptron (MLP) are used as classifiers to recognize different emotional states and compare these techniques' effectiveness on each dataset. SVM finds a hyperplane that best separates classes in the feature space of the data. SVM can work efficiently with both linear and nonlinear data by using kernel functions. Some common kernel functions are linear, radial basis function (rbf) and polynomial (poly). RF is a classification technique that uses a combination of decision trees to create a large classifier through the selection mechanism between trees. RF is very effective for classification problems because it simultaneously mobilizes hundreds of smaller models inside it, with different rules, to make a final decision. This technique has a fast training and prediction speed. It works well on many types of data and has a high ability of noise resistance. MLP neuron networks use successive neuron layers to make predictions corresponding to their input data. This kind of neural network works well on both linear and nonlinear data.

In our works, the rbf kernel is chosen for the SVM classifier. This is the most commonly used kernel function because of its fast calculation and high efficiency on nonlinear data. The appropriate set of hyperparameters of both SVM and RF classifiers were selected using the grid search CV. For neural networks, two hidden layers of the MLP network are used and evaluated. The rectified linear activation unit (ReLU) function in Figure 3 - a non-linear activation function that allows the network to approximate a nonlinear relationship between input features and output labels - was applied to each layer. For selecting the optimal number of neurons in hidden layers, Bayesian Optimization - an optimization algorithm based on the Bayes theorem - was used. To find the minimum value of the target function, Bayesian optimization constructs a function called surrogate function (a probability model) based on the previous evaluation results of the target function. This allows the algorithm to spend more time evaluating promising parameter values, therefore, less time-consuming and computational costs will be spent on areas in which the model shows poor performance. This optimization algorithm is now commonly used for the problem of optimizing the structure of deep neural networks. In addition, we use dropout - a technique that helps avoid the over-fitting phenomenon - for training our neural networks.

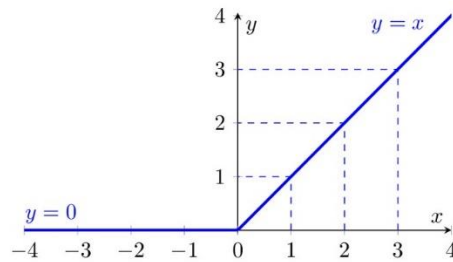


Figure 3. Rectified linear unit

4. RESULTS AND DISCUSSION

After training SVM, RF, and MLP classifiers using features vectors (13 features for the non-overlapped dataset and 19 features for the overlapped dataset), the recognition results are summarized in Table 2.

Table 2. Classification results

Classifier (%)	No emotion (%)	Anger (%)	Hate (%)	Grief (%)	P-love (%)	R-love (%)	Joy (%)	Reverence (%)	Total (%)
Non-overlapped data									
SVM	77.5	77.5	72.5	62.5	80.0	62.5	60.0	85.0	72.19
RF	80.0	90.0	75.0	65.0	75.0	65.0	67.5	95.0	76.56
MLP	77.5	77.5	75.0	67.5	80.0	57.5	62.5	87.5	73.12
Overlapped data									
SVM	93.55	94.62	97.85	94.62	93.55	86.02	93.55	98.92	94.09
RF	98.92	97.85	100	97.85	97.85	90.32	98.92	100	97.72
MLP	97.85	97.85	97.85	88.17	97.85	82.80	96.77	96.77	94.49

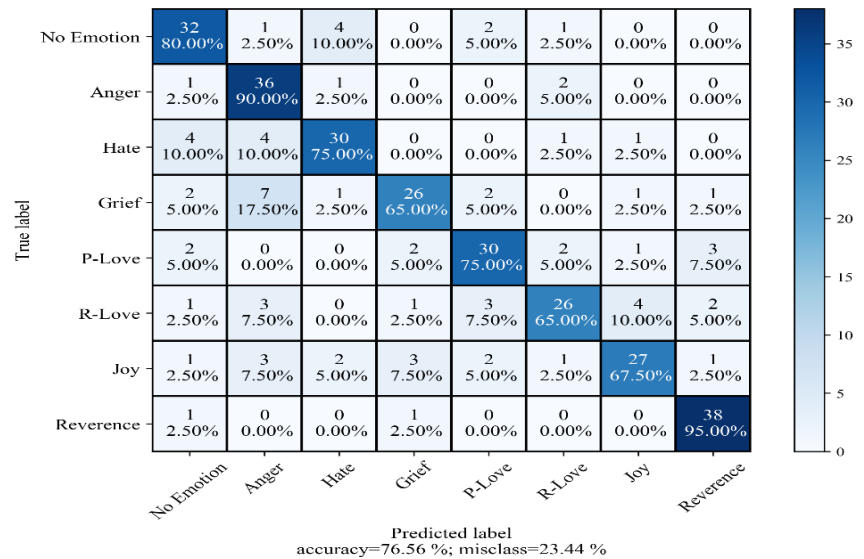
Using the RF classifier on the non-overlapped data, the highest accuracy achieved on the test-set is 76.56%. By using SVM with the same test-set, the accuracy is 72.19%. Testing MLP (with two hidden layers in which there are 128 neurons in the first layer, 96 neurons in the second layer, and the dropout ratio for each layer is 0.2) on this dataset, the obtained accuracy is 73.12%.

For the overlapped dataset, the obtained results are much better than that of the non-overlapped dataset. By using the RF classification technique, the accuracy of recognition results is 97.72%. By using SVM and MLP (with two hidden layers, in which the first layer consisting of 128 neurons, the second layer consisting of 80 neurons, and the dropout ratios for each layer are 0.2 and 0.0, respectively), the accuracy of obtained results are 94.09% and 94.49% respectively, as shown in Table 2.

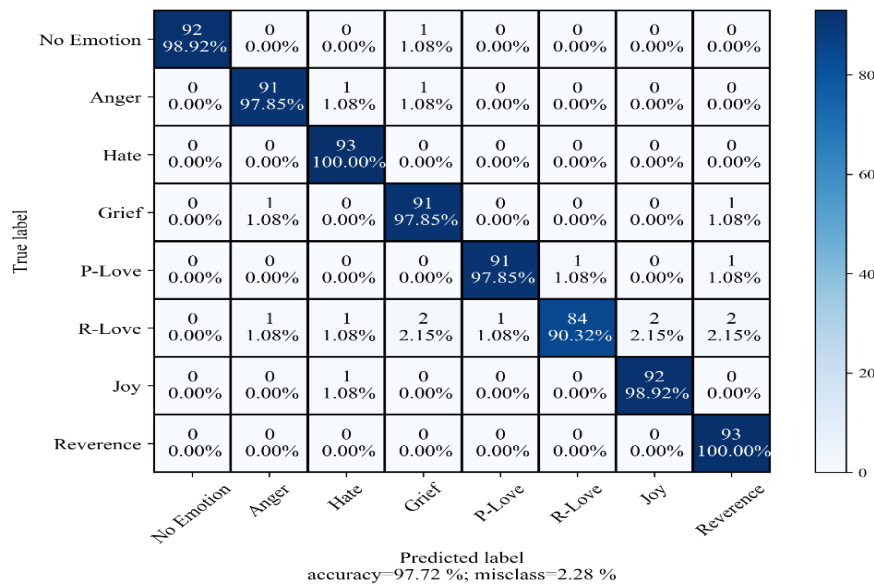
Table 2 also showed that the lowest accuracy of recognition results is with the R-Love state for both overlapped and non-overlapped data (except the case of using SVM with non-overlapped data, where the classification of Joy state has the lowest accuracy). The highest accuracy is always presented on the Reverence state in both ways of data enrichment (95% for non-overlapped data and 100% for overlapped data).

From the confusion matrices in Figure 4, for non-overlapped data, R-Love and Grief states have the lowest recognition accuracy. For R-Love, 50% of recognition mistakes fall on two emotion states, Joy and P-Love. Negative states (Anger, Hate, Grief) are mostly misclassified to each other or No emotion state. For example, 71% (10/14) of the total misclassifications are of Grief to Anger, Hate, and No emotion; 80% (8/10) are of Hate state to Anger and No emotion).

The result of 76.56% with the non-overlapping data over eight emotional states is almost equivalent to the result of the previous works using the MIT dataset (for example, 81.25% in [24]). Meanwhile, when using overlapping data, our recognition accuracy is over 90% for all used classification techniques - RF, SVM, and MLP. The highest accuracy was obtained by using RF to recognize all eight different emotional states, with an average rate of 97.72%.



(a)



(b)

Figure 4. Confusion matrices of RF classifier; (a) non-overlap data, (b) overlapped data

5. CONCLUSIONS

In this paper, we presented a method for re-organizing the MIT dataset by splitting and overlapping data. The obtained result is a much bigger dataset (more than 600 sets of 4 signals per emotional state), with a length of 10 seconds for each signal segment. Using this new dataset with 19 extracted features selected by the SFFS algorithm and applying the RF technique, the recognition results showed high accuracy (97.72%) on all eight emotional states. This indicates that, with an appropriate organization of input dataset, the combination of some physiological signals, such as BVP, EMG, GSR, RESP can also give more accurate results in emotion recognition without EEG, ECG signals.

REFERENCES

[1] D. Meng, X. Peng, K. Wang and Y. Qiao, "Frame attention networks for facial expression recognition in videos," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, doi: 10.1109/ICIP.2019.8803603.
 [2] A. Kumar, N. Garg and G. Kaur, "An Emotion Recognition Based on Physiological Signals," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 95, pp. 2278-3075, 2019, doi: 10.35940/ijitee.I1054.0789S19.

- [3] J. Umamaheswari and A. Akila, "An Enhanced Human Speech Emotion Recognition Using Hybrid of PRNN and KNN," *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 177-183, doi: 10.1109/COMITCon.2019.8862221.
- [4] H. Kusuma, M. Attamimi and H. Fahrudin, "Deep learning based facial expressions recognition system for assisting visually impaired persons," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 3, pp. 1208-1219, Jun. 2020, doi: 10.11591/eei.v9i3.2030.
- [5] M. F. Alghifari, T. S. Gunawan, M. A. bin W. Nordin, S. A. A. Qadri, M. Kartiwi, Z. Janin, "On the use of voice activity detection in speech emotion recognition," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 4, pp. 1324-1332, 2019, doi: <https://doi.org/10.11591/eei.v8i4.1646>.
- [6] S. Mirsamadi, E. Barsoum and C. Zhang, "Automatic speech emotion recognition using recurrent neural networks with local attention," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 2227-2231, doi: 10.1109/ICASSP.2017.7952552.
- [7] S. Wu, X. Xu, L. Shu and B. Hu, "Estimation of valence of emotion using two frontal EEG channels," *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 1127-1130, doi: 10.1109/BIBM.2017.8217815.
- [8] S. Oh, J.-Y. Y. Lee and D. K. Kim, "The Design of CNN Architectures for Optimal Six Basic Emotion Classification Using Multiple Physiological Signals," *Sensors*, vol. 20, no. 3, p. 866, 2020, doi: 10.3390/s20030866.
- [9] M. Monajati, S. H. Abbasi, F. Shanania and S. Shamekhi, "Emotions states recognition based on physiological parameters by employing of fuzzy-adaptive resonance theory," *International Journal of Intelligence Science*, vol. 2, no. 4, pp. 166-175, 2012, doi: 10.4236/ijis.2012.224022.
- [10] M. Murugappan, "Electromyogram signal based human emotion classification using KNN and LDA," *2011 IEEE International Conference on System Engineering and Technology*, 2011, pp. 106-110, doi: 10.1109/ICSEngT.2011.5993430.
- [11] O. Bazgir, Z. Mohammadi and S. A. H. Habibi, "Emotion Recognition with Machine Learning Using EEG Signals," *2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME)*, 2018, pp. 1-5, doi: 10.1109/ICBME.2018.8703559.
- [12] S. Alhagry, A. A. Fahmy and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Emotion*, vol. 8, no. 10, pp. 355-358, 2017, doi: 10.14569/IJACSA.2017.081046.
- [13] S. Salari, A. Ansarian and H. Atrianfar, "Robust emotion classification using neural network models," *2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, 2018, pp. 190-194, doi: 10.1109/CFIS.2018.8336626.
- [14] Y. Cimtay and E. Ekmekcioglu, "Investigating the Use of Pretrained Convolutional Neural Network on Cross-Subject and Cross-Dataset EEG Emotion Recognition," *Sensors*, vol. 20, no. 7, p. 2034, Apr. 2020, doi: 10.3390/s20072034.
- [15] M. N. Dar, M. U. Akram, S. G. Khawaja and A. N. Pujari, "CNN and LSTM-Based Emotion Charting Using Physiological Signals," *Sensors*, vol. 20, no. 16, p. 4551, Aug. 2020, doi: 10.3390/s20164551.
- [16] N. Thammasan, K. Moriyama, K. Fukui and M. Numao, "Familiarity effects in EEG-based emotion recognition," *Brain Informatics*, vol. 4, no. 1, pp. 39-50, Jan. 2017, doi: 10.1007/s40708-016-0051-5.
- [17] V. Gupta, M. D. Chopda and R. B. Pachori, "Cross-Subject Emotion Recognition Using Flexible Analytic Wavelet Transform From EEG Signals," in *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2266-2274, 15 March 2019, doi: 10.1109/JSEN.2018.2883497.
- [18] T. Song, W. Zheng, P. Song and Z. Cui, "EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks," in *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532-541, 1 July-Sept. 2020, doi: 10.1109/TAFFC.2018.2817622.
- [19] Z. Cheng, L. Shu, J. Xie and C. L. P. Chen, "A novel ECG-based real-time detection method of negative emotions in wearable applications," *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, 2017, pp. 296-301, doi: 10.1109/SPAC.2017.8304293.
- [20] P. Das, A. Khasnobish and D. N. Tibarewala, "Emotion recognition employing ECG and GSR signals as markers of ANS," in *2016 Conference on Advances in Signal Processing (CASP)*, pp. 37-42, 2016, doi: 10.1109/CASP.2016.7746134.
- [21] P. Sarkar and A. Etemad, "Self-supervised ECG Representation Learning for Emotion Recognition," *IEEE Transactions on Affective Computing*, vol. 3045, no. c, pp. 1-13, 2020, doi: 10.1109/TAFFC.2020.3014842.
- [22] H. Ferdinando, T. Seppänen and E. Alasaarela, "Enhancing Emotion Recognition from ECG Signals using Supervised Dimensionality Reduction," in *ICPRAM*, 2017, pp. 112-118, doi: 10.5220/0006147801120118
- [23] Y. Hsu, J. Wang, W. Chiang and C. Hung, "Automatic ECG-Based Emotion Recognition in Music Listening," in *IEEE Transactions on Affective Computing*, vol. 11, no. 1, pp. 85-99, 1 Jan.-March 2020, doi: 10.1109/TAFFC.2017.2781732.
- [24] R. W. Picard, E. Vyzas and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175-1191, Oct. 2001, doi: 10.1109/34.954607.
- [25] J. Wagner, Jonghwa Kim and E. Andre, "From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification," *2005 IEEE International Conference on Multimedia and Expo*, 2005, pp. 940-943, doi: 10.1109/ICME.2005.1521579.

BIOGRAPHIES OF AUTHORS

Van-Dung Pham has completed his B.Eng degree in Electrical Engineering at Hanoi University of Science and Technology in 2020 and is currently pursuing his Masters in Electrical Engineering. His research interests are in signal processing, artificial intelligence and affective computing.



Thanh-Long Cung received his B.Eng degree in Measurement and Automatic Control, in 2000, and his M.Sc degree in Measurement and Control Systems, in 2002, at Hanoi University of Science and Technology, Vietnam. He received his PhD degree in Electronics - Electrotechnique - Automation, in 2012, at Ecole Normale Supérieure Paris - Saclay, France. He is currently a researcher/lecturer at the School of Electrical Engineering, Hanoi University of Science and Technology. His research interests include electromagnetic non-destructive testing/evaluation, human emotion recognition, sensors, and signal processing.