Emotion Recognition Based on DEAP Database using EEG Time-Frequency Features and Machine Learning Methods

T D Kusumaningrum¹, A Faqih¹ and B Kusumoputro¹

¹ Dept of Electrical Engineering Universitas Indonesia

Abstract. In recent years, research of the human emotional state is becoming importance, especially in its application for patient monitoring and in the treatment management system of that patient. In this paper, an EEG based emotion recognition system is developed that consists of a feature extraction subsystem and a classifier subsystem. As better performance of the feature extraction subsystem may produce higher recognition accuracy, nine features derived from the time and frequency domain from the EEG signal is used and analyzed. We have utilized support vector machine and Random Forest methods for classifying the emotional state of the subject, and compare its results with other machine learning methods. Using two-fold data validation model, the experiment result shows that the highest recognition accuracy is produced by using Random Forest method, i.e., 62.58%.

1. Introduction
Recently, research trend on recognizing the human emotion is mostly changed into the use of EEG signal over the conventionally used face image, voice signal or other physiological signals. As human may manage to falsify the outlook emotion through those conventional signals, it is difficult to falsify the emotion behind the EEG data taken. However, this signal is very complex and highly susceptible to noise, makes the data acquisition and its processing systems are more complicated.

The importance of recognizing the emotion of a patient, for example, is recently known as an important parameter in the illness treatment management system. It is well-known that maintaining the emotional response of a patient may increase the healthy progress and lowering the amount of medication [1]. Although more experiments are still necessary for studying the influence of the positive emotional state to the survival of patients with life threatening illness (e.g. cancer) [2], in general, however, positive emotions such as happiness causes better health and longevity [3].

The device for EEG research, such as the Emotive Epoch [4], which could measure the signal wirelessly with easy to set-up and inexpensive, has greatly increased the used of EEG signal for human emotion recognition. The emotion classification usually divided into two class of systems, i.e., discrete system and dimensional system. Based on discrete system, Plutchik [4] proposed eight basic emotion states based on discrete system: anger, fear, sadness, disgust, surprise, anticipation, acceptance, and joy. While for the dimensional system, the most widely used classification is the bipolar model - valence and arousal dimensions proposed by Russell [5].

As the EEG based emotion recognition research is still in the initial stage, looking for a prominent feature extraction of EEG signal and the classifier that could provide higher recognition accuracy on determining the emotion directly from EEG signal has been the main objectives of this research. Valence represents the quality of an emotion, ranging from unpleasant to pleasant, while Arousal denotes the quantitative activation level, from not aroused to aroused. Mehrabian then proposed an
additional dimension besides the arousal and valence dimensions, called dominance, which is also named as the control dimension of emotion that makes the dimensional models more complete [6], [7]. It is expected, therefore, the accuracy of the recognition rate of the emotion recognition system could be improved significantly. However, the recent recognition rate of using various feature extraction and the classifier could not be high as it is expected. Using the DEAP database [8] as the reference for comparison, recognition rate of two-dimensional Russel’s emotion from numerous researches is just about 55% to 60% in average when using two-folds method that could be increased to about 90% when using ten-folds [9]-[12]. Using their own database, deep learning using LSTM-RNN is utilized for classifying the emotion of the participants resulting about 59% accuracy [13]. Various kind of features extraction methods are also explored by numerous researchers, however, comparison of their results are difficult to analyzed, since the experiments are not using the same database. Using their own database, Zhu et.al used a differential entropy as the emotion feature extraction system and linear SVM as classifier shows that the recognition accuracy is 64.82% [14]. Other researchers have analyzed the utilization of ERD/ERS features and shows that the gamma band was the optimal frequency of happiness and sadness [15], while others prove that the good choice of features can improve the recognition rate result even though with back-propagation neural network [16]. In this paper, we would like to explore a different type of feature extraction method, i.e., time and frequency domain features of EEG signal. This paper is structured as follows. Section 2 explains the used DEAP database and its preprocessing of the EEG signal. Section 3 explains in detail the feature extraction system and the classifier method that have been used in this experiment, while results and its analysis are clearly presented in Section 4 and Section 5 concludes the paper.

2. Material and Method
Please follow these instructions as carefully as possible so all articles within a conference have the same style to the title page. This paragraph follows a section title so it should not be indented.

2.1. Experimental Data
We conduct the analysis of our system based on DEAP dataset, which publicly accessible, and the experiments are conducted according to block diagram shown in Fig. 1. The DEAP dataset consists of EEG recording of 32 subjects, where the data from each subject is taken through 32 EEG sensors and the other 8 sensor consists of EMG and EOG sensors, respectively. Brain signal of the subject is recorded when the subject is watching 40 types of videos, where each video induced different types of emotions. The subject rated their emotional experienced using two-dimensional emotion scale after the stimulus is activated.

Through out of the experiments, the 40 types of videos are utilized to induce 4 types of emotion on each subject, with video #1 to video #10 used for High Arousal High Valence, video #11 to video #20 for Low Arousal High Valence, video #21 to video #30 for High Arousal Low Valence, and video #31 to video #40 for High Arousal High Valence, respectively.

2.2. Preprocessing Data
To remove the 8 channels of EMG and EOG signal, preprocessing the EEG data before used are necessary to be conducted. We have also separated the brain signals according to their frequency range followed by normalizing within the entire range. The four types of brain signals that are extracted from the preprocessing system are alpha rhythm (8-15 Hz), beta rhythm (16-31 Hz), gamma rhythm (>32 Hz), and theta rhythm (4-7 Hz), respectively. As the human brain signals is affected by emotion state of the subjects, the importance of these rhythm types of the brain signal to be calculated separately is take into consideration.
However, we have excluded the delta rhythm in our developed system, because emotion recognition should be done when people are conscious, while the delta signals are produced when people are sleeping.

In the preprocessing stage, we also label the data by looking for the already defined class of each state based on the previously stated videos type. The emotion recognition accuracy is then evaluated using binary classification.

2.3. Experimental Procedure

In this paper we have used n-fold data validation in all of our experiments, because n-fold cross validation represents all the recognition rate of the data-fraction better than other splitting methods, and as the consequence, able to make the prediction of all data used in the experiments better. In this experiment, we try different kind of n-fold cross validation to consider the best recognition result. Using n-fold data validation procedure, we have firstly determining the parameter n which refers to the number of data sample that has to be divided into, and each fold will be used as a testing dataset, while remaining is used as the training dataset. When we determined to use a 5-fold cross validation procedure, for example, the DEAP database is then divided into 5 groups of data or folds. In the first iteration, the last fold is used as the testing dataset, while the first four fractions are used as the training dataset. In the second iteration, the fourth fold is used as the testing dataset, while the other fractions are used as the training dataset. This process is continued until all of the five folds have been used as the testing dataset.

3. Feature Extraction Process and The Classifier

The developed automatic emotion recognition system consists of a feature extraction subsystem and a neural network as a classifier subsystem. In the feature extraction subsystem, we have determined to use a time-frequency domain features, while for the classifier subsystem, we would like to analyze various machine learning methods such as SVM, Random Forest and compare its recognition rate with k-NN and weighted k-NN methods.

3.1. Time Frequency Domain Feature Extraction

In this work, after preprocessing the data, we calculated the features for each of the four rhythms with a 4-s sliding window and a 2-s overlap, and then, the mean of the feature values extracted from those sliding windows was adopted as the trial’s feature. By using this procedure, the number of features extracted for one trial is: \((9 \times 32 + 14) \times 4 = 1208\).
The extracted time features of the EEG signal that used in this experiments are: peak to peak mean, mean squared value, variance, respectively, while for frequency features are: Hjorth [17] parameters, i.e., complexity, mobility and activity, and 4 other characteristics of frequency domain signal such as, maximum power spectral frequency, power spectral density and power sum. The importance of these features is analyzed clearly in [12] which is adopted in our experiments. Firstly, the arithmetic mean of the vertical length from the very top to the very bottom of the time series is calculated, followed by the arithmetic mean of the squares of the time series and determined as the peak to peak mean and mean squared value, respectively. The next feature is variance, which measures the degree of dispersion of the time series. The EEG time series is then converted into the frequency domain using Fourier transform, and we calculated the sum of the power spectral, and extracted further the maximum power spectral density along with its corresponding frequency value and the three Hjorth parameters as have been explained earlier.

3.2. Support Vector Machine (SVM)

A support vector machine is supervised machine learning algorithm that usually used for classification and regression model. The SVM technique offered better classification results compare with that of artificial neural networks because it has wide range of generalization which prevents the model from over fitting. Using a proper kernel function, SVM is a powerful machine learning in handling a non-linear data for solving the regression and classification problem. The goal of SVM is to build the optimal hyper-plane that maximizes the margin between classes. In mathematical expression, this hyper plane can be written by minimizing the cost function:

$$J = \sum_{i=1}^{N} a_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j y_i y_j k(x_i, x_j)$$

(1)

There are two kinds of kernel type that usually used in SVM technique, i.e., linear and non-linear kernels, such as RBF or Gaussian kernel. In this experiment, however, we have used linear kernel because this kernel has less prone to over fitting compare with that of the non-linear kernel. As we concerned with the best accuracy of the classifier, we have not limit the iteration of the learning mechanism of the SVM.

3.3. Random Forest

Random Forest is a type of supervised classifier that uses an ensemble learning technique through multiple decision trees in its training stages. The number of the trees determined the accuracy of the classifier, with the more trees in the system increases the system accuracy. Decision tree is a rule-based system with a supervised training dataset together with the its target, and during training, those trees creates the rules for each of certain classes. These rules are then utilized as the parameter to the new testing data that may differs from the training dataset.

In the beginning of the algorithm, the random forest selected some features from all of the features given for each class, while in the testing stage, testing dataset is passed through all the tree that already defined during learning. By considering each prediction target, the targets votes are then calculated. The Random Forest algorithm has advantage characteristic that never been producing over fitting classification results. As the Random Forest always finds the least error of the problem-model, this technique also could be used as a regression task as also for classification and feature extraction tasks.

In this experiment, we have used the Random Forest for classifying 2 classes of the dataset, where 100 trees are utilized as the parameter, which is approximately 10% of the total data features.

3.4. k-NN and Weighed k-NN (Wk-NN)
Nearest Neighbor (NN) is the easy and common used of classifier for pattern recognition problems. k-NN classifier could be classified as a non-parametric type classification, where the new variable is placed in the training dataset, and the new variable’s class is then decided using the nearest distance calculation to the existing learning dataset. The k shows how many nearest neighbors are used as data comparison to new variable data.

The number of k in k-NN classifier has shown high contribution to the classification result, however, defining the optimum value of k is very difficult to determine. It is then weighted k-NN is developed to minimize the disadvantage of k-NN. The concept of the weighted k-NN algorithm is constructed by introducing the neighbor weight, that varied exponentially based on square Euclidean distance of the neighbor. In this experiment we have used k = 7, after numerous trying the k number from zero to fifteen, where this number of k has mostly given the best classification result.

4. Experimental Results And Analysis

In this experiment, we used the computer with the specification of 7th generation Intel Core i5 with NVIDIA GPU 930MX. The overall experiments are performed by using four n-fold cross validation technique, i.e., 2-fold, 3-fold, 5-fold and 10-fold, respectively. The accuracy of the classification results is calculated by averaging the accuracy from each fold of data validation using true positive, true negative, false positive and false negative calculation through:

\[ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \] (2)

As in this research we have only concern with the arousal dimension of the Russel bipolar system [5], i.e., High Arousal and Low Arousal, respectively, the accuracy confusion matrix could be calculated and depicted in Fig. 2. The recognition accuracies of the used classifiers are depicted in Fig. 3, that consists of the average accuracy of SVM, Random Forest, k-NN and Weighted k-NN, respectively. As shown in this figure, the recognition accuracy from each n-fold for SVM classifier are almost the same, while these phenomena are also happened with that of Random Forest classifier. For these classifiers, the highest recognition accuracy is achieved, when using 5-fold cross validation technique, while Random Forest shows just about 2.5% higher. This might be happened because the random forest classifier has characteristic that could not be in over-fitting condition, while the SVM classifier has to relate their maximum performance to the critical parameter adjustment that need to set in advance. When we further analyzed these comparison results with the BPNN method, Random Forest and SVM methods have performed with lower computational costs.

![Confusion matrix](image-url)
Figure 3. Comparison of the recognition accuracy using SVM and Random Forest classifier, k-NN and Weighted k-NN.

Compare with that of Nearest Neighbor classifiers, however, the recognition accuracy difference is considerably high with about 7% to 9%. This higher recognition accuracy for SVM and Random Forest has shown that the classifier using learning mechanism is performed better. When we compare those recognition accuracies with that of BPNN classifier from our previous experiment results, it can be concluded that classifier performs the learning mechanism during its process has higher recognition accuracy compare with the family of Nearest Neighbor classifiers.

More detail analysis of the SVM and Random Forest classifiers can be seen from their confusion matrixes that are depicted in Figure 4. The total number of the confusion matrix for each of the used classifiers shows the number of the n fold cross validation that we have used in this research. The upper part of this figure shows the confusion matrix for the SVM classifier for 2-fold, 3-fold, 5-fold and 10-fold, respectively, while the lower part shows the same confusion matrixes for Random Forest classifier. As can be clearly seen from this figure, the percentage of the actual recognition accuracy is just about 31% to 32% for High Arousal and 29% to 30% for Low Arousal, respectively. These values are considered as low recognition accuracy, showing that those classifiers are not appropriate for determining the emotional state of the human.

Figure 4. Confusion matrix of SVM and Random Forest classifier.
5. Conclusions
In this work, we have presented the classification of human emotional state using EEG signal through its time–frequency domain feature extraction system. Various classifiers using machine learning methods such as Random Forest, SVM, k-NN and Weighted k-NN are utilized, and compared its recognition accuracy through four n-fold cross validation method. Results show that recognition accuracy of the classifier is not so high, that is about 60% in average, as the same with that already published by other researchers that using the same DEAP database. We have found that the problem is not lies in the classifier nor the feature extraction subsystem, but may lied in how we could make the normalized extracted value from the feature extraction subsystem is always the same for the same emotional state for different human subjects.

6. References
[1] E. Mumford, H. J. Schlesinger, G. V. Glass, “The Effects of Psychological Intervention on Recovery from Surgery and Heart Attacks: An Analysis of the Literature,” American Journal of Public Health, 1982.
[2] J. C. Coyne, H. Tennen, A. V. Ranchor, “Positive Psychology in Cancer Care: A Story Line Resistant to Evidence,” Annals of Behavioral Medicine, vol. 39 no. 1, pp 35-42, 2010.
[3] E. Diener, M. Y. Chan, “Happy People Live Longer: Subjective WellBeing Contributes to Health and Longevity,” Applied Psychology: Health and Well Being, vol. 3 no. 1, pp 1-43, 2011.
[4] R. Plutchik, Emotions and life : perspectives from psychology, biology, and evolution, 1st ed. Washington, DC: American Psychological Association, 2003.
[5] Kuhn T 1998 Density matrix theory of coherent ultrafast dynamics Theory of Transport Properties of Semiconductor Nanostructures (Electronic Materials vol 4) ed E Schöll (London: Chapman and Hall) chapter 6 pp 173–214
[6] A. Mehrabian, “Framework for a comprehensive description and measurement of emotional states,” Genetic, social, and general psychology monographs, vol. 121, pp. 339-361, 1995.
[7] A. Mehrabian, "Pleasure-Arousal-Dominance: A general framework for describing and measuring individual differences in temperament,” Current Psychology, vol. 14, pp. 261-292, 1996.
[8] Koelstra, S. C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras,”DEAP: A Database for Emotion Analysis using Physiological Signals,” IEEE Trans. Affective Computing. pp.1-15, 2012.
[9] B. Krisandhika, A. Faqih, P. D. Purnamasari, B. Kusumoputro, “Emotion Recognition System Based on EEG Signals using Relative Wavelet Energy Features and A Modified Radial Basis Function Neural Network,” International Conference on Consumer Electronics and Devices. p.50-54, 2017.
[10] P. D. Purnamasari, A. A. P. Ratna, B. Kusumoputro, "EEG Based Patient Emotion Monitoring using Relative Wavelet Energy Feature and Back Propagation Neural Network. IEEE ISN: 978-1-4244-9270-1. p.280-2823, 2015.
[11] P. D. Purnamasari, A. A. P. Ratna, B. Kusumoputro, “Development of filtered bispectrum for EEG signal feature extraction in automatic emotion recognition using artificial neural networks,” Algorithm vol. 10 no. 2, 63, 2017.
[12] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou and B. Hu, “Exploring EEG Features in Cross-Subject Emotion Recognition,” Front. Neurosci. 12.162 doi: 10.3389/fnins.2018.00162, 2018
[13] Reddy K. et.all, “EEG-based emotion recognition using LSTM-RNN machine learning algorithm,” 1st International Conference on Innovations in Information and Communication Technology, 2019.
[14] J. Zhu, S. Rosset, R. Tibshirani, and T. J. Hastie, “1-norm support vector machines,” Proc. 16th Intern. Conf. on Neural Information Processing Systems (Cambridge: MIT Press), pp. 49–56, 2004.
[15] M. Li, and B.L. Lu, “Emotion classification based on gamma-band EEG,” Proc. 31st Ann. Intern. Conf. of the IEEE Engineering in Medicine and Biology Society (New York, NY: IEEE Press), pp. 1223–1226, 2009.

[16] I. Belakhdar et al. 2016. A comparison Between ANN and SVM Classifier for Drownside Detection Based on Single EEG Channel. IEEE

[17] B. Hjorth. 1970. EEG analysis based on time domain properties. Electroencephalogr. Clin. Neurophysiol. 29, 306–310