Feature extraction comparison for facial expression recognition using adaptive extreme learning machine

Muhammad Wafi[1], Fitra A. Bachtiar[1], Fitri Utaminingrum[2]
[1]Intelligent System Laboratory, Faculty of Computer Science, Brawijaya University, Malang, Indonesia
[2]Computer Vision Research Group, Faculty of Computer Science, Brawijaya University, Malang, Indonesia

Article Info

ABSTRACT

Facial expression recognition is an important part in the field of affective computing. Automatic analysis of human facial expression is a challenging problem with many applications. Most of the existing automated systems for facial expression analysis attempt to recognize a few prototypes emotional expressions such as anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. This paper aims to compare feature extraction methods that are used to detect human facial expression. The study compares the gray level co-occurrence matrix, local binary pattern, and facial landmark (FL) with two types of facial expression datasets, namely Japanese female facial expression (JFFE), and extended Cohn-Kanade (CK+). In addition, we also propose an enhancement of extreme learning machine (ELM) method that can adaptively select best number of hidden neurons adaptive ELM (aELM) to reach its maximum performance. The result from this paper is our proposed method can slightly improve the performance of basic ELM method using some feature extractions mentioned before. Our proposed method can obtain maximum mean accuracy score of 88.07% on CK+ dataset, and 83.12% on JFFE dataset with FL feature extraction.

Corresponding Author:

Fitra A. Bachtiar
Faculty of Computer Science, Brawijaya University
8 Veteran Road, Malang, East Java, Indonesia
Email: fitra.bachtiar@ub.ac.id

1. INTRODUCTION

Humans in their daily lives have an instinct to be able to interact with one another in order to achieve a goal. In carrying out these interactions the interpretation of one's emotional state becomes very important in creating good communication. The emotional state of a person in interacting can be reflected in words, gestures and facial expressions [1]. Among those reflections, the face is a complex and difficult element to understand, even in an idle position the face can provide information about several emotional states or moods [2]. Facial expressions recognition is an important part of studying how humans react to the environment in the field of affective computing which aims to diagnose and measure someone's emotional expression explicitly and connect implicitly [3]. Facial expression recognition has been widely applied in several fields, including facial expression recognition for the detection of system errors in smart environments [4], enhancing the gaming experience [5], intelligent tutoring system [6], cooking experience [7], and so on. Those examples show that facial expressions play an important role in decision making, help the learning process, and also provide an assessment based on an action.

Feature extraction is one of the important processes in recognizing facial expressions. Several feature extraction models on faces have been developed and have weaknesses and strengths in each model [8]. For example, template-based models that have a good level of accuracy. However, the template-based

models the computational process has more complex computation and the images must have the same size, orientation, and level of illumination. Similar to template-based models, color segmentation-based models, and appearance-based models, the level of illumination, color and image quality also play an important role to get maximum results. Geometry-based models have a good level of accuracy and the amount of data needed is less than the three models mentioned before [8], and the level of illumination does not really affect this model. Therefore, selecting the suitable feature extraction is important prior to implementing the data to the classification model.

There are several algorithms that can be used to recognize facial expressions which range from machine learning to neural networks models. Machine learning methods such as k-nearest neighbor (KNN) or support vector machine (SVM) have been widely used to solve this problem. However, previous study shows that neural network algorithm could achieved reliable accuracy even not the highest among hidden Markov model, AdaBoost, and SVM [9]. In other several studies found that the level of accuracy possessed by these algorithms is classified as lower when compared to the neural network (NN) model [10]–[12]. NN is an emulation of the human biological nervous system consisting of several artificial neurons that are connected to each other in a group and can process information through a computational process [13]. NN itself has been widely used in solving several problems related to classification, some of which are the recurrent neural network (RNN) algorithm which is used as a spam detection with affect intensities [14]. Convolutional neural network (CNN) which is used as a food image classification [15], [16] and deep belief network (DBN) for recognition of emotions on human faces [17]. Lastly, extreme learning machine (ELM) which is used in real-world classification problems, including breast cancer, authentication on banknotes, [18], detecting eye fixation point [19], and so on. Those previous study show that the NN architectural model is able to perform various classification processes with good results. One of the NN algorithms that has a relatively simple model and only has one hidden layer consisting of several neurons or what is often called Single layer feedforward neural networks (SLFNNs) is the ELM algorithm, where the weight between input and hidden neurons set randomly and are constant during the learning and prediction phases [20]. Therefore, ELM's simple architecture is considered capable of performing the classification process with a faster computation time when compared to traditional gradient-based algorithms such as backpropagation [21].

In this study, we propose an enhancement for the ELM method to classify facial expressions such as anger, contempt, disgust, fear, happy neural, sadness and surprised. We also compare texture-based and geometry-based feature extraction to find the best feature extraction method in terms of facial expression detection. The feature extraction method that we used to compare are gray level co-occurrence matrix (GLCM), local binary pattern (LBP) and facial landmark (FL).

2. FEATURE EXTRACTION COMPARISON USING AELM
2.1. Datasets
The datasets used to compare feature extraction method in this paper are extended Cohn-Kanade (CK+) [22] and Japanese female facial expression (JAFFE) [23]. The facial expression on CK+ datasets as shown in Figure 1, consists of 654 images and divided into several number of basic human expressions, such as anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise. Meanwhile, JAFFE dataset consists of 213 images of facial expressions and has some basic human expressions like anger, disgust, fear, happiness, neutral, sadness, and surprise as shown in Figure 2.
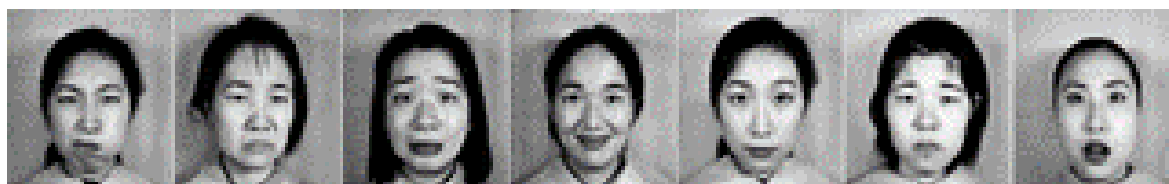

Figure 1. Extended Cohn-Kanade (CK+) dataset

Figure 2. Japanese female facial expression (JAFFE) dataset

2.2. Feature extraction

In this paper, we use three types of feature extraction to compare the extraction performance. The three feature extractions are GLCM, LBP and FL. Each of the extraction method represents three different extractions. The GLCM is used as a texture extraction method, the LBP is a visual descriptor extraction method, and the FL is a landmark that represents facial area.

2.2.1. Grey level co-occurrence matrix

Grey level co-occurrence matrix (GLCM) is an image texture analysis technique. GLCM represents the relationship between 2 neighboring pixels which has a gray intensity (grayscale intensity), distance and angle. There are 8 angles that can be used on GLCM, including corners 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°. Meanwhile, the distance usually used are 1, 2, and 3. The steps are used to calculate the gray level co-occurrence matrix features of the grayscale image are described as follows: i) creating a matrix work area from the image input, ii) doing the formation of the initial GLCM matrix of two-pixel pairs which are aligned according to the direction of 0°, 45°, 90°, or 135°, iii) symmetrizing matrix by summing the initial GLCM matrix with the transpose value, iv) normalizing the GLCM matrix by dividing each matrix element by the number of pixel pairs, and v) extracting feature of GLCM.

The GLCM feature is calculated using a square matrix based on region of interest dimension in the facial expression images. Total six numbers of features that extracted in this paper, those six features are contrast, angular second-moment feature (ASM), energy, dissimilarity, homogeneity, and correlation features. Refer to [24] and [25] for further details.

2.2.2. Local binary pattern (LBP)

LBP encodes the grayscale of an image by comparing selected pixel value to its neighbor. In the LBP approach for texture classification, the occurrences of the LBP codes in an image are collected into a histogram. LBP encodes the grayscale of an image by comparing selected pixel value to its neighbor. In the LBP approach for texture classification, the occurrences of the LBP codes in an image are collected into a histogram. The classification is then performed by computing simple histogram similarities. However, considering a similar approach for facial image representation results in a loss of spatial information and therefore one should codify the texture information while also retaining their locations. One way to achieve this goal is to use the LBP texture descriptors to build several local descriptions of the face and combine them into a global description. Such local descriptions have been gaining interest lately, which is understandable given the limitations of the holistic representations. These local feature-based methods are more robust against variations in pose or illumination than holistic methods. For further details please see the basic methodology for LBP based face description proposed by Ahonen et al. [26].

2.2.3. Facial landmark (FL)

The first step of the facial landmarks (FL) is to localize the face area or namely the region of interest (ROI) of the input image. It is important to mention that we can use different methods for face detection like HaaR cascades or Viola Jones, but Dlib library itself has its own detector to select the ROI and get the coordinates of a face bounding box coordinates of a key facial points. Specifically, we used Dlib facial landmark which was implemented by Kazemi and Sullivan [27]. Facial landmarks are defined as the detection and localization of certain points characteristics on the face. Commonly used landmarks are the eye corners, the nose tip, the nostril corners, the mouth corners, the end points of the eyebrow arcs, ear lobes, and chin [28]. Landmarks such as eye corners or nose tip are known to be little affected by facial expressions, hence they are more reliable and are in fact referred to as fiducial points or fiducial landmarks in the face processing literature.

2.3. Proposed adaptive extreme learning machine (ELM)

Single layer feedforward neural networks (SLFNs) are extreme learning machines (ELM) which was first introduced by Huang et al. [29] with the aim of filling one of the weaknesses in other feedforward neural

network algorithms, which has slow learning speed [29]. The slow learning process performance can occur because the used of parameters, namely the weight and bias of the feedforward network, need to be adjusted iteratively. In addition, the use of the gradient-based learning method used in the learning process has an impact on the process being slow, or very vulnerable to reaching local minimums [21]. In general, ELM algorithm can be done in two stages, namely the training process where ELM will learn the input data by doing some calculation, and then it will continue to the testing stage to get the results. At the training process ELM start to the output weights values, where this value is the result of the training process that will be used in the testing process later. It started by initializing random matrix values of input weights (W_{jk}) and bias values (b) and calculating the initial matrix H' then add the result of calculation to one of the activation functions. After that, the result of activation function is calculated using Moore-Penrose generalized inverse, then we can start to proceed to the testing process [29]. In the ELM process it is difficult to determine the exact number of hidden neurons because the input weights and bias value has been initialized randomly as well as the number of hidden neurons, the random number of hidden neurons makes quite inefficient for the performance of ELM itself. ELM has to generate a great number of hidden neurons to achieve great performance too.

In this paper we modify basic ELM method to find the best number of hidden neurons automatically by defining the minimum hidden neurons Nmin and maximum hidden neurons N_{max} that can be incrementally updated based on the interval values $N_{interval}$, the idea was adapted and modified from self-adaptive extreme learning machine [30], by removing some iteration step that can takes more time to finish the process. The steps of the proposed ELM method can be represented in algorithm 1. By selecting the best hidden neurons through its accuracy result on ELM method, technically it will increase the performance of the method, we called it as an adaptive extreme learning machine (aELM). The adaptive state of hidden neurons is controlled by its minimum and maximum number of hidden neurons and will incrementally adjusted based on interval values until reach the maximum state.

Algorithm 1. Adaptive ELM algorithm
Result: Best hidden neuron number and its result accuracy
Begin
   Step 1: Initialize the minimum (N_{min}), maximum (N_{max}), and interval ($N_{interval}$) of hidden neurons
   Step 2: while hidden neuron is not equal to N_{max} do
     Perform the basic ELM method with number of hidden neurons = N_{min},
     Evaluate the prediction results,
     Update $N_{min} = N_{min} + N_{interval}$
   Step 3: Select best hidden neuron number based on its accuracy.
   Step 4: Output the best hidden neuron number and its result accuracy
End

3. RESULTS AND DISCUSSION

In this section, the results and discussion are described. We compared the three features extraction which are GLCM, LBP, and FL. We used variation distance and angle on GLCM feature extraction, while on LBP we analyze about number of point and radius. Lastly, for the FL we use all the 68 points extracted from the images. We did not conduct the experiments of FL due to FL having no parameters that would affect the recognition result. In this research we used our proposed ELM method to classify basic emotions like anger, contempt, disgust, fear, happiness, neutral, sadness, and surprise on the datasets mentioned before and also compare the recognition rate with the basic ELM methods. Before we continue to the results of the comparison for the two feature extraction methods (LBP and GLCM), first, we evaluate the best parameters used by each feature extraction methods. Next, the evaluation steps are conducted to measure the performance of the basic ELM method compared to our proposed method.

3.1. Experiments on GLCM parameters

GLCM parameters which are distance and angle parameters using several variations of pair parameters as: *search distance (d)={1, 3, 5, 8, 10}, angles (θ)={0°, 45°, 90°, 135°}, GLCM properties = {Contrast, ASM, Energy, Homogeneity, Correlation, Dissimilarity}*. To evaluate the best accuracy and select the best pairs of GLCM parameters we used the basic ELM method with number of hidden neurons of 12 neurons and the activation function used sigmoid activation function. The results of the best GLCM parameters can be seen in Figure 3(a). The best GLCM score is about 51.16% where distance $d = \{1, 5\}$, and angles $\theta = \{0°; 45°; 90°\}$ on JAFFE dataset. Meanwhile, on CK+ dataset the best parameters can be seen in Figure 3(b). The GLCM achieved 54.19% with $d = 1$ and angles $\theta = 90°$.
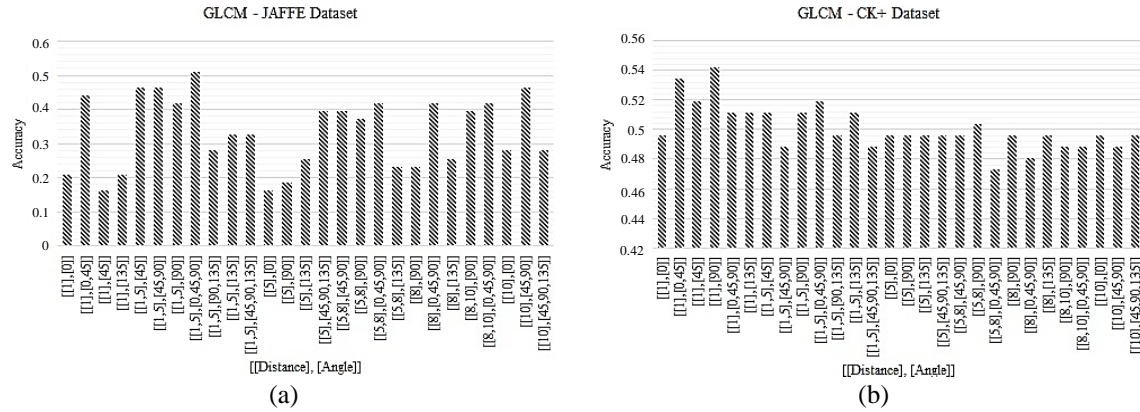
Figure 3. Experiments on GLCM feature extraction parameters for (a) JAFFE dataset and (b) CK+ dataset

3.2. Experiments on LBP parameters

Evaluation of the LBP parameters is needed to find the best parameters before we tested in our proposed method. LBP has two parameters that can affect the classification results which are number of points and radius. The values of the number of points and radius used in this paper are as: number of *points={8, 12, 16, 24, 32}, radius={1, 2, 4, 6, 8}*. The result for the best LBP parameters shows that the best LBP number of points is 12 and for the radius is 8 with the accuracy score of those combinations of the LBP parameters is about 41.86% using JAFFE dataset as shown in Figure 4(a). Meanwhile, on CK+ dataset LBP is able to achieve highest score at 51.9% with some pair combination of points and radius as seen in Figure 4(b). In this step, we choose the 12 number of points and 8 radius of LBP parameters to use in the next experiments.
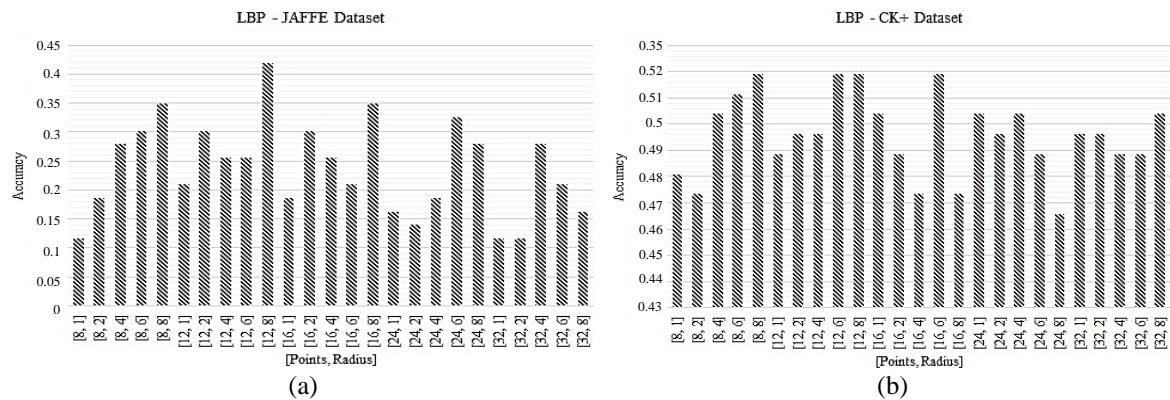

Figure 4. Experiments on LBP feature extraction parameters using (a) JAFFE dataset and (b) CK+ dataset

3.3. Experiments on basic ELM method

In this paper, to compare the ELM method with our proposed ELM method, we use the best parameters of the feature extractions mentioned before, which are GLCM, LBP and FL. For the first experiments, we test both of our methods parameters such as number of hidden neurons and activation function to obtain maximum performance through the K-Fold cross validation. After that we obtained the evaluation metric which consists of F1 score, precision and recall score for the basic ELM methods itself. After we found the best hidden neurons and activation function by using the best parameters of GLCM, LBP and FL, then the best mean F1 score is used to compare the recognition results. Lastly, the best fold accuracy is used to show the confusion matrix of the final results. The basic ELM parameters setup to test the performance for each feature extractions mentioned before is as: *number of hidden neurons={2, 4, 8, 12, 16, 20, 24, 32, 64, 120}, activation function={Sigmoid, Cosine, Hyperbolic Tangent, Linear}*. Figure 5(a) shows that basic ELM method is tested using FL, LBP and GLCM feature extractions with its best parameters which tested before. The results of FL feature extraction are reaching the maximum mean accuracy about 68.07% with number of hidden neurons 32 when using the Sigmoid activation function, while the GLCM feature extractions

maximum mean accuracy score at 51.69% with 64 hidden neurons. Finally, LBP highest mean accuracy score is 31.36% with 64 hidden neurons using JAFFE dataset. On the other hand, using CK+ dataset all feature of FL method can reached recognition rate of 86.23% with 120 hidden neurons, LBP achieved 51.37% recognition rate with 20 hidden neurons and GLCM is reached recognition rate of 52.3% with 16 hidden neurons as shown in Figure 5(b).
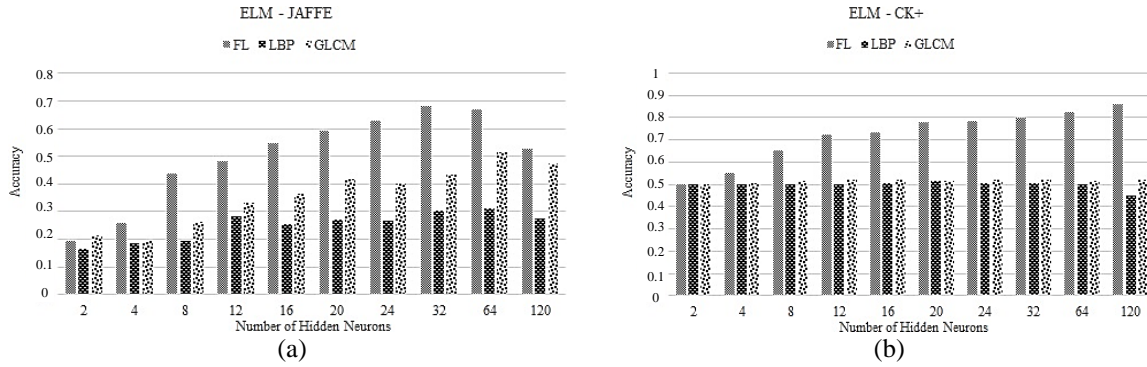

Figure 5. Experiments on ELM number of hidden neurons using (a) JAFFE dataset and (b) CK+ dataset

After getting the best hidden neurons for both datasets, activation function is tested on basic ELM method by using number of hidden neurons obtained in previous step. Figure 6(a) shows that the best activation function for FL feature extraction is hyperbolic tangent and its maximum mean accuracy score is about 73.22%, and for LBP feature extraction the best activation function is cosine for 33.74% mean accuracy score, and best activation function for GLCM is on hyperbolic tangent for 54.54% mean accuracy score on JAFFE dataset. Meanwhile, on CK+ dataset the best activation function for FL feature extraction is hyperbolic tangent with 86.84% mean accuracy score, on LBP best mean accuracy score is when using hyperbolic tangent with 51.98%, and GLCM best activation function is on Sigmoid with 52.3% mean accuracy score as seen in Figure 6(b).
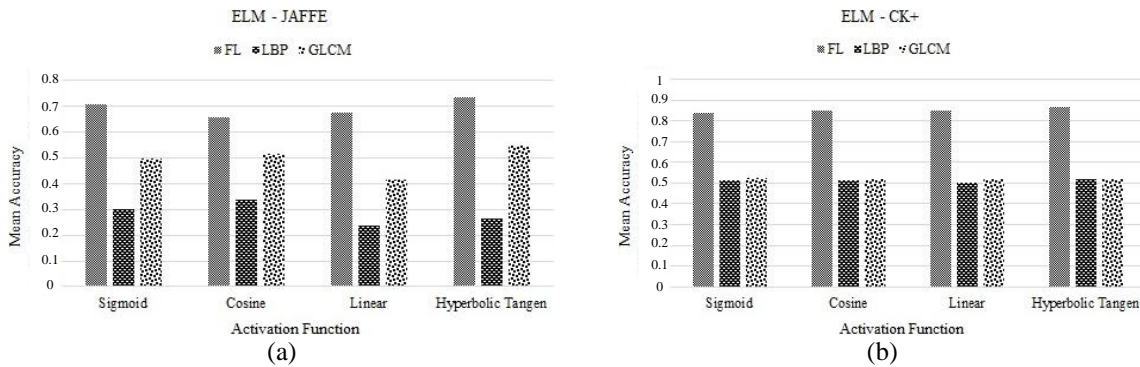

Figure 6. Experiments on ELM activation function using (a) JAFFE dataset and (b) CK+ dataset

3.4. Experiments on adaptive ELM method

Before we compare aELM and basic ELM, experiments should be carried out to find the best parameters on the aELM such as, the combination of N_{min}, N_{max} and $N_{interval}$. The main purpose of doing some experiments on aELM parameters is to find the best combination of its parameters like described in Table 1. If the values of $N_{interval}$ is too large the chance of finding the best performance may decrease. On the other hand, if the value is too small, the time computation may be increased. aELM parameters are tested by using best parameters of FL, LBP, and GLCM similar to experiments performed in the basic ELM. Figure 7(a) shows that the best pair hidden neurons parameters for our proposed method with default Sigmoid activation function using FL feature extraction is N_{min}=2, N_{max}=100 and $N_{interval}$=2 for mean accuracy score 82.7%, for GLCM and LBP the aELM parameters are N_{min}=3, N_{max}=150 and $N_{interval}$=3

with mean accuracy score 63.85% for GLCM and 51.16% for LBP on JAFFE Dataset, respectively. Meanwhile, using CK+ dataset FL can reach about 88.98% recognition rate with N_{min}=10, N_{max}=300 and $N_{interval}$=10. As shown in Figure 7(b), LBP and GLCM feature extraction able to reach recognition rate of 54.59% with N_{min}=2, N_{max}=100 and $N_{interval}$=2, and recognition rate of 53.37% with N_{min}=3, N_{max}=150 and $N_{interval}$=3, respectively.

The activation function for our proposed method is also tested because the variety of hidden neurons will affect the performance of our proposed method in both datasets. Figure 8(a) shows that the best activation function for our proposed method on JAFFE dataset is cosine for all of the feature extraction. The mean accuracy for FL, LBP, and GLCM feature extraction are 83.59%, 45.99%, and 66.23%, respectively. For CK+ dataset the best activation function is sigmoid for all of the feature extraction. The mean accuracy for FL, LBP, and GLCM are 89.59%, 54.74%, and 53.36%, respectively. These results can be seen in Figure 8(b).

Table 1. Adaptive ELM parameters

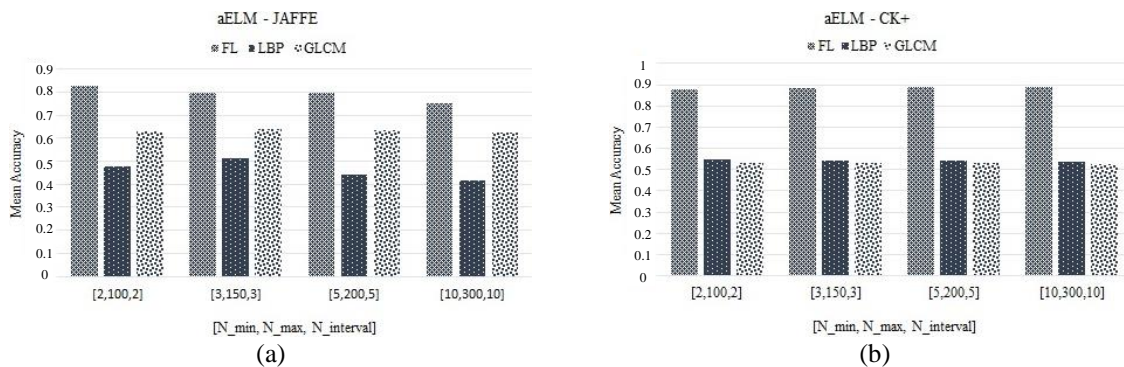| N_{min}	N_{max}	$N_{interval}$	Parameters
2	100	2	Sigmoid, Cosine, Hyperbolic Tangent, Linear
3	150	3	Sigmoid, Cosine, Hyperbolic Tangent, Linear
5	200	5	Sigmoid, Cosine, Hyperbolic Tangent, Linear
10	300	10	Sigmoid, Cosine, Hyperbolic Tangent, Linear

Figure 7. Experiments on aELM number of hidden neurons using (a) JAFFE dataset and (b) CK+ dataset
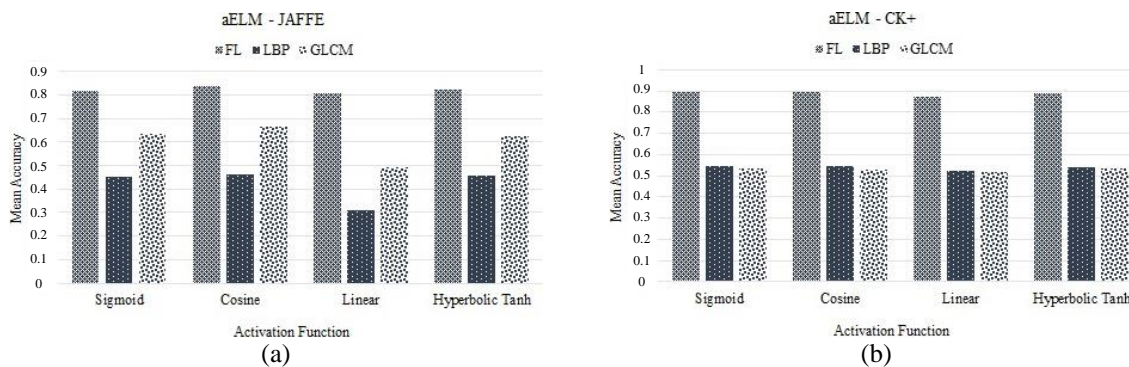

Figure 8. Experiments on aELM activation function using (a) JAFFE dataset and (b) CK+ dataset

3.5. Comparison results

After getting all the best parameters for feature extraction, basic ELM, and aELM method, the final step is to run the results for every method's configuration using its best parameters from the experiments performed in previous step. The confusion matrix result is used to compare the recognition rate of both models. Final result described on Figure 9(a), where the FL extraction feature shows the best performance compared to the other extraction features LBP and GLCM. Facial landmark feature extraction can achieve a mean accuracy score of 71.75% on the JAFFE dataset using the basic ELM method. Meanwhile, using our proposed aELM method FL can achieve its best on 83.12%. Using LBP feature extraction, basic ELM can only get the mean accuracy score of 30.93%, while using our proposed aELM can slightly improve the

performance of LBP about 48.33%. As for GLCM feature extraction, our proposed method also outperforms the basic ELM. GLCM can achieve its best accuracy score in 69.01% compared with basic ELM method that can only be able to get 54.5% recognition rate. For CK+ dataset, FL feature extraction is way better than the other feature extraction too. Using FL feature extraction, it was able to achieve accuracy score of 88.07% while using our proposed method and can slightly increase all of feature extraction results as shown in Figure 9(b).

(a)   (b)
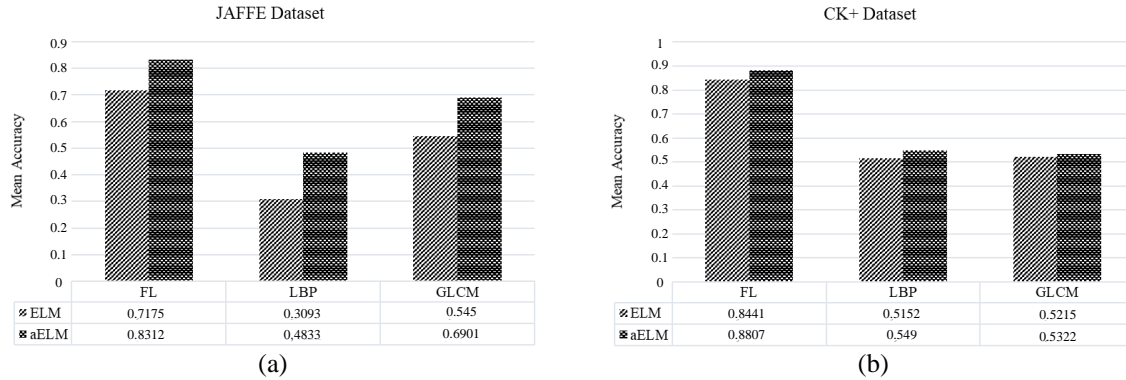
Figure 9. Comparison results for FL, LBP and GLCM with basic ELM and aELM using both (a) JAFFE dataset and (b) CK+ dataset

In this paper, we only show the confusion matrix of the best feature extraction and method used for both JAFFE and CK+ dataset. Confusion matrix result is obtained from the best K-Fold accuracy score of our experiments. Confusion matrix is shown in Figure 10(a) for JAFFE dataset and Figure 10(b) for CK+ dataset. There are some misclassified prediction labels by aELM method which is neutral, contempt, happy, and sadness. Misclassification class of data can happen due to unequal distribution between its majority and minority classes, which is called class imbalance. In CK+ dataset the distribution between class neutral, anger, contempt, disgust, fear, happy, sadness, and surprised is imbalanced. For example, in CK+ the majority class is neutral, where there are 327 images from a total of 654 images which is 50% of total images. Meanwhile, the minority class from CK+ is contempt, which only have 18 images total. This imbalance class problem in classifying object can dramatically skew the performance of classifiers [31]. As a result, our proposed aELM method cannot recognize some labels expression from CK+ datasets. Besides that, the confusion matrix result from JAFFE dataset with FL feature extraction can be seen on Figure 10(a), which is shown that aELM method sometimes fail to recognize happy, and surprised expression, we assume this problem is caused by the number of datasets is too small which can affect the learning process of classifier. Furthermore, the JAFFE dataset is also imbalanced, although the distribution between class is not extreme compared to CK+ dataset.

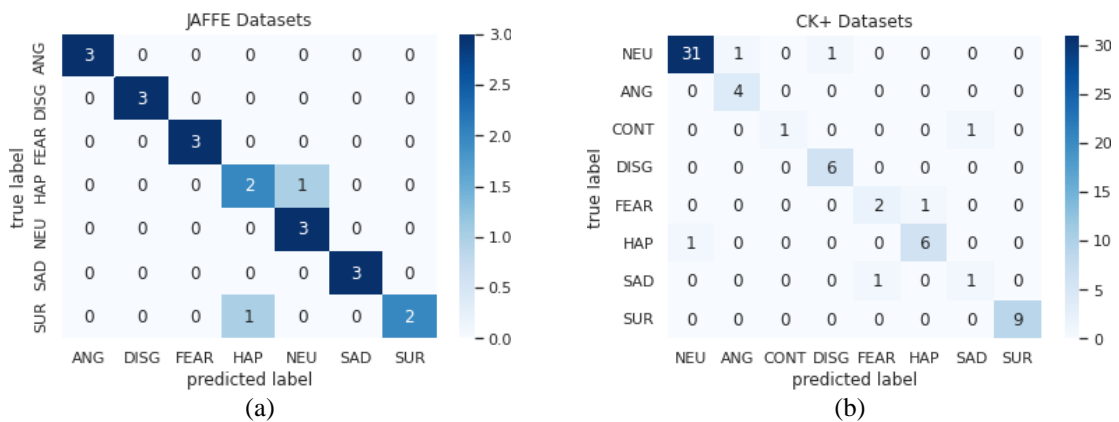(a)   (b)

Figure 10. Confusion matrix results comparison with aELM method with FL feature extraction using (a) JAFFE dataset and (b) CK+ dataset

Based on the experiment results, our method can perform better because the number of hidden neurons is automatically adjusted based on N_{min}, N_{max} and $N_{interval}$ values, after all hidden neurons are being tested, then aELM will select its best accuracy score from the results and set as the output of aELM. Even the number of hidden neurons is adaptively selected, the aELM parameter itself is used to find best pairs of N_{min}, N_{max} and $N_{interval}$ to determine the correct initial value so that it is expected that it does not affect the computation time that takes too long. Meanwhile, in terms of feature extraction, FL feature extraction is way better than GLCM and LBP in case of recognize expression from human face using ELM method in this paper.

4.   CONCLUSION

In this paper, the comparison for different types of feature extraction methods in terms of recognizing human facial expression has been done using neural network algorithm, namely ELM. Feature extraction comparison has been done using FL, GLCM, and LBP, from the results discussed before shows that FL has outperformed the other feature extractions. Apart from comparing feature extraction, we also propose an aELM method. Our proposed method can slightly improve the performance of the basic ELM method itself. The result shows that aELM highest mean accuracy score is 88.07% on CK+ dataset, and 83.12% on JAFFE dataset using FL feature extraction. Although aELM can improve the performance of basic ELM in general, there are some problems to solve in future, which is from the input weight that still randomly generated that also can affect the classification performance. Further research may consider to used Nguyen-Widrow method to minimize the randomness of the generated weight.

REFERENCES

[1]   J. M. Garcia-Garcia, V. M. R. Penichet, M. D. Lozano, J. E. Garrido, and E. L.-C. Law, "Multimodal affective computing to enhance the user experience of educational software applications," *Mobile Information Systems*, vol. 2018, pp. 1–10, Sep. 2018, doi: 10.1155/2018/8751426.

[2]   J. F. Cohn and P. Ekman, "Measuring facial action," in *The New Handbook of Methods in Nonverbal Behavior Research*, Oxford., J. A. Harrigan and R. K. R. S. Rosenthal, Eds. New York, 2008, pp. 9–64. doi: 10.1093/acprof:oso/9780198529620.003.0002.

[3]   H. C. K. Lin, C. H. Wu, and Y. P. Hsueh, "The influence of using affective tutoring system in accounting remedial instruction on learning performance and usability," *Computers in Human Behavior*, vol. 41, pp. 514–522, 2014, doi: 10.1016/j.chb.2014.09.052.

[4]   Y. Yaddaden, M. Adda, A. Bouzouane, S. Gaboury, and B. Bouchard, "User action and facial expression recognition for error detection system in an ambient assisted environment," *Expert Systems with Applications*, vol. 112, pp. 173–189, Dec. 2018, doi: 10.1016/j.eswa.2018.06.033.

[5]   M. T. Akbar, M. N. Ilmi, I. V Rumayar, J. Moniaga, T.-K. Chen, and A. Chowanda, "Enhancing game experience with facial expression recognition as dynamic balancing," *Procedia Computer Science*, vol. 157, pp. 388–395, 2019, doi: 10.1016/j.procs.2019.08.230.

[6]   R. Z. Cabada, M. L. B. Estrada, F. G. Hernández, R. O. Bustillos, and C. A. Reyes-García, "An affective and web 3.0-based learning environment for a programming language," *Telematics and Informatics*, vol. 35, no. 3, pp. 611–628, 2018, doi: 10.1016/j.tele.2017.03.005.

[7]   A.-M. Brouwer, M. A. Hogervorst, J. B. F. van Erp, M. Grootjen, E. van Dam, and E. H. Zandstra, "Measuring cooking experience implicitly and explicitly: Physiology, facial expression and subjective ratings," *Food Quality and Preference*, vol. 78, Dec. 2019, doi: 10.1016/j.foodqual.2019.103726.

[8]   E. Bagherian and R. W. O. K. Rahmat, "Facial feature extraction for face recognition: a review," in *International Symposium on Information Technology*, 2008, vol. 2, pp. 1–9. doi: 10.1109/ITSIM.2008.4631649.

[9]   P. Kulkarni and R. T. M., "Analysis on techniques used to recognize and identifying the Human emotions," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 3, pp. 3307–3314, Jun. 2020, doi: 10.11591/ijece.v10i3.pp3307-3314.

[10]   Y. Kumar and S. Sharma, "A systematic survey of facial expression recognition techniques," *Proceedings of the International Conference on Computing Methodologies and Communication*, pp. 1074–1079, 2018, doi: 10.1109/ICCMC.2017.8282636.

[11]   Y. Chen, H. Wu, T. Wang, and Y. Wang, "A comparison of methods of facial expression recognition," *WRC Symposium on Advanced Robotics and Automation*, pp. 96–101, 2018, doi: 10.1109/WRC-SARA.2018.8584202.

[12]   Y. Huang, F. Chen, S. Lv, and X. Wang, "Facial expression recognition: a survey," *Symmetry*, vol. 11, no. 10, Sep. 2019, doi: 10.3390/sym11101189.

[13]   A. Sinkov, G. Asyaev, A. Mursalimov, and K. Nikolskaya, "Neural networks in data mining," in *2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, 2016, vol. 18, no. 6, pp. 1–5. doi: 10.1109/ICIEAM.2016.7911596.

[14]   N. A. Farahisya and F. A. Bachtiar, "Spam email detection with affect intensities using recurrent neural network algorithm," in *2nd International Conference on Information Technology and Education*, Jan. 2022, pp. 206–211. doi: 10.1109/ICITE54466.2022.9759865.

[15]   G. Ciocca, P. Napoletano, and R. Schettini, "CNN-based features for retrieval and classification of food images," *Computer Vision and Image Understanding*, vol. 176–177, pp. 70–77, 2018, doi: 10.1016/j.cviu.2018.09.001.

[16]   R. P. Prasetya and F. A. Bachtiar, "Indonesian food items labeling for tourism information using convolution neural network," in *International Conference on Sustainable Information Engineering and Technology (SIET)*, Nov. 2017, vol. 2018-Janua, pp. 327–331. doi: 10.1109/SIET.2017.8304158.

[17]   M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Information Fusion*, vol. 51, pp. 10–18, Nov. 2019, doi: 10.1016/j.inffus.2018.10.009.

[18]   V. Christou, M. G. Tsipouras, N. Giannakeas, A. T. Tzallas, and G. Brown, "Hybrid extreme learning machine approach for heterogeneous neural networks," *Neurocomputing*, vol. 361, pp. 137–150, Oct. 2019, doi: 10.1016/j.neucom.2019.04.092.

[19] F. A. Bachtiar, G. Pangestu, F. Pradana, I. Arwani, and D. Syauqy, "Eyeball movement detection using sector division approach and extreme learning machine," in *International Seminar on Intelligent Technology and Its Application: Intelligent Systems for the New Normal Era*, Jul. 2021, pp. 184–190. doi: 10.1109/ISITIA52817.2021.9502211.

[20] N.-D. Hoang and D. Tien Bui, "Slope stability evaluation using radial basis function neural network, least squares support vector machines, and extreme learning machine," in *Handbook of Neural Computation*, 1st ed., Elsevier, 2017, pp. 333–344. doi: 10.1016/B978-0-12-811318-9.00018-1.

[21] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006, doi: 10.1016/j.neucom.2005.12.126.

[22] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, Jun. 2010, pp. 94–101. doi: 10.1109/CVPRW.2010.5543262.

[23] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 200–205. doi: 10.1109/AFGR.1998.670949.

[24] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[25] P. K. Mall, P. K. Singh, and D. Yadav, "GLCM based feature extraction and medical X-RAY image classification using machine learning techniques," in *IEEE Conference on Information and Communication Technology*, Dec. 2019, pp. 1–6. doi: 10.1109/CICT48419.2019.9066263.

[26] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006, doi: 10.1109/TPAMI.2006.244.

[27] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874, 2014, doi: 10.1109/CVPR.2014.241.

[28] H. Ouanan, M. Ouanan, and B. Aksasse, "Facial landmark localization: Past, present and future," in *4th IEEE International Colloquium on Information Science and Technology (CiSt)*, Oct. 2016, vol. 0, pp. 487–493. doi: 10.1109/CIST.2016.7805097.

[29] G-B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *IEEE International Joint Conference on Neural Networks*, 2004, vol. 2, pp. 985–990. doi: 10.1109/IJCNN.2004.1380068.

[30] G.-G. Wang, M. Lu, Y.-Q. Dong, and X.-J. Zhao, "Self-adaptive extreme learning machine," *Neural Computing and Applications*, vol. 27, no. 2, pp. 291–303, Feb. 2016, doi: 10.1007/s00521-015-1874-3.

[31] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *Journal of Big Data*, vol. 5, no. 1, Dec. 2018, doi: 10.1186/s40537-018-0151-6.

BIOGRAPHIES OF AUTHORS

Muhammad Wafi 🔵 SC ○ is a student of Brawijaya University, 2018-2021. He is pursuing master's degree in Faculty of Computer Science, Brawijaya University, Indonesia, and his research interests are affective computing, intelligent system, and computer vision. He completed his bachelor's degree in 2017 from Faculty of Computer Science, Brawijaya University. He can be contacted at mwafiez@gmail.com.

Fitra A. Bachtiar 🔵 SC ○ is a lecturer in the Faculty of Computer Science at Brawijaya University. He is a graduate of Electrical Engineering, Brawijaya University, 2008, and Ritsumeikan University Graduate School of Science and Engineering, 2011. He received the Dr. Eng from Ritsumeikan University in 2016. His research interests are affective computing, affective engineering, intelligent system, and data mining. He can be contacted at fitra.bachtiar@ub.ac.id.

Fitri Utaminingrum 🔵 SC ○ was born in Surabaya, East Java, Indonesia. She is an associate professor in the Faculty of Computer Science, Brawijaya University, Indonesia. Currently, her focus research is about the smart wheelchair, especially on the development of Image algorithms. She is also a coordinator in computer vision research groups and a full-time lecturer at Brawijaya University, Indonesia. She has published her work in several reputable journals and conferences indexed by Scopus. She obtained a Doctor of Engineering in the field of Computer Science and Electrical Engineering from Kumamoto University, Japan. She can be contacted at f3_ningrum@ub.ac.id.