ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(25), DOI: 10.17485/ijst/2015/v8i25/80978, October 2015* ISSN (Online) : 0974-5645

Facial Expression Recognition using Dual Stage MLP with Subset Pre-Training

Kang-Won Lee, Tai-Hyung Kim, Sung-Hoon Kim, Seung-Hyung Lee and Hyon-Soo Lee*

Depatment of Computer Engineering, Kyung Hee University, Korea; kw_lee@khu.ac.kr, kimth04@khu.ac.kr, shoonkim@khu.ac.kr, shlee7@khu.ac.kr, leehs@khu.ac.kr

Abstract

This paper suggests a Dual Stage MLP, a new ensemble structure for facial expression recognition, and Subset Pre-Training which supports the structure as an effective learning method. Subset Pre-Training is the asymmetrical pre-training method in which 1) some parts of training data are trained to MLPs in the first stage, 2) then the whole training data are utilized to train MLP in the second stage. Facial expression recognition was carried out based on Extended Cohn Kanade database to demonstrate the effectiveness of the proposing method, and the Local Binary Pattern was set as the feature for an experiment. The result of the experiment confirmed the bigger asymmetry guarantees the higher classification rates. Compared with the entire time needed in learning the whole training data, the proposing method reduced the time by one eighth while showing the improved classification accuracy rate 99.25%.

Keywords: Dual Stage MLP, Ensemble Structure, Facial Expression, Pre Training

1. Introduction

In recent years, the recognition technology has been widely utilized in many sectors ranging from smart phone, smart car and smart home system. These recognition technologies are transforming various devices into more human-friendly and intelligent devices. Facial expression recognition is the core technology that enables emotional communication between humans and smart devices like computers, and many experiments have been actively carried out in various sectors including commercial system, entertainment, intelligent robot to name a few[1–3].

The changes in human emotions can be detected from local image based features mostly around eyebrows, eyes, and mouth. Accordingly, it is generally accepted that local image based approaches show more stable and improved results than holistic image based approaches which make use of the entire facial features[4–8].

This research adopted the ensemble learning to classify the facial expressions. The ensemble structure seeks Divide-and-Conquer through the task decomposition. In the first stage, MLP (Multi-Layer Perceptron) is generated for ensemble structure on individual landmark offered by CK+ (Extended Cohn Kanade) DB[6] then LBP (Local Binary Pattern)[9] features of each landmark is utilized for MLP's training. Also, in the second stage, another MLP is generated to learn new features set by the results of first stage's MLP for the execution of the final facial expression recognition. In the CK+, the total of 68 partial landmark information from face is offered based on the AAM (Active Appearance Model)[10]. This paper adopted both the 36 major landmarks and LBP features from the previous research[7] for training and testing of the proposing structure.

2. Previous Work

2.1 Selection of Major Landmark

In this work, 36 major landmarks selected from previous research[7] are employed. Of the total number of 68

** Author for correspondence*

landmarks offered by the existing CK+ data, 51 landmarks are selected by isolating 17 less influential appearance feature-based landmarks on outline of the face. Among 51 landmarks, the superior 36 landmarks with less dispersion rates are selected for every 7 expression classes. In the previous research, MLPs were created on 36 chosen landmarks, and then the LBP feature of each landmark was utilized to train MLP. The ensemble learning method, which carries out voting by applying weighting values on the result of 36 MLPs, was employed for the research. The selected major landmarks are illustrated in the following Figure 1.

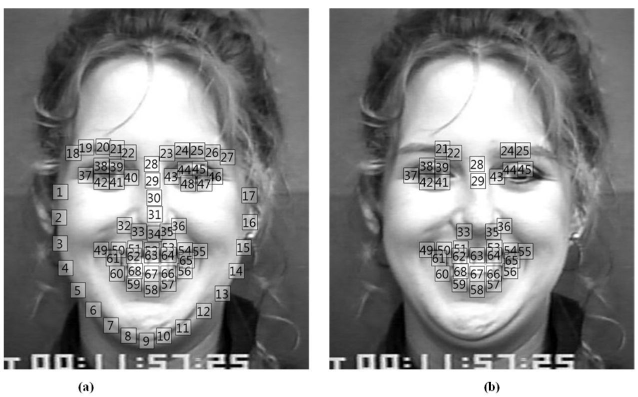

Figure 1. Major landmark selection. (a) Original 68 landmarks. (b) Selected 36 major landmarks.

2.2 Ensemble Learning Method

Bagging and Boosting are the representative methods used in drawing results from the ensemble structure[11–14]. Bagging is the multiple voting method of weak classifiers used to select the most frequently voted class. Boosting is the voting method which applies weighting values on misclassified data, then weighting is carried out depending on the credibility of the weak classifier[13,14].

Facial expression recognition method shows the issues of differences in landmarks which show diversified features depending on facial classes as it can be shown in the Figure 1. Since Bagging is unlikely to solve the weighting issues for each classifier, it is recognized as inappropriate. Sensitivity to the outlier is regarded as a disadvantage of boosting, and this approach is far from ideal when it comes to cases like applying weighting values both on inter-classifiers and inter-classes as it can be confirmed by the facial expression recognition issues.

3. Subset Pre-Training on Dual Stage MLP

3.1 System Overview

In this section, one solution is devised to overcome limitations of existing Bagging and Boosting methods

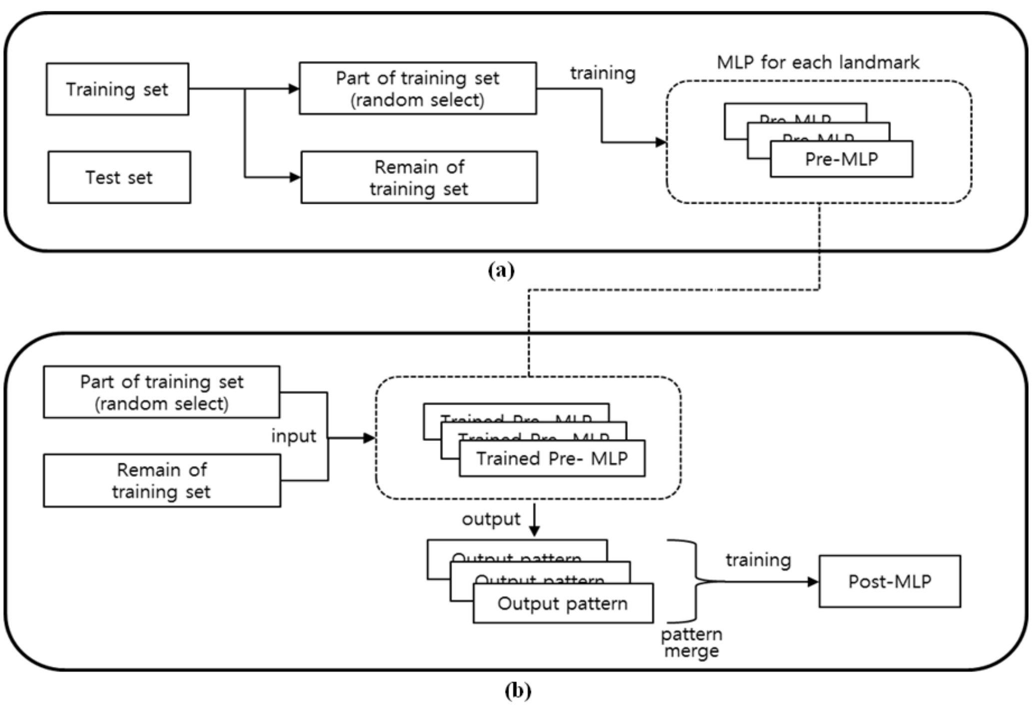

Figure 2. Structure of Dual Stage MLP and Subset Pre-Training. (a) First stage. (b) Second stage.

in the facial expression recognition. It creates new Post-MLP learning results patterns of Pre-MLPs for each landmark. In doing so, weighting issues spanning from inter-weak classifiers to inter-classes can be subjugated. In the step (a), subsets from entire training data are randomly selected to train Pre-MLPs for each landmark. New patterns are created by merging outcomes resulted from putting entire training data with trained Pre-MLPs for each landmark. In the following step (b), Post-MLP is trained with using patterns created in the previous step.

3.2 Training of First Stage MLP

In the Figure 2, Pre-MLPs for each landmark of step (a) sets the uLBP (uniform Local Binary Pattern)[9,15] features for each landmark as input values in which major 58 patterns out of 59 are utilized to acquire 58 input neurons. From the experimental result, 30 hidden neurons are selected, and the number of output neuron is identical to the number of CK+ expression classes as 7. The totals of 36 Pre-MLPs for each landmark are generated as the numbers of designated landmarks are the same. 90% out of the entire dataset is randomly selected and set as the training data. Among training data, 5 data from each class or the total of 35 data are utilized to train Pre-MLPs for individual landmark.

3.3 Training of Second Stage MLP

In the step (b) illustrated in the Figure 2, the entire training data is entered into the trained Pre-MLPs. Then, the Post-MLP is trained with derived patterns generated

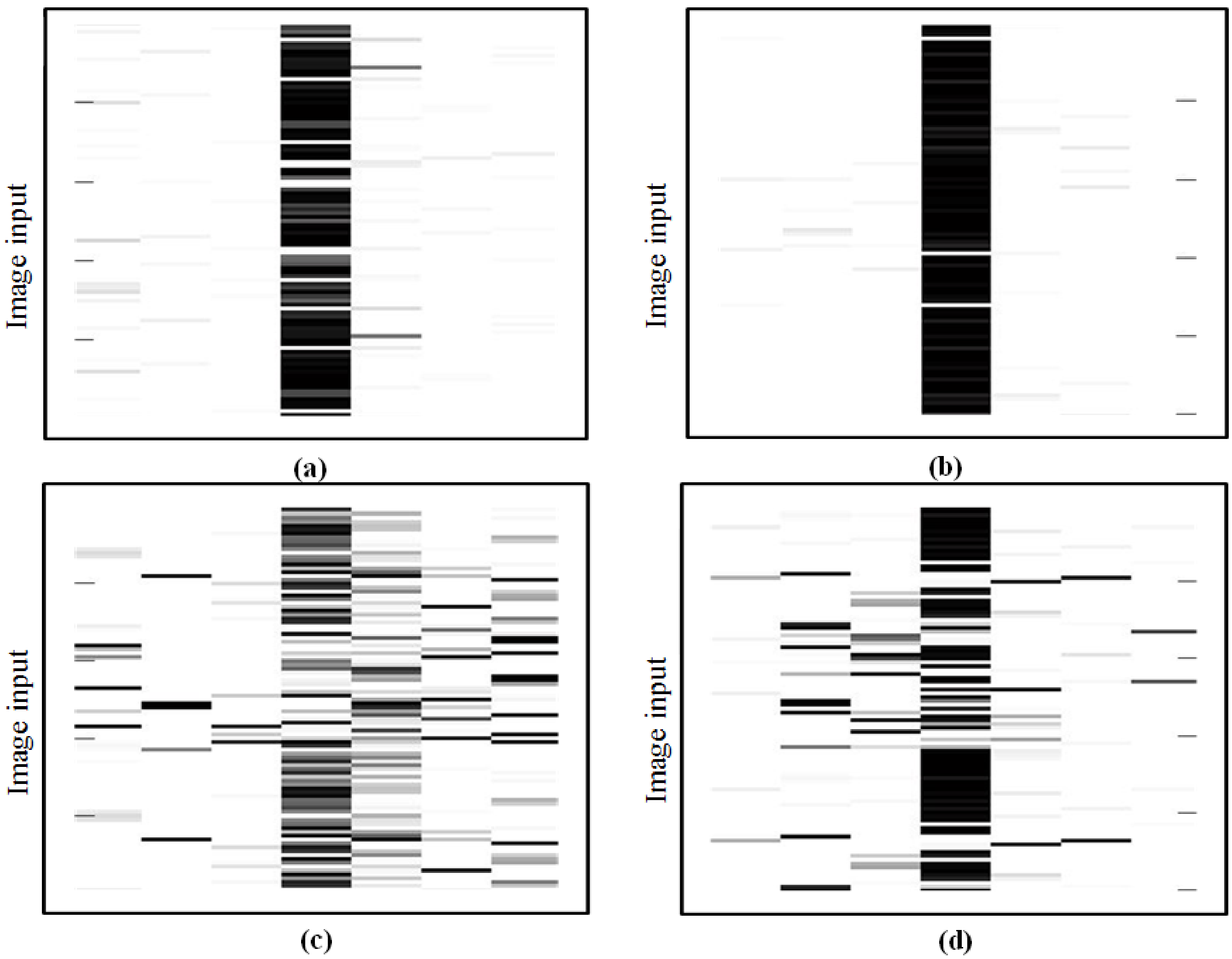

Figure 3. Result pattern of Pre-MLP with and without Subset Pre-Training. (a) Class result of landmark 21. (b) Class result of landmark 68. (c) Class result of landmark 21. (d) Class result of landmark 68.

from merging output results between Pre-MLPs. The input neuron of Post-MLP has the number of 252 input neurons and it can be easily computed by multiplying the number of output neurons of Pre-MLP to the number of Pre-MLPs. The number of output neuron is identical to the number of expression classes as 7. As Pre-MLPs were trained with only a few data in step (a), the outcome has certain dispersion rates to some degree. In the following step (b), the outcome with dispersion rates are re-trained through the supervised learning. In spite of the dispersed results in the first stage, step (b) featured speedy learning pace as it generates clustered results through Pre-MLPs in the first stage.

3.4 Subset Pre-Training

As the common MLP with dual stage form prints thoroughly aligned result in the first stage MLP, the MLP in the second stage is trained with aligned data. This gives rise to overfitting symptoms in which training error is low while the test error is high.

Both (a) and (b) in the Figure 3 shows the output result of Pre-MLP on the landmark 21 and 68, which all are trained by the entire training data without going through the Subset Pre-Training. It can be confirmed that output results for every data are identical. Subset Pre-Training was applied to (c) and (d) of Figure 3, and it is the output result of Pre-MLP on landmark 21 and 68 which went through the pre-training with 5 data from each facial expression from the entire training data. The landmark 21 shows high dispersion rates while the landmark 68 shows relatively regular results. As it can be shown above, the data exempted from pre-training can have either high dispersion rates or vice versa depending on each landmark. By putting those dispersed results into the supervised learning again in Post-MLP training, it can recognize the dispersed results generated from the actual testing environment as the class it belongs to. Such asymmetrical training method contributed to preventing the overfitting, and it led to the improved classification performances in the actual testing. Also, the amount of data needed for the training of Pre-MLPs in the first stage can be shortened by one over forty five, which in turn helps to reduce the training time of proposed structure.

4. Experiment Result

4.1 CK+ Database Explanation

The CK+ data, dataset for facial expression recognition, was adopted to verify the validity of both the Dual Stage MLP structure and Subset Pre-Training method proposed in this paper. CK+ is a database widely used in the field of facial expression recognition and it is comprised of 7 emotion classes expressing anger, contempt, disgust, fear, happiness, sadness, and surprise. A total of 327 sequences are providing emotion labels. Although the end frame of each sequence has the emotion label, the last 3 frames or the total of 981 data were collected to acquire more image samples and to guarantee the fair comparison of research conditions with previous studies. The composition of each class is as follows Table 1.

Table 1. The composition of each facial expression data for experiment

Anger	Con	Dis	Fear	Happy	Sad	Sup	Total
135	54	177	75	207	84	249	981

CK+ data composition per class is unbalanced as it can be shown in the Table 1. Except for 10% of data from each expression for test, the classifier was trained by repeatedly selecting training data until the number of it reaches 225 which is 90% equal to the number of the most prevalent data, surprise, for easiness of training[16,17]. Configuring the training set and test set are shown in the Table 2 below.

Table 2. The composition of training and test set by repeatedly selecting

	Anger	Con	Dis	Fear	Happy	Sad	Sup	Total
Original data set	135	54	177	75	207	84	249	981
Training set	122	49	160	68	187	76	225	887
Training set by repetition	225	225	225	225	225	225	225	1575
Test set	13	5	17	7	20	8	24	94

4.2 The Classification Result using Subset Pre-Training

Proposing Dual Stage MLP structure differentiated the number of training data each for Pre-MLP and Post-

MLP, and then repeated test processing with test set for 10 times. The average value of the experiment results was illustrated in the Table 3 below.

Table 3. The recognition result of asymmetric training (Subset Pre-Training)

Average Recognition Result of Asymmetric Training (%)							
	Pre-MLP 5	Pre-MLP 40	Pre-MLP 80	Pre MLP 120	Pre-MLP 160	Pre-MLP 200	Pre-MLP 225
Post-MLP 5	62.91	71.26	79.16	83.42	87.42	92.13	92.27
Post-MLP 40	78.15	72.41	79.88	81.59	88.16	91.07	93.32
Post-MLP 80	86.23	84.72	80.33	82.10	88.22	92.33	94.71
Post-MLP 120	91.60	90.45	87.32	83.67	88.09	92.65	94.82
Post-MLP 160	93.16	93.20	91.07	88.28	87.85	91.99	93.92
Post-MLP 200	97.47	96.04	95.99	94.45	95.43	92.23	94.53
Post-MLP 225	97.96	96.33	96.18	95.76	95.92	94.80	94.86

As it can be shown from the experiment results, the outcome of the facial expression recognition was improved when less data is trained in Pre-MLP while more data is trained in Post-MLP. Thus, applying Subset Pre-Training on the Dual Stage MLP structure can improve performance by reduce the overfitting.

4.3 Computational Complexity of Subset Pre-Training

Subset Pre-Training can help to reduce the amount of time need to training Pre-MLP per landmark. In general, considering the both cases of training Pre-MLPs in the first stage with 225 data from each expression and the other case of training them with 5 data from each expression through Subset Pre-Training, the training time can be shortened by one over forty five.

The Post-MLP in the second stage shows fast learning pace as it is trained with already clustered results of the Pre-MLPs in the first stage, so it hardly affects the entire training time. The experiment was carried out through matlab with the 4core 3.2GHZ CPU computing environment. The entire training time under Subset Pre-Training application, error convergence graph per epoch of Pre-MLP in the first stage, and error convergence graph per epoch in the second stage Post-MLP are illustrated in the following Table 4 and Figure 4.

4.4 Performance and Comparison

Subset pre-training shows deviations in the results depending on the subset of training data which selected for first stage though it offers higher learning speed and more accurate classification results. Subset Pre-Training repeatedly selected for five times inside the training data to cope with the aforementioned issue. Pre-MLP for individual landmark in the first stage goes through learning with five data per expression for five times, while the Post-MLP in the second stage learns via merging

Table 4. Training time under Subset Pre-Training

Training Time of Proposed Method				
Number of training dat per each class (ea)	Training time(sec)	Number of training data per each class (ea)	Training time(sec)	Entire training time(sec)
Pre-MLP 5	566.88	Post-MLP 5	23.19	590.07
Pre-MLP 40	4447.02	Post-MLP 40	23.26	4470.28
Pre-MLP 80	9052.25	Post-MLP 80	23.33	9075.58
Pre-MLP 120	13306.63	Post-MLP 120	22.84	13329.47
Pre-MLP 160	18010.73	Post-MLP 160	23.01	18033.74
Pre-MLP 200	22692.27	Post-MLP 200	23.41	22715.68
Pre-MLP 225	25249.15	Post-MLP 225	23.06	25272.21

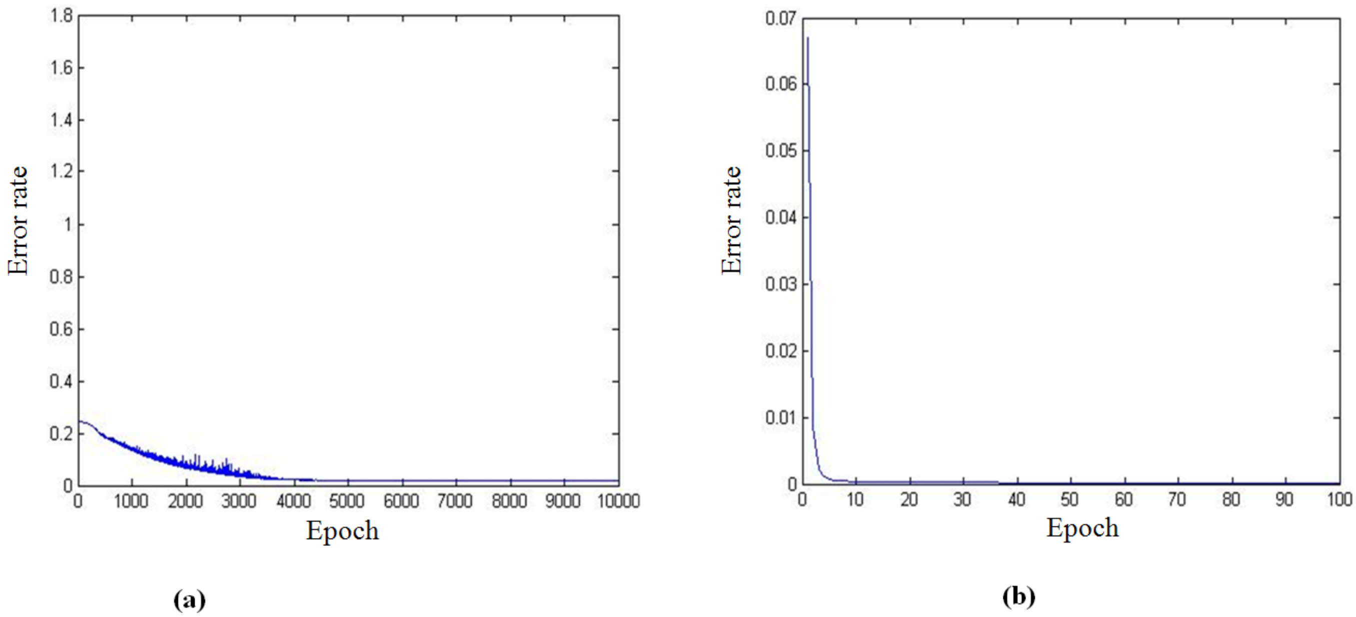

Figure 4. Error convergence graph. (a) Error per epoch of Pre-MLP. (b) Error per epoch of Post-MLP

output results from the first stage. Both the training data and test data are randomly selected, and the average result from 30 times repeated facial expression classification testing, and the entire training time can be shown in the Table 5 below.

Table 5. Confusion matrix of 7-class expression recognition using 5 time Subset Pre-Training on Dual Stage MLP and entire training time

Confusion Matrix of 7-class Expression Recognition (%)							
	Anger	**Con**	**Dis**	**Fear**	**Happy**	**Sad**	**Sup**
Anger	97.7	0.0	1.5	0.0	0.0	0.0	0.8
Con	0.0	100	0.0	0.0	0.0	0.0	0.0
Dis	0.6	0.0	99.4	0.0	0.0	0.0	0.0
Fear	0.0	1.4	0.0	98.6	0.0	0.0	0.0
Happy	0.0	0.0	0.0	0.5	99.5	0.0	0.0
Sad	0.0	0.0	0.0	0.0	0.0	100	0.0
Sup	0.0	0.0	0.0	0.4	0.0	0.0	99.6
Entire Training Time and Average of 7-class Facial Expression Recognition rate							
Entire training time (sec)		3149.35		Average of Facial Recognition Rate (%)			99.25

Also, the comparative analysis was made between classification results of the proposing method and the results from latest research papers[7,8,18]. Every result was derived from CK+ data, and the experiment results from using last 3 frames of each sequence were compared to guarantee the fair experiment conditions. The experiment result from those state-of-the-art researches was used as data for comparison for this research.

As it can be shown in the Table 6, this research was able to attain the highest classification rates from the facial expression recognition experiment. It was able to attain the higher rates 99.25% of facial expression recognition with less shortened training time 3250 sec by employing proposing method while the most current research of others based on BDBN (Boosted Deep Belief Network)[8] took 8 days for training.

5. Conclusions and Future Works

In this paper, Dual Stage MLP, a new classifier structure, and an effective learning method for the structure called Subset Pre-Training are presented. The CK+ DB based

Table 6. Performance comparison

Methods	LBP-SVM[18]	LBP-weighted MLP[7]	BDBN[8]	Proposed Dual Stage MLP Without Subset Pre-Training	Proposed Dual Stage MLP With Subset Pre-Training
Performance (%)	95.1	95.6	96.7	94.86	99.25

on facial expression recognition experiment was utilized to guarantee the validity. The result of the research indicates that the proposing method shows 1) the highest classification rate of 99.25% which is higher than other recent research outcomes, and 2) 8 times faster learning pace compared with the existing method of learning entire training data.

The proposing method can help to remove the possible weighting issues arise from local image based approaches of appearances by employing ensemble method. Accordingly, the research outcome can be possibly applied not only to previous approach based on landmark, but to block image based facial expression recognition method. For the future work, proposing method will be applied to diverse facial expression recognition databases such as Cohn Kanade Action Unit code, and JAFFE (Japanese Female Facial Expression)[19] DB. Also, applying of diverse features on learning will be continued for the generalization of the proposing method.

6. References

1. Fasel B, Luettin J. Automatic facial expression analysis: A survey. Pattern Recog. 2003 Jan; 36(1): 259–75.
2. Pantic M, Rothkrantz LJM. Auto analysis of facial expression: The state of art. IEEE Trans Pattern Analysis and Machine Intelligence. 2000 Dec; 22(12): 1424–45.
3. Bettadapura V. Face expression recognition and analysis: The state of the art tech report; 2012 Apr.
4. Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition. IEEE Trans Pattern Anal Mach Intell. 2006; 28(12): 2037–41.
5. Gundimada S, Asari VK. Facial recognition using multisensor images based on localized kernel eigen spaces. IEEE Trans Image Process. 2009; 18(6): 1314–25.
6. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression. Computer Vision and Pattern Recognition Workshops (CVPRW); 2010. p. 94–101.
7. Kim TH, Kim SH, Lee SH, Lee HS. Facial expressions recognition using LBP feature based on major landmarks. Workshop on Image Processing and Image Understanding; Korea. 2014.
8. Ping L, et al. Facial expression recognition via a boosted deep belief network. Conference on IEEE Computer Vision and Pattern Recognition (CVPR); 2014. p. 1805–12.
9. Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition. 1996 Jan; 29(1): 51–9.
10. Cootes T, Edwards G, Taylor C. Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001; 23(6): 681–5.
11. Leo B. Bagging predictors. Machine Learning. 1996; 24(2): 123–40.
12. Leo B. Arcing classifier (with discussion and a rejoinder by the author). The Annals of Statistics. 1998; 26(3):801–49.
13. Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine Learning. 2000; 40(2): 139–58.
14. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer And System Sciences. 1997; 55(1): 119–39.
15. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002 Jul; 24(7): 971–87.
16. He H, Garcia EA. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering. 2009; 21(9): 1263–84.
17. Alejo R, et al. Improving the classification accuracy of RBF and MLP neural networks trained with imbalanced samples. Intelligent Data Engineering and Automated Learning–IDEAL. Berlin Heidelberg: Springer; 2006. p. 464–71.
18. Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing. 2009; 27(6):803–16.
19. Lyons MJ, Julien B, Shigeru A. Automatic classification of single facial images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1999; 21(12): 1357–62.