Sensors and Materials, Vol. 36, No. 11 (2024) 4927–4938
MYU Tokyo

S & M 3840

A Feature-fusion-based Convolutional Neuro-fuzzy Classifier for Facial Emotion Recognition

Cheng-Jian Lin* and Xue-Qian Lin

Department of Computer Science & Information Engineering, National Chin-Yi University of Technology,
Taichung 411, Taiwan

(Received April 30, 2024; accepted October 10, 2024)

Keywords: emotion recognition, convolutional neural network, feature fusion, fuzzy system

In the application of face recognition, emotion recognition has gradually received attention. The main reason is that human emotions can best reveal human behaviors, feelings, thoughts, and intentions. By analyzing and interpreting the characteristics of human faces, we can learn about a person's current emotional state. To effectively find out the facial expression feature information and classify expressions, we use image sensors to capture facial expressions and propose a Feature-fusion-based Convolutional Neuro-fuzzy Classifier (FF-CNFC) to implement facial emotion recognition. In the FF-CNFC model, the neuro-fuzzy network classifier replaces the traditional fully connected neural network classifier for reducing the number of adjustable parameters. In addition, different fusion methods, including channel global maximum/average pooling fusion, global maximum/average pooling fusion, and network feature mapping methods, were used for the comparison of expression classification. In our experiment, we used the Multi Pose, Illumination, Expressions (Multi-PIE) face data set. The confusion matrix was used as the evaluation standard, and the accuracy, sensitivity, precision, and F1-score were calculated to evaluate performance of the model and judge it's quality. Experimental results indicated that the accuracy, sensitivity, precision, and F1-score of the proposed FF-CNFC model with global maximum pooling fusion are 99.60, 99.58, 99.58, and 99.58%, respectively, and are higher than those of other similar models. In addition, the proposed FF-CNFC model has a smaller number of parameters than the other models.

1. Introduction

Emotions are human physiological reactions or the most subjective manifestation. Facial emotion recognition (FER) can detect people's attention, such as their behavior, personality, mental status, and whether they are lying. Regardless of gender, culture, nationality, and race, the facial emotions of most people can be recognized by image sensors. FER methods could be divided into geometry- and appearance-based methods. Valstar and Pantic(1) used a geometric feature-based method recognizing facial muscle action units to obtain facial features. Zhang *et al.*(2) used an appearance-based method that applied Gabor filters to the face to extract the

*Corresponding author: e-mail: cjlin@ncut.edu.tw

https://doi.org/10.18494/SAM5113

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

appearance changes of the face. Owing to the differences in viewing angles and complex background information in face and expression recognition, previous image-processing methods have not been able to improve the recognition accuracy, and most of them can only be applied to frontal and close-range FER. By combining AI, FER technology can solve the above-mentioned problems such as viewing angle and light. At present, FER is widely used in human–computer interaction, medical care, psychology, and transportation.

In FER methods, machine learning (ML) technology is widely utilized and divided into three steps. The first step is to preprocess the image (such as improving image resolution, contrast adjustment, or grayscale) for improving the accuracy of FER. The second step is to extract the image of the face and then detect landmark features on the face (such as nose, eyes, mouth, and eyebrows).(3,4) Image preprocessing and facial feature extraction can be divided into global and local features. Global features include Eigenface, Fisherfaces, and principal component analysis.(5,6) Local features include local binary pattern and local discriminant analysis.(7–12) The third step is that ML classification methods (such as support vector machine, K-nearest neighbors, AdaBoost, and random forest) are used to classify the standard six emotions, namely, smiling, sadness, surprise, anger, fear, and disgust, and neutral. In the above-mentioned methods, the facial features still need to be selected by the user and are a very important factor affecting the recognition results.

Currently, the convolutional neural network (CNN) has become the mainstream of image recognition and has achieved good classification results. In FER applications, CNN has a simpler feature discrimination process than traditional ML(13–17) and need not preprocess feature labels for images. Instead, it directly extracts features from the image through the convolution kernel and effectively retains the required facial expression features. Lin *et al.*(18) proposed a multi-CNN based on an improved fuzzy integral to recognize facial emotions. They combined multiple CNNs, namely, AlexNet, GoogLeNet, and LeNet, to produce better results. To sum up, CNN can directly extract image features without specifically marking the details of expressions and effectively learn the landmark features of facial emotions. Huang *et al.*(19) combined the residual neural network and the squeeze-and-excitation network to implement FER applications. Chen *et al.*(20) combined fuzzy rough set theory and CNN to perform FER, and specifically removed noise samples from the original data to reduce uncertainty in fuzziness and indiscernibility. In the above-mentioned methods, these models still have too many parameters resulting in slow learning.

In this paper, we propose a Feature-fusion-based Convolutional Neuro-fuzzy Classifier (FF-CNFC) model for performing FER in the Multi Pose, Illumination, Expressions (Multi-PIE) face data set. The proposed FF-CNFC model is different from the traditional CNN model, especially in the feature fusion and fully connected layers. The main contributions of this study are as follows.

In the proposed FF-CNFC model, the neuro-fuzzy classifier replaces the traditional fully connected neural network classifier for reducing the number of adjustable parameters.

Five feature fusion methods in the FF-CNFC model were used for the comparison of expression classification, such as channel global maximum/average pooling fusion (CGMPF and CGAPF), global maximum/average pooling fusion (GMPF and GAPF), and network feature mapping fusion (NFMF) methods.

Through the feature fusion method, feature information can be compressed and dimensionally reduced to improve model computing efficiency and reduce the total number of network parameters.

Experimental results showed that the FER accuracy of the proposed FF-CNFC model is 99.60% and is higher than those of the LeNet (98.71%) and AlexNet (98.91%) models.

The rest of this paper is organized as follows. In Sect. 2, we provide a detailed introduction of the proposed FF-CNFC model. In Sect. 3, we present the experimental results obtained using the proposed FF-CNFC model on the Multi-PIE face data set. In Sect. 4, we provide the conclusions of this study and recommendations for future research.

2. Methods

In this section, we describe the proposed FF-CNFC model architecture for FER, as shown in Fig. 1. The FER process is divided into the following: (1) the collection of facial emotion images, (2) the training and testing of the data set through FF-CNFC and various feature fusion methods, and (3) the evaluation of the best FER results using various feature fusion methods.

2.1 Collection of facial emotion images

We used the seven emotions of the Multi-PIE face data set and adopted k-fold cross validation.[21] Five-fold cross validation is used to achieve FER, as shown in Fig. 2. Each facial emotion category is randomly divided into five sets of data. In each facial emotion category, four sets of data are used for training, and the remaining set is used for testing. Finally, the average accuracy is obtained by averaging the five accuracy results.

2.2 Proposed FF-CNFC model

The structure of the proposed FF-CNFC model is shown in Fig. 3. The FF-CNFC model consists of the convolution layer, the FER feature fusion layer, and the neuro-fuzzy network (NFN). The convolution layer includes convolution and maximum pooling operations. That is, facial emotional features are extracted through convolutional layers. In addition, maximum

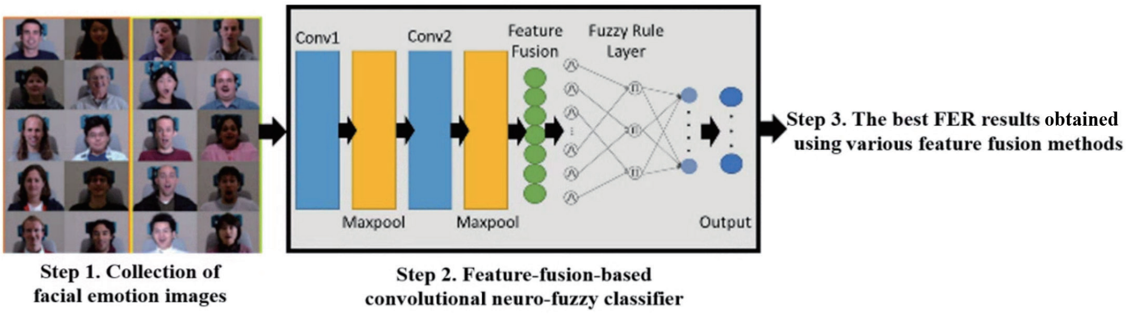
Conv1
Conv2
Feature Fusion
Fuzzy Rule Layer
Maxpool
Maxpool
Output
Step 1. Collection of facial emotion images
Step 2. Feature-fusion-based convolutional neuro-fuzzy classifier
Step 3. The best FER results obtained using various feature fusion methods

Fig. 1. (Color online) Proposed FER system.

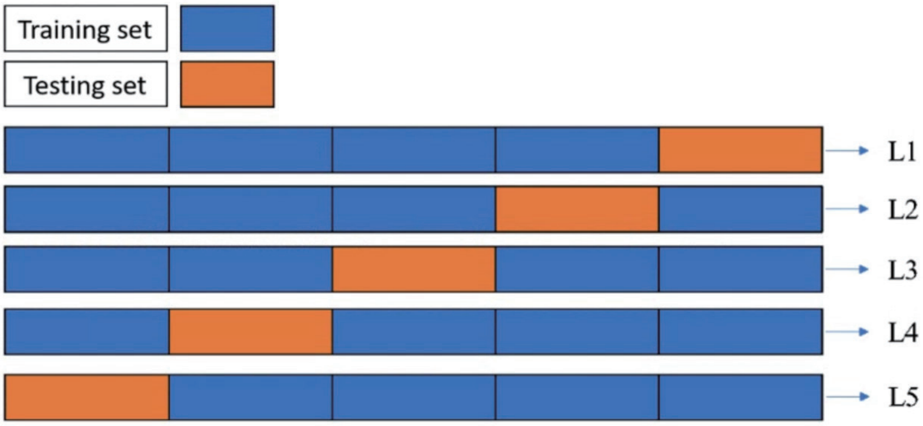

Fig. 2. (Color online) K-fold cross validation.

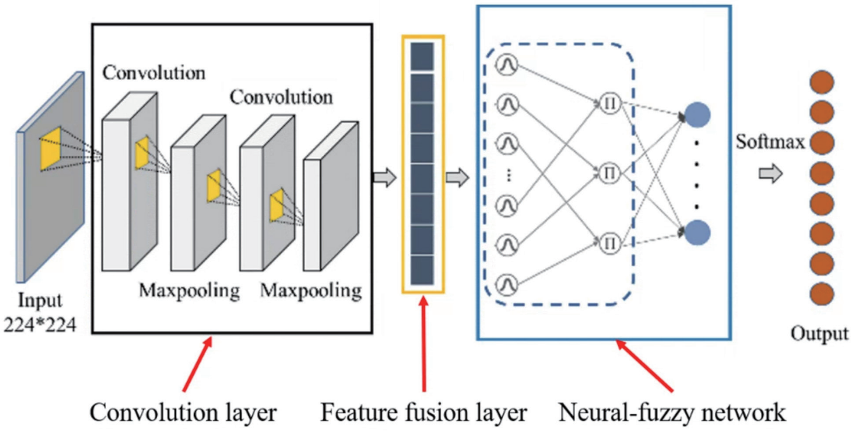

Fig. 3. (Color online) Architecture of FF-CNFC model.

pooling operation is used for feature compression to preserve more texture and speed up the operation.(22) The feature fusion layer fuses the features obtained after the convolutional layer operation to reduce the feature dimension. The NFN replaces the fully connected neural network classifier to obtain a smaller number of parameters and improve the recognition ability. Finally, softmax is used to calculate seven facial emotion categories.

2.2.1 Convolution layer

In this study, we used convolution layers including convolution and maximum pooling operations to capture the emotional features of facial images. The image was set to 224 × 224. A 3 × 3 convolution kernel was used for sliding stride. During the stride, interactive stacking and inner product operations were performed to obtain new feature values. The maximum pooling operation was used to compress each layer of convolution to speed up the operation process and reduce the computational load of deep networks.

The convolution operation performs a product operation on the overlapping positions according to the sliding step of the convolution kernel. By sliding from left to right and from top to bottom, the feature map of the new matrix is obtained as shown in Fig. 4. The formula is

$$w_o = \frac{\left(w_i - k + 2p\right)}{s} + 1, \tag{1}$$

where w_i is the size of the input image, w_o is the size of the output feature map, p is the padding, and s is the sliding stride.

2.2.2 Feature fusion layer

Table 1 shows that five different feature fusion methods are used to merge the facial features of the convolutional layer and obtain more useful features. The two global pooling fusion and channel global pooling fusion methods are divided into maximum and average pooling operations, respectively. According to different operations on features, five different fusion methods are obtained and shown in Fig. 5.

The calculation formula of the NFMF method is

$$f_z = \sum\nolimits_{i=1}^{n} w_{zi} * x_i, \tag{2}$$

where f_z is the z-th fusion result output, n is the number of input features, x_i is the i-th input feature, and w_{zi} is the i-th input weight used in the z-th fusion result.

2.2.3 NFN

The NFN combines the human-like reasoning method of fuzzy theory with the learning ability of neural networks. The architecture of NFN is divided into input, fuzzification, rule base, and defuzzification. Fuzzification is to fuzzify the input signal, and its value is between 0

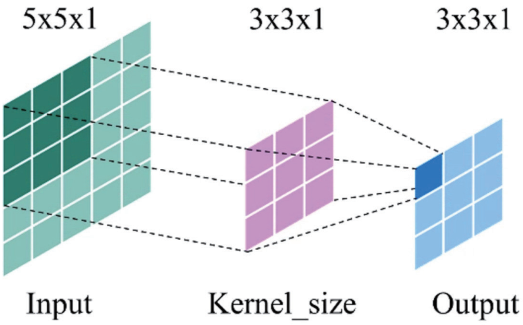

Fig. 4. (Color online) Convolution operation.

Table 1
Various facial feature fusion methods.

Methods	Brief description
GMPF GAPF	Fuse the height and width of the entire feature
CGMPF CGAPF	Maximum/average operation is used for each channel of the feature map
NFMF	Each feature is fused by using different weights

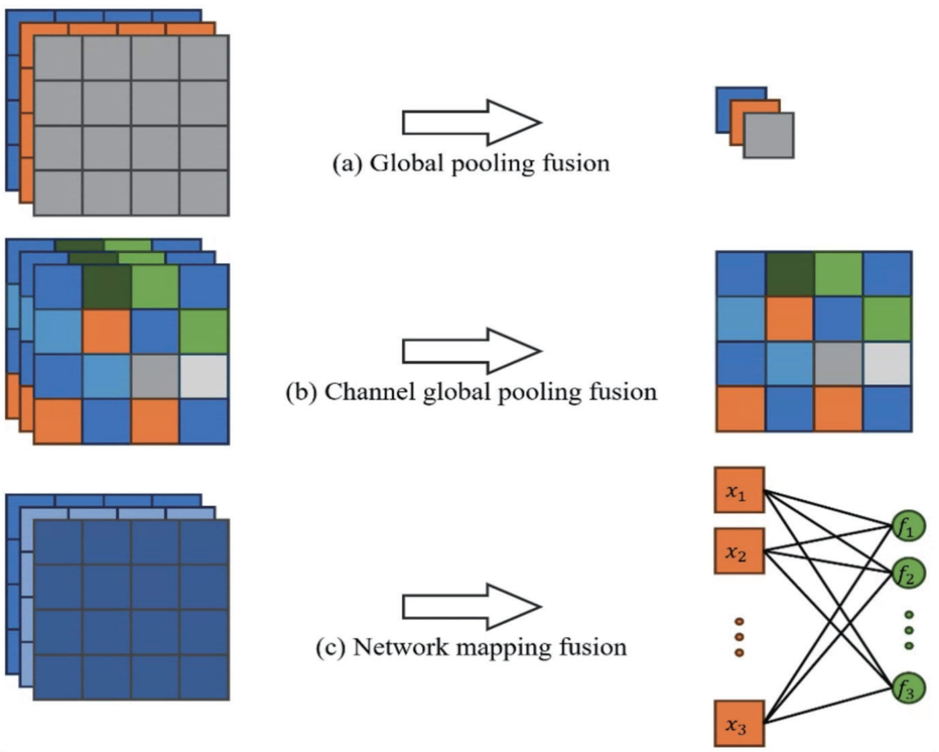
(a) Global pooling fusion
(b) Channel global pooling fusion
(c) Network mapping fusion
x_1
x_2
x_3
f_1
f_2
f_3

Fig. 5. (Color online) Schematic diagram of various fusion methods: (a) GMPF and GAPF, (b) CGMPF and CGAPF, and (c) NFMF.

and 1. In this study, we used the Gaussian function as the membership function. The output of the feature fusion layer was used as the input of the NFN classifier. Fuzzy rules are represented by If~Then~ and expressed as

$$\text{Rule}_j\text{: IF } x_1 \text{ is } A_{1j} \text{ and } x_2 \text{ is } A_{2j} \text{...and } x_i \text{ is } A_{ij} \text{ and...and } x_n \text{ is } A_{nj} \text{ THEN } y_i \text{ is } w_j\,, \tag{3}$$

where x_1 is the input, A is the fuzzy set, w_j is the output weight, and n is the input dimension.

In fuzzification, each input is fuzzified using a Gaussian function to obtain the degree of membership function. The formula is

$$\mu_{ij}(x) = \exp\left(-\frac{\left[x_i - m_{ij}\right]^2}{\sigma_{ij}^2}\right), \tag{4}$$

where x_i is the input, m_{ij} is the average, and σ_{ij} is the standard deviation.

In the rule base, the fuzzy intersection operation is performed on the fired strengths of the membership functions corresponding to each input. The formula of the fired strength of each fuzzy rule is

$$F_j = \prod_{i=1}^{n} \mu_{ij}\,. \tag{5}$$

Finally, the fired strength of each fuzzy rule is used as input for defuzzification. The crisp output is calculated as

$$y_k = \sum_{j=1}^{R} F_j w_{jk}\,, \tag{6}$$

where y_k is the k-th output, R is the number of fuzzy rules, F_j is the fired strength of the j-th rule, and w_{jk} is the output weight.

3. Experiments

To verify the accuracy and stability of the experiment, we used the FF-CNFC model with four convolution layers and compared it with different feature fusion methods. We also used the Multi-PIE face data set as verification data. We adopted the confusion matrix to analyze the model performance.

3.1 Data set

Figure 6 illustrates facial expressions at different angles and brightness for the Multi-PIE face data set. In this experiment, we selected 24912 facial images as training data and 6228 facial images as testing data, and divided them into seven expressions, namely, normal, squinting, happy, disgusted, surprised, smiling, and shouting.

3.2 Setting environment and model parameters

The model training environment uses TensorFlow and Keras as the development tools for the deep learning environment. We used the FF-CNFC model architecture with four convolution layers to implement FER, as shown in Table 2.

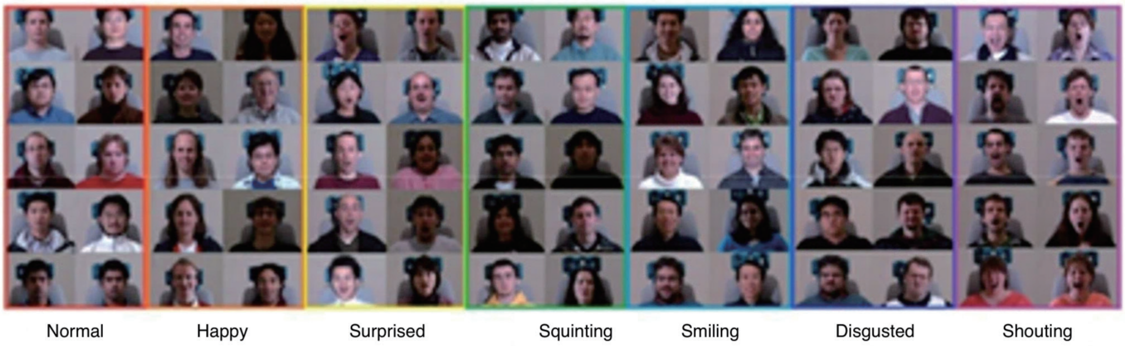

Fig. 6. (Color online) Multi-PIE face data set.

Table 2
Parameter settings of FF-CNFC model with four convolutional layers.

Layer	Image size	Kernel size	Number of filters	Stride
Input image	224 × 224 × 3			
Convolution layer 1		3 × 3	32	2
Maximum pooling layer 1		2 × 2		2
Convolution layer 2		3 × 3	64	1
Maximum pooling layer 2		2 × 2		2
Convolution layer 3		3 × 3	128	1
Maximum pooling layer 3		2 × 2		2
Convolution layer 4		3 × 3	64	1
Maximum pooling layer 4		2 × 2		2
Facial feature fusion			64	
Fuzzy rule			64	
Defuzzification			Number of categories	

3.3 Confusion matrix

In this study, we evaluated the FF-CNFC model performance by using accuracy, sensitivity, precision, and $F1$-*score* (Table 3). The categorical cross-entropy loss function was used as an evaluation indicator, and the formulas are as follows.

$$Accuracy = \frac{\left(TP + TN\right)}{\left(TP + FP + TN + FN\right)} \tag{7}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Precision = \frac{TN}{TN + FP} \tag{9}$$

Table 3
Evaluation indicator.

Real	Prediction	
	Positive	Negative
Positive	*TP*	*FN*
Negative	*FP*	*TN*

$$Loss = -\sum_{i=1}^{n} \hat{y}_{i1} \log y_{i1} + y_{i2} \log y_{i2} + \ldots + y_{im} \log y_{im} \tag{10}$$

TP represents a positive sample that was successfully predicted correctly, *FP* represents a positive sample that was incorrectly predicted, *TN* represents a negative sample that was successfully predicted correctly, *FN* represents a negative sample that was incorrectly predicted, *n* is the number of samples, and *m* is the number of categories.

When the sensitivity and precision of each category are obtained, $Average_{Sensitivity}$ and $Average_{Precision}$ can be obtained through macro-average. The formulas are as follows.

$$Average_{Sensitivity} = \frac{S_1 + S_2 + \ldots S_n}{n} \tag{11}$$

$$Average_{Precision} = \frac{P_1 + P_2 + \ldots P_n}{n} \tag{12}$$

$$F1\text{-}score = 2 \times \frac{\left(Average_{Precision} \times Average_{Sensitivity}\right)}{\left(Average_{Precision} + Average_{Sensitivity}\right)} \tag{13}$$

Here, S_1 and P_1 represent the sensitivity and accuracy of the first category, and *n* represents the number of total categories.

3.4 Experimental results using FF-CNFC model

In our experiments, the results obtained using FF-CNFC models with four convolutional layers and five different fusion methods were compared. The experimental results are shown in Table 4. In Table 4, the *accuracy*, *sensitivity*, *precision*, and *F*1-*score* of the proposed FF-CNFC model with GMPF are 99.60, 99.58, 99.58, and 99.58%, respectively, and are higher than those of other feature fusion methods.

We compared the performance of our FF-CNFC model with those of the other CNN models such as LeNet,(23) AlexNet,(24) GoogleNet,(25) multiple CNNs with improved fuzzy integral (MCNNs-IFI),(18) coupled generative adversarial network (cpGAN),(26) LS-SIFT,(27) and VGG-16 using facial representation learning (VGG-16-FRL).(28) The results are shown in Table 5. The accuracies of LeNet, AlexNet, GoogLeNet, MCNNs-IFI, cpGAN, LS-SIFT, VGG-16-FRL, and

Table 4
Comparison results of various feature fusion methods.

Methods	Fusion methods	*Accuracy* (%)	*Sensitivity* (%)	*Precision* (%)	*F*1-*score* (%)	*Loss*	*Parameter*
Channel pooling fusion	Average	98.70	98.68	98.70	98.69	0.0399	175687
	Maximum	99.45	99.43	99.44	99.43	0.0202	175687
Global pooling fusion	Average	98.84	98.79	98.79	98.79	0.0375	175687
	Maximum	99.60	99.58	99.58	99.58	0.0201	175687
NFMF	Weighted product	98.96	98.93	98.91	98.92	0.0321	585927

Table 5
Performance comparison of various deep learning networks.

Models	*Accuracy* (%)	Total parameters [million (M)]
LeNet[23]	98.71	1
AlexNet[24]	98.91	16
GoogLeNet[25]	99.04	21
MCNNs-IFI[18]	99.64	3
cpGAN[26]	96.43	8
LS-SIFT[27]	94.80	1
VGG-16-FRL[28]	98.95	0.3
Proposed model	99.60	0.175

FF-CNFC models are 98.71, 98.91, 99.04, 99.64, 96.43, 94.80, 98.95, and 99.60%, respectively. In addition, the numbers of parameters of LeNet, AlexNet, GoogLeNet, MCNNs-IFI, cpGAN, LS-SIFT, VGG-16-FRL, and FF-CNFC models are 1, 16, 21, 3, 8, 1, 0.3, and 0.175 M, respectively. In this paper, we proposed FF-CNFN, which combines four convolutional layers and a fuzzy neural network through a GMPF layer. The *accuracy* of our model is 99.60%, which is similar to that of MCNNs-IFI (99.64%), but the total number of parameters of our model is 0.175 M, which is less than that of MCNNs-IFI (3 M).

4. Conclusions

In this study, we used image sensing to capture facial expressions and proposed the FF-CNFC model to implement FER. In the proposed FF-CNFC model, the NFN classifier replaces the traditional fully connected layer network classifier to reduce the number of adjustable parameters. In addition, five fusion methods, including CGMPF and CGAPF, GMPF and GAPF, and NFMF methods, were used for the comparison of FER classification. In our experiment, we used the Multi-PIE face data set and also the confusion matrix as the evaluation standard to judge the quality of the model. Experimental results indicated that the *accuracy*, *sensitivity*, *precision*, and *F*1-*score* of the proposed FF-CNFC model with GMPF are 99.60, 99.58, 99.58, and 99.58%, respectively, and are higher than those of other feature fusion methods. In addition, the proposed FF-CNFC model has a smaller number of adjustable parameters than the other models.

In future research work, we expect to incorporate the learning rate and optimizer into adjustable parameters for improving the accuracy of model recognition. In addition, we expect to implement the proposed FF-CNFC model using a field-programmable gate array to facilitate future real-time FER applications.

References

1 M. Valstar and M. Pantic: 2006 Conf. Computer Vision and Pattern Recognition Workshop (2006) 149.

2 Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu: Proc. Third IEEE Int. Conf. Automatic Face and Gesture Recognition (1998) 454−459.

3 V. Upadhyay and D. Kotak: 2020 Fourth Int. Conf. Inventive Systems and Control (2020) 15.

4 S. Shojaeilangari, W.-Y. Yau, K. Nandakumar, J. Li, and E. K. Teoh: IEEE Trans. Image Process. **24** (2015) 2140. https://doi.org/10.1109/TIP.2015.2416634

5 C. V. R. Reddy, U. S. Reddy, and K. V. K. Kishore: Traitement du Signal **36** (2019) 13. http://doi.org/10.18280/ts.360102

6 D. Wahyuningsih, C. Kirana, R. Sulaiman, Hamidah, and Triwanto: 2019 7th Int. Conf. Cyber and IT Service Management (2019) 1.

7 A. M. Jagtap, V. Kangale, K. Unune, and P. Gosavi: 2019 Int. Conf. Intelligent Sustainable Systems (2019) 219.

8 J. Wang, J. Zheng, S. Zhang, J. He, X. Liang, and S. Feng: 2016 9th Int. Symp. Computational Intelligence and Design (2016) 303.

9 C. Shan, S. Gong, and P. W. McOwan: Image Vision Comput. **27** (2009) 803. https://doi.org/10.1016/j.imavis.2008.08.005

10 M. M. Ahsan, Y. Li, J. Zhang, M. T. Ahad, and K. D. Gupta: Technologies **9** (2021) 31. https://doi.org/10.3390/technologies9020031

11 L. Sun, J. Dai, and X. Shen: 2021 2nd Int. Conf. Artificial Intelligence and Education (2021) 64.

12 S. Wang, Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang: IEEE Trans. Multimedia **12** (2010) 682. https://doi.org/10.1109/TMM.2010.2060716

13 S. B. Sukhavasi, S. B. Sukhavasi, K. Elleithy, A. El-Sayed, and A. Elleithy: Int. J. Environ. Res. Public Health **19** (2022) 3085. https://doi.org/10.3390/ijerph19053085

14 I. Oliveira, J. L. Silva, F. P. Quispe, and A. B. Alvarez: 2021 IEEE Engineering Int. Research Conf. (2021) 1.

15 R. Gill and J. Singh: 2021 10th Int. Conf. System Modeling & Advancement in Research Trends (2021) 497.

16 H. Xiao, W. Li, G. Zeng, Y. Wu, J. Xue, J. Zhang, C. Li, and G. Guo: Appl. Sci. **12** (2022) 807. https://doi.org/10.3390/app12020807

17 J.-C. Kim, M.-H. Kim, H.-E. Suh, M. T. Naseem, and C.-S. Lee: Appl. Sci. **12** (2022) 5493. https://doi.org/10.3390/app12115493

18 C.-J. Lin, C.-H. Lin, S.-H. Wang, and C.-H. Wu: Appl. Sci. **9** (2019) 2593. https://doi.org/10.3390/app9132593

19 Z.-Y. Huang, C.-C. Chiang, J.-H. Chen, Y.-C. Chen, H.-L. Chung, Y.-P. Cai, and H.-C. Hsu: Sci. Rep. **13** (2023) 8425. https://doi.org/10.1038/s41598-023-35446-4

20 X. Chen, D. Li, P. Wang, and X. Yang: IEEE Access **8** (2020) 2772. https://doi.org/10.1109/ACCESS.2019.2960769

21 S. Yadav and S. Shukla: 2016 IEEE 6th Int. Conf. Advanced Computing (2016) 78.

22 C.-J. Lin and J.-Y. Jhang: IEEE Access **10** (2022) 14120. https://doi.org/10.1109/ACCESS.2022.3147866

23 Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner: Proc. IEEE **86** (1998) 2278. https://doi.org/10.1109/5.726791

24 A. Krizhevsky, I. Sutskever, and G. E. Hinton: Commun. ACM **25** (2012) 1097. https://doi.org/10.1145/3065386

25 C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich: 2015 IEEE Conf. Computer Vision and Pattern Recognition (2015) 1.

26 F. Taherkhani, V. Talreja, J. Dawson, M. C. Valenti, and N. M. Nasrabadi: 2020 IEEE Int. Joint Conf. Biometrics (2020) 1.

27 S. D. Lin and P. E. Linares Otoya: IEEE Access **12** (2024) 76648. https://doi.org/10.1109/ACCESS.2024.3406911

28 J. Xin, Z. Wei, N. Wang, J. Li, and X. Gao: IEEE Trans. Inf. Forensics Secur. **19** (2024) 934. https://doi.org/10.1109/TIFS.2023.3329686

About the Authors

Cheng-Jian Lin received his B.S. degree in electrical engineering from Ta Tung Institute of Technology, Taipei, Taiwan, R.O.C., in 1986 and his M.S. and Ph.D. degrees in electrical and control engineering from National Chiao Tung University, Taiwan, R.O.C., in 1991 and 1996, respectively. Currently, he is a chair professor of the Computer Science and Information Engineering Department, National Chin-Yi University of Technology, Taichung, Taiwan, R.O.C. His current research interests are in machine learning, pattern recognition, intelligent control, image processing, intelligent manufacturing, and evolutionary robots. (cjlin@ncut.edu.tw)

Xue-Qian Lin received his B.S. degree from the Computer Science and Information Engineering Department of National Chin-Yi University of Technology, Taichung, Taiwan, in 2021. Currently, he is a graduate student in the same department. His current research interests are in fuzzy neural network, image processing, and machine learning. (th0rnlin1412@gmail.com)