University of Baghdad
College of Engineering
JOURNAL OF ENGINEERING

Journal of Engineering

journal homepage: www.jcoeng.edu.iq

Volume 32 Number 6 June 2026

A Multi-CNN Fusion Approach for Improved Facial Expression Recognition

Ahmed Ahmed [1], **Yahya Ahmed** [2], **Sara Raed** [3, *]

[1]Artificial Intelligence Department, College of Information Technology, Ninevah University, Mosul, Iraq
[2]Presidency University, Northern Technical University, Mosul, Iraq
[3]Department of Computer and Information Engineering, College of Electronic Engineering, Ninevah University, Mosul, Iraq

ABSTRACT

Facial expression recognition (FER) is crucial in expressing a human's emotional state. Emotions and expressions on the human face are information that computers and deep learning can recognize. FER is a current research topic due to the recent advancement and use of human-computer interaction systems. The recognition of facial expressions is challenging for current deep learning models due to the variable brightness, background, pose, etc. of the face images. This work presents an improved learning method based on a feature fusion convolutional neural network (CNN) to recognize seven facial expressions. First, we trained three different base model structures and then fused the features of the pre-trained base models 1 and 2 from the final fully connected layers to obtain a fusion network. Second, the fusion network was trained and used with the third pre-trained base model for performance evaluation. We applied the max-score fusion and mean-score fusion techniques between the fusion network and the third pre-trained base model to predict the output class. Our results indicated that the proposed method outperforms the base models in all metrics and achieves a classification accuracy of 69.03% on the facial expression dataset.

Keywords: Fusion network, Deep learning, Convolutional neural network, Feature fusion, Facial expression recognition.

1. INTRODUCTION

In daily interactions and communication with others, facial expressions play a significant role. According to research conducted by psychologists **(Cavallo et al., 2018)**, the components of voice and language represent only 38% and 7%, respectively, of the complete emotional data shared by individuals in verbal communication. On the other hand, up to 55% of all information is due to facial expressions in social communication. Recently, facial recognition has become vital because it is used in many applications, such as robotics,

*Corresponding author
Peer review under the responsibility of University of Baghdad.
https://doi.org/10.31026/j.eng.2026.06.10

Article received: 01/03/2026
Article revised: 23/05/2026
Article accepted: 31/05/2026
Article published: 01/06/2026

healthcare, interactive games, and driver fatigue monitoring **(Revina and Emmanuel, 2021)**. FER is the main key to expressing the emotional experiences and feelings of humans **(Alkababji and Abd, 2021; Ge et al., 2022; Lan et al., 2025; Tang et al., 2025; Aly, 2025; Salloum et al., 2025)**.

FER analyzes a specific facial expression through a static image or video and uses the results of the analysis to classify a person's feelings **(Wang et al., 2020)**. Facial emotions can be categorized into seven different classes. Each of these emotions is characterized by specific facial muscle movements and configurations that express a particular emotional state **(Sharma et al., 2018)**. At the same time, recognizing facial expressions accurately remains a difficult process due to the effect of irrelevant facial information on the FER. For example, different poses, partial obstruction (e.g., hair, glasses), and background interference are the main causes of irrelevant information **(Niu et al., 2021)**.

The FER model consists of feature extraction and expression recognition. Features are extracted from the convolutional layers. The low-level features of the face (e.g., lines, shapes, edges, corners, etc.) are extracted from the first layers, and the advanced-level face recognition features are obtained from the last layers **(Ahmed et al., 2019)**. A CNN classifier is trained on input samples for facial expression recognition.

There are two categories of feature extraction methods for FER, i.e., features based on geometric attributes and others based on visual appearance **(Xu et al., 2018)**. The appearance feature-based approach involves discriminative features that are extracted by the Gabor filter **(Liu et al., 2012)**, local binary patterns (LBP) descriptor **(Shan et al., 2009)**, and a pyramid of histogram of gradients (PHOG) **(Dhall et al., 2011)**. The appearance feature-based approach is robust to noise. It reserves detailed information on facial expressions. The features based on geometric attributes depend on the measurement of geometric attributes of the face to identify facial expressions. The geometric feature-based approach is convenient, but it is sensitive to noise and hard to describe some subtle changes in the face **(Sun et al., 2019)**.

CNNs have made great successes in the fields of computer vision applications and machine learning tasks **(Russakovsky et al., 2015; Krizhevsky et al., 2017; Szegedy et al., 2015)**. CNNs extract learned features that transition from low-level visual attributes to high-level features. A deep neural network was used to generate a collection of facial features to increase classification accuracy on several facial recognition datasets **(Mollahosseini et al., 2016)**. A learning method was presented using a fusion convolutional neural network to classify human facial expressions. We train three models with different structures and then concatenate the features of the pre-trained models 1 and 2 from the final fully connected layers to build a fusion CNN. The fusion CNN is trained and used with the third base model for performance evaluation. We apply the max-score fusion and mean-score fusion techniques between the fusion network and the third base model as a final prediction to classify different expressions.

FER using CNNs that combine multiple models has been an important area of research focused on improving the accuracy of emotion detection. Several studies have proposed innovative techniques to enhance the recognition rates of these systems **(Chouhayebi et al., 2024)**. Zhu and Wen introduced a multi-channel convolutional neural network model that uses attention-based fusion, achieving a recognition accuracy of 93.56% on the FER2013 dataset **(Zhu and Wei, 2024)**. Cao et al. developed a method based on a multi-level, multi-model fusion convolutional neural network, reaching 71.78% accuracy on the challenging FER2013 dataset **(Cao et al., 2023)**. Avanija et al. emphasized the value of facial expression recognition systems in detecting emotions like anger, happiness, and surprise, which can aid

in human behavior analysis **(Avanija et al., 2022)**. **(Ni et al., 2022)** proposed a cross-modality attention-based convolutional neural network for recognizing expressions with subtle intensities, demonstrating high accuracy across various databases. These studies collectively demonstrate the effectiveness of fusion-based convolutional neural network models in improving the accuracy of facial expression recognition across different datasets and challenges.

(Yan et al., 2018) presented three different models for video sequences called the differential geometric fusion network (DGFN), deep-facial-sequential network (DFSN), and DFSN-1. DGFN depends on the geometric features of psychological and physiological rules and uses a conventional deep learning network. DFSN is constructed based on CNN. Then, DFSN-1 combines DGFN and DFSN to increase the facial expression recognition accuracy **(Tang et al., 2018)**. **(Ning et al., 2019)** proposed a multi-channel deep spatial-temporal feature fusion neural network (MDSTFN) for recognizing static image facial expressions. The model of MDSTFN is based on extracting information from the variations between the emotional face image and the neutral face image. To improve the performance, they investigated three feature-based fusion schemes: score average, SVM, and neural network **(Sun et al., 2019)**. A face expression recognition function fusion network (FFN-FER) was proposed by **(Yanli et al., 2019)** based on a cross-domain learning algorithm for FER. The network consists of a common feature, an intra-category (IC) channel, and a distinction feature inter-category (ID) channel. The common features are learned by the IC channel, while the characteristic features are learned by the ID channel. Finally, a fusion network combines the learning features of the two channels for FER **(Ji et al., 2019)**.

(Jie and Yongsheng, 2019) proposed three different CNN models for FER. The first model is called the shallow network to deal with the complicated topology and overfitting. The second model, a dual-branch CNN, utilizes two separate branches to extract both traditional LBP features and deep learning features. The third model, a pre-trained CNN, is designed to handle scenarios with a limited number of training images. They extracted three different features from a shallow network, a dual-branch CNN, and a CNN, and then they fused these features for classification **(Shao and Qian, 2019)**. A method to learn temporal dynamics and spatial features for FER was proposed in **(Liang et al., 2020)**. A deep network is utilized to extract spatial features from each frame, and a convolutional network, which accepts two consecutive frames as input, is used to model the temporal dynamics. The system gathers information from combined features by a BiLSTM network. **(Xie and Hu, 2018)** presented a method for solving the FER problem known as deep comprehensive multipatches aggregating convolutional neural networks (CNNs). They used a deep-based structure that includes two CNN branches. The relevant features are derived from individual image patches by the first branch, while the holistic features are captured from the whole expressive image in the second branch. A multi-task global-local Network was suggested for facial expression recognition **(Yu et al., 2020)**. They used different modules to extract features of various expressions. Finally, they fused the two modules to capture variations of different expressions.

We propose an improved learning method by fusing the features of the pre-trained models 1 and 2 to generate a fusion network, which is used with the third pre-trained model to improve the accuracy for FER. The contributions of this study are outlined as follows:

1. We introduce an improved approach to classify human facial expressions using a fusion CNN.

2. We train three different model structures and fuse the features of the pre-trained models 1 and 2 to produce a fusion CNN. The fusion CNN is then trained and utilized with the third pre-trained model for classification.

2. PROPOSED METHOD

We present our method to classify human facial expressions. The proposed method's framework with two steps is depicted in **Fig. 1**. First, we train three different base model structures and then concatenate the features of the pre-trained base models 1 and 2 to build a fusion network.
Second, the fusion network is trained and used with the third pre-trained base model for evaluation. Since the accuracies of the first and second base CNNs are less than the accuracy of the third model, they have been selected to be fused to build a fusion network. The max-score fusion and mean-score fusion techniques between the fusion network and the third pre-trained base model are applied as a final prediction to classify facial expressions into different categories.
Fig. 1 shows that the first base model has 5 convolutional layers and 4 fully connected layers (fcs), while the second base model consists of 4 convolutional layers and 3 fc layers. Batch normalization (BN), rectified linear unit (ReLU), max-pooling, and dropout layers follow each convolutional layer. Similarly, the three fc layers for base model 1 and the two fc layers for base model 2 are followed by BN, ReLU, and dropout layers. The third base model includes 6 convolutional layers and 2 fc layers. The first and the fifth convolutional layers are followed by BN and ReLU layers, while the other convolutional layers are followed by BN, ReLU, max-pooling, and dropout layers. The first fc layer of the third base model is followed by ReLU. Finally, a softmax layer is applied to the last fc layer in each model, generating the predicted class label.

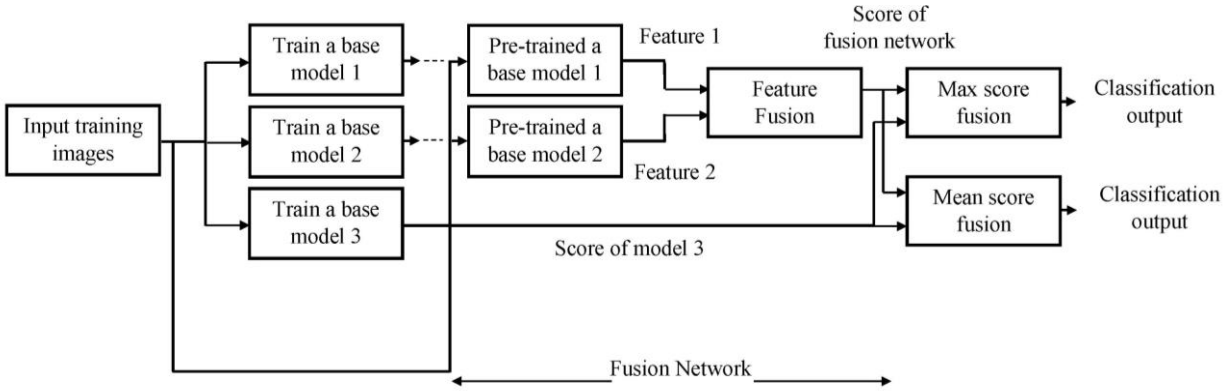


Figure 1. Framework of the proposed method with two stages

2.1 Fusion CNN

The objective of this task is to fuse the pre-trained base models 1 and 2. The architecture of pre-trained base model 1 differs from that of pre-trained base model 2. It has 5 convolutional layers and 4 fc layers, while the structure of base model 2 has 4 convolutional layers and 3 fc layers. The fusion network is trained on the input training samples and is used with base model 3 for performance evaluation, as shown in **Fig. 2.**

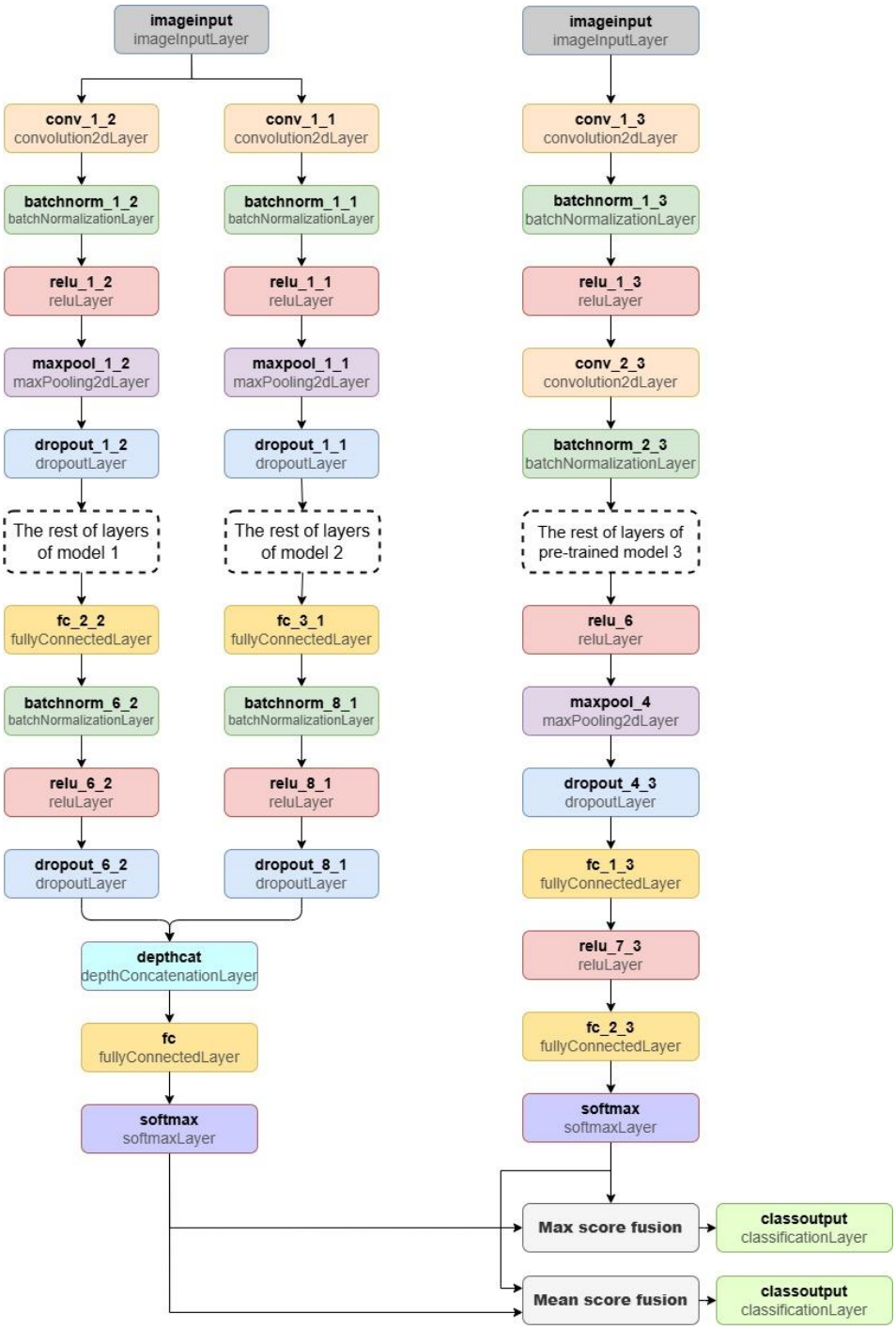


Figure 2. Prediction by fusion convolutional neural network and the third model

Assume the concatenated feature vector for base models 1 and 2 is represented as
$x_{cat} = x_1 \,||\, x_2$
where || denotes the concatenation operator. The score function of x_{cat} is then computed as in Eq. (1)

$$S_{cat} = f(x_{cat}, W, b) = W x_{cat} + b \quad (1)$$

Adam optimizer is employed to update the weights W and bias b as shown in Eqs. (2) to (7)

$$W_{t+1} = W_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \quad (2)$$

$$b_{t+1} = b_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\varepsilon} \quad (3)$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \quad (4)$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \quad (5)$$

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \quad (6)$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \quad (7)$$

where η is the learning rate, m_t and v_t are the first and second moment estimates, t is the time step, β_1 and β_2 are the exponential decay rates, ε is a small constant for numerical stability, and $g_t = \nabla_W L_t$ is the gradient of the loss function with respect to the model parameters. To assign a probability to each class, we use the softmax function for the fusion network at the final layer, as shown in Eq. (8)

$$P_{fusion\_Network}(Y = l|X = x_{cat}) = \frac{e^{S_{cat,l}}}{\sum_r e^{S_{cat,r}}} \quad (8)$$

Here, $P_{fusion\_Network}(Y = l|X = x_{cat})$ represents the probability of the fusion network for the class l given the input x_{cat}. $S_{cat,l}$ represents the score of class l. $\sum_r e^{S_{cat,r}}$ represents the sum of the scores of all classes. The loss function of the fusion network can be calculated in Eq. (9)

$$L_{fusion\_network} = -\log P(Y = l|X = x_{cat}) = -\log\left(\frac{e^{S_{cat,l}}}{\sum_r e^{S_{cat,r}}}\right) \quad (9)$$

After training the fusion network, we apply the max-score fusion and mean-score fusion techniques between the probabilities of the fusion network and the third base model to predict the output label, as shown in Eqs. (10) and (11)

Let $P_{fusion\_Network} = [P_{f1}, P_{f2}, P_{f3}, \dots, P_{fn}]$
and
$P_{third\_model} = [P_{tm1}, P_{tm2}, P_{tm3}, \dots, P_{tmn}]$
Then,

$$P_{max} = max(P_{fusion\_Network}, P_{third\_model}) \quad (10)$$

$$P_{mean} = \frac{P_{fusion\_Network} + P_{third\_model}}{2} \quad (11)$$

The final predicted class is obtained using:

$$\hat{y}_{max} = \text{argmax}(P_{max}) \quad (12)$$

$$\hat{y}_{mean} = \text{argmax}(P_{mean}) \quad (13)$$

where, P_{max} and P_{mean} represent the probability vectors obtained using the max-score fusion and mean-score fusion techniques, respectively. P_f and P_{tm} represent the probabilities predicted by the fusion network and the third model, respectively. The number of classes is denoted by n.

2.2 Prediction by Fusion CNN and Third Base Model

The features of the first and second pre-trained base models are concatenated to build a fusion network. This network is trained and used with the third base model for final prediction. In this work, we used two techniques to predict the output label, called max-score

fusion and mean-score fusion techniques. The max-score fusion technique is an ensemble learning technique that combines the predictions from the fusion network and the third base model for final label prediction. For each test sample, the max-score fusion technique is applied between the predictions of the fusion CNN and the third base model by selecting the highest probability.

The mean-score fusion technique, which is known as averaging, is an ensemble learning method that combines the predictions from the fusion network and the third base model to determine the final prediction. For each test image, the average is applied between the scores or the probabilities of the fusion CNN and the third base model. Then, the final predicted label is obtained by applying the argmax function to the averaged probability scores. Algorithm 1 summarizes the proposed method to build a fusion CNN and utilize it with the third CNN to predict the output.

Algorithm 1: Prediction by fusion CNN and CNN3.
Input: Training dataset *D*, CNN1, CNN2, and CNN3.
Output: Learned models
1: Initialize: $i \leftarrow 1$, Epoch $\leftarrow N$, Batch size, and learning rate
2: repeat
3: Train CNN1, CNN2, and CNN3 on *D*
4: Create a fusion CNN by concatenating pre-trained CNN1 and CNN2
5: Train Fusion CNN on *D*
6: Use a fusion CNN to predict the output
7: Use the max-score fusion technique between the fusion CNN and CNN3 to predict the output
8: Use the mean-score fusion technique between the fusion CNN and CNN3 to predict the output
9: $i \leftarrow i + 1$
10: until $i <$ Epoch

3. EXPERIMENTAL WORK

This section describes the dataset description and the FER results of our proposed approach over seven different categories. We employ a facial expression dataset **(Pierre and Aaron, 2013),** including seven classes (Angry, Disgust, Fear, Happy, Neutral, Sad, and Surprise) to evaluate our approach. The training set contains 28,709 images, while the test set includes 7,178 images. Each image has a resolution of 48 × 48 pixels. The provided training and testing sets included with the dataset were directly used in this study. The distribution of training and testing samples for each class is presented in **Table 1**. A visual representation of the seven classes is shown in **Fig. 3**.

Table 1. The training and test sets of the facial expression dataset.

Label	Training Set	Test Set
Angry	3995	958
Disgust	436	111
Fear	4097	1024
Happy	7215	1774
Neutral	4965	1233
Sad	4830	1247
Surprise	3171	831

Figure 3. Samples of the facial expression dataset.

4. RESULTS AND DISCUSSION

This section evaluates the efficacy of our method on the facial expression dataset. The proposed method was compared to three different base model architectures. These models are trained separately using the Adam optimizer with a learning rate of 0.001, a batch size of 256, and a maximum epoch of 40. Different metrics, such as Accuracy (ACC), Precision (P), Recall (R), and F1 score (F1), have been used to evaluate our method. These metrics can be computed in Eqs. (14 to 17)

$$ACC = \frac{M1+M2}{M1+M2+N1+N2} \tag{14}$$

$$P = \frac{M1}{M1+N1} \tag{15}$$

$$R = \frac{M1}{M1+N2} \tag{16}$$

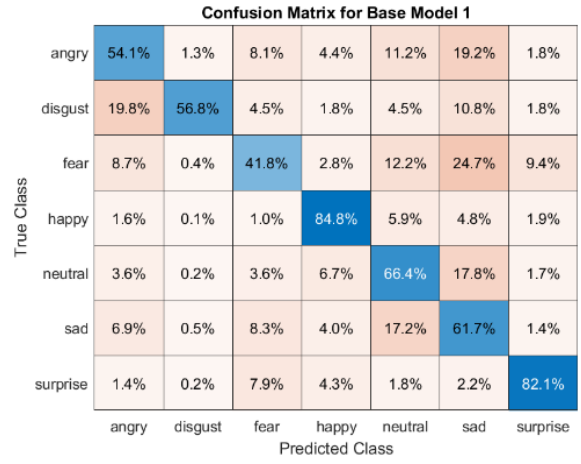
$$F1 = \frac{2 \cdot P \cdot R}{P+R} \tag{17}$$

Where M1, M2, N1, and N2 are the true positive, true negative, false positive, and false negative, respectively. **Table 2** demonstrates the effectiveness of the proposed method compared with baseline CNNs, standard deep learning architectures, and previous FER studies. The proposed method using the mean-score fusion technique outperformed the base CNN1, base CNN2, base CNN3, ResNet-50, GoogleNet, and EfficientNet-B0 by 3.6%, 4.11%, 2.63%, 7.27%, 7.17%, 8.23%, respectively. Furthermore, the proposed method achieved higher accuracy compared with the methods of Zahara et al. (2020) and Díaz et al. (2024), which reported accuracies of 65.97% and 63.0%, respectively, on the FER-2013 dataset. The confusion matrices for three base CNNs, fusion CNN, max-score fusion, and mean-score fusion are shown in **Fig. 4**.

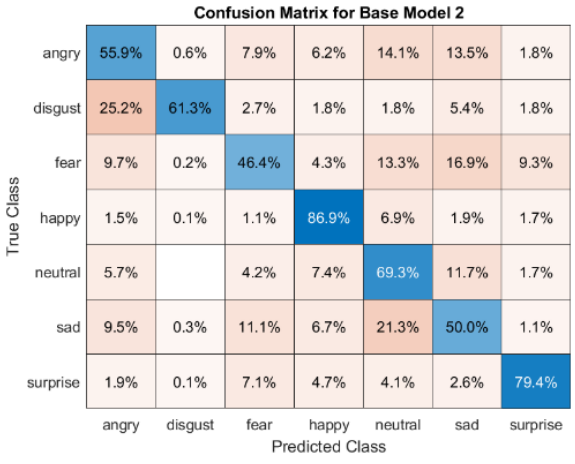
Table 2. Performance comparison on the facial expression dataset.

Method	Accuracy [%]
Base CNN1	66.63
Base CNN2	66.30
Base CNN3	67.26
ResNet-50 **(He et al., 2016)**	64.35
GoogleNet **(Szegedy et al., 2015)**	64.41
EfficientNet-B0 **(Tan and Le, 2019)**	63.78
(Zahara et al., 2020)	65.97
(Díaz et al., 2024)	63.0
Our Fusion CNN	67.53
Our Max-Score Fusion	68.57
Our Mean-Score Fusion	69.03

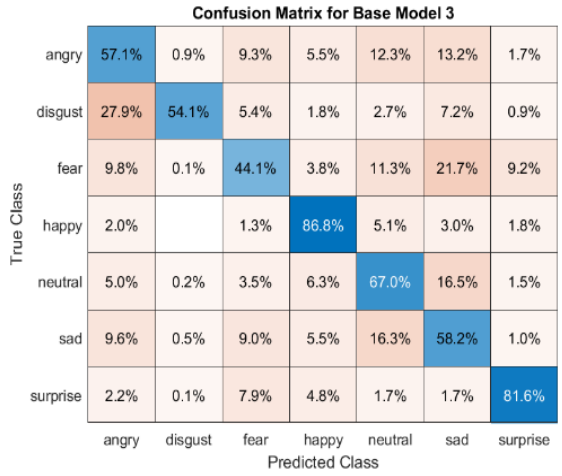
In **Fig. 5**, the P, R, and F1 metrics are computed for each class to show the efficacy of our method compared to the base CNNs. **Table 3** illustrates that our improved FER method outperforms the base CNNs in the metrics of average P, R, and F1.



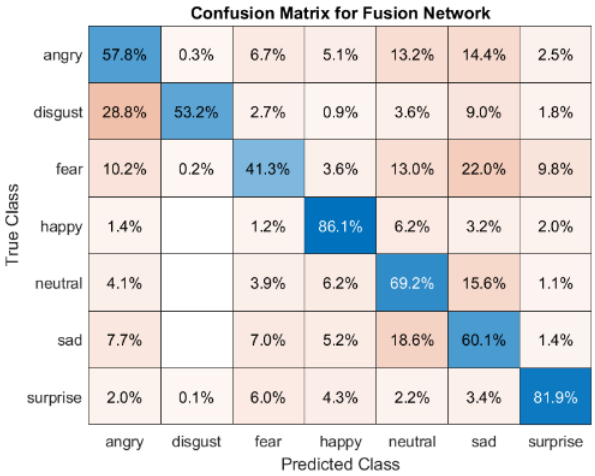
(a)



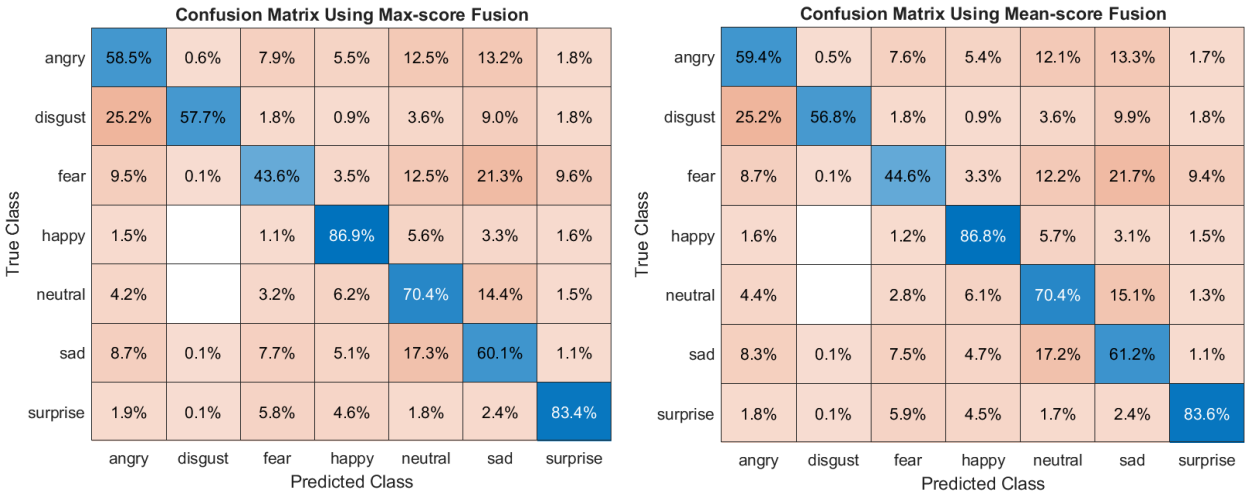
(b)



(c)



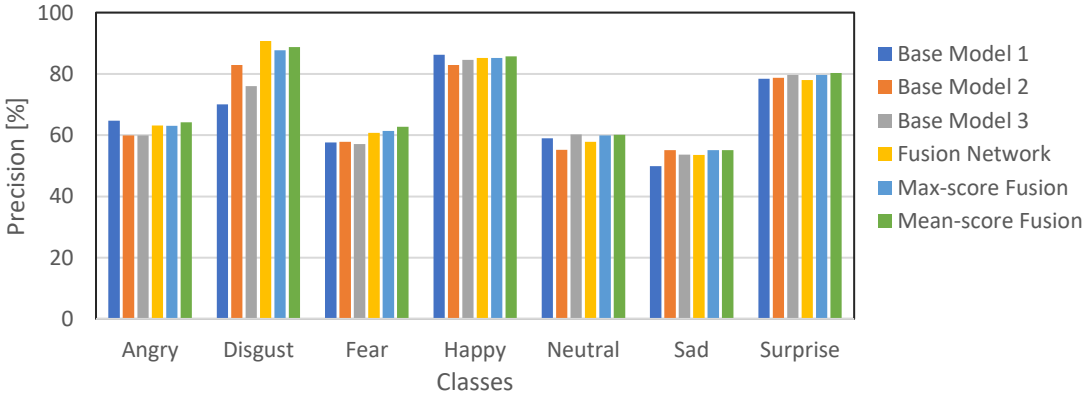
(d)



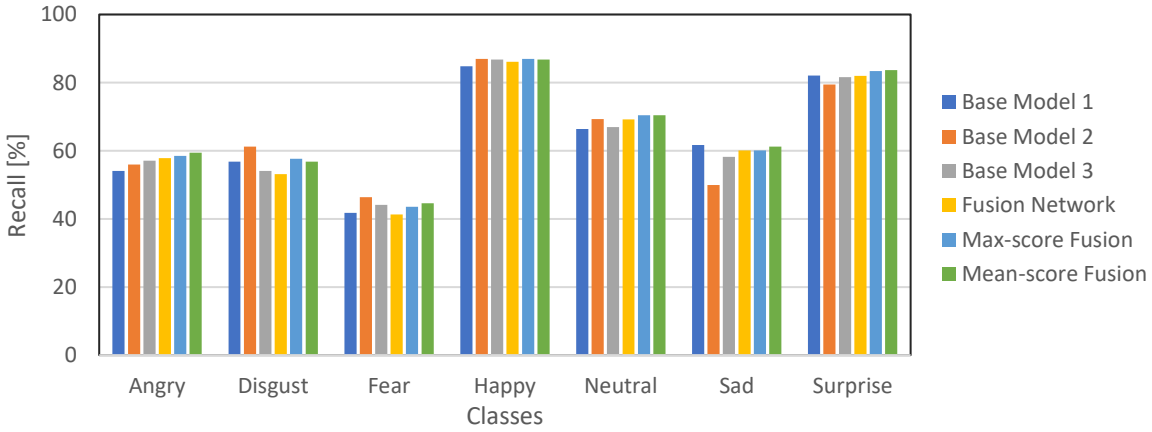
(e) (f)

Figure 4. Confusion matrices. (a) Base model 1. (b) Base model 2. (c) Base model 3. (d) Fusion network. (e) Max-score fusion. (f) Mean-score fusion.

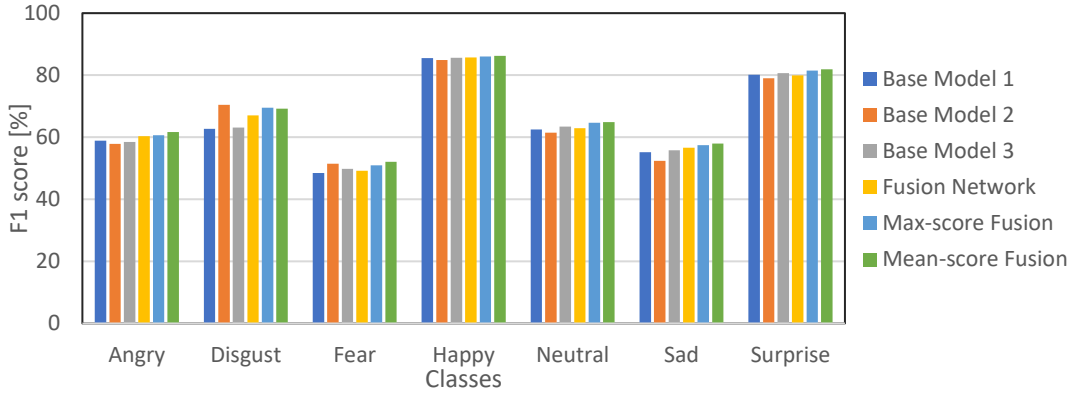
In this work, we use the support vector machine (SVM) and k-nearest neighbor (KNN) classifiers with a number of neighbors k =5 to compare with our method. For each pre-trained base model, we use the last fc layer to extract features, then the SVM and KNN classifiers are used to recognize facial expressions. **Table 4** shows the high accuracy obtained by our method compared to different classifiers.



(a)



(b)



(c)

Figure 5. The values of metrics P (a), R (b), and F1 (c) in the facial expression dataset.

Table 3. The values of average P, R, and F1 on the facial expression dataset.

Method	Average P	Average R	Average F1
Base CNN1	66.54	63.94	64.77
Base CNN2	67.50	64.18	65.36
Base CNN3	67.31	64.12	65.28
Our Fusion CNN	69.90	64.23	65.98
Our Max-Score Fusion	70.27	65.78	67.27
Our Mean-Score Fusion	70.97	66.12	67.72

Table 4. Performance comparison on the facial expression dataset with SVM and KNN classifiers.

Method	Accuracy [%]
Base CNN1 + SVM	66.69
Base CNN2 + SVM	65.83
Base CNN3 + SVM	66.29
Our Fusion Network + SVM	66.72
Base CNN1 + KNN	65.91
Base CNN2 + KNN	65.62
Base CNN3 + KNN	66.29
Our Fusion Network + KNN	66.12
Our Max-Score Fusion	68.57
Our Mean-Score Fusion	69.03

5. CONCLUSIONS

We present an improved learning method that leverages three convolutional neural networks with different architectures to enhance classification accuracy. By combining the feature maps from the two pre-trained CNNs, a fusion convolutional neural network is constructed, which is used with the third base CNN for the final prediction. Our proposed method achieves favorable performance by employing the mean-score fusion technique to combine the probabilities of the fusion CNN and the third base CNN, followed by selecting the class with the highest averaged probability to predict the output label. Experimental results indicate that our method achieved a classification accuracy of 69.03% on the facial expression dataset compared to the base models.

Credit Authorship Contribution Statement

Ahmed Ahmed: Experimental work, Writing, Validation, Methodology. Yahya Ahmed: Review & editing, Validation, Proofreading. Sara Raed: Writing –review and editing, Writing –original draft, Validation, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Ahmed, A., Yousif, H., Kays, R. and He, Z., 2019. Semantic region of interest and species classification in the deep neural network feature domain. *Ecological Informatics*, 52, pp. 57–68. https://doi.org/10.1016/j.ecoinf.2019.05.006

Alkababji, A. M. and Abd, S. R., 2021. Half-face-based recognition using principal component analysis. *Indonesian Journal of Electrical Engineering and Computer Science*, 22, pp. 1404–1410. https://doi.org/10.11591/ijeecs.v22.i3.pp1404-1410

Aly, M., 2025. Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model. *Multimedia Tools and Applications*, 84, pp. 12575–12614. https://doi.org/10.1007/s11042-024-19392-5

Avanija, J., Madhavi, K. R., Sunitha, G., Sangapu, S. C., and Raju, S., 2022. Facial expression recognition using convolutional neural network. *In Proceedings of 2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR).* pp. 1–7. IEEE. https://doi.org/10.1109/icaitpr51569.2022.9844221

Cao, J., Hu, C., Kong, L. and Yu, Z., 2023. Expression recognition based on a multi-level multi-model fusion deep convolutional neural network. *Highlights in Science, Engineering and Technology,* 34, pp. 232–237. https://doi.org/10.54097/hset.v34i.5477

Cavallo, F., Semeraro, F., Fiorini, L., Magyar, G., Sinčák, P. and Dario, P., 2018. Emotion modelling for social robotics applications: a review. *Journal of Bionic Engineering,* 15, pp. 185–203. https://doi.org/10.1007/s42235-018-0015-y

Chouhayebi, H., Mahraz, M. A., Riffi, J. and Tairi, H., 2024. A dynamic fusion of features from deep learning and the HOG-TOP algorithm for facial expression recognition. *Multimedia Tools and Applications*, 83, pp. 32993–33017. https://doi.org/10.1007/s11042-023-16779-8

Dhall, A., Asthana, A., Goecke, R. and Gedeon, T., 2011. Emotion recognition using PHOG and LPQ features. *In Proceedings of 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG).* pp. 878–883. IEEE. https://doi.org/10.1109/fg.2011.5771366

Díaz, A. A. O., Tamayo, S. C., De Oliveira, D. N. and Abensur, G. A., 2024. Models for real-time emotion classification: FER-2013 dataset. *In Proceedings of Intelligent Systems Conference. Springer Nature Switzerland.* pp. 289–304. https://doi.org/10.1007/978-3-031-66431-1_19

Ge, H., Zhu, Z., Dai, Y., Wang, B. and Wu, X., 2022. Facial expression recognition based on deep learning. *Computer Methods and Programs in Biomedicine*, 215, P. 106621. https://doi.org/10.1016/j.cmpb.2022.106621

He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).* pp. 770–778. https://doi.org/10.1109/cvpr.2016.90

Ji, Y., Hu, Y., Yang, Y., Shen, F. and Shen, H. T., 2019. Cross-domain facial expression recognition via an intra-category common feature and inter-category distinction feature fusion network. *Neurocomputing*, 333, pp. 231–239. https://doi.org/10.1016/j.neucom.2018.12.037

Krizhevsky, A., Sutskever, I. and Hinton, G. E., 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60, pp. 84–90. https://doi.org/10.1145/3065386

Lan, X., Xue, J., Qi, J., Jiang, D., Lu, K. and Chua, T.-S., 2025. Expllm: Towards chain of thought for facial expression recognition. *IEEE Transactions on Multimedia*, 27, pp. 3069-3081. https://doi.org/10.1109/tmm.2025.3557704

Liang, D., Liang, H., Yu, Z. and Zhang, Y., 2020. Deep convolutional BiLSTM fusion network for facial expression recognition. *The Visual Computer,* 36, pp. 499–508. https://doi.org/10.1007/s00371-019-01636-3

Liu, W., Song, C., Wang, Y. and Jia, L., 2012. Facial expression recognition based on gabor features and sparse representation. *In Proceedings of 2012 12th International Conference on Control Automation Robotics & Vision (ICARCV).* pp. 1402–1406. IEEE. https://doi.org/10.1109/ICARCV.2012.6485394

Mollahosseini, A., Chan, D. and Mahoor, M. H., 2016. Going deeper in facial expression recognition using deep neural networks. *In Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision (WACV).* pp. 1–10. IEEE. https://doi.org/10.1109/WACV.2016.7477450

Ni, R., Yang, B., Zhou, X., Cangelosi, A. and Liu, X., 2022. Facial expression recognition through cross-modality attention fusion. *IEEE Transactions on Cognitive and Developmental Systems*, 15, pp. 175–185. https://doi.org/10.1109/tcds.2022.3150019

Niu, B., Gao, Z. and Guo, B., 2021. Facial expression recognition with LBP and ORB features. *Computational Intelligence and Neuroscience*, 2021, P. 8828245. https://doi.org/10.1155/2021/8828245

Pierre, C. and Aaron, C., FER-2013 Dataset. Available at: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data (Accessed: 17 May 2026).

Revina, I. M. and Emmanuel, W. S., 2021. A survey on human face expression recognition techniques. *Journal of King Saud University-Computer and Information Sciences*, 33, pp. 619–628. https://doi.org/10.1016/j.jksuci.2018.09.002

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A. and Bernstein, M., 2015. Image net large scale visual recognition challenge. *International Journal of Computer Vision*, 115, pp. 211–252. https://doi.org/10.1007/s11263-015-0816-y

Salloum, S. A., Alomari, K. M., Alfaisal, A. M., Aljanada, R. A. and Basiouni, A., 2025. Emotion recognition for enhanced learning: using AI to detect students' emotions and adjust teaching methods. *Smart Learning Environments,* 12, P. 21. https://doi.org/10.1186/s40561-025-00374-5

Shan, C., Gong, S. and Mcowan, P. W., 2009. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 27, pp. 803–816. https://doi.org/10.1016/j.imavis.2008.08.005

Shao, J. and Qian, Y., 2019. Three convolutional neural network models for facial expression recognition in the wild. *Neurocomputing*, 355, pp. 82–92. https://doi.org/10.1016/j.neucom.2019.05.005

Sharma, P., Esengönül, M., Khanal, S. R., Khanal, T. T., Filipe, V. and Reis, M. J., 2018. Student concentration evaluation index in an e-learning context using facial emotion analysis. *In Proceedings of International Conference on Technology and Innovation in Learning, Teaching and Education.* pp. 529–538. Springer. https://doi.org/10.1007/978-3-030-20954-4_40

Sun, N., Li, Q., Huan, R., Liu, J. and Han, G., 2019. Deep spatial-temporal feature fusion for facial expression recognition in static images. *Pattern Recognition Letters*, 119, pp. 49–61. https://doi.org/10.1016/j.patrec.2017.10.022

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).* pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Tan, M. and Le, Q., 2019. EfficientNet: rethinking model scaling for convolutional neural networks. *In Proceedings of International Conference on Machine Learning (ICML).* pp. 6105–6114. PMLR. https://doi.org/10.48550/arXiv.1905.11946

Tang, X., Gong, Y., Xiao, Y., Xiong, J. and Bao, L., 2025. Facial expression recognition for probing students' emotional engagement in science learning. *Journal of Science Education and Technology*, 34, pp. 13–30. https://doi.org/10.1007/s10956-024-10143-7

Tang, Y., Zhang, X. M. and Wang, H., 2018. Geometric-convolutional feature fusion based on learning propagation for facial expression recognition. *IEEE Access*, 6, pp. 42532–42540. https://doi.org/10.1109/access.2018.2858278

Wang, W., Xu, K., Niu, H. and Miao, X., 2020. Emotion recognition of students based on facial expressions in online education based on the perspective of computer simulation. *Complexity*, 2020, P. 4065207. https://doi.org/10.1155/2020/4065207

Xie, S. and Hu, H., 2018. Facial expression recognition using hierarchical features with deep comprehensive multipatches aggregation convolutional neural networks. *IEEE Transactions on Multimedia,* 21, pp. 211–220. https://doi.org/10.1109/TMM.2018.2844085

Xu, L., Yang, A., Fei, M. and Zhou, W., 2018. Brief technical analysis of facial expression recognition. *In Proceedings of International Conference on Intelligent Computing for Sustainable Energy and Environment.* pp. 302–310. Springer. https://doi.org/10.1007/978-981-13-2384-3_28

Yu, M., Zheng, H., Peng, Z., Dong, J. and Du, H., 2020. Facial expression recognition based on a multi-task global-local network. *Pattern Recognition Letters*, 131, pp. 166–171. https://doi.org/10.1016/j.patrec.2020.01.016

Zahara, L., Musa, P., Wibowo, E. P., Karim, I. and Musa, S. B., 2020. The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based Raspberry Pi. *In Proceedings of 2020 Fifth International Conference on Informatics and Computing (ICIC).* pp. 1–9. *IEEE.* https://doi.org/10.1109/ICIC50835.2020.9288560

Zhu, M. and Wei, L., 2024. A multi-channel convolutional neural network based on attention mechanism fusion for facial expression recognition. *Applied Mathematics and Nonlinear Sciences*, 9, P. 2. https://doi.org/10.2478/amns.2023.1.00084

نهج دمج الشبكات العصبية التلافيفية المتعددة لتحسين التعرف على تعبيرات الوجه

احمد احمد[1]، يحيى احمد[2]، سارة رائد[3]، *

[1]قسم الذكاء الاصطناعي، كلية تكنولوجيا المعلومات، جامعة نينوى، الموصل، العراق
[2]الجامعة التقنية الشمالية، الموصل، العراق
[3*]هندسة الحاسوب والمعلوماتية، كلية هندسة الالكترونيات، جامعة نينوى، الموصل، العراق

الخلاصة

يعد التعرف على تعبيرات الوجه أمراً بالغ الأهمية في التعبير عن الحالة العاطفية للإنسان. إن العواطف والتعبيرات التي تظهر على وجه الانسان هي معلومات يمكن لأجهزة الكمبيوتر والتعلم العميق التعرف عليها. يعتبر التعرف على تعبيرات الوجه موضوع بحث حالي بسبب التقدم الحاصل واستخدام أنظمة التفاعل بين الإنسان والحاسوب. ان التعرف على تعبيرات الوجه يمثل تحدياً للنماذج الحالية للتعلم العميق بسبب التغييرات في السطوع والخلفية والوضع وما إلى ذلك لصور الوجه. يقدم هذا البحث طريقة تعلم محسنة تعتمد على دمج ميزات الشبكة العصبية التلافيفية للتعرف على سبعة تعبيرات للوجه. أولاً، يتم تدريب ثلاث شبكات اساسية مختلفة ثم يتم دمج الميزات او السمات للشبكة الاولى والثانية المدربتين مسبقًا من الطبقات النهائية الكاملة الاتصال للحصول على الشبكة الاندماجية. ثانيًا، يتم تدريب الشبكة الاندماجية واستخدامها مع الشبكة الاساسية الثالثة المُدربة مسبقًا لتقييم الأداء. يتم تطبيق تقنيات الدرجة القصوى والمتوسط بين الشبكة الاندماجية والشبكة الأساسية الثالثة المدربة مسبقًا للتنبؤ بفئة الإخراج. تشير النتائج إلى أن الطريقة المقترحة تتفوق على النماذج الأساسية في جميع المقاييس وتحقق دقة تصنيف تبلغ 69.03٪ في مجموعة بيانات تعبيرات الوجه.

الكلمات المفتاحية: دمج الشبكات، التعلم العميق، الشبكات العصبية التلافيفية، دمج الميزات، التعرف على تعابير الوجه.