



Facial Expression Classification Using Convolutional Neural Network on the FER2013 Dataset

Muhammad Farrel Adynata^{1*}, Verrel Mhd. Andrifan²,
Hadi Defri³, Azmil Hasim⁴

^{1,2}Department of Information System, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

³Department of Chemical Engineering, Gifu University, Japan

⁴Department of Human Sciences, Sultan Idris University of Education, Malaysia

E-Mail: ¹12350312932@students.uin-suska.ac.id, ²12350313914@students.uin-suska.ac.id,
³hadidefri@gmail.com, ⁴azmilhasil@gmail.com

Received Aug 4th 2026; Revised Sep 15th 2026; Accepted Sep 28th 2026; Available Online Oct 05th 2026

Corresponding Author: Muhammad Farrel Adynata

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Facial expression classification is an important task in computer vision that enables systems to identify human emotions from facial images. This study aims to evaluate the performance of a Convolutional Neural Network (CNN) for classifying facial expressions using the FER2013 dataset. The dataset consists of grayscale facial images categorized into seven emotion classes: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The images were normalized and augmented using rotation, shifting, zooming, and horizontal flipping to improve model generalization. A CNN architecture consisting of three convolutional layers with 32, 64, and 128 filters, followed by max-pooling, a fully connected layer, dropout, and a softmax output layer, was trained using the Adam optimizer for 20 epochs with a batch size of 32. Model performance was evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The model achieved a test accuracy of 60%, with weighted precision, recall, and F1-score of 60%. Among the seven classes, Happy achieved the highest F1-score of 79%, followed by Surprise at 73%, while Fear, Sad, and Disgust achieved F1-scores of 45%, 46%, and 47%, respectively. These results indicate that the CNN can effectively classify several facial expressions, particularly Happy and Surprise, while performance is lower on visually similar and less-represented classes. The contribution of this study is an evaluation and class-level analysis of a CNN-based facial expression classification approach on the FER2013 dataset, providing insights into performance differences across emotion categories and potential directions for improving classification performance.

Keywords: Convolutional Neural Network, Deep Learning, FER2013, Image Classification, Facial Expression Recognition

1. INTRODUCTION

Facial expression recognition (FER) has become an important research area in computer vision because facial expressions provide meaningful non-verbal information about human emotional states and can support applications such as human-computer interaction, healthcare, security, education, and intelligent systems. Recent developments in deep learning have significantly advanced FER by enabling models to automatically learn discriminative facial features directly from image data. In particular, convolutional neural networks (CNNs) remain widely used for facial expression classification because their hierarchical feature-extraction mechanism captures spatial patterns associated with different facial expressions. However, reliable FER remains challenging under practical conditions due to variations in illumination, head pose, occlusion, image resolution, individual characteristics, and visual similarity across emotional categories. In addition, limitations in the quantity and distribution of labeled facial-expression data can lead to overfitting and imbalanced classification performance across emotion classes. [13], [15], [16], [17].

The development of Artificial Intelligence (AI) technology, especially in Deep Learning, has led to tremendous progress in many digital image processing tasks. One of the most widely used techniques is the Convolutional Neural Network (CNN), a network architecture designed to automatically extract visual features from images. CNNs can learn important attributes of an object without manual feature extraction, resulting in superior performance over conventional techniques across various image classification tasks [4], [9], [18]. Recent studies have also shown that deep learning remains the primary method for developing facial expression recognition systems due to its ability to understand complex facial feature representations [16], [17].



The Facial Expression Recognition 2013 (FER2013) dataset is a widely used benchmark dataset for facial expression classification. It contains 35,887 facial images collected from Internet-based image searches and processed into grayscale images with a resolution of 48×48 pixels. The dataset consists of seven emotion categories: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The distribution of samples across these categories is highly imbalanced, with Disgust containing 547 images and Happy containing 8,989 images. This imbalance poses a challenge for classification models, as performance may vary considerably across emotion categories. Recent studies continue to use FER2013 as a benchmark for evaluating facial expression recognition methods, including CNN-based and more advanced deep learning approaches. [17], [19].

Nonetheless, clustering facial expressions remains challenging due to variations in lighting, head position, image clarity, and the resemblance among certain expressions. Certain emotions, such as fear and surprise, or sadness and neutrality, frequently exhibit very similar visual patterns, thereby heightening the likelihood of clustering errors [14], [19]. Additionally, variations in data volume across emotion categories can affect the model's ability to adapt to new data [16], [19].

Recent studies have demonstrated the continued development of deep learning approaches for facial expression classification using the FER2013 dataset. Dada et al. (2023) compared CNN-10 with VGG19, InceptionV3, and the Vision Transformer (ViT), achieving 84.3% accuracy on the FER2013 dataset. The result demonstrates the effectiveness of a CNN architecture specifically designed for facial emotion classification, although the model is more complex than a conventional CNN architecture. Mejia-Escobar et al. (2023), meanwhile, investigated a data-centric approach that progressively filtered and reclassified FER2013 samples. Their CNN achieved 66.26% accuracy on the original dataset and 86.71% after reclassifying the dataset, indicating that data quality and label consistency can have a substantial effect on facial expression classification performance.

Other recent studies have explored lightweight CNNs, transformer architectures, and hybrid models to address the limitations of conventional CNN-based facial expression recognition. A lightweight CNN study published in IEEE Access in 2024 reported 63% accuracy on the FER2013 dataset, compared with 58% for MobileNetV2 and 65% for ShuffleNetV2, highlighting the trade-off between computational efficiency and classification performance. In the same year, the Swin-FER model achieved 71.11% accuracy on FER2013 by incorporating a modified Swin Transformer architecture to capture broader spatial relationships in facial images. However, transformer-based models generally require greater computational resources and sufficient training data. Hybrid approaches have subsequently combined CNNs and transformers to exploit both local facial features and global contextual information, as demonstrated by the CL-TransFER framework. More recently, Anand and Babu (2026) reported 91.6% accuracy using a complex hybrid architecture involving a FER Vision Transformer, depthwise convolution, LSTM, attention mechanisms, and optimization on the FER2013 dataset. These developments indicate that increasing architectural complexity can improve reported performance, but may also introduce higher computational requirements and greater methodological complexity.

2. MATERIAL AND METHOD

This study aims to categorize facial expressions using a Convolutional Neural Network (CNN) on the FER2013 dataset. The research process begins with data collection, initial data processing, CNN structure development, model training, and model performance assessment. The research process implemented is shown in Figure 1.

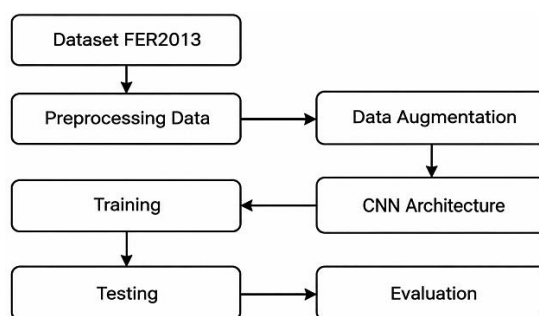


Figure 1. Research Methodology

2.1. Dataset Description

This research employs the Facial Expression Recognition 2013 (FER2013) dataset introduced by Goodfellow et al. [1]. This collection consists of over 35,000 grayscale facial images, each measuring 48×48 pixels, categorized into seven emotional groups: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral.

The dataset is divided into three components: a training set, a validation set, and a test set. This data split aims to train the model, validate it during training, and assess its performance on unseen data.

2.2. Preprocessing Data

The preprocessing stage is performed to improve data quality and prepare images for CNN model training. The initial step is to normalize the image pixel values from the range 0–255 to 0–1 using a rescaling method. This normalization process aims to accelerate model convergence and increase stability during training. In addition to normalization, this study employs data augmentation on the training dataset. The methods applied include 15-degree image rotation, vertical and horizontal shifts of up to 10%, zooming up to 10%, and horizontal flipping. The purpose of applying data augmentation is to increase the diversity in the training data and minimize the possibility of overfitting during model training [7], [8], [7], [19].

2.3. CNN Architectures

The classification model applied in this study is a Convolutional Neural Network (CNN). CNN was chosen for its superior ability to automatically extract visual features from facial images [4], [9], [10], [11]. The implemented CNN architecture consists of three convolutional layers, followed by a max-pooling layer to reduce the feature size. The feature extraction results are flattened via a flatten layer and passed to a fully connected layer for classification.

Table 1. Shows the CNN architecture used in this study

Layer	Configuration
Input Layer	$48 \times 48 \times 1$
Conv2D	32 Filter, 3×3 , ReLU
MaxPooling2D	2×2
Conv2D	64 Filter, 3×3 , ReLU
MaxPooling2D	2×2
Conv2D	128 Filter, 3×3 , ReLU
MaxPooling2D	2×2
Flatten	-
Dense	128 Neuron, ReLU
Dropout	0.5
Output Layer	7 Neuron, Softmax

The implemented CNN architecture has three convolutional layers, accompanied by a max pooling layer to reduce the feature size. The extracted features are flattened into a one-dimensional vector and sent to the fully connected layer for classification.

2.4. Model Training

The model is trained using the Adam optimizer because it can adjust the learning rate for various image classification problems. [4]. The loss function used is categorical cross-entropy because this research involves multi-class classification.

Table 2. Training Parameters

Parameter	Value
Epoch	20
Batch Size	32
Optimizer	Adam
Loss Function	Categorical Cross-Entropy
Learning Rate	0.001

Model training was conducted on Google Colab, which provides GPU support to accelerate computation. Training parameters were selected to balance learning speed and model effectiveness for ideal convergence. The Adam optimizer was chosen because it can automatically adjust the learning rate during training, making it suitable for image classification problems. Furthermore, a batch size of 32 helped maintain the stability of model weight updates, while a set number of 20 epochs was deemed sufficient to enable the model to learn facial expression patterns without excessively increasing the risk of overfitting.

2.5. Performance Evaluation

Model performance evaluation was conducted using test data. Model performance was assessed using various evaluation metrics, including accuracy, precision, recall, and F1 Score. Additionally, a confusion matrix was used to assess the model's ability to classify each emotion type.

3. RESULTS AND DISCUSSION

This study aims to classify seven facial expression types in the FER2013 dataset using a Convolutional Neural Network (CNN). After completing preprocessing, data enhancement, model training, and evaluation, the results show the model's ability to identify various facial expressions with a fairly high level of accuracy. See Table 3. The assessment was carried out using accuracy, precision, recall, and F1-score metrics, as well as confusion matrix analysis, to evaluate the model's performance for each emotion category. In this section, we discuss model training results, classification evaluation, and model performance analysis based on the confusion matrix and comparisons with previous studies [16], [18], [20], [22], [23].

Table 3. Classification Report

Emotion	Precision	Recall	F1-score
Angry	56%	48%	52%
Disgust	86%	33%	47%
Fear	48%	42%	45%
Happy	76%	82%	79%
Neutral	53%	61%	56%
Sad	43%	50%	46%
Suprise	80%	60%	73%
Weighted Average	60%	60%	60%

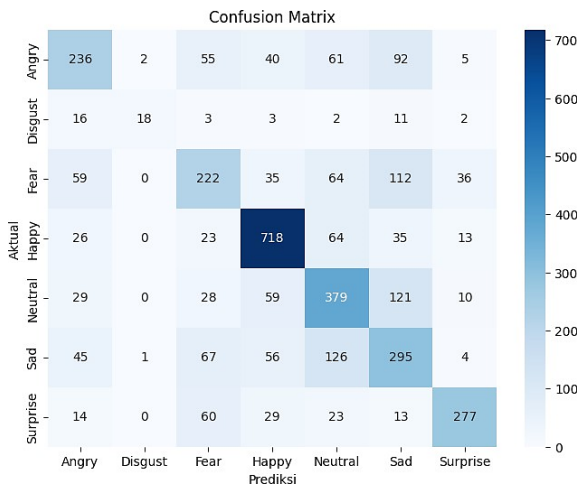


Figure 2. Confusion Matrix

The confusion matrix illustrates the distribution of correct and incorrect predictions across the seven facial expression classes. The model demonstrates relatively stronger classification performance for Happy and Surprise, while greater confusion occurs among classes with visually similar facial characteristics.

3.1. Model Training Result

Figures 3 and 4 show the evolution of accuracy during CNN model training. The accuracy values gradually increase with each epoch until they reach a steady state. This indicates that the model is capable of learning facial expression patterns well on the FER2013 dataset.

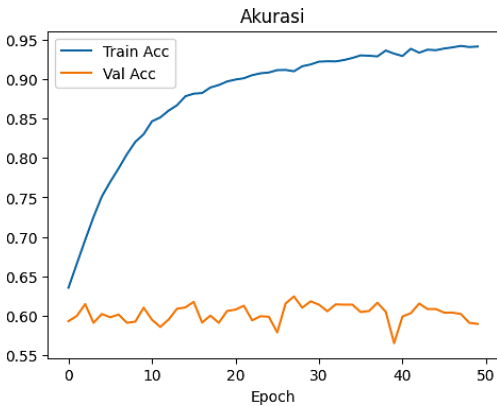


Figure 3.1 Graphic Accuracy Model CNN

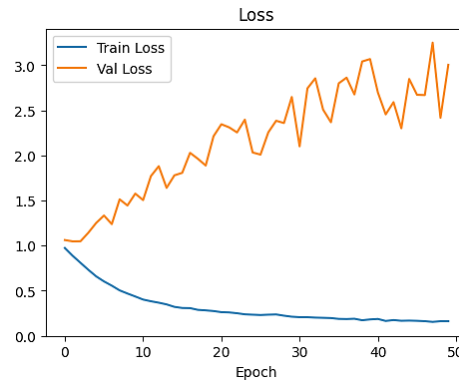


Figure 4. CNN Model Loss Graph

The continuous increase in accuracy scores in each epoch indicates that the learning process is progressing well and that the model parameters have been updated in a more optimal direction. The accuracy graph, which begins to stabilize at the end of each epoch, indicates that the model has nearly reached a point of convergence, where additional epochs do not provide significant performance improvements. This situation indicates that the CNN architecture successfully extracted important features from facial images to aid in the emotional expression classification process.

Overall, the accuracy and loss graphs show that the CNN model trained on the FER2013 dataset has undergone a consistent learning process and reached convergence. The continuous increase in training accuracy values accompanied by a decrease in training loss values indicates that the model has succeeded in gradually understanding facial feature patterns. Although there are slight differences between the training and validation graphs, the model does not show signs of severe overfitting, so the results from this training can be considered quite representative for the next evaluation stage.

3.2. Model Evaluation

Based on the evaluation results in Table 3, the CNN model achieved 60% accuracy on the test data. The Happy category performed best with an F1-score of 0.79, followed by the Surprise category with an F1-score of 0.73. On the other hand, the Fear and Sad categories received lower F1-scores of 0.45 and 0.46, respectively. This indicates that the model still struggles to distinguish among several types of emotions with similar visual properties.

The differences in performance across emotion categories seen in these evaluation results reflect an inherent challenge in the FER2013 dataset, where the uneven distribution of data across categories impacts the model's ability to generalize. For example, the Disgust category has significantly fewer samples than other categories, such as Happy, resulting in a high precision score (0.86), but a very low recall score (0.33). This suggests that the model tends to overestimate minority categories and that techniques such as oversampling or class weighting may be necessary to address the class imbalance and improve overall performance, as shown in Table 4.

Table 4. Classification Record

Emotion Class	Precision	Recall	F1-Score
Angry	0.56	0.48	0.52
Disgust	0.86	0.33	0.47
Fear	0.48	0.42	0.45
Happy	0.76	0.82	0.79
Neutral	0.53	0.61	0.56
Sad	0.43	0.50	0.46
Surprise	0.80	0.67	0.73
Macro Average	0.63	0.54	0.57
Weighted Average	0.60	0.60	0.60

The results of the assessment indicate that the CNN model has a fairly good ability to detect facial expressions, although performance varies across emotion categories. The Happy and Surprise categories achieved fairly high precision, recall, and F1-score because these two expressions have clearer, more easily recognized visual characteristics, such as a smile in a happy expression and more open eyes and mouth in a surprised expression. Meanwhile, the Angry, Fear, Sad, and Neutral categories showed lower performance due to similar facial expression patterns, making it difficult for the model to distinguish among these classes.

Compared with several previous studies using the FER2013 dataset, the 60% accuracy achieved in this study is still considered good for a simple CNN with a small number of convolutional layers. Although this

model is not yet able to achieve results comparable to more complex architectures such as ResNet, EfficientNet, or Vision Transformer, which typically achieve accuracies above 70%, this finding indicates that the CNN architecture used has successfully recognized important features from facial images. Therefore, the developed model can serve as a basis for further research to improve performance through architectural modifications, hyperparameter tuning, or transfer learning.

3.3. Analysis Confusion Matrix

Based on the confusion matrix, the Happy and Surprise categories had better classification rates than the other categories. Conversely, the Fear and Sad categories showed higher misclassification rates due to their relatively similar visual characteristics to those of the other emotion categories (Figure 6).

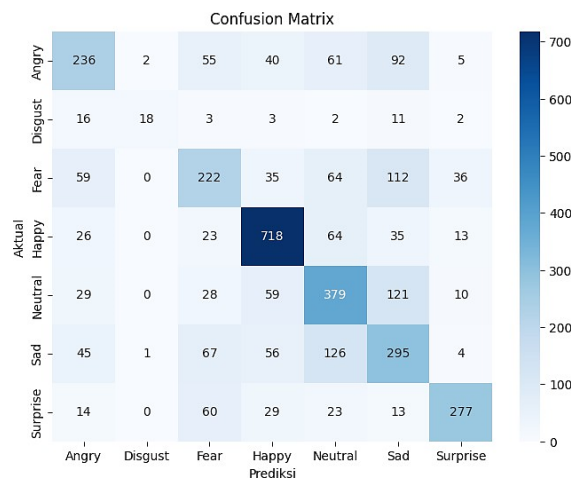


Figure 5. Confusion Matrix

The results of this confusion matrix analysis also show a pattern of misclassification errors that aligns with previous research on the FER2013 dataset. Facial expressions that exhibit overlapping facial muscle movements, such as Fear and Surprise, which both involve wide eye opening, and Sadness and Neutral, which tend to show a flat expression, are the main factors causing misclassification. These results emphasize the importance of exploring more complex CNN architectures or implementing attention mechanisms to enable models to recognize more subtle features and distinguish between similar emotions.

In addition to displaying classification effectiveness for each emotion category, the confusion matrix provides insight into the distribution of prediction errors during the testing phase. This data is crucial for identifying model weaknesses and formulating appropriate improvement strategies. The few remaining misclassification errors indicate that the model's features are not yet fully effective in distinguishing expressions with similar visual characteristics. Therefore, future research could focus on enhancing feature extraction by adding convolutional layers, applying transfer learning, or using attention mechanisms to improve classification accuracy for highly similar categories.

3.4. Overall Model Performance

The experimental results show that the proposed Convolutional Neural Network (CNN) achieved 60% accuracy on the FER2013 test set. The weighted precision, recall, and F1-score also reached 60%, indicating that the model was able to classify facial expressions across the seven emotion categories with moderate performance. These results demonstrate that the CNN architecture used in this study learned discriminative visual features from 48 × 48 grayscale facial images. However, the performance also indicates that facial expression classification remains challenging with a relatively simple CNN architecture, particularly for emotion categories that share visually similar facial characteristics.

The differences in the performance of individual classes provide additional information that cannot be captured by accuracy alone. Therefore, the class-level precision, recall, and F1-score are important for understanding how the model performs across different emotional expressions.

3.5. Class-Level Classification Performance

In contrast, Fear and sadness showed lower classification performance. Of the 528 Fear samples, 222 were correctly classified, while 112 samples were incorrectly classified as Sad, 64 as Neutral, and 59 as Angry. Similarly, among 594 Sad samples, 295 were correctly classified, while 126 were predicted as Neutral, 67 as Fear, and 56 as Happy. These results indicate that the main classification errors occurred among emotion categories with relatively similar visual characteristics. In particular, the substantial confusion between Fear

and sadness suggests that the learned visual features were insufficiently discriminative to clearly distinguish these two expressions. The Angry class also showed considerable confusion with other categories. Of the 491 Angry samples, 236 were correctly classified, while 92 were classified as Sad, 61 as Neutral, and 55 as Fear. This pattern further indicates that the model had difficulty distinguishing Angry from expressions that overlap in facial features. The Disgust class produced a different pattern. Only 18 of 55 Disgust samples were correctly classified, resulting in a recall of 32.73%. However, its precision reached 85.71%. This indicates that although most of the model's Disgust predictions were correct, a large proportion of actual Disgust samples were classified into other emotion categories. The relatively small number of Disgust samples compared with the other classes may contribute to this behavior.

3.6. Analysis of the Confusion Matrix

The confusion matrix provides a more detailed explanation of the model's overall accuracy, which is approximately 60%. A total of 2,145 out of 3,589 test samples were classified correctly. The matrix shows that the model performed most consistently for Happy, while substantial misclassification occurred among Fear, Sad, Angry, and Neutral. The largest individual misclassification occurred in the Fear-to-Sad category, where 112 Fear samples were misclassified as Sad. Another substantial error occurred for Sad samples classified as Neutral, with 126 cases. Angry samples were also frequently classified as Sad, with 92 cases. These patterns suggest that the model's errors were not randomly distributed but were concentrated among several visually similar emotion categories.

The results may be associated with the characteristics of the FER2013 dataset, which consists of low-resolution 48×48 grayscale facial images. Such representations provide less detailed visual information than high-resolution color images, potentially making subtle differences in facial muscle movements more difficult to distinguish. Furthermore, the unequal sample counts across emotion categories may limit the model's ability to learn robust representations for minority classes, particularly Disgust. The use of data augmentation through rotation, shifting, zooming, and horizontal flipping was intended to increase the variability of training samples. Nevertheless, the confusion matrix indicates that these techniques alone were not sufficient to eliminate class-level classification errors. More targeted approaches, such as class weighting, oversampling, or feature-learning architectures with attention mechanisms, could therefore be investigated in future research.

3.7. Implications of the Results

The confusion matrix demonstrates why accuracy alone is insufficient to describe the performance of the facial expression classification model. Although the overall accuracy reached approximately 60%, the recall values varied substantially across classes, from 32.73% for Disgust to 81.68% for Happy. This variation indicates that the model does not provide uniform recognition performance across all emotion categories.

Therefore, the combination of accuracy, precision, recall, F1-score, and confusion matrix provides a more comprehensive evaluation of the proposed CNN. In particular, the confusion matrix identifies specific error patterns that cannot be observed from the overall accuracy alone. These findings can be used as a basis for improving the model in subsequent research.

4. CONCLUSION

This study successfully used a Convolutional Neural Network (CNN) to classify facial expressions on the FER2013 dataset, which contains seven emotion classes: Anger, Disgust, Fear, Happiness, Neutral, Sadness, and Surprise. The test results showed that the model achieved 60% accuracy, with 60% weighted precision, 60% weighted recall, and 60% weighted F1-score. Based on the assessment results, the Happy category achieved the best performance with an F1-score of 0.79, followed by the Surprised category with an F1-score of 0.73. Meanwhile, the Fear and Sad categories showed poorer performance because they share visual characteristics with other emotion categories, making them more difficult to group accurately.

The results of this study show that CNN can automatically detect facial expressions with satisfactory results on the FER2013 dataset. For future research, model efficiency can be improved by using more complex deep learning structures such as ResNet, EfficientNet, or Vision Transformer, as well as through hyperparameter optimization and more varied data augmentation methods.

In addition to demonstrating CNNs' ability to classify facial expressions, this study underscores the crucial role of preprocessing and data augmentation in improving the model's learning quality. Data normalization and image augmentation help the model acquire a more diverse feature representation, which in turn improves its ability to generalize to previously unseen data. Therefore, the combination of an appropriate CNN architecture, efficient training parameters, and robust data processing techniques is key to creating an effective facial expression recognition system that can be further developed for various AI-based applications.

REFERENCES

- [1] I. J. Goodfellow et al., Challenges in Representation Learning: A Report on Three Machine Learning Contests, 2013.

- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [3] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701–1708.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [5] C. Pramerdorfer and M. Kampel, "Facial Expression Recognition Using Convolutional Neural Networks: State of the Art," 2016.
- [6] Y. Khairuddin and Z. Chen, "Facial Emotion Recognition: State of the Art Performance on FER2013," 2021.
- [7] O. C. Oguine, C. N. Nwankwo, and E. N. Onwuka, "Hybrid Facial Expression Recognition (FER2013) Model for Real-Time Emotion Classification and Prediction," 2022.
- [8] M. Al-Shabi, W. P. Cheah, and T. Connie, "Facial Expression Recognition Using a Hybrid CNN-SIFT Aggregator," 2016.
- [9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014.
- [12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
- [13] M. Pantic and L. J. M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.
- [14] M. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The First Facial Expression Recognition and Analysis Challenge," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2011, pp. 921–926.
- [15] M. H. Yap, N. Ugail, and R. Zwigelaar, "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets," *Information*, vol. 15, no. 3, 2024.
- [16] S. Minaee, M. Minaei, A. Abdolrashidi, and Y. Wang, "Deep Facial Expression Recognition: A Survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, 2022.
- [17] B. B. Yap et al., "A Review of Facial Emotion Recognition Approaches for Human–Robot Interaction," *Robotics*, vol. 12, no. 3, p. 72, 2023.
- [18] M. Bie, H. Xu, Y. Gao, K. Song, and X. Che, "Swin-FER: Swin Transformer for Facial Expression Recognition," *Applied Sciences*, vol. 14, no. 14, 2024.
- [19] Y. Wang et al., "A Survey on Facial Expression Recognition of Static and Dynamic Emotions," 2024.
- [20] C. L. Kim and B.-G. Kim, "Few-shot Learning for Facial Expression Recognition: A Comprehensive Survey," *Journal of Real-Time Image Processing*, 2023.
- [21] H. Li et al., "Knowledge-Enhanced Facial Expression Recognition with Emotional-to-Neutral Transformation," 2024.
- [22] S. Li et al., "Attention Mechanism-Based CNN for Facial Expression Recognition," *Comput. Electr. Eng.*, vol. 106, p. 108541, 2023.
- [23] Z. Wang et al., "EfficientNet-Based Facial Expression Recognition Using FER2013 Dataset," *IEEE Access*, vol. 11, pp. 50012–50024, 2023.
- [24] C. Fan et al., "Label Distribution Learning for Facial Expression Recognition with Class-Imbalance Problem," *Pattern Recognit.*, vol. 141, p. 109607, 2023.
- [25] A. Dosovitskiy et al., "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," in *Proc. ICLR*, 2021.
- [26] T. Zhao et al., "Vision Transformer for Facial Expression Recognition," *Sensors*, vol. 23, no. 4, p. 2176, 2023.
- [27] P. Liu et al., "Facial Expression Recognition via Disentangled Representation Learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 2, pp. 1145–1157, 2024.
- [28] Y. Chen et al., "Multimodal Facial Expression Recognition Using CNN and Temporal Attention," *Neurocomputing*, vol. 575, p. 127296, 2024.