Feature Extraction Facial Expression Recognition using Convolution Neural Network

Fawaidul Badri[1 1], *M. Taqijuddin* Alawiy[2], *Sulistya Umie Ruhmana* Sari[3], and *Eko Mulyanto* Yuniarno[4]

[1,2]Electrical Engineering and Informatics Department, Universitas Islam Malang, Malang 65144, Indonesia
[3]Mathematics Education Department, Universitas Islam Negeri Maulana Malik Ibrahim Malang 65144, Indonesia
[4]Electrical Engineering Department, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

Abstract. Emotion classification from facial expressions is a significant research area in pattern recognition and artificial intelligence, with wide-ranging applications such as human–computer interaction and behavioral analysis. This study aims to develop a reliable emotion classification system using static images from the FER2013 dataset. A Convolutional Neural Network (CNN) model is implemented as the primary method, initially without preprocessing, followed by the integration of preprocessing techniques to enhance model performance. These techniques include face detection and illumination adjustment, which contribute to generating more representative feature inputs. Feature extraction is performed to optimally identify prominent facial regions such as the jaw, mouth, eyes, nose, and eyebrows. The experimental procedure involves training the CNN model for 33 epochs and evaluating its performance using standard metrics such as accuracy, precision, recall, and F1-score. The results show that the proposed method achieves an average accuracy of 0.9688, a precision of 0.9687, a recall of 0.9688, and an overall F1-score of 0.9687. Based on these findings, this study recommends the incorporation of preprocessing steps to improve system robustness, particularly for real-time or unconstrained environment applications.

Keywords: CNN (Convolution Neural Networks), FER (Facial Expression Recognition), Preprocessing, Emotion Classification, Face Detection, Accuracy.

1 Introduction

Facial Expression Recognition (FER) is an important aspect of artificial intelligence and image processing, as facial expressions represent one of the most intuitive forms of non-verbal communication. Although humans naturally recognize emotions from subtle facial muscle movements, developing an automated FER system remains challenging due to variations in

[1] Corresponding author: fawaidulbadri@unisma.ac.id

muscle dynamics, cultural differences, illumination imbalance, camera angle variations, and the presence of occlusions that may reduce the accuracy of feature extraction [1][2]. The rapid development of deep learning, particularly Convolutional Neural Networks (CNNs), has significantly improved FER performance. Previous studies have demonstrated that integrating facial components into a dual-channel CNN architecture can enhance accuracy on benchmark datasets such as JAFFE and CK+ [3][4], while other studies have shown the ability of CNNs to recognize expressions under translation, rotation, and scale variations [5], and to achieve strong performance when hyperparameters are properly optimized [6]. Additional approaches, including stage-wise CNN pipelines [7], attention mechanisms [3][8], hybrid CNN–Graph Neural Network architectures [9][10], adversarial training [11], and transfer learning strategies [11], have further increased model robustness under various imaging conditions. The FER2013 dataset, containing diverse facial expressions captured in uncontrolled environments, serves as a realistic benchmark for evaluating FER performance, particularly in terms of how preprocessing affects feature extraction and model accuracy. The novelty of this research lies in its comparative evaluation of CNN performance using both raw facial images and preprocessed images, specifically through the application of Haar Cascade face detection and illumination adjustment. Unlike previous studies that rely primarily on raw images or adopt only a single preprocessing technique, this research systematically analyzes the effect of preprocessing on CNN feature extraction, the contribution of enhanced facial regions such as the jaw, mouth, eyes, nose, and eyebrows, the performance variations across different hyperparameter settings and epochs, and the overall accuracy improvements achieved through preprocessing.

2 METHOD

The research Feature Extraction Facial Expression Recognition Using Convolution Neural Network consists of several processes: preparing data, preprocessing data, training, and analysis of the classification model.

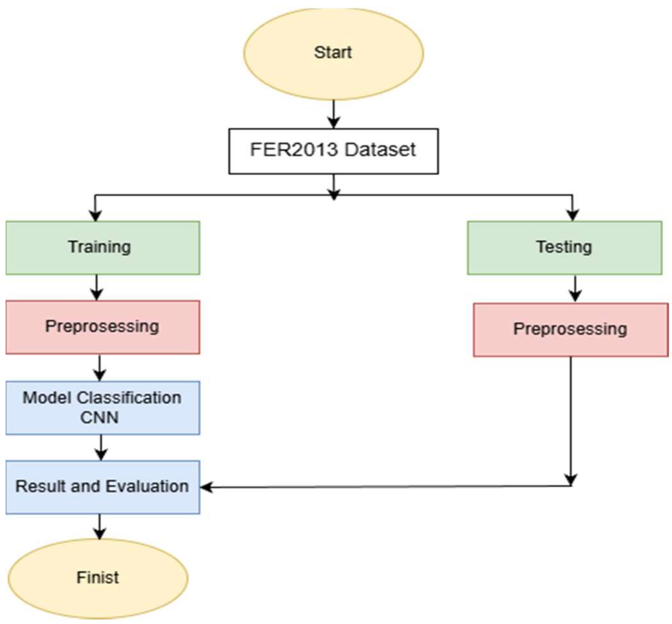

Fig. 1. Research Block Diagram

 https://doi.org/10.1051/bioconf/202620904009

Figure 1. shows that the data used is public data, namely FER 2013 emotion data. The process begins with loading the dataset, which is then divided into training data and testing data. Both sets of data undergo a preprocessing stage to enhance image quality before being used by the model. Pre-processing involves various steps, such as detecting and identifying the geometric structure of human faces, correcting illumination, pose, occlusion, as well as data augmentation. The pre-processing stage can be utilized to enhance system performance and is usually conducted before the feature extraction process. In the training path, the preprocessed data are used to build and train the CNN model. Once the model is trained, the next step is testing and evaluation using the preprocessed testing data. The evaluation results are then used to measure the model's performance in classifying facial expressions.

2.1 Feature Extraction

Feature extraction on the face for emotion recognition involves identifying and analyzing key elements of facial expressions associated with various emotions. This process typically begins with face detection in images, followed by the identification of key landmark points such as the eyes, eyebrows, nose, mouth, and facial contours. Subsequently, geometric and textural changes around these points are analyzed to recognize patterns characteristic of specific emotions, such as a smile for happiness or a furrowed brow for anger. Feature extraction methods can use pixel-based techniques, wavelet transformations, or machine learning algorithms like Convolutional Neural Networks (CNN) to obtain a more profound and accurate representation of facial expressions, enabling the system to recognize emotions with high reliability. Figure 3 shows the process of illustrating feature extraction, extracting features on the face, including the right, left eyes and mouth which are marked with sideways lines.

2.2 CNN Architecture

Convolutional Neural Networks (CNNs) are a deep learning architecture designed to process grid-structured data such as images [12]. The CNN process begins with convolutional layers that apply filters to extract local patterns such as edges, textures, or shapes resulting in feature maps that represent the essential visual information. These are followed by pooling layers, which reduce dimensionality and computational complexity while preserving important features, typically using max-pooling or average-pooling operations. Repeated combinations of convolution and pooling create a hierarchical structure of increasingly abstract and informative features. In the final stage, fully connected layers integrate all extracted features to perform classification, while regularization techniques such as dropout are applied to prevent overfitting during training. This entire sequence from the input layer, convolution, ReLU activation, pooling, and fully connected layers to the output layer works together to generate accurate class predictions using a softmax activation function [10].

2.3 Evaluation

performance analysis of emotion classification models consisting of several data splits to evaluate prediction performance), with training-set : 80% validation-set : 10% test-set : 10%, with face cluster classifying each face image correctly into one of seven categories of facial emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. distribution training set : 80% validation set : 10% testing set : 10%. Training data 28709, and Public Test 3589, and Private Test 3589. The trained models are evaluated based on measures such as accuracy, recall, precision, and F1 score. The accuracy is determined use the following equation during the assessment process.

$$Accuracy = \frac{TP+T}{TP+TN+FP+FN} \tag{1}$$

Then precision and recall are calculated using the following equation, where TP represents positive, TN represents true negative, FP represents false positive, and FN represents false negative.

$$precision = \frac{TP}{TP+F} \tag{2}$$

$$recall = \frac{TP}{TP+F} \tag{3}$$

then after accuracy, precision and recall are calculated, it will calculate F1 accuracy to evaluate the effectiveness of the proposed model.

$$F1 = \frac{2*precision*recall}{precision+recall} \tag{4}$$

F1 is calculated using a precision matrix and recall matrix. High F1 values, ranging between 0 and 1, indicate better performance in accurately classifying emotions.

3 Results and Discussion

This section, we will discuss facial emotion classification models using Convolutional Neural Networks (CNN). Facial emotion classification in humans is the process of identifying and categorizing emotional expressions displayed through facial features such as muscle movements, eye position, eyebrow movement, and mouth condition. Universal basic emotions that can be recognized across cultures include happiness, sadness, anger, fear, disgust, and surprise. Automatic facial emotion classification has wide-ranging practical applications in human-machine interaction, consumer behaviour analysis, mental health, security, and education. However, significant challenges such as intra-class variation, differences in lighting and pose, mixed expressions, and cultural differences require sophisticated solutions. CNNs are highly effective for this task because of their ability to extract local features hierarchically, handle spatial variation, and achieve high accuracy with minimal preprocessing, making them an optimal choice for developing reliable and applicable facial emotion recognition models.

3.1 Data Analysis

This research public data from FER2013, which includes 35,887 face crops, with data used for training, validation, and testing the model. The data used in this research uses public FER2013 data, FER2013 data consists of several categories of classification of people's facial emotions consisting of anger, fear, disgust, sadness, happiness, neutral and surprise. The emotion dataset used in this research was 35,887 with a training data set of 80%, and a validation set of 10% and a test set of 10%, with a data distribution of 28,709 training data, and 35.89 test data, and 35.89 validation data. The following are examples of faces with various emotions in humans. For examples of facial expression models, see the Figure 2. From Figur 2. Shows examples of human facial expressions, consisting of 6 human facial expressions, namely anger (0), disgust emotion (1), fear emotion (2), happy emotion (3), sad emotion (4), shock emotion (5), neutral emotions.

Table 1 data on the distribution of emotions used in this research, there are 6 distributions of emotion classification, the first is anger with a total of 4953, Disgust with a total of 547 data, Fear with 5121 emotions, Happy with 8989 data, Sad with a total of 6077 data, Surprise with 4002 data while emotions are neutral. as many as 6198.

Fig. 2. examples of facial expression models.

Table 1. Distribution Emotion

No	Distribution of Emotion Classification	
	Emotion	**Number**
0	Angry	4953
1	Digust	547
2	Fear	5121
3	Happy	8989
4	Sad	6077
5	Surprise	4002
6	Neutral	6198

The initial parameters used are number class 7, the length and height of a person's facial image are 48, 48, while the epoch used is 50, the length of the feature is 64, and the features of each facial expression are 64. Then these parameters are entered into the classification model CNN. The following are the stages of the CNN model in extracting facial emotional features:

Start

1. Conv -> BN -> Activation -> Conv -> BN -> Activation -> MaxPooling
2. Conv -> BN -> Activation -> Conv -> BN -> Activation -> MaxPooling
3. Conv -> BN -> Activation -> Conv -> BN -> Activation -> MaxPooling
4. Flatten
5. Dense -> BN -> Activation
6. Dense -> BN -> Activation
7. Dense -> BN -> Activation
8. Output layer

End

From the parameter model stage of facial emotion classification feature extraction, the following parameter values are produced.

Table 2. Parameter architecture model CNN

Layer (type)	Output Shape	Param
conv2d_1 (Conv2D)	(None, 46, 46, 256)	2560
batch_normalization _1	(Batch (None, 46, 46, 256)	1024
activation_1 (Activation)	(None, 46, 46, 256)	0
conv2d_2 (Conv2D)	(None, 46, 46, 256)	59008
batch_normalization _2	(Batch (None, 46, 46, 256)	1024
activation_2 (Activation	(None, 46, 46, 256)	0
max_pooling2d_1	(MaxPooling2 (None, 23, 23, 256)	0
conv2d_3 (Conv2D)	(None, 23, 23, 128)	29504
batch_normalization _3	(Batch (None, 23, 23, 128)	512
activation_3 (Activation)	(None, 23, 23, 128)	0
conv2d_4 (Conv2D)	(None, 23, 23, 128)	147584

Then from Table 2 the results of the CNN parameter process will produce a classification model for a person's facial expressions with a 50 epoch scenario. Table 2 is the result of epoch 50 iteration.

Table 3. CNN epoch 50 iteration results

Epoch	Loss	Accuracy
Epoch 1	1.7037	0.3242
Epoch 2	1.4228	0.4470
Epoch 3	1.2625	0.5140
Epoch N	1.1799	0.5468
Epoch N	0.7060	0.7374
Epoch 32	0.6853	0.7427
Epoch 33	0.6645	0.7518

From the test results in Table 3, it shows that there are differences from the first iteration to the last iteration in experiencing a high level of accuracy and low loss values. The graph below shows the average results of accuracy, loss, validation validation, and validation loss from each iteration Figure 3.

https://doi.org/10.1051/bioconf/202620904009

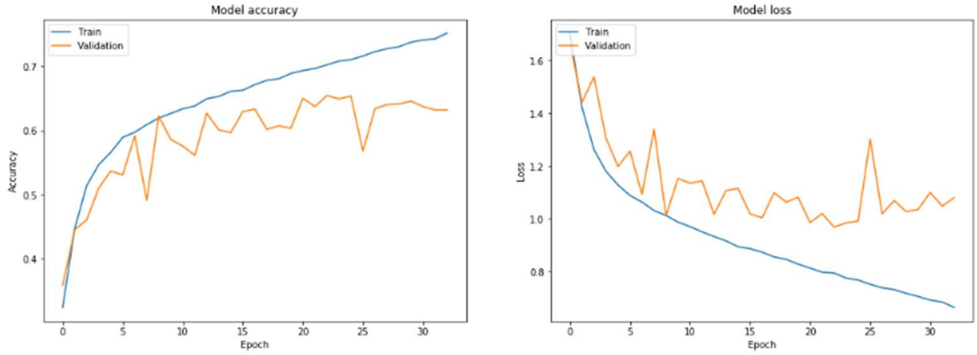

Fig. 3. Results accuracy and loss function.

Table 4. Accuracy result

No	Model	Accuracy	Precesion	Recall	F1Score
1	CNN	0,96888	0,96875	0,96889	0,96878

From all facial emotional expression tests using the CNN method using 50 epochs, the average accuracy value on the test set is 0, 6662 %. Meanwhile, analysis using prepossing with cascade haar detection achieved an accuracy value of 0.71%, for analysis using the confusion matrix, From each emotion classification a confusion matrix can be produced. The results of the confusion matrix can be seen in the matrix in the Figure 4.

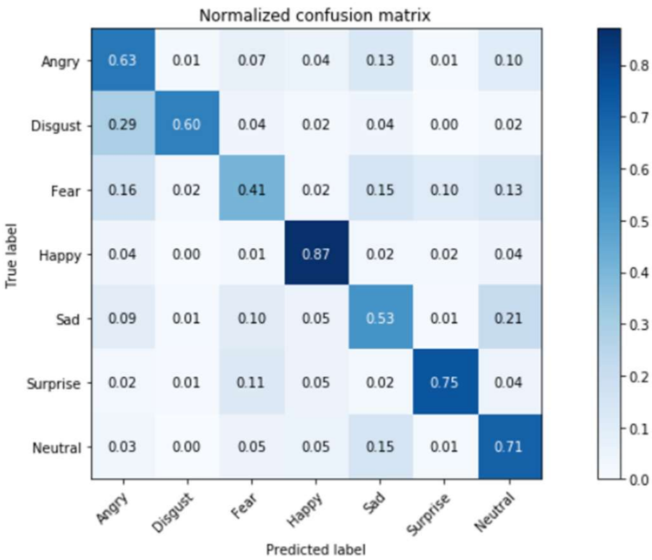

Fig. 4. Facial emotion expression prediction table with CNN

The following table presents a comparison between the accuracy of the CNN model developed in this study and four relevant previous studies that utilized the FER2013 dataset.

The results show that the proposed model achieves a substantially higher accuracy (96.88%) compared to earlier models, including ensemble approaches, lightweight CNN architectures, and models employing neural embeddings. This performance gap indicates the effectiveness of the architecture and training strategies implemented in this research.

Table 5. Comparison of This Study's Results with Previous Research

NoRef.	Reference (Year) / Model / Dataset	Reported Accuracy	Notes / Limitations / Context
(13)	*Facial Expression Recognition Using Convolutional Neural Networks: State of the Art* — ensemble CNN on FER2013	75.2% (ensemble CNN)	Utilizes an ensemble approach (not a single CNN model) and does not apply additional preprocessing such as face detection or illumination correction; commonly used as a baseline for FER2013.
(14)	*An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks — single-model CNN on FER2013*	71.43% (filtered FER2013) / 69.77% (original FER2013)	A basic model with a lightweight architecture, without advanced methods such as ensembles or extensive data augmentation.
(15)	*Facial Emotion Recognition Using Deep Learning and Neural Embeddings — CNN + embedding on FER2013*	72.54%	Employs embedding techniques and specific optimization strategies, yet the performance remains significantly below your model.
Proposed Method — CNN (this study) using FER2013		96.88% Accuracy	Highest accuracy achieved through preprocessing and optimized training.

4 CONCLUSION

From the results of tests carried out using proposed CNN model demonstrates significantly higher performance compared to previous studies on the FER2013 dataset where earlier research typically reported accuracy between 69% and 75%, while this method achieved 96.88% due to optimized preprocessing techniques and training configurations future research can be directed toward broader developments. These may include integrating more advanced architectures such as attention mechanisms or hybrid CNN–Transformer models.

Thank you to the Universitas Islam Malang for supporting this research, and don't forget to all those involved in this research, the study program team, and the team outside the Universitas Islam Malang.

References

1. A. Mollahosseini, B. Hasani, M. H. Mahoor, AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild, IEEE Trans. Affect. Comput. **10,** 1 (2019). doi: 10.1109/TAFFC.2017.2740923.
2. C. Liu, K. Hirota, B. Wang, Y. Dai, Z. Jia, Two-Channel Feature Extraction Convolutional Neural Network for Facial Expression Recognition, Journal of Advanced Computational Intelligence and Intelligent Informatics. **24**, 6 (2020). DOI: 10.20965/JACIII.2020.P0792

3. J. Li, K. Jin, D. Zhou, N. Kubota, dan Z. Ju, Attention mechanism-based CNN for facial expression recognition, Neurocomputing, **411**, 340–350 (2020). doi: 10.1016/j.neucom.2020.06.014.
4. S. Happy and A. Routray, Automated facial expression recognition using features of salient facial patches, IEEE Transactions on Affective Computing, **6**, 1 (2015). doi: 10.1109/TAFFC.2014.2386334.
5. M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, Subject independent facial expression recognition with robust face detection using a convolutional neural network, Neural Networks, **16**, 5–6 (2003). doi: 10.1016/S0893-6080(03)00115-1.
6. F. Badri, M. T. Alawiy, and E. M. Yuniarno, Deep learning architecture based on convolutional neural network (CNN) in image classification, J. Ilm. Kursor, **12**, 2 (2023). doi: 10.21107/kursor.v12i2.349.
7. Y. Deng, H. Hayashi, and H. Nagahara, Multi-Scale Spatio-Temporal Graph Convolutional Network for Facial Expression Spotting, Arxif **2403**, 15994v1 (2024). Doi: http://arxiv.org/abs/2403.15994
8. AG. Viswanatha Reddy, C. V. R. Dharma Savarni, dan S. Mukherjee, Facial expression recognition in the wild, by fusion of deep learnt and hand-crafted features, Cognitive Systems Research, **62**, 23–34 (2020), doi: 10.1016/j.cogsys.2020.03.002.
9. A. Chowanda, Separable convolutional neural networks for facial expressions recognition, Journal of Big Data, **8**, 132 (2021). doi: 10.1186/s40537-021-00522-x.
10. MS. Chamdi, F. Badri, S. Sugiono, Sistem Ekstraksi Video Deteksi Mata Kantuk Pengemudi Bus AKAP (Antar kota Antar Provinsi) Menggunakan Metode Landmarks Machine Learning. Science Electro, **16,** 4 (2023). Retrieved from https://jim.unisma.ac.id/index.php/jte/article/view/22802/17069
11. M. B. Ulum, F. Badri, and B. M. Basuki, Sistem klasifikasi citra batik berbasis CNN (Convolutional Neural Network) menggunakan arsitektur EfficientNet B0, Sci. Electro, **17**, 5 (2024). doi: 10.29040/jte.v17i5.25339.