Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 10, No. 2, June 2024, pp. 279-289
ISSN: 2338-3070, DOI: 10.26555/jiteki.v10i2.28863

Development of Convolutional Neural Network Models to Improve Facial Expression Recognition Accuracy

Fatimatuzzahra, Lindawati, Sopian Soim
Department of Electrical Engineering, Telecommunication Engineering Study Program, Sriwijaya State Polytechnic

ARTICLE INFO

Article history:

Received May 22, 2024
Revised June 22, 2024
Published July 01, 2024

Keywords:

Convolutional Neural Network;
Deep Learning;
Facial Expression;
Classification

ABSTRACT

Advancements in information and computer technology, particularly in machine learning, have significantly alleviated human tasks. One of the current primary focuses is facial expression recognition using deep learning methods such as Convolutional Neural Network (CNN). Complex models like CNNs often encounter issues such as gradient vanishing and overfitting. This study aims to enhance the accuracy of CNN models in facial expression recognition by incorporating additional convolutional layers, dropout layers, and optimizing hyperparameters using Grid Search. The research utilizes the FER2013 public dataset sourced from the Kaggle website, trained and evaluated using CNN models, hyperparameter tuning, and downsampling methods. FER2013 comprises thousands of facial images representing various human expressions, with a specific focus on four facial expression categories (angry, happy, neutral, and sad). Through the addition of convolutional and dropout layers, as well as hyperparameter optimization, the developed model demonstrates a significant improvement in accuracy. Findings reveal that the refined CNN model achieves a highest accuracy of 98.89%, with testing accuracy at 89%, precision 78%, recall 78%, and F1-score 78%. This research contributes by enhancing facial expression recognition accuracy through optimized CNN models and providing a framework beneficial for the social-emotional development of children with special needs and aiding in the detection of mental health conditions. Additionally, it identifies avenues for future research, including exploring advanced data augmentation techniques and integrating multimodal information. Furthermore, this study paves the way for applications across diverse fields like human-computer interaction and mental health diagnostics.

Corresponding Author:

Lindawati, Department of Electrical Engineering, Telecommunication Engineering Study Program, Sriwijaya State Polytechnic, South Sumatera, Indonesia
Email: lindawati@polsri.ac.id

1. INTRODUCTION

Facial Expression Recognition (FER) has become a crucial aspect in various application contexts, including human-computer interaction and medical diagnostics [1]-[7]. This technology also has a significant impact in fields such as security, education, medical rehabilitation, and driver safety [8-13], as well as improving the effectiveness of human communication by enabling service providers to deliver more accurate responses [14-18]. However, the main challenge in generalizing FER across various datasets and real-world scenarios is highly complex [19]. The variability of facial expressions between individuals and situations requires systems to handle subtle and complex variations.

Additionally, the diversity of datasets in terms of size, resolution, and demographic representation adds to the difficulty of training reliable models without compromising accuracy. Addressing potential biases in expression types and population groups necessitates a careful approach to ensure the fairness and precision of the models [20-22].

The importance of facial expressions in human communication has become a significantly growing subject of research [23]. Various methods, especially those involving convolution processes on images, have been developed for facial expression recognition using deep learning, such as CNN [24].

CNN have proven effective in automatically extracting features from images, enabling the training of models to recognize facial emotions with high accuracy [25-28]. Although the use of Convolutional Neural Network (CNN) in FER has proven effective, current methodologies often overlook specific innovations crucial for improving accuracy.

The International Conference on Machine Learning (ICML) in 2013 introduced Facial Expression Recognition (FER) as a new challenge in the field of emotion recognition. FER provides a diverse dataset with complex natural challenges. The use of this dataset has tested the ability of CNN models to recognize facial emotions [29].

Further research has optimized CNN models using methods such as Multi Branch and Convolutional Block Attention Module, achieving significant accuracy for various datasets, such as FER2013 with 69.49%, FERPLUS with 84.63%, and CK+ with 99.39% [18]. Although Multi Branch and Convolutional Block Attention Module are believed to enhance the performance of CNN, the risk of increasing model complexity may reduce recognition accuracy.

In other research, dataset generalization was achieved using model architectures such as MobileNet and ResNet-18, which produced satisfactory accuracy for various types of datasets, including Thermal and RGB [30]. The use of CNN architectures like Multiple Branch Cross-Connected on datasets such as FER2013, CK, FER, and RAF has also shown significant progress in facial expression recognition, despite facing challenges related to architectural complexity and the limitations of available training datasets [31]. Although models like MobileNet and ResNet-18 have shown satisfactory results across various datasets, including thermal and RGB variations, challenges like overfitting remain an issue, especially with small datasets like FER2013.

Based on previous research, this indicates that the CNN model for facial expression recognition with the FER2013 dataset still needs improvement. Current complex CNN models still face new challenges, including gradient vanishing and overfitting [32-38]. Therefore, improvements to CNN architecture are necessary to enhance the accuracy of CNN models in facial expression recognition. Some improvements, such as adding convolutional layers, can be used to extract information from input images in CNN [39-42], helping to reduce the potential loss of feature information from the input [43].

Although the proper hyperparameters are key to building CNN model architecture, especially in the convolution layers, clear guidelines for determining hyperparameters are still not fully available and are often determined through experimentation or based on previous researchers' experience [44-45]. One approach to improving CNN performance during the learning phase is hyperparameter optimization [46-47]. The grid search method is an effective approach to identifying optimal hyperparameter combinations by systematically evaluating various hyperparameter value combinations from a grid.

In addition to improvements in CNN models, preparing the dataset to reduce overfitting is also necessary. Research [48] conducted by Ding-Xuan Zhou explains that high convolution usage can cause models to struggle in analyzing networks with increased width. To address this, downsampling should be used to reduce the network width resulting from convolution. Downsampling, which involves data cleaning techniques by removing less useful samples in the majority class rather than incrementally adding informative samples [49].

This research aims to refine CNN architecture by optimizing convolutional layers, hyperparameters, and data preprocessing to improve facial expression detection accuracy. The role of convolution in extracting features from input images is crucial to minimizing feature information loss during the early stages of data processing [31-32]. This research contributes by enhancing facial expression recognition accuracy through optimized CNN models and providing a framework beneficial for the social-emotional development of children with special needs and aiding in the detection of mental health conditions.

2. METHODS

In Fig. 1, the methodology used in this research is depicted. The research begins with input data, downsampling the dataset, data preprocessing, model training, model testing, detection, and model evaluation.

Vol. 10, No. 2, June 2024, pp. 279-289

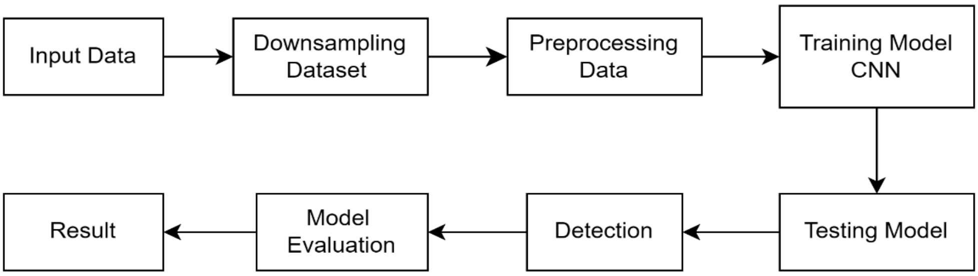


Fig. 1. Research Methodology

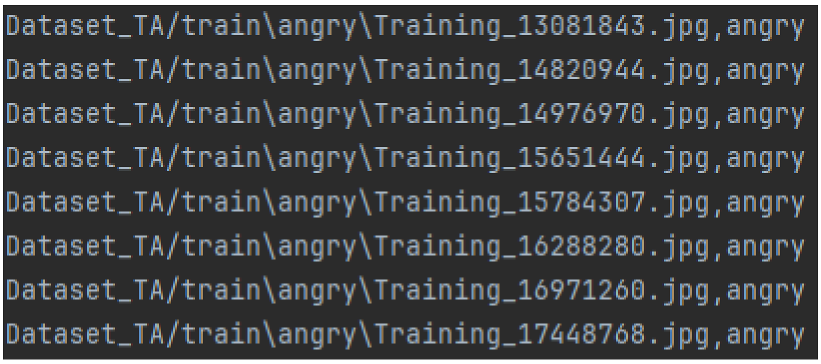
2.1. Input Data

To distinguish between various facial expressions, data was collected from the "Facial Expression Recognition 2013" (FER2013) dataset. The facial expression dataset was obtained from the website https://www.kaggle.com/datasets/msambare/fer2013?resource=download. This dataset consists of approximately 35,887 images divided into training and test categories. The dataset includes 28,709 training images and 7,178 test images, featuring human faces displaying various emotional expressions such as happy, sad, angry, fearful, disgusted, surprised, and neutral. Each image is 48x48 pixels in size and is in grayscale. This research focuses on the common expressions of angry, happy, neutral, and sad, with the entire dataset containing approximately 26,217 images. Fig. 2 shows a sample from the FER2013 dataset for each facial expression category.

Fig. 2. Example of FER2013 dataset

The dataset used in this research was obtained in a zip folder format from Kaggle. The initial process involved extracting the zip folder, resulting in a folder structure consisting of 'train' and 'test' folders. The dataset was then imported into the PyCharm application for model processing.

To facilitate the program in reading and processing the dataset, the data was converted into CSV (Comma-Separated Values) format according to the original dataset categories, namely 'train' and 'test'. The CSV format simplifies the model's access to labels, thereby streamlining the training process on the data. Fig. 3 shows the contents of the dataset that has been converted into CSV format.



```
Dataset_TA/train\angry\Training_13081843.jpg,angry
Dataset_TA/train\angry\Training_14820944.jpg,angry
Dataset_TA/train\angry\Training_14976970.jpg,angry
Dataset_TA/train\angry\Training_15651444.jpg,angry
Dataset_TA/train\angry\Training_15784307.jpg,angry
Dataset_TA/train\angry\Training_16288280.jpg,angry
Dataset_TA/train\angry\Training_16971260.jpg,angry
Dataset_TA/train\angry\Training_17448768.jpg,angry
```

Fig. 3. Example of dataset in CSV format

2.2. Downsampling Dataset

Fig. 4 shows the total number of the four facial expressions in the dataset. It is evident from the dataset that there is an imbalance. To prevent overfitting due to this imbalance, a downsampling process was conducted. Downsampling involves reducing the sample size in the dataset to match the number of samples in the smallest class. As seen in Fig. 4, the 'angry' category has the lowest count, so the downsampling process adjusts the other categories to match the size of the 'angry' category. Fig. 5 shows the amount of data after the downsampling process, with the number of data points in the angry, happy, neutral, and sad classes being 5,000 each, resulting in balanced data.

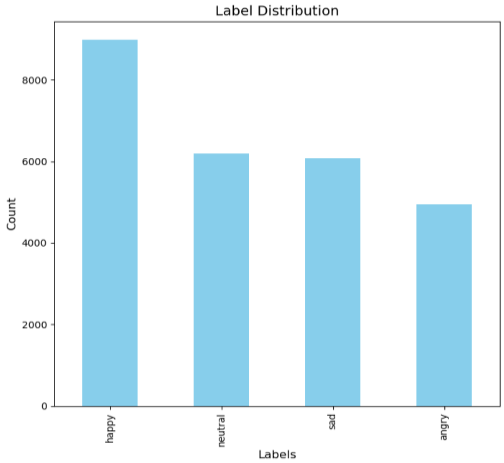


Fig. 4. Data visualization

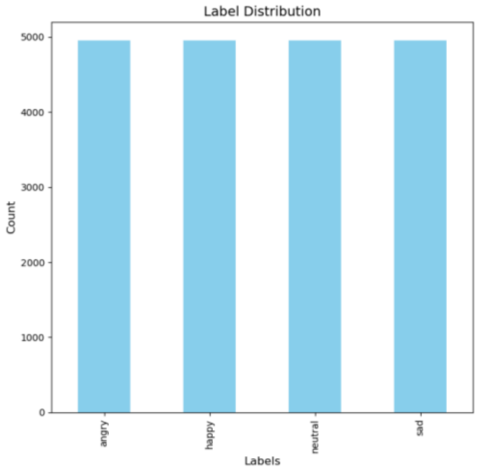


Fig. 5. Visualization of data downsampling

2.3. Preprocessing Data

Before conducting training on the model, data preprocessing is necessary to enhance data quality and model performance. In this study, which falls under supervised learning, preprocessing steps include Train-Test Split, where the dataset is split into 80% training data and 20% testing data for model validation. Additionally, linking and labeling processes are performed on the dataset. Labeling is based on facial expression categories present. Labels in the dataset are aligned with the location in the image files, while links in the dataset are created to indicate the source or location of the image files used to associate images with labels.

2.4. Training Model CNN

In Fig. 6, the basic CNN architecture before improvements consists of several key components in the feature learning and classification processes. Feature learning includes convolutional layers, pooling layers, and ReLU activation, while classification involves flattening, fully connected layers, and softmax. This architecture allows the CNN to learn hierarchical representations from input images and classify them based on the learned features.

In Fig. 7, the CNN architecture after several improvements to address overfitting is presented. In this study, two convolutional layers were added, followed by batch normalization and dropout layers.

Convolutional layers can add new features to capture more information from the dataset, improving the model's ability to learn from the data. The addition of batch normalization enhances the stability and performance of CNN training. Furthermore, dropout layers were added to help prevent overfitting by randomly "dropping out" (disabling) a number of neurons during training.

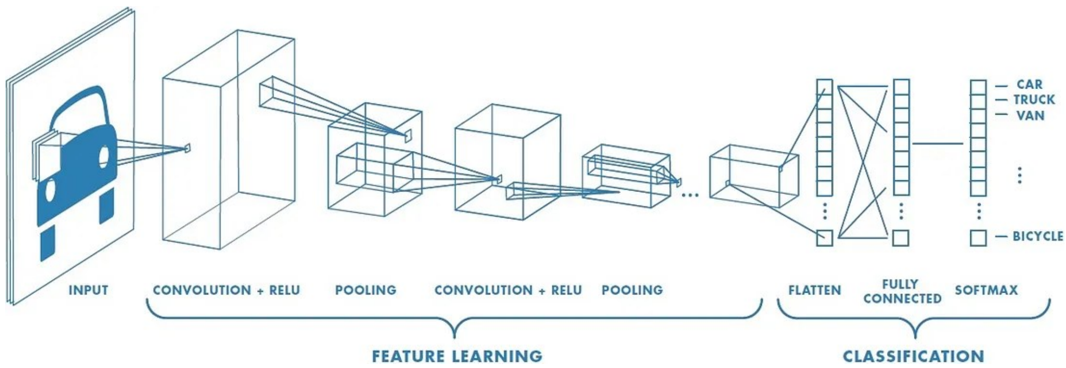


Fig. 6. Basic CNN Architecture

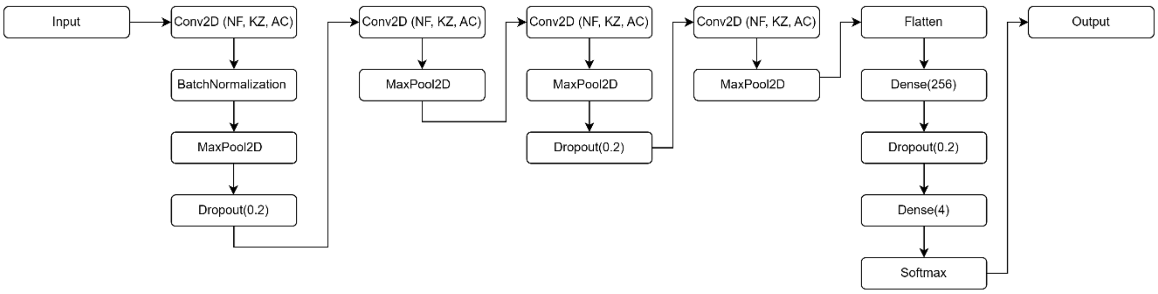


Fig. 7. CNN Architecture After Improvements

Additionally, hyperparameter tuning was employed to find the best combination of hyperparameters using the Grid Search method, which involves testing every combination from a predefined set of hyperparameters. The Adam optimizer was employed to adjust each parameter using estimates of the first moment (mean of gradients) and the second moment (mean of squared gradients), helping the model converge faster and more stably.

Table 1 presents the model hyperparameters used in this study. In this study, four convolutional layers are employed, each followed by a ReLU activation function and pooling layer, to enhance the detection of intricate features within the input image data. A solitary dense layer in the fully connected network is utilized to classify the features extracted by these convolutional layers. Furthermore, a dropout rate of 0.2 is applied, which means that 20% of the neurons in this layer are deactivated during each training cycle. Since the dataset encompasses four distinct class categories (more than two classes), the Categorical Cross-Entropy loss function is utilized.

Table 1. Hyperparameter in the proposed model

Parameter	*Value*
Convolution Layer	4
Dense Layer	1
Dropout	0.2
Loss Function	*Categorical cross entropy*
Activation function	ReLu
Optimizer	Adam
Learning rate	0.001
Number of epochs	100
Batch size	16

To address the issue of gradient vanishing, the ReLU activation function is used. ReLU activates neurons only if they have positive inputs, which accelerates the convergence process. The model training process utilizes a learning rate to balance between fast convergence and stability. With 100 epochs, the model will see each sample in the dataset 100 times, and the model will be updated after every 16 data samples according to the batch size.

2.5. Testing Model

The model testing process aims to obtain facial expression detection results, specifically by calculating the number of detected facial expressions from the test data. The evaluation of accuracy utilizes a Confusion Matrix, which incorporates metrics such as accuracy, precision, recall, and F1-score. Accuracy measures the model's ability to classify different expressions effectively across the board. Precision shows how often the model correctly classifies a facial expression as positive when it predicts positive (measuring the model's accuracy in detecting a specific expression like "Angry" without misidentifying other expressions as "Angry" too often). Recall measures the model's ability to detect all instances of a particular expression, ensuring the model does not miss many expressions present in the data. The F1-score provides a balance between precision and recall, allowing for a more comprehensive evaluation of the model. Equations (1) to (4) represent the formulas for calculating accuracy, precision, recall, and F1-score [50].

According to (1), accuracy can be calculated by dividing the percentage of accurate positive and negative predictions by the total amount of data. The accuracy value can be obtained using the following formula.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{1}$$

Equation (2) states that precision is the ratio of all positive data that is deemed positive to the total number of correct positive forecasts. The following formula yields the Precision value.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

Recall is defined as the fraction of positive predictions to all positive true data (3). The following equation yields the recall value.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \tag{3}$$

The F1-Score is defined as the harmonic mean of recall and precision, as shown in (4). The following formula yields the F1-Score value.

$$\text{F1-Score} = 2\left(\frac{Precision \times Recall}{Precision + Recall}\right) \times 100\% \tag{4}$$

Explanation:
1. True Positive (TP) represents an image sample that is classified as positive and is indeed positive.
2. True Negative (TN) represents an image sample that is classified as negative and is indeed negative.
3. False Positive (FP) represents an image sample that is classified as negative but is actually positive.
4. False Negative (FN) represents an image sample that is classified as positive but is actually negative.

2.6. Detection

Detection is performed to assess the model's precision in identifying human facial expressions in real-time. How well the CNN model detects and distinguishes the four facial expressions is a crucial factor in assessing the model's accuracy in facial expression recognition. This is the main component in evaluating the CNN model's accuracy in facial expression detection. The model's correctness is the evaluation metric used in this study.

3. RESULTS AND DISCUSSION

In Fig. 8, the performance graph of the model using the training dataset shows training accuracy indicated by the red line and validation accuracy indicated by the blue line. Fig. 8. demonstrates that the training accuracy of the CNN model before improvements reached 100%, whereas the validation accuracy was only 70%. This 30% gap indicates substantial overfitting, where the model excels at recognizing patterns in the training data but struggles to apply these patterns to new data (validation set).

In Fig. 9, the training accuracy of the CNN model is shown after modifications to the model architecture, including adding convolutional layers, batch normalization, and dropout layers. This improved CNN model demonstrates a better balance between training and validation accuracies, indicating that the model is no longer overfitting and can generalize well. The training accuracy of this model reaches 98.89%, while the validation accuracy achieves 89.9%.

In Table 2, the evaluation results for 200 human facial expression images are presented. Overall, 178 facial expression images were correctly detected, 22 images were falsely identified as facial expressions when they were not (false positives), and 22 facial expression images were missed (false negatives). Table 2 shows that the accuracy, precision, recall, and f1-score of the test results each reached 89%, 78%, 78%, and 78%, respectively.

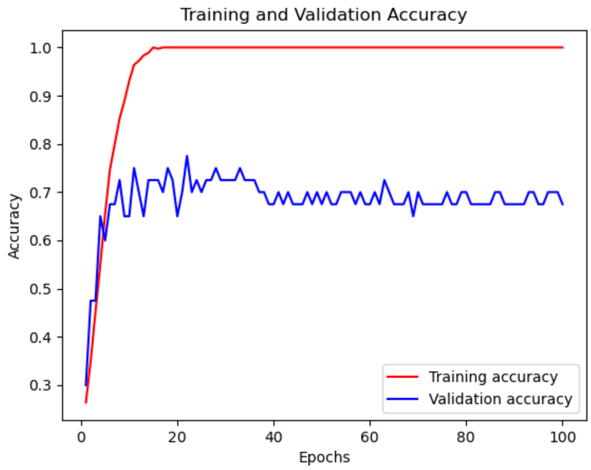


Fig. 8. The accuracy results of the basic CNN model

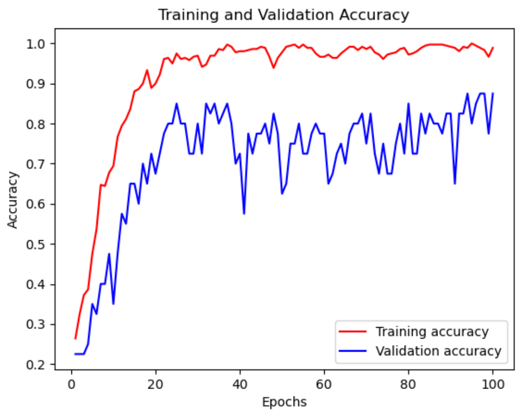


Fig. 9. The accuracy results of the CNN model after improvements

Table 2. The Evaluation Results of the CNN Model After Improvement

Data	TP	FP	FN	TN	Accuracy	Precision	Recall	F1-Score
Data Testing	156	44	44	556	89%	78%	78%	78%

TP represents the number of cases where the model correctly detects existing facial expressions; a high TP value indicates that the model is proficient at recognizing facial expressions. FP indicates the number of cases where the model erroneously detects facial expressions that do not exist, and a low FP value suggests that the model tends to accurately detect expressions. The number of cases where the model fails to detect actual expressions is represented by FN. FN indicates the model's weakness in detecting specific expressions and can reduce confidence in the model's detection results.

The study evaluates the performance of the enhanced CNN model and compares it with that of the basic CNN model (before improvement) using the same dataset. Table 3 displays the performance comparison between the two models for facial expression detection.

From Table 3, which contrasts the model's performance, it's apparent that the enhanced CNN model has shown improved performance. Specifically, accuracy increased by 0.25%, recall increased by 1.3%, and f1-score increased by 1.3%. The performance of the enhanced CNN model suggests that this CNN architecture outperforms the basic CNN model, especially across multiple performance metrics including accuracy, precision, recall, and f1-score.

Table 3. Model Performance Comparison

Algorithm	TP	FN	FP	TN	A	P	R	F1-Score
Basic CNN	155	47	43	555	88,75%	78,2%	76,7%	76,7%
Improved CNN	156	44	44	556	89%	78%	78%	78%

Real-Time Facial Expression Detection Results

Fig. 10. displays real-time outcomes for detecting human facial expressions. In the image, the improved model successfully detects all tested categories of facial expressions: angry, happy, neutral, and sad. This demonstrates that the modified model is capable of identifying facial expressions with high accuracy in real-time conditions.

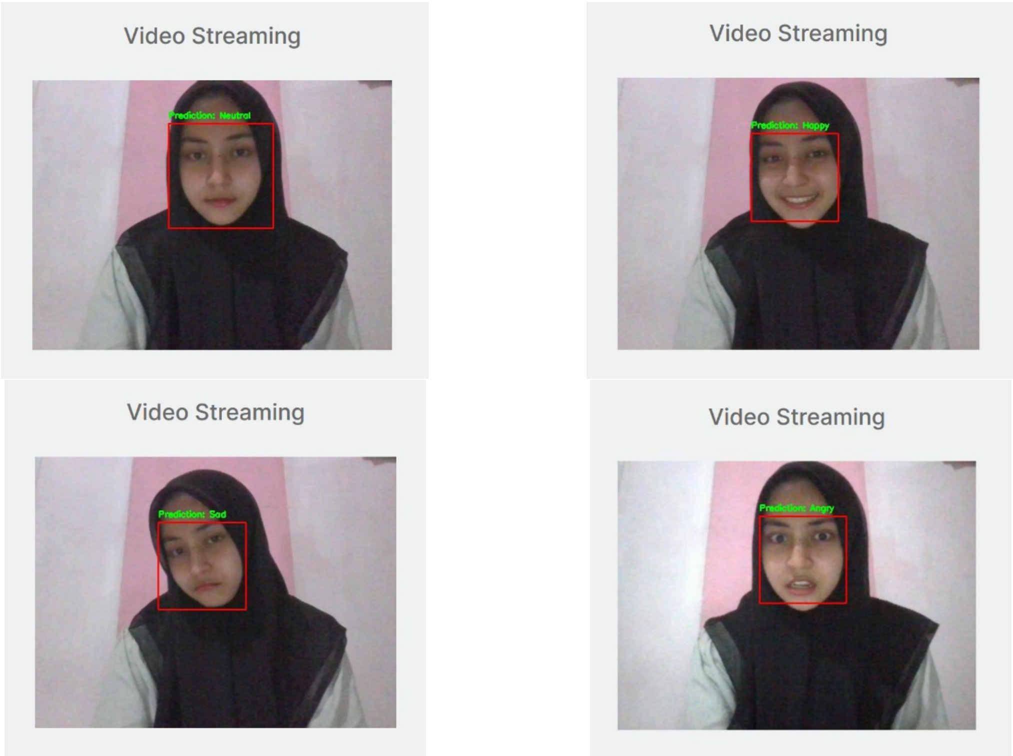


Fig. 10. Real-Time Detection Outcomes of the Enhanced CNN Model

4. CONCLUSION

Based on the conducted research, the improved CNN model has proven to be quite effective in detecting human facial expressions. Comprehensive testing yielded satisfactory results, with a training accuracy of 98.89%, testing accuracy of 89%, precision of 78%, recall of 78%, and f1-score of 78%. Additionally, the improved CNN model successfully detected facial expressions in real-time with good accuracy. These results indicate significant progress in the model's performance in recognizing human facial expressions. They provide evidence that the applied improvements can enhance the model's ability to accurately monitor the emotional states of children with special needs and detect mental health issues.

Although this study shows how CNN models are effective in identifying facial expressions in humans, there is significant potential for further research in developing more advanced model architectures. One intriguing area is the exploration of integrating state-of-the-art techniques such as Faster R-CNN or YOLO for facial expression detection. Research could focus on optimizing detection performance while considering factors like speed, accuracy, and adaptability to variations in facial expressions.

REFERENCES

[1] J. Yu, M. Tan, H. Zhang, Y. Rui, and D. Tao, “Hierarchical Deep Click Feature Prediction for Fine-Grained Image Recognition,” *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 2, pp. 563–578, Feb. 2022, https://doi.org/10.1109/TPAMI.2019.2932058.

[2] A. Chen, H. Xing, and F. Wang, "A Facial Expression Recognition Method Using Deep Convolutional Neural Networks Based on Edge Computing," *IEEE Access*, vol. 8, pp. 49741–49751, 2020, https://doi.org/10.1109/ACCESS.2020.2980060.

[3] H. V, S. T, and M. Siddappa, "Facial Emotion Recognition Method Based on Canny Edge Detection Using Convolutional Neural Network," in *International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), IEEE,* pp. 425–430, Nov. 2023, https://doi.org/10.1109/ICCCIS60361.2023.10425079.

[4] F. Makhmudkhujaev, M. Abdullah-Al-Wadud, M. T. Bin Iqbal, B. Ryu, and O. Chae, "Facial expression recognition with local prominent directional pattern," *Signal Process Image Commun*, vol. 74, pp. 1–12, May 2019, https://doi.org/10.1016/j.image.2019.01.002.

[5] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognit Lett*, vol. 120, pp. 69–74, Apr. 2019, https://doi.org/10.1016/j.patrec.2019.01.008.

[6] K. Rengasamy, P. Joshi, and V. V. S. Raveendra, "Real-time hybrid facial expression analysis model using quantum distance-based classifier and classical artificial neural networks: QUADCANN," *Quantum Inf Process,* vol. 23, no. 6, p. 220, Jun. 2024, https://doi.org/10.1007/s11128-024-04437-3.

[7] C. Dewi, L. S. Gunawan, S. G. Hastoko, and H. J. Christanto, "Real-Time Facial Expression Recognition: Advances, Challenges, and Future Directions," *Vietnam Journal of Computer Science*, vol. 11, no. 02, pp. 167–193, May 2024, https://doi.org/10.1142/S219688882330003X.

[8] W. Wu, Y. Yin, X. Wang, and D. Xu, "Face Detection With Different Scales Based on Faster R-CNN," *IEEE Trans Cybern*, vol. 49, no. 11, pp. 4017–4028, Nov. 2019, https://doi.org/10.1109/TCYB.2018.285948.

[9] S. M. Saleem Abdullah and A. M. Abdulazeez, "Facial Expression Recognition Based on Deep Learning Convolution Neural Network: A Review," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, Apr. 2021, https://doi.org/10.30880/jscdm.2021.02.01.006.

[10] G. R. Alexandre, J. M. Soares, and G. A. Pereira Thé, "Systematic review of 3D facial expression recognition methods," *Pattern Recognit*, vol. 100, p. 107108, Apr. 2020, https://doi.org/10.1016/j.patcog.2019.107108.

[11] M. Cesarelli, F. Martinelli, F. Mercaldo, and A. Santone, "Emotion Recognition from Facial Expression using Explainable Deep Learning," in *IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), IEEE*, Sep. 2022, pp. 1–6. https://doi.org/10.1109/DASC/PiCom/CBDCom/Cy55231.2022.9927951.

[12] M. Sajjad *et al.,* "A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines," Alexandria Engineering Journal, vol. 68, pp. 817–840, Apr. 2023, https://doi.org/10.1016/j.aej.2023.01.017.

[13] D. Bhagat, A. Vakil, R. K. Gupta, and A. Kumar, "Facial Emotion Recognition (FER) using Convolutional Neural Network (CNN)," *Procedia Comput Sci*, vol. 235, pp. 2079–2089, 2024, https://doi.org/10.1016/j.procs.2024.04.197.

[14] G. Yolcu *et al.,* "Facial expression recognition for monitoring neurological disorders based on convolutional neural network," *Multimed Tools Appl*, vol. 78, no. 22, pp. 31581–31603, Nov. 2019, https://doi.org/10.1007/s1104201907959-6.

[15] I. M. Revina and W. R. S. Emmanuel, "A Survey on Human Face Expression Recognition Techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 6, pp. 619–628, Jul. 2021, https://doi.org/10.1016/j.jksuci.2018.09.002.

[16] Q. Jaleel and I. Hadi, "Facial Action Unit-Based Deepfake Video Detection Using Deep Learning," *in 4th International Conference on Current Research in Engineering and Science Applications (ICCRESA), IEEE*, Dec. pp. 228–233, 2022, https://doi.org/10.1109/ICCRESA57091.2022.10352085.

[17] Y. Nan, J. Ju, Q. Hua, H. Zhang, and B. Wang, "A-MobileNet: An approach of facial expression recognition," *Alexandria Engineering Journal*, vol. 61, no. 6, pp. 4435–4444, Jun. 2022, https://doi.org/10.1016/j.aej.2021.09.066.

[18] Y. He, "Facial Expression Recognition Using Multi-Branch Attention Convolutional Neural Network," *IEEE Access*, vol. 11, pp. 1244–1253, 2023, https://doi.org/10.1109/ACCESS.2022.3233362.

[19] C.-L. Kim and B.-G. Kim, "Few-shot learning for facial expression recognition: a comprehensive survey," *J Real Time Image Process*, vol. 20, no. 3, p. 52, Jun. 2023, https://doi.org/10.1007/s11554-023-01310-x.

[20] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets," *Information*, vol. 15, no. 3, p. 135, Feb. 2024, https://doi.org/10.3390/info15030135.

[21] J. A. Brooks *et al.,* "Deep learning reveals what facial expressions mean to people in different cultures," *iScience*, vol. 27, no. 3, p. 109175, Mar. 2024, https://doi.org/10.1016/j.isci.2024.109175.

[22] A. S. Redfern and C. P. Benton, "Representation of facial identity includes expression variability," *Vision Res*, vol. 157, pp. 123–131, Apr. 2019, https://doi.org/10.1016/j.visres.2018.05.004.

[23] M. Karnati, A. Seal, D. Bhattacharjee, A. Yazidi, and O. Krejcar, "Understanding Deep Learning Techniques for Recognition of Human Emotions Using Facial Expressions: A Comprehensive Survey," *IEEE Trans Instrum Meas*, vol. 72, pp. 1–31, 2023, https://doi.org/10.1109/TIM.2023.3243661.

[24] M. Karnati, A. Seal, D. Bhattacharjee, A. Yazidi, and O. Krejcar, "Understanding Deep Learning Techniques for Recognition of Human Emotions Using Facial Expressions: A Comprehensive Survey," *IEEE Trans Instrum Meas*, vol. 72, pp. 1–31, 2023, https://doi.org/10.1109/TIM.2023.3243661.

[25] J. Li, K. Jin, D. Zhou, N. Kubota, and Z. Ju, "Attention mechanism-based CNN for facial expression recognition," *Neurocomputing*, vol. 411, pp. 340–350, Oct. 2020, https://doi.org/10.1016/j.neucom.2020.06.014.

[26] M. A. K. Raiaan *et al.,* "A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks," *Decision Analytics Journal*, vol. 11, p. 100470, Jun. 2024, https://doi.org/10.1016/j.dajour.2024.100470.

[27] E. Ayan, H. Erbay, and F. Varçın, "Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks," *Comput Electron Agric*, vol. 179, p. 105809, Dec. 2020, https://doi.org/10.1016/j.compag.2020.105809.

[28] Y. Liu, H. Pu, and D.-W. Sun, "Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices," *Trends Food Sci Technol*, vol. 113, pp. 193–204, Jul. 2021, https://doi.org/10.1016/j.tifs.2021.04.042.

[29] Y. Wang *et al.,* "Deep leaning in food safety and authenticity detection: An integrative review and future prospects," *Trends Food Sci Technol*, vol. 146, p. 104396, Apr. 2024, https://doi.org/10.1016/j.tifs.2024.104396.

[30] M. C. Gursesli, S. Lombardi, M. Duradoni, L. Bocchi, A. Guazzini, and A. Lanata, "Facial Emotion Recognition (FER) Through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets," *IEEE Access*, vol. 12, pp. 45543–45559, 2024, https://doi.org/10.1109/ACCESS.2024.3380847.

[31] D.-H. Lee and J.-H. Yoo, "CNN Learning Strategy for Recognizing Facial Expressions," *IEEE Access*, vol. 11, pp. 70865–70872, 2023, https://doi.org/10.1109/ACCESS.2023.3294099.

[32] C. Shi, C. Tan, and L. Wang, "A Facial Expression Recognition Method Based on a Multibranch Cross-Connection Convolutional Neural Network," *IEEE Access*, vol. 9, pp. 39255–39274, 2021, https://doi.org/10.1109/ACCESS.2021.3063493.

[33] G. Habib and S. Qureshi, "Optimization and acceleration of convolutional neural networks: A survey," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 7, pp. 4244–4268, Jul. 2022, https://doi.org/10.1016/j.jksuci.2020.10.004.

[34] Z. Hu, J. Zhang, and Y. Ge, "Handling Vanishing Gradient Problem Using Artificial Derivative," *IEEE Access*, vol. 9, pp. 22371–22377, 2021, https://doi.org/10.1109/ACCESS.2021.3054915.

[35] H. L. Dawson, O. Dubrule, and C. M. John, "Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification," *Comput Geosci*, vol. 171, p. 105284, Feb. 2023, https://doi.org/10.1016/j.cageo.2022.105284.

[36] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 2, pp. 757–774, Feb. 2023, https://doi.org/10.1016/j.jksuci.2023.01.014.

[37] C. R. Edwin Selva Rex, J. Annrose, and J. Jenifer Jose, "Comparative analysis of deep convolution neural network models on small scale datasets," *Optik (Stuttg)*, vol. 271, p. 170238, Dec. 2022, https://doi.org/10.1016/j.ijleo.2022.170238.

[38] D. Palma-Ramírez *et al.,* "Deep convolutional neural network for weld defect classification in radiographicimages," *Heliyon*, vol. 10, no. 9, p. e30590, May 2024, https://doi.org/10.1016/j.heliyon.2024.e30590.

[39] K. Presannakumar and A. Mohamed, "Deep learning based source identification of environmental audio signals using optimized convolutional neural networks," *Appl Soft Comput*, vol. 143, p. 110423, Aug. 2023, https://doi.org/10.1016/j.asoc.2023.110423.

[40] P. Singh, P. Mazumder, and V. P. Namboodiri, "Context extraction module for deep convolutional neural networks," *Pattern Recognit*, vol. 122, p. 108284, Feb. 2022, https://doi.org/10.1016/j.patcog.2021.108284.

[41] A. Ghosh, N. D. Jana, S. Das, and R. Mallipeddi, "Two-Phase Evolutionary Convolutional Neural Network Architecture Search for Medical Image Classification," *IEEE Access*, vol. 11, pp. 115280–115305, 2023, https://doi.org/10.1109/ACCESS.2023.3323705.

[42] W. Ming *et al.,* "Review: Application of Convolutional Neural Network in Defect Detection of 3C Products," *IEEE Access*, vol. 9, pp. 135657–135674, 2021, https://doi.org/10.1109/ACCESS.2021.3116131.

[43] S. Raziani and M. Azimbagirad, "Deep CNN hyperparameter optimization algorithms for sensor-based humanactivity recognition," *Neuroscience Informatics*, vol. 2, no. 3, p. 100078, Sep. 2022, https://doi.org/10.1016/j.neuri.2022.100078.

[44] Y. Quan, Z. Li, C. Zhang, and H. Ma, "Object Detection Model Based on Deep Dilated Convolutional Networks by Fusing Transfer Learning," *IEEE Access*, vol. 7, pp. 178699–178709, 2019, https://doi.org/10.1109/ACCESS.2019.2958817.

[45] T. Zhou, Q. Hu, Z. Hu, and R. Zhen, "An adaptive hyper parameter tuning model for ship fuel consumption prediction under complex maritime environments," *Journal of Ocean Engineering and Science*, vol. 7, no. 3, pp. 255–263, Jun. 2022, https://doi.org/10.1016/j.joes.2021.08.007.

[46] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020, https://doi.org/10.1016/j.neucom.2020.07.061.

[47] D. Elhani, A. C. Megherbi, A. Zitouni, F. Dornaika, S. Sbaa, and A. Taleb-Ahmed, "Optimizing convolutional neural networks architecture using a modified particle swarm optimization for image classification," *Expert Syst Appl*, vol. 229, p. 120411, Nov. 2023, https://doi.org/10.1016/j.eswa.2023.120411.

[48] M. A. K. Raiaan *et al.,* "A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks," *Decision Analytics Journal*, vol. 11, p. 100470, Jun. 2024, https://doi.org/10.1016/j.dajour.2024.100470.

[49] D.-X. Zhou, "Theory of deep convolutional neural networks: Downsampling," *Neural Networks*, vol. 124, pp. 319 327, Apr. 2020, https://doi.org/10.1016/j.neunet.2020.01.018.

[50] W. Lee and K. Seo, "Downsampling for Binary Classification with a Highly Imbalanced Dataset Using Active Learning," *Big Data Research*, vol. 28, p. 100314, May 2022, https://doi.org/10.1016/j.bdr.2022.100314.

[51] S. Kilicarslan, M. Celik, and Ş. Sahin, "Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification," *Biomed Signal Process Control*, vol. 63, p. 102231, Jan. 2021, https://doi.org/10.1016/j.bspc.2020.102231.

BIOGRAPHY OF AUTHORS

Fatimatuzzahra, is currently a student in the Electrical Engineering Department, Diploma 4 Telecommunication Engineering Study Program at Politeknik Negeri Sriwijaya, where she started studying in 2020. As a dedicated student, Fatimatuzzahra has a great interest in technology development, especially in the field of AI and machine learning. Alongside her studies, she also actively participates in the LDK Karisma organization at Politeknik Negeri Sriwijaya, involved in various extracurricular activities and events. Email: fazzahrah125@gmail.com.

Lindawati, is an Assistant Professor and Lecturer in the Department of Electrical Engineering, Telecommunication Engineering Study Program, State Polytechnic of Sriwijaya, Indonesia. She received a Bachelor of Engineering Degree in Electrical Engineering from the University of Sriwijaya, Indonesia, and a Master of Information Technology Degree in Information Technology from the University of Indonesia, Indonesia. His research interests include information technology, IoT, and robotics. Email: lindawati@polsri.ac.id

Sopian Soim, is an Assistant Professor and Lecturer in the Department of Electrical Engineering, Telecommunication Engineering Study Program, State Polytechnic of Sriwijaya, Indonesia. He received a Bachelor of Engineering Degree in Electrical Engineering from the University of Sriwijaya, Indonesia, and a Master of Engineering Degree in Electrical Engineering from the Sepuluh Nopember Institute of Technology, Indonesia. His research interests include wireless communications, IoT, and data science. Email: sopiansoim@gmail.com.