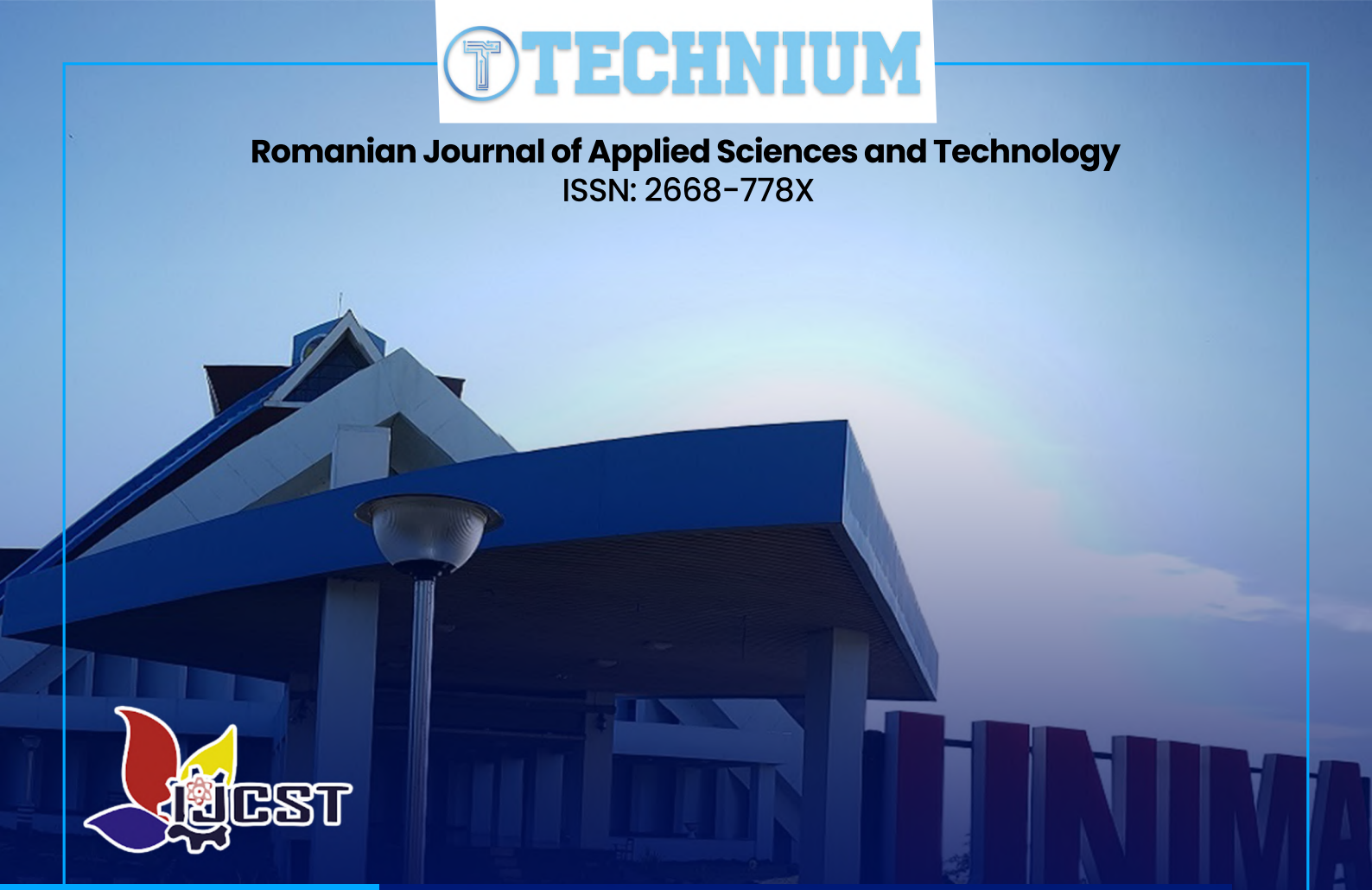
TECHNIUM

Romanian Journal of Applied Sciences and Technology
ISSN: 2668-778X



vol. 17 / 2023

The 7th International Conference on Science Technology

organized by
Faculty of Social Science and
Law Universitas Negeri Manado and
Consortium of International Conference
on Science and Technology

The Innovation Breakthrough in Digital and Disruptive Era

PLUS COMMUNICATION

International Communication & PR

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 17, 173-178, November, 2023
ISSN: 2668-778X
www.techniumscience.com

Detecting Face Expressions in Real-Time Using Convolutional Neural Network (CNN) Algorithm

Muhammad Haris Irham[1*], *Abdul Mubarak*[2], *Munazat Salmin*[3] and *Rosihan*[4]

[1,2,3,4]Informatics Department, Faculty of Engineering, Khairun University, Ternate, Indonesia

Abstract. This research discusses the use of a Convolutional Neural Network (CNN) with the MobileNetV2 model in the real-time detection of human facial expressions. This research aims to develop a human face expression detection system using deep learning algorithms. This study used the observation data collection method and obtained secondary data from the FER2013 data set which contains 28,709 training samples, 3,859 validation data sets, and 3,859 test samples, for a total of 35,887 images with a resolution of 48x48 and seven categories of facial expressions. The training results showed that the CNN model using MobileNetV2 achieved an accuracy of 57% in the training process and 51% in the validation process. Based on the analysis of these results, testing using a confusion matrix with an accuracy of 51% concluded that the model was unable to properly recognize patterns of data with disgust and fear categories, leading to low accuracy. Some factors contributing to the system's inability to recognize expressions were due to similarities between facial expressions such as sad and fearful, or sad and disgusted. This study provides new insights into the development of technology for detecting human facial expressions using deep learning and the MobileNetV2 model.

Keywords: *Face Expression Detection, Deep Learning, Convolutional Neural Network (CNN), MobileNetV2, Real-time.*

* Corresponding author: mharisirham35@gmail.com

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 17, 173-178, November, 2023
ISSN: 2668-778X
www.techniumscience.com

1 Introduction

In communication and social interaction, humans use two forms of communication: verbal and nonverbal. One form of nonverbal communication is facial expression, which is used to convey a person's emotional state to their interlocutor. Humans can quickly discern when their conversation partner is angry, happy, disgusted, or displaying a neutral expression. However, a study conducted by [1] "Body language is more effective in perceiving someone's emotions than just facial expressions," conducted at a therapy hospital in London. Nevertheless, by discerning the interlocutor's expression, humans can already gain enough assistance in making the communication flow or topic of discussion more appropriate and focused. But what about computers? Can computers automatically discern a person's emotional state through facial expressions?

In the current era of the industrial revolution, technology is advancing rapidly, particularly in terms of interaction between humans as users and computers as providers of technological services. Therefore, automatic detection of human facial expressions has become highly important to be applied in various areas, such as image understanding, healthcare, human-computer interaction, video games, and data-driven animation. This has prompted many application developers to compete in designing algorithmic methods for more accurate, lightweight, and user-friendly facial expression detection, aiming to assist in the interaction between humans and computers. In the application of artificial intelligence (AI), particularly in the field of computer vision, we can train computers to perceive, detect, and classify human facial expressions through digital images using a machine or deep learning techniques.

One previous study on facial expression detection was conducted by [2], titled "Facial Expression Detection Using Gabor and Haar Features." In her research, face detection was performed using the Viola-Jones algorithm to enable computers to read and recognize human faces. The initial step involved contrast enhancement using histogram equalization, followed by feature calculation using Gabor and Haar features. In the classification process, the One vs. All SVM method was utilized. The achieved results were satisfactory; however, it was noted that increasing the number of training samples for the expression data model would be beneficial. Additionally, [3] on facial expression recognition titled "A Study of Facial Expression Recognition Using PCA and CNN." Facial expression recognition is crucial in computer systems and human-computer interactions.

2 Literature Review

2.1 Facial Expressions

Research on human facial expressions can be explained through the theory of Emotion Expression [4]. This theory states that the emotional expressions related to the human face are the result of evolutionary programs inherited by humans from their ancestors. Ekman asserts that six basic emotions can be identified through facial expressions: anger, fear, sadness, happiness, surprise, and disgust.

2.2 Deep Learning

Deep learning is a subfield within the field of artificial intelligence and machine learning. The deep learning method utilizes neural networks with multiple layers to detect objects, recognize speech, translate languages, and perform other tasks. Deep learning enables the computation of models with layers of processing that can learn data representations at different levels of abstraction. The backpropagation algorithm is used to adjust parameters in each layer to calculate each layer's output. Deep convolutional networks have also made significant advancements in processing images, audio, video, text, and more [5].

2.3 Convolutional Neural Network

Convolutional Neural Network (CNN) is an algorithm that applies Multilayer Perceptron (MLP) to process two-dimensional data effectively. CNN falls under the category of Deep Neural Networks due to its high depth and is commonly used in image data processing. When performing image data classification, MLP is not suitable as it does not retain the spatial information of the image and treats each pixel as an independent feature, resulting in suboptimal performance [6]

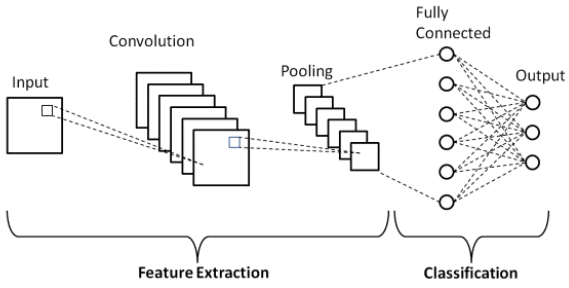


Fig. 1. CNN Architecture

2.4 Transfer Learning

Deep learning has become popular thanks to the availability of large-scale datasets like ImageNet. However, acquiring such datasets can be challenging. As a solution, transfer learning has emerged, which involves training a model using a pre-trained model on a different dataset. While transfer learning helps address limited dataset availability, the dataset size remains crucial for maintaining or improving model accuracy. Research [7] highlights the significance of dataset size in transfer learning, using tiny-magnet and multiplaces2 datasets. Experiments showed that freezing the initial layers of the network resulted in better performance for small target datasets. Smaller dataset sizes also yielded superior outcomes.

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 17, 173-178, November, 2023
ISSN: 2668-778X
www.techniumscience.com

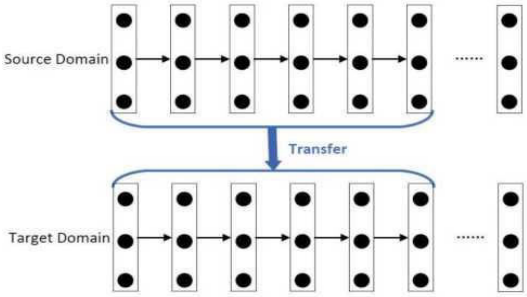


Fig. 2. Diagram Network-based Deep Transfer Learning

2.5 MobilenetV2

MobileNetV2 is a convolutional neural network (CNN) architecture designed specifically for mobile devices, aiming to optimize computational resources. It improves upon the previous MobileNet by utilizing techniques such as depthwise convolution, pointwise convolution, linear bottlenecks, and shortcut connections. In image classification experiments, MobileNetV2 achieves higher accuracy compared to MobileNetV1 while using fewer parameters. The architecture consists of two types of blocks: residual blocks with stride 1 and stride 2. These blocks are combined to form the MobileNetV2 architecture, enabling efficient and resource-friendly computations on mobile devices. [8]

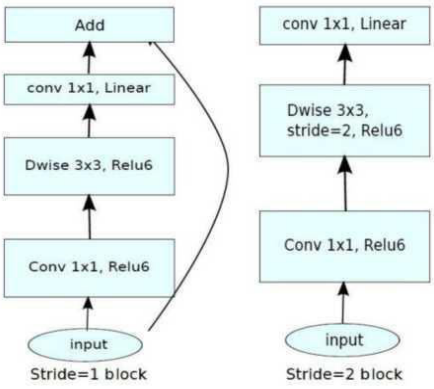


Fig. 3. *Residual Blocks on MobileNetV2*

3 Results and Discussion

3.1 Data Analysis

The data used in this research is obtained from the Kaggle website database, specifically the FER2013 dataset, with a training data size of 21,535 and a testing data size of 7,174. The images in the dataset have a size of 48x48 pixels and are in grayscale. The data consists of 7 categories: angry, disgusted, fearful, happy, neutral, sad, and surprised.

3.2 Training Model

In the training process, the model was trained for 30 epochs with a learning rate of 0.001 and an output layer of 7. The batch size used in this research was 128.

3.3 Model Results and MobileNetV2 Testing

Research Results with the implementation of the MobileNetV2 CNN model showed a training accuracy of 57% with a loss value of 1.4711, and a validation accuracy of 51% with a validation loss value of 1.3168, using a maximum of 30 epochs. However, the training process stopped at epoch 9 to prevent overfitting, as the validation loss value started to increase in the previous 3 epochs. The average values of precision, recall, and F-measure in Table 4.1 yielded 51% with a testing accuracy of 51%. Based on the accuracy graph, which shows consistent values for both accuracy and validation accuracy, and the relatively stable average testing values, the model did not exhibit overfitting. However, the obtained loss values were still high, resulting in the model not reaching its maximum accuracy.

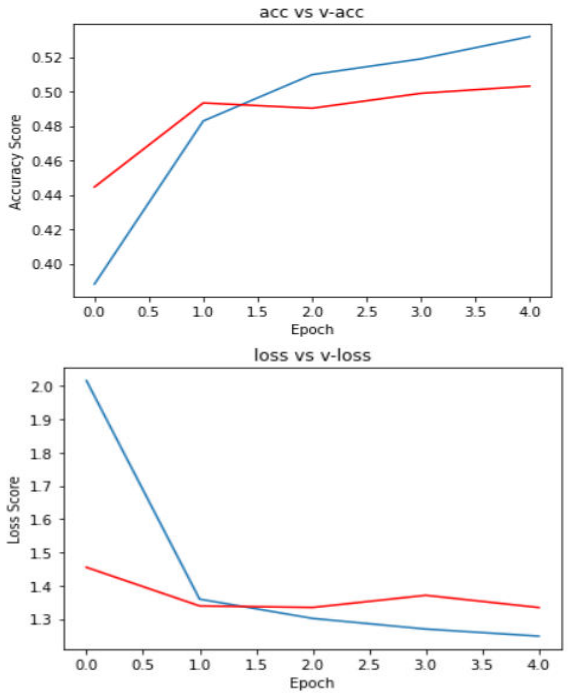


Fig. 4. The comparison of Accuracy and Loss

Next, the testing results were evaluated using a confusion matrix.

Technium: Romanian Journal of Applied Sciences and Technology
Vol. 17, 173-178, November, 2023
ISSN: 2668-778X
www.techniumscience.com

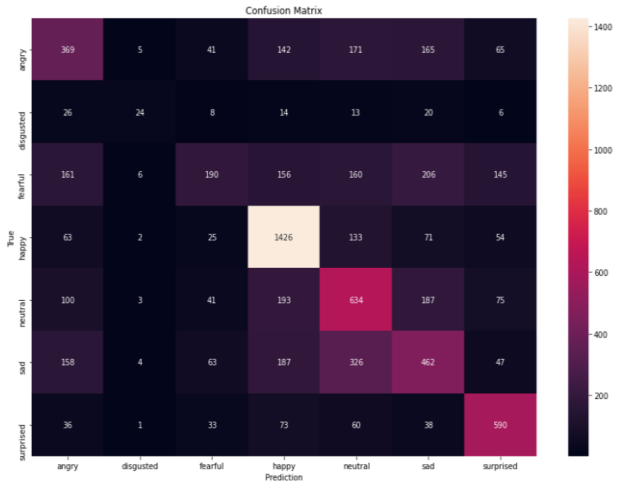


Fig. 5. Confusion Matrix.

Based on the results of the confusion matrix in Figure 5, it can be concluded that darker colors with smaller numbers indicate low accuracy, while brighter colors with larger numbers indicate better accuracy. The obtained results still struggle to recognize patterns in the "disgust" and "fear" categories, leading to a high level of misclassification for these categories. For more details, the percentage values of precision, recall, and F-measure for each class can be found in Table 1.

Table 1. Performance Evaluation Results of the Model Testing

Label	*Precision*	*Recall*	*F-Measure*	Data
Angry	0,40	0,39	0,39	958
Disgusted	0,53	0,22	0,31	111
fear	0,47	0,19	0,27	1024
Happy	0,65	0,80	0,72	1774
Netral	0,42	0,51	0,46	1233
Sad	0,40	0,37	0,39	1247
Suprise	0,60	0,71	0,65	831
Accuracy	51%			

Based on the performance evaluation results of applying the MobileNetV2 model with a learning rate of 0.001 and epoch = 30, as shown in Table 4.1, the average precision, recall, and F-measure values indicate a performance of 51%.

3.4 Accuracy Analysis

Furthermore, the analysis of the CNN model using MobileNetV2 indicates that it falls into the low-performance category. This is attributed to the training process yielding an accuracy of 57% with a loss value of 1.4711, and a validation accuracy of 51% with a validation loss of 1.3168. The model was trained for 30 epochs, and early stopping was applied at epoch 9 to prevent overfitting, as the model started to exhibit signs of overfitting from epochs 10 to 30. The low loss and accuracy in the CNN model can be attributed to several factors:

1. Face Expressions There are similarities between facial expressions, such as anger and surprise, sadness and disgust. This is due to the system recognizing similar facial features, leading to difficulties in accurately classifying images based on their expression. Proper categorization of facial expressions is crucial to avoid classification errors. The detection of facial expression similarities and similarities between different facial expressions can be seen in Table 2.

Table 2. Result of Similarity of Face Expressions

No	Expression	Detection Results	Expression Results
1.	Happy		The characteristic of the happy expression includes wide-open eyes, a smile on the lips, and sometimes puffed-up cheeks. The classification results for the happy class achieved an accuracy of 66%. However, the system also predicts similarities with other facial expressions, such as 11% for sad expressions and 11% for disgusted expressions.
2.	Angry		The characteristic of the angry expression includes narrowed eyes, as well as furrows on the forehead and around the lips. The classification results for the angry class achieved an accuracy of 50%. However, the system also predicts similarities with other facial expressions, such as 35% for happy expressions and 13% for surprised expressions.
3.	Sad		The characteristic of the sad expression includes tired and baggy eyes, drooping lips, as well as furrows on the forehead and around the lips. The classification results for the sad class achieved an accuracy of 40%. However, the system also predicts similarities with other facial expressions, such as 43% for the disgusted expression.
4.	Fear		The fearful expression is characterized by wide, open eyes, pulled-back lips, and furrows on the forehead. The accuracy of classifying the fearful expression was 59%. However, the system also predicted similarities with the surprised expression, accounting for 30% of the predictions.

Technium: Romanian Journal of Applied Sciences and Technology
Vol. 17, 173-178, November, 2023
ISSN: 2668-778X
www.techniumscience.com

5.	Suprise		The surprised expression is characterized by wide-open eyes, an open mouth, and furrows on the forehead. The accuracy of classifying the surprised expression was 50%. However, the system also predicted similarities with the fearful expression, accounting for 41% of the predictions.
6.	Disgust		The disgusted expression is characterized by narrowed eyes, curled lips, and wrinkles on the nose and forehead. The accuracy of classifying the disgusted expression was 68%. However, the system also predicted similarities with the fearful expression, accounting for 18% of the predictions.
7.	Neutral		The neutral expression is characterized by normal open eyes, flat lips, and no wrinkles on the face. The accuracy of classifying the neutral expression was 31%. However, the system also predicted similarities with the happy expression, accounting for 35% of the predictions, and with the disgusted expression, accounting for 25% of the predictions.

2. The MobileNetV2 model has too many layers, causing some information from the image to be lost after undergoing image filtering.

3. For real-time facial expression detection, the detection results will be better if the face captured by the camera is clear and the background does not interfere with the camera.

3.5 Interface Display

The system's results are implemented into a website interface using the Python platform, Flask. The website system can detect facial expressions by inputting an image or by detecting facial expressions in real time.

3.6 Home Display

The Home display is the initial page when users access the system. There are two buttons at the top of the website bar to access the system's features.

Fig. 6. Home Display.

3.7 Image Prediction Display

In this display, users can make predictions on random images. The system will automatically classify the images using the pre-trained model. The system predicts whether the image belongs to the classes (sad, angry, scared, disgusted, neutral, happy, or surprised).

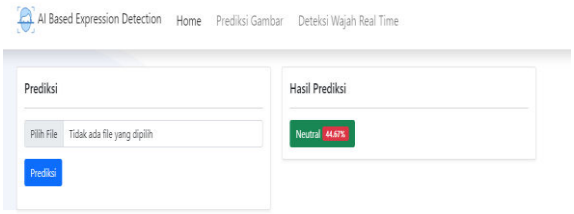


Fig. 7. Image Prediction Display

3.8 Real-time Display

The last page is the real-time face detection page. On this page, users can detect their faces in real-time. The system will perform the detection and classify the detected face (sad, angry, scared, disgusted, neutral, happy, or surprised).

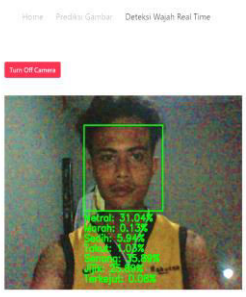


Fig. 8. Display Realtime

4 Conclusion

From the research conducted using the CNN MobileNetV2 model, it can be concluded that the model has low accuracy. This is indicated by the training results, which achieved an accuracy of 57% with a loss value of 1.4711, and a validation accuracy of 51% with a validation loss of 1.3168. The training process stopped at epoch 9 to prevent overfitting. The precision, recall, and F-Measure results also showed low values, at 51%. Analysis of these results indicates

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 17, 173-178, November, 2023
ISSN: 2668-778X
www.techniumscience.com

that the model struggles to recognize patterns in the disgusted and scared categories, leading to a high level of misclassification.

Several factors contribute to the low accuracy and loss of this model. These include similarities between facial expressions, such as anger and surprise, sadness and disgust, as well as the excessive number of layers in the MobileNetV2 model, causing some image information to be lost after undergoing filtering. Additionally, the environmental conditions during real-time facial expression detection may not be conducive to accurate detection.

The system's results were implemented into a website interface using the Flask platform in Python. The website allows users to detect facial expressions by inputting an image or performing real-time detection.

References

1. A. L. Seandrio, A. H. Pratomo, and M. Yanu, “Implementation of Convolutional Neural Network (CNN) in Facial Expression Recognition Implementasi Convolutional Neural Network (CNN) Pada Pengenalan Ekspresi Wajah,” vol. 18, no. 2, pp. 211–221, 2021.
2. R. Wiryadinata, U. Istiyah, R. Fahrizal, P. Priswanto, and S. Wardoyo, “Sistem Presensi Menggunakan Algoritme Eigenface dengan Deteksi Aksesoris dan Ekspresi Wajah,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 6, no. 2, 2017.
3. Hendra Son Simon, “Penentuan Posisi Objek Berbasis Image Processing Dengan Menggunakan Metode Convolutional Neural Network,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2020.
4. P. A. Nugroho, I. Fenriana, and R. Arijanto, “Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia,” *Algor*, vol. 2, pp. 12–21, 2020.
5. I. Azhari, A. R. Sanjaya, A. R. Sanjaya, D. Wajah, D. Learning, and C. N. Network, “Implementasi Algoritma Convolutional Neural Network Dalam Deteksi Emosi Manusia,” vol. 1, no. 1, pp. 112–118, 2020.
6. M. V. Overbeek, “Histogram of Oriented Gradient Untuk Deteksi Ekspresi Wajah Manusia,” *High Educ. Organ. Arch. Qual. J. Teknol. Inf.*, vol. 10, no. 2, pp. 81–86, 2018.
7. M. Altun, H. Gürüler, O. Özkaraca, F. Khan, J. Khan, and Y. Lee, “Monkeypox Detection Using CNN with Transfer Learning,” *Sensors*, vol. 23, no. 4, 2023.
8. L. Hu and Q. Ge, “Automatic facial expression recognition based on MobileNetV2 in Real-time,” *J. Phys. Conf. Ser.*, vol. 1549, no. 2, 2020.