Uncontrolled Environments Face Recognition based on Transfer Learning Technique for Secure Automatic Door Access System

Anggi Rinaldi[1], Muhammad Ikram Andrinur Akbar[2], Iman Fahruzi[3]

anggirinaldi1207@gmail.com[1]

Department of Mechatronics Engineering, Politeknik Negeri Batam, Batam, Indonesia

Abstract— Over the past four decades, artificial intelligence technology, particularly in artificial neural networks and related methods, has advanced rapidly. Deep learning, a major branch of artificial intelligence, has proven its effectiveness in addressing various problems, especially those involving large-scale data such as images, text, and sound. One notable application of deep learning is in developing automated door systems. These systems offer several benefits, including reducing direct contact with door handles, which is increasingly important for cleanliness and health concerns. This research proposes using deep learning, specifically transfer learning techniques, to detect facial expressions of individuals approaching the door. By recognizing these facial expressions, the system can automatically activate a motor to open the door if the input matches the system's criteria. During the development phase, we employed the MobileNetV2 architecture for facial expression detection. Testing was conducted with the ESP32 device, and the model was trained and validated over 25 epochs. The experiments revealed that the model achieved a maximum accuracy of 53%. This research contributes to creating more efficient and user-friendly automated door systems. By leveraging deep learning technology, we aim to enhance safety and comfort for users.

Keywords: deep learning, automated door, transfer learning, facial expression, MobileNetV2.

1. Introduction

Neural networks and other artificial intelligence methods have advanced significantly over the past four decades, with deep learning emerging as a key component. Deep learning utilizes multilayer Artificial Neural Networks (ANNs) and has proven highly effective in addressing various problems involving large datasets, such as images, text, and sound. This effectiveness has led to its widespread adoption in both research and industry [1].

One interesting application of deep learning is in developing automated door systems that use computer vision to detect human facial expressions. Previous research has investigated various methods for automatic door opening, including e-KTP, fingerprints, SMS, and Telegram bots. However, these approaches have limitations, such as accuracy issues and constraints related to the datasets used [2-5].

Facial detection provides a more universal solution, as the face is central to both expression and recognition. Facial expressions can be categorized into seven types: happy, sad, disgusted, fearful, angry, surprised, and neutral. Recognizing these emotions is essential for various applications, including human-computer interaction and behavior monitoring [10-13].

The seven facial expression categories can be used to create a model for facial expression detection through transfer learning, using a dataset of labeled grayscale images corresponding to these expressions.

The goal of designing a door-opening system that detects facial expressions in real-time using transfer learning is to enhance convenience and provide a more comfortable way to access the door.

2. Methodology

A. Research Design

The following is a diagram of the research workflow to be conducted. For more details, please refer to Figure 1 below.

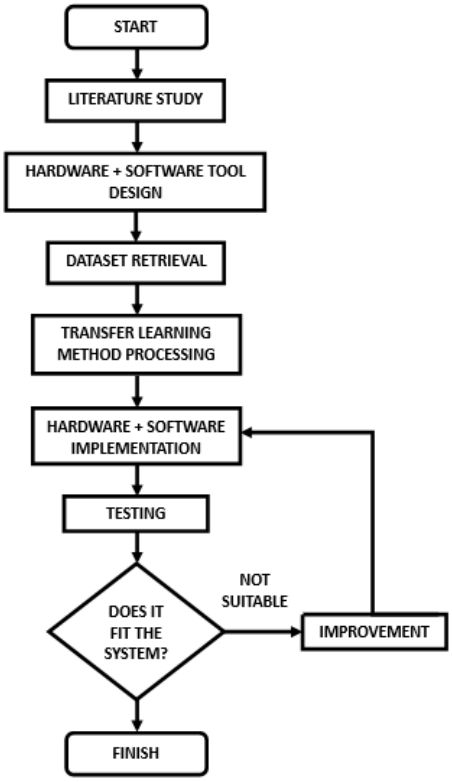

Figure 1. Research Flow

B. Mechanical Design

The mechanical design involves several stages, including designing the door and assembling various basic components. The design uses materials such as plywood and acrylic to create a sliding door similar to those at the Polibatam information center. For more details, please refer to Figures 2 and 3.

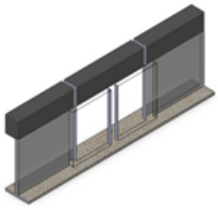

Figure 2. Back Door Design

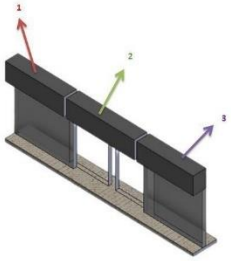

Figure 3. Front Door Design

Description:
1. DC Motor + Driver location
2. Location of Infrared Sensor 1
3. Location of Infrared Sensor 2

At the top of the door, the controller, DC motor, and driver can be installed to optimize space usage. The door will move according to the motor's direction. Motors are placed at the upper left and right ends, connected to pulleys via belts. The door is divided into two parts: the left door engages with the upper belt, and the right door engages with the lower belt. This setup ensures that the door moves in the opposite direction each time the belts rotate.

C. Electrical Design

In the electrical section, several key components are used, including the NodeMCU ESP32, two infrared sensors, an L298N motor driver, a 12V DC motor, a 12V battery or power supply unit (PSU), resistors, a webcam, and a PC or laptop. For more details, please refer to Figure 4 below.

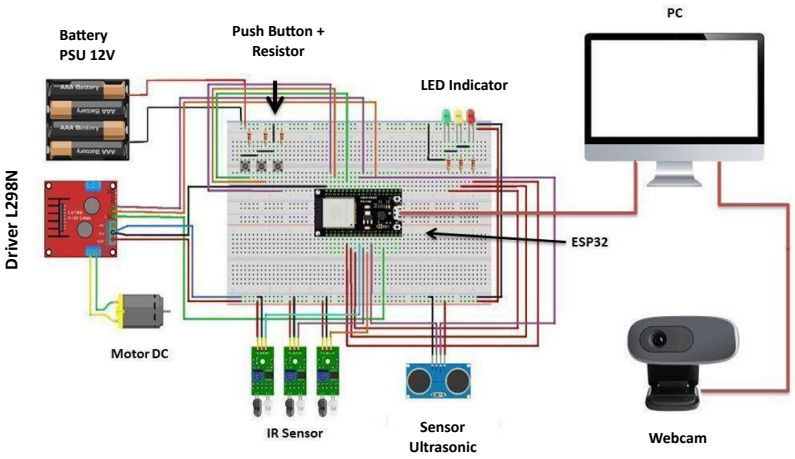

Figure 4. Electrical Wiring

D. Software Design

The dataset generated from the augmentation process can be used as training data for the transfer learning method. Transfer learning involves using a model that has been trained on one dataset as a starting point to solve a similar problem. This model is then adapted and updated to fit the new dataset [14].

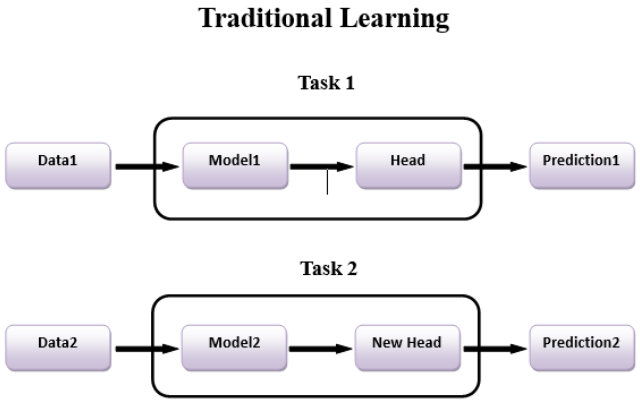

Figure 5. Transfer Learning [14]

Transfer learning can improve the accuracy or adjust the output of a machine learning model. Its benefits include the ability to develop models with high accuracy and to speed up the training process by utilizing a pre-trained model [15].

Transfer learning has two types, namely feature extractor and fine tuning.
a) Feature Extractor
The feature extractor is a transfer learning method that works by freezing the layers of the pre-trained model. This method only modifies the last layer of the pre-trained model to fit the new dataset. The final layer is then trained to produce output that aligns with the new dataset 14]. For more details can be seen in the picture 6 below.

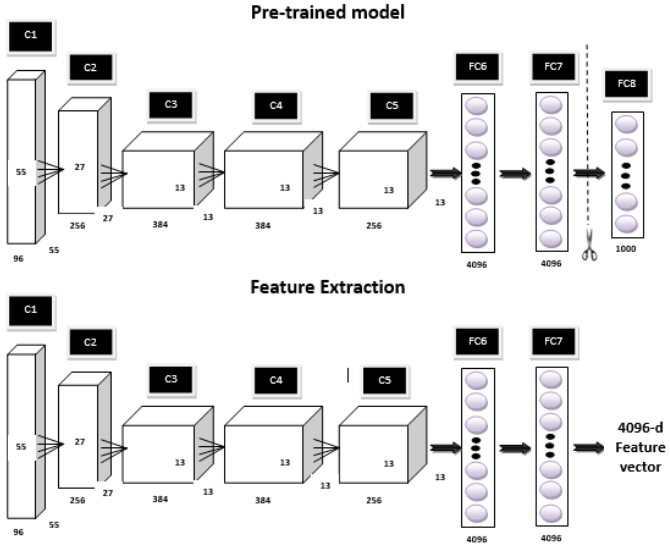

Figure 6. Feature extractor process

b) Fine Tuning

Fine-tuning is a transfer learning method that uses a pre-trained model without freezing its neural network layers. Instead, the fine-tuning method retrains the model on the unfrozen layers, allowing it to adapt and produce output according to the new dataset [14]. In this research, the type of transfer learning used is fine tuning. For more details can be seen in the picture 7 below.

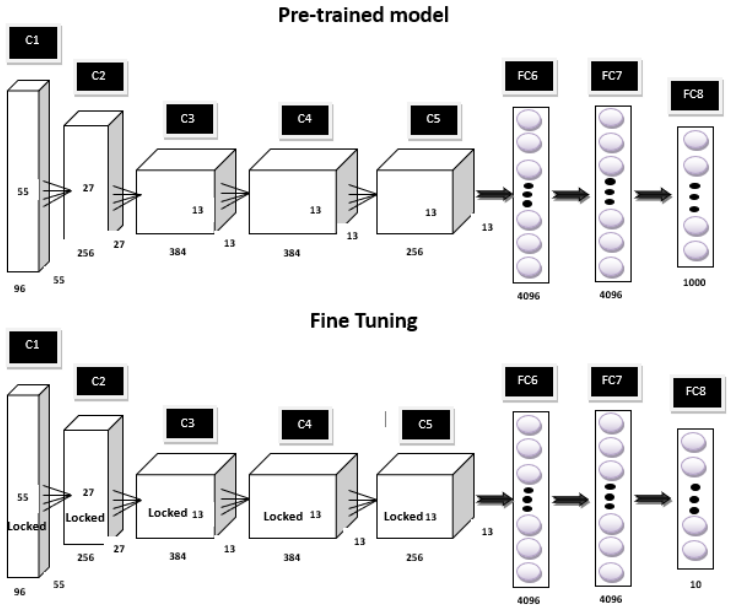

Figure 7. Fine Tuning Process

The following is the process of transfer learning used for training the model and detecting facial expressions overall.

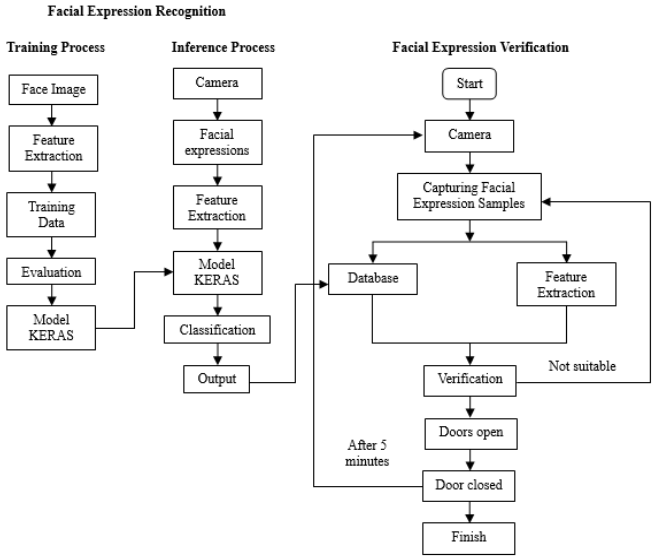

Figure 8. Transfer Learning Process

1) Facial Expression Design

This design represents the initial stage of the process for training models to recognize facial expressions. The model development is done in Jupyter Notebook using Anaconda and the TensorFlow library. This stage starts with data acquisition, followed by data training, evaluation, and finally model conversion, which is saved as a Keras model file. The formatted model is then integrated into the inference program, which is used to recognize facial expressions from the given data

a) Data Acquisition

Data acquisition is performed in real-time using a webcam. The table below shows the number of facial expression images saved in (.jpg) format, which will be used to classify seven facial expressions: angry, disgusted, fearful, happy, sad, neutral, and surprised. During the training

process, the data will be split into two sets: training data and testing data. For more details, please refer to the table 1 below.

Facial Expressions	Data Training (image)	Data Testing (image)
angry	3.993	960
disgust	436	111
fear	4.103	1,018
happy	7,164	1,825
sad	4.982	1,216
netral	4,938	1,139
surprise	3.205	797
Totals	**28.821**	**7.066**

Table 1. Image Dataset

Based on table 1, it can be concluded that the dataset with the largest number of images is the happy expression, with 7,164 images for training data and 1,825 images for testing data. The dataset with the fewest images is the disgust expression, with 436 images for training data and 111 for testing data.

For training this Transfer Learning method, the dataset used is the Facial Expression by AffecNet, which can be obtained directly from the author or downloaded from the Kaggle.com website.

1. Feature Extraction

Feature extraction is the process of obtaining facial images for detection using the HaarCascade Classifier provided by the OpenCV library. This classifier can also be downloaded from GitHub in XML format, which includes a pre-trained algorithm for detecting human faces. For more details, please refer to the image 9 below.

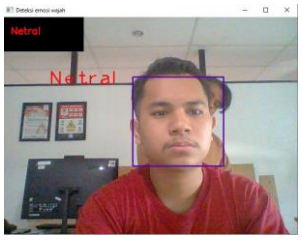

Figure 9. Face ROI and Bounding Box

2. Training Data

In the data training stage, the model architecture used is MobileNetV2 with the transfer learning method. The MobileNetV2 architecture consists of 157 layers, and the input size for the grayscale facial expression images is 224 x 224 pixels. For more details, please refer to the table 2 below.

Table 2. The following is a summary of the architecture of the MobileNetV2 model

No	Layer	Shape	Params
1	Input Layer	224,224,3	0
2	MobileNetV2 (*base model*)	(7,7,1280)	409.600
3	*Dense*	128	163.968
4	*Dense* 1	64	8.256
Total Parameters			**2.430.663**

3. Results and Discussion

In the model, training and validation are conducted over 25 epochs. The results of these epochs are presented in the table 3 below.

Table 3. Accuracy & Loss

EPOCH	TIME	LOSS	ACCURACY
1/25	406s 441ms/step	0.1572	0.2567
2/25	396s 441ms/step	0.0752	0.1986
3/25	395s 440ms/step	0.1951	0.2294
4/25	391s 436ms/step	0.3411	0.3513
5/25	391s 435ms/step	0.2571	0.2730
6/25	391s 436ms/step	0.2405	0.2897
7/25	391s 435ms/step	0.1886	0.3096
8/25	389s 434ms/step	0.1362	0.3300
9/25	389s 433ms/step	0.1915	0.3494
10/25	392s 436ms/step	0.1354	0.3698
11/25	389s 433ms/step	0.1778	0.3893
12/25	388s 433ms/step	0.1351	0.4069
13/25	5810s 6s/step	0.1887	0.4256
14/25	397s 442ms/step	0.1254	0.4502
15/25	391s 435ms/step	0.1022	0.4565
16/25	55928s 62s/step	0.1663	0.4701
17/25	393s 438ms/step	0.3287	0.4851
18/25	393s 438ms/step	0.3104	0.4898
19/25	392s 437ms/step	0.2713	0.5055
20/25	395s 440ms/step	0.2574	0.5085
21/25	2450s 3s/step	0.2376	0.5168
22/25	393s 438ms/step	0.2179	0.5243
23/25	392s 436ms/step	0.2037	0.5291
24/25	1896s 2s/step	0.1930	0.5333
25/25	392s 436ms/step	0.1832	0.5365

Based on the table above, the following graph has been created, representing 25 epochs of training. For more details, please refer to the figure 10 graph below.

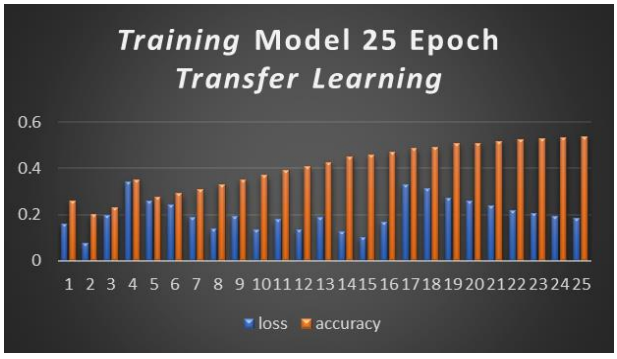

Figure 10. Graph of Training Loss and Accuracy Results on MobileNetV2

As shown in the figure above, the highest training loss value is 0.3411 at epoch 4, while the lowest is 0.0752 at epoch 2. Analyzing the graph, the highest training accuracy value reaches 0.5365, as indicated by the orange line at epoch 25, and the lowest accuracy value is 0.1986 at epoch 2. Therefore, based on this analysis, epoch 25 is considered the best epoch due to its high accuracy.

Following 25 epochs of model training, accuracy can be evaluated using a confusion matrix and a classification report from the transfer learning method. For more details, please refer to the figure 11 below.

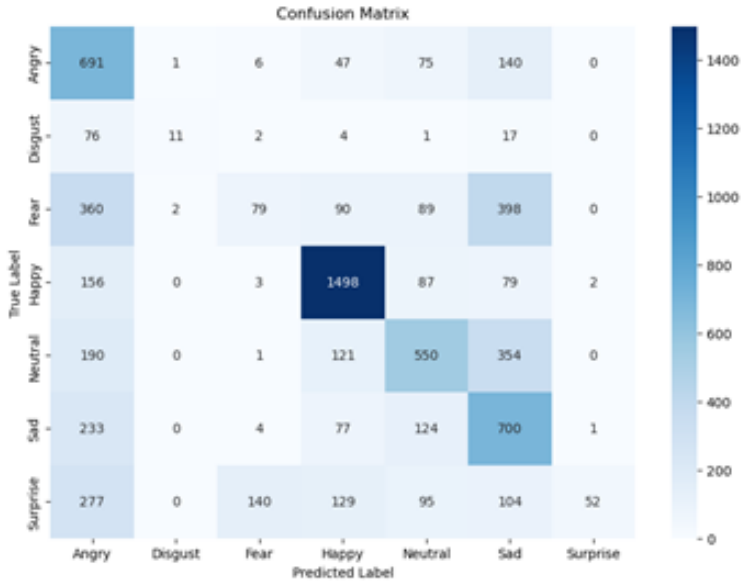

Figure 11. Confusion Matrix Transfer Learning

The confusion matrix is shown in the image above. From a total of 7,066 images, the data is as follows: 691 "angry" expressions were correctly identified as "angry," 11 "disgust" expressions were identified as "disgust," 79 "fear" expressions were identified as "fear," 1,498 "happy" expressions were identified as "happy," 550 "neutral" expressions were identified as "neutral," 700 "sad" expressions were identified as "sad," and 52 "surprise" expressions were identified as "surprise." These values are used to create the classification report, which is illustrated in the figure 12 below.

	precision	recall	f1-score	support
Angry	0.35	0.72	0.47	960
Disgust	0.79	0.10	0.18	111
Fear	0.34	0.08	0.13	1018
Happy	0.76	0.82	0.79	1825
Neutral	0.54	0.45	0.49	1216
Sad	0.39	0.61	0.48	1139
Surprise	0.95	0.07	0.12	797
accuracy			0.51	7066
macro avg	0.59	0.41	0.38	7066
weighted avg	0.57	0.51	0.46	7066

Figure 12. Classification Report Transfer Learning

From the figure above, we can see the accuracy, precision, recall, and F1-score values for the transfer learning method using the MobileNetV2 model. The accuracy is 51%, the average precision is 59%, the average recall is 41%, and the average F1-score is 38%.

Based on these results, it can be concluded that transfer learning performs relatively well in detecting facial expressions.

I. Testing

Testing was conducted using an HP ProBook 430 G1 laptop with the following specifications:

Table 4. Laptop Specifications

No	Spesifikasi Laptop	
1	Device name	WINDOWS-66A5858
2	Processor	Intel Core (TM) i7-4500U CPU @ 1.80GHz - 2.40 GHz
3	RAM	8.00 GB
4	System type	64-bit operating system, x64-based processor
5	Cam Resolution	1280 x 720
6	Horizontal resolution	96 dpi
7	Vertical resolution	96 dpi

3.1 Sample Reading with Conditions

The tests conducted involved reading facial expressions under different environmental conditions. The results of these tests are shown in the table 5 below.

Table 5. Test Results of Sample Reading Under Conditions

No	Expression Input	Condition	Results
1	Angry	Indoor + Lamp	Detected
2	Disgust	Indoor + Lamp	Undetected
3	Fear	Indoor + Lamp	Detected
4	Happy	Indoor + Lamp	Detected
5	Neutral	Indoor + Lamp	Detected
6	Sad	Indoor + Lamp	Detected
7	Surprise	Indoor + Lamp	Detected
8	Angry	Indoor + No Lights	Detected
9	Disgust	Indoor + No Lights	Undetected
10	Fear	Indoor + No Lights	Undetected
11	Happy	Indoor + No Lights	Detected
12	Neutral	Indoor + No Lights	Detected
13	Sad	Indoor + No Lights	Detected
14	Surprise	Indoor + No Lights	Undetected
15	Angry	Outdoor + Daytime	Detected
16	Disgust	Outdoor + Daytime	Undetected
17	Fear	Outdoor + Daytime	Detected
18	Happy	Outdoor + Daytime	Detected
19	Neutral	Outdoor + Daytime	Detected
20	Sad	Outdoor + Daytime	Detected
21	Surprise	Outdoor + Daytime	Detected
22	Angry	Outdoor + Night	Detected
23	Disgust	Outdoor + Night	Undetected
24	Fear	Outdoor + Night	Undetected
25	Happy	Outdoor + Night	Detected
26	Neutral	Outdoor + Night	Detected
27	Sad	Outdoor + Night	Detected
28	Surprise	Outdoor + Night	Detected

Based on the testing results in the table above, several expressions are clearly detected, including happy, sad, neutral, and angry, in both bright and dark environments. In bright conditions, expressions like surprised and scared are sometimes detected but are not recognized in dark environments. The disgust expression can be detected in bright conditions, though the results are inconsistent, with some instances of it being mistaken for fear or sadness.

The analysis indicates that lighting conditions significantly affect the detection results. In bright environments (indoors with lights and outdoors during the day), six out of seven expressions can be detected. In dark environments (indoors without lights and outdoors at night), the disgust and fear expressions are not well detected.

The light level used was 40 lux, measured with an Android app called Lux Light Meter Pro and a light meter. The light level used was 40 lux can be explained further depending on the environment. For indoor use the light level was 40 lux, which is usually suitable for dim indoor spaces, like rooms or hallways with low lighting. For outdoor use the light level was 40 lux, which is quite low for outdoor settings, similar to the light during early morning or at dusk. This means the statement describes the lighting level, whether for a low-light indoor environment or an outdoor setting with very little light.

3.2 Reading Facial Expressions at Different Distances

The testing involved reading facial expressions at various distances. The results of these tests are shown in the table 6 below.

Table 6. Test Results of Facial Expression Reading with Distance

No	Distance to Camera	Expression Input	Results
1	30 cm	Happy	Detected
2	50 cm	Happy	Detected
3	80 cm	Happy	Detected
4	100 cm	Happy	Detected
5	130 cm	Happy	Detected
6	150 cm	Happy	Detected
7	180 cm	Happy	Detected
8	200 cm	Happy	Detected
9	230 cm	Happy	Detected
10	250 cm	Happy	Detected

Based on the results in the table above, this method is effective at detecting facial expressions even at significant distances. The maximum distance tested was 250 cm from the camera to the person. In this test, only the happy expression was evaluated. This was because the motor was used directly during testing, simplifying the process. Additionally, the happy expression is used as the input for the motor actuator or automatic door opener.

3.3 Facial Expression Reading of Motorbike Movement

The test involved reading facial expressions in relation to motor movement. This testing was conducted to ensure that the detected expression corresponds to the expected system input. The results are shown in the table 7 below.

Table 7. Test Results of Facial Expression Reading against Motor Movement

No	Sample	Expression Input	Motor condition
1		Angry	Off
2		Disgust	Off
3		Fear	Off
4		Happy	On
5		Neutral	Off
6		Sad	Off
7		Surprise	Off

This test was conducted to confirm that the desired expression (happy) can successfully activate the motor or open the door. As shown in the testing table above, the happy expression successfully triggers the motor as intended. Conversely, if the expression is not happy, the motor does not move, and the door remains closed.

3.4 Testing the Tool

1. Face Detection Accuracy
   - Face detection accuracy reaches 51%.
   - Using transfer learning with the MobileNetV2 model which can provide quite good performance.
   - Transfer learning is less good at detecting given facial expressions if it is in a room that has less or dim lighting.
2. Door opening/closing practice

- The open/close accuracy of this method is quite good and precise because it is supported by the model that is owned by
- If the environment is poorly lit, errors in detecting facial expressions may cause the door to open or close incorrectly.

The results of this study align closely with the original questions and objectives outlined in the Introduction. This research aims to assess the effectiveness of the transfer learning method for detecting facial expressions in an ESP32-based door opener system. The findings provide insight into how well transfer learning can be applied to facial expression detection for such applications.

Discussing the research results facilitates drawing clear conclusions. Evaluating the performance of the transfer learning method in facial expression detection allows us to make well-supported conclusions about its effectiveness for door opener systems using ESP32. A thorough discussion aids in formulating concrete and measurable conclusions.

Each result should be accompanied by a detailed interpretation, particularly regarding the application of transfer learning methods. These interpretations help explore the implications of the research findings for using transfer learning in facial expression detection for door opener systems. A deep understanding of the method's effectiveness enhances confidence in the research outcomes.

Comparing the results with previous studies is crucial for assessing the contribution of transfer learning methods in facial expression detection. Are our findings consistent with other research? Focusing on the transfer learning method allows us to identify the unique contributions of this study and any differences in its implementation.

While transfer learning offers numerous advantages, it also has limitations that should be acknowledged. For instance, data or computational limitations associated with the use of ESP32. Recognizing these limitations provides an honest perspective on the constraints and opportunities for further development in facial expression detection using transfer learning methods.

4. Conclusion

This research represents a significant advancement in facial expression detection for automatic door opening systems, utilizing transfer learning methods specifically designed for the ESP32 platform. The successful implementation of this method enhances user interaction by providing a more seamless and intuitive experience, while introducing a new security-focused solution through facial expression detection.

The study demonstrated that transfer learning methods, particularly using the MobileNetV2 model, yielded satisfactory results. The model achieved an accuracy of 51%, with an average precision of 59%, an average recall of 41%, and an average F1-score of 38%.

Based on these results, it can be concluded that transfer learning is effective in detecting facial expressions. Testing across various environmental conditions and distances showed that the system performs well in different situations, although some limitations were observed, particularly in dark environments.

The PC, which relies on the classification method as its main processor, responds more effectively when the visitor directs their face towards the webcam. The accuracy of detection is further enhanced by using a high-resolution webcam, a PC with robust specifications, and adequate lighting.

In conclusion, this research confirms that the use of transfer learning for facial expression detection in an ESP32-based door opener system is effective. However, there is potential for further improvement, especially in adapting to more complex environmental conditions and expanding the system's applications. Future research should focus on optimizing the model and evaluating its performance in broader usage scenarios.

This research is far from perfect. Departing from this problem, the author provides suggestions to researchers with the same theme as follows:

1. Improve safety of the door system

This research aims to identify the optimal method for the system's future use. However, the current hardware design has several shortcomings, particularly regarding security. One potential enhancement is the addition of an ultrasonic sensor to prevent the door from closing abruptly while a visitor is still standing in the doorway.

2. Make the system integrated with IoT

Integrating IoT technology can enhance the system by incorporating advanced features. The author suggests adding an IoT system to the tool to monitor visitor numbers, perform troubleshooting on components, and more. This enhancement would transform the tool from a mere automatic door into a comprehensive system capable of evaluating visitor traffic at specific locations.

Reference

[1] Nugroho, P. A., Fenriana, I., & Arijanto, R. (2020). Implementation of deep learning using convolutional neural networks (CNN) for human expression.

[2] Broto, S., Muqod, S., & Fath, N. (2023). Electronic lock access control system for home security using RFID-based e-KTP. Techno.Com, 22(1), 167–175. https://doi.org/10.33633/tc.v22i1.6964

[3] Tahir, A. (2023). Design and development of teaching media for fingerprint implementation on sliding library doors. Volume 06(Issue 01).

[4] Wivanius, N., Wijanarko, H., & Ramadhan Novian, T. (2019). Locker security system based on GSM module, Bluetooth module, and reed sensor. Journal of Electrical and Mechanical Engineering, 5(1), 38–47. https://doi.org/10.35143/elementer.v5i1.2513

[5] Heranof, M. R., & Yendri, D. (2023). Automatic cat cage device based on microcontroller with Telegram monitoring. CHIPSET, 4(01), 71–79. https://doi.org/10.25077/chipset.4.01.71-79.2023

[6] Nuraeni, N., et al. (2021). Door access system based on face recognition using ESP32 module and Telegram application. Jurnal Mediatika, 4(3), 115. https://doi.org/10.26858/jmtik.v4i3.23700

[7] Goleman, D. (2002). Emotional intelligence.

[8] P. Ekman, “Universals and Cultural Differences in Facial Expressions of Emotion BT - Nebraska Symposium on Motivation,” Nebraska Symposium on Motivation, vol. 19. pp. 207– 282, 1972, [Online]. Available: papers3://publication/uuid/FDC5E29A-0E28-4DDF-B1A4- F53FEE0B4F70.K. Elissa, “Title of paper if known,” unpublished.

[9] C. Dalvi, M. Rathod, S. Patil, S. Gite, and K. Kotecha, “A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets and Future Directions,” IEEE Access, vol. 9, pp. 165806–165840, 2021, doi:

[10] S. D. Walsh et al., “Physical and Emotional Health Problems Experienced by Youth Engaged in Physical Fighting and Weapon Carrying,” PLoS One, vol. 8, no. 2, p. e56403, Feb. 2013, doi: 10.1371/journal.pone.0056403.
[11] World Health Organization : WHO, “Youth violence,” Oct. 11, 2023. Accessed: Dec. 07, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/youth-violence 10.1109/ACCESS.2021.3131733.
[12] N. K. Benamara, M. Val-Calvo, J. R. Alvarez-Sanchez, A. DiazMorcillo, J. M. F. Vicente, E. Fernandez-Jover dan T. B. Stambouli, “Real-Time Emotional Recognition for Sociable Robotics Based on Deep Neural Networks Ensemble,” 2019.
[13] A. Karpathy, “Convolutional Neural Networks for Visual Recognition,” Stanford University, [Online]. Available: http://cs231n.github.io/. [Diakses 30 March 2020].
[14] Wajah, D., Suhu, P., Berbasis, T., Termal, K., Dinan, S., & Fauzan, A. (2023). Transfer learning approach for the system. Volume 10(Issue 5).
[15] N. Donges, <What Is Transfer Learning? Exploring the Popular Deep Learning Approach.,= Builtin, 25 Agustus 2022. [Online]. Available: https://builtin.com/data-science/transfer-learning.