The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi

Conference Paper · November 2020

DOI: 10.1109/ICIC50835.2020.9288560

Lutfiah Zahara
Gunadarma University

Purnawarman Musa
University of Gunadarma, Depok, Indonesia

Eri Prasetyo
Gunadarma University

Saiful Musa
Politeknik Gorontalo

The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi

1st Lutfiah Zahara
Departmen of Computer Science
Gunadarma University
Depok, Indonesia
lutfiahzahara@student.gunadarma.ac.id

2nd* Purnawarman Musa
Departmen of Computer Science
Gunadarma University
Depok, Indonesia
p_musa@staff.gunadarma.ac.id

3rd Eri Prasetyo Wibowo
Departmen of Computer Science
Gunadarma University
Depok, Indonesia
eri@staff.gunadarma.ac.id

4th Irwan Karim
Departmen of Informatic Engineering
Gorontalo Polytechnic
Gorontalo, Indonesia
irwankarim@poligon.ac.id

5th Saiful Bahri Musa
Departmen of Informatic Engineering
Gorontalo Polytechnic
Gorontalo, Indonesia
saiful.bm@poligon.ac.id

*Corresponding author: Purnawarman Musa (p_musa@staff.gunadarma.ac.id, http://p_musa.staff.gunadarma.ac.id)

***Abstract*—One of the ways humans communicate is by using facial expressions. Research on technology development in artificial intelligence uses deep learning methods in human and computer interactions as an effective system application process. One example, if someone does show and tries to recognize facial expressions when communicating. The prediction of the expression or emotion of some people who see it sometimes does not understand. In psychology, the detection of emotions or facial expressions requires analysis and assessment of decisions in predicting a person's emotions or group of people in communicating. This research proposes the design of a system that can predict and recognize the classification of facial emotions based on feature extraction using the Convolution Neural Network (CNN) algorithm in real-time with the OpenCV library, namely: TensorFlow and Keras. The research design implemented in the Raspberry Pi consists of three main processes, namely: face detection, facial feature extraction, and facial emotion classification. The prediction results of facial expressions in research with the Convolutional Neural Network (CNN) method using Facial Emotion Recognition (FER-2013) were 65.97% (sixty-five point ninety-seven percent)**

***Index Terms*—Micro-expression, Facial Emotion Prediction, Convolutional Neural Network (CNN), FER-2013 Dataset, Artificial Intelligence**

I. INTRODUCTION

Information technology is generating along development of applications that facilitate human life and work. Today's trend in the development of modern information technology is artificial intelligence technology or commonly known as Artificial Intelligence (AI). Technological developments in the modern era like today have changed when a new form of interaction is implemented such as the use of buttons, monitor screens, touch screens, to voice commands that function to help social work, as well as in applications such as security control systems, interfaces between humans and computers, is also the first step in human face recognition (Human Recognition Face). The research [1] proposes an EEG signal filtering process using the Empirical Mode Decomposition (EMD) and Wavelet Packet Decomposition (WPD) methods in shaping human emotional feature data. One approach to understanding user responses is a facial expression recognition system to understand users' responses [2].

Facial expression is the result of a facial gesture or expression that shows the position of the muscles on the human face as a form of non-verbal communication as well as an essential way of expressing one's emotions as a form of feelings, intentions, goals, and opinions to others [3]. Currently, the process of predicting facial expression scoring systems for services and services has been widely used by computers by manually selecting emotions for the level of satisfaction of a service displayed on a computer screen. Sometimes a person's emotional level by selecting on the screen is considered inaccurate in showing an expression of customer satisfaction [4]. The use of real-time facial expression detection using artificial intelligence in a system can increase accuracy. So that it can directly detect facial expressions from consumers when served at the cashier, for example.

The authors propose the research aims to test the accuracy of the successful predictions of 7 (seven) facial micro-expressions, namely happy, sad, angry, scared, surprise, contempt, and disgust in real-time by utilizing the Facial Emotion Recognition dataset or (FER-2013).

II. State of the Art

A. Computer Vision and Face Object Detection

Computer Vision is a process of transformation or change from data originating from a video camera or photo or image into a decision result or a new presentation, where the results of the transformation have an interest in achieving a goal. The image results obtained on the camera device must meet the standards to get the image quality obtained, including the speed and performance of embedded processing [5]. One of the applications of computer vision concepts is to perform image segmentation techniques for the classification process of an image object, such as the classification of the physical shape and size of chicken eggs [6]. Another computer vision concept is the Facial Recognition system, which allows algorithms in Computer Vision to detect features on a human face and compare them with the detected face's self-profile, starting from the name, gender, and age.
Computer Vision, which is in charge of Facial Recognition, can be used to authenticate the identity of the face owner in order to maintain data security [7]–[9].

The face is one part of the face with a unique shape from the shape of the human body—starting from raising eyebrows, squinting, forming curves on the lips, the direction of the gaze with the eyes, etc. The expression itself in KBBI means disclosure or the process of stating (i.e., showing or stating intentions, ideas, feelings, and so on) [10]. In this modern era, many methods are used to read human facial expressions for various purposes, one of which is psychology, to read the facial expressions of patients accurately. As well as in assessing consumer satisfaction with a given product or service, only with the expressions that humans show the system will detect the facial expressions of consumers.

The face detection process in this study uses the Haar Cascade Classifier method. The term Haar denotes a mathematical function (Haar Wavelet) in the form of a box. Initially, image processing was only by looking at each pixel's RGB value, but this method turned out to be ineffective. Viola and Jones developed an image processing that formed the Haar-Like feature. The Haar-like feature processes images in squares, wherein one box there are several pixels. The per box is then processed and produces different values that indicate dark and light areas. These values will use as the basis for image processing. The image part of the face will automatically be detected so that the face position will adjust to the face data in each image.

B. Facial Expression Recognition 2013 (FER-2013)

The 2013 Facial Expression Recognition dataset (FER-2013) is a dataset provided by Kaggle, introduced at the International Conference on Machine Learning (ICML) in 2013 [11] introduced by Pierre-Luc Carrier and Aaron Courvill.

In this dataset, each face has been categorized based on

Fig. 1: Image in Dataset FER-2013 (source: [12])

emotion categories, where the FER-2013 dataset is a grayscale image measuring $48_{pixel} \times 48\ _{pixel}$ for each image. The total FER-2013 dataset is 35,887 consisting of 7 (seven) different types of micro expression, and marked with labels based on 7 (seven) different classifications starting from the index label 0 to 6, which is described in Table I.

TABLE I: Number of data in the FER-2013 dataset

micro-expression (Classification)	Validation Data		Training Data	Dataset Total
	Public	Private		
Angry	467	491	3995	4953
Disgust	56	55	436	547
Fear	496	528	4097	5121
Happy	895	879	7215	8989
Sadness	653	594	4830	6077
Surprise	415	416	3171	4002
Contempt	607	626	4965	6198
	3589	3589	28709	**35887**

C. Micro Classification of Facial Expressions

A microexpression is a facial expression that can easily observe and distinguish it as a communication method in social psychology [13]. Facial expressions provide information about emotions, convey our intentions and goals, and play an essential role in human interaction. The ability to recognize and understand facial expressions automatically facilitates the intended communication.

The process in the classification of human facial expressions consists of three stages: face detection, feature extraction, and facial expression classification. In this study, the authors applied a system that could classify facial expressions at a macro level, consisting of 7 (seven) basic human expressions:

1) Happy
   A smile expression is an expression that can show that someone is feeling happiness or liking something. The happy expression is on the upward movement of the cheek muscles and the sides or edges of the lips to form a smile.

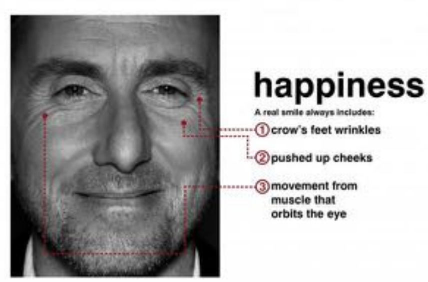


Fig. 2: Characteristic of happy expressions (source: [14])

2) Anger
Anger facial expressions arise from the match between what is expected and a reality. The expression is shown on both sides of the inner eyebrows which are merging and leaning down, while the lips are narrowing, and the way the eyes are sharp when looking.

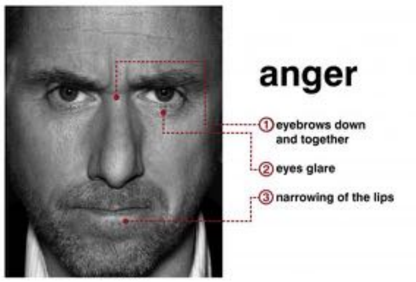


Fig. 3: Characteristic of anger expression (source: [14])

3) Sadness
A face that shows sadness appears when disappointment or a feeling of missing something. Based on the characteristics of a sad facial expression when the eye loses focus, the lips are pulled downwards, and the upper eyelid droops.

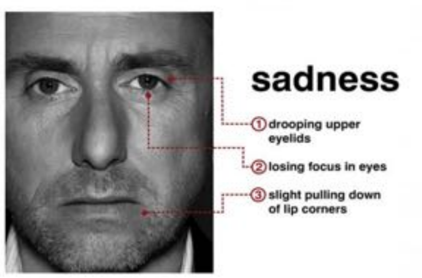


Fig. 4: Characteristic of sad expression (source: [14])

4) Fear
The form of expression that appears when someone experiences an inability to cope with any event or in a scary atmosphere, then that person is said to be afraid. The expression of fear on a person's face is seen from the two eyebrows that rise at the same time, the eyelids tighten, and the lips that are open horizontally.

Fig. 5: Characteristic expression of fear (source: [14])

5) Disgust
A person who expresses his face in a state of disgust due to seeing something not common or listening to information that is not worth hearing. An expression of disgust will be read when a person's face in the area of the nose bridge is wrinkled and the upper lip rises.

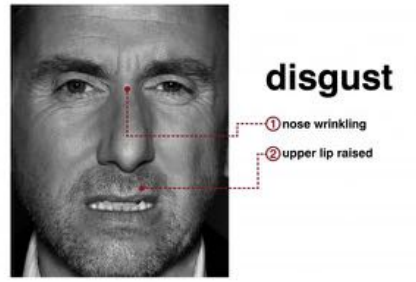


Fig. 6: Expression of disgust (source: [14])

6) Surprise
Expressions of surprise are obtained when someone does not know beforehand an event or message received that is sudden, unexpected or important. Expression is a shocked face represented by the raised eyebrows, the eyes wide open, and the mouth opening reflex.

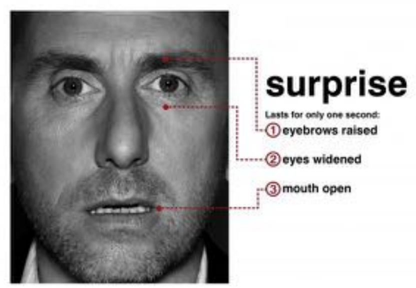


Fig. 7: Characteristic expression of surprise (source: [14])

7) Contempt The facial expression of a person who is identified as being arrogant and has no sense of respect at the other person often underestimates other people. The expression is shown from a movement that raises one corner of the lips.

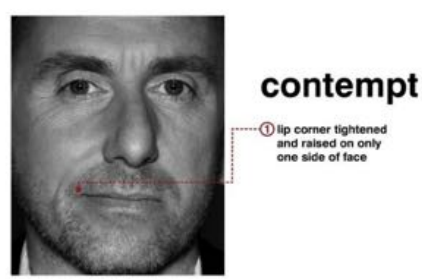


Fig. 8: Characteristics of contempt expression (source: [14])

D. Artificial Intelligence (AI)

Artificial Intelligence, commonly known as Artificial Intelligence (AI), is a computer science that studies how a computer can do work and done by humans. AI is a computer system that forms to identify and model human thought processes & design machines to mimic human behavior [15]. Humans can cleverly solve problems that arise because humans have knowledge and experience that can help in solving problems. So that computers can act like and as well as humans, computers given the knowledge and ability to reason to get experiences like humans. There are three AI goals: to make computers smarter, understand intelligence, and make machines more useful. Intelligence is the ability to learn or understand from experience, understand contradictory and ambiguous messages, respond quickly and well to new situations, use reasoning to solve problems, and solve them effectively [16]. There are two main parts needed in order to be able to carry out an AI [15] application, namely:

1) Knowledgebase, which contains facts, theories, thoughts, and relationships with one another.
2) Inference engine, which is the machine's ability to draw conclusions based on experience.

E. Convolutional Neural Network and Deep Learning

Convolutional Neural Network (CNN) is an algorithm included in the deep neural network or Deep Learning family due to the high network depth and is significantly superior when implemented on [17] image data. Deep Learning (DL) is a Neural Networks technique that uses specific techniques such as Restricted Boltzmann Machine (RBM) to accelerate the learning process in Neural Network, which uses multiple layers or more than seven layers. With the existence of Deep Learning, the time required for training will be less because the problem of losing the gradient in backpropagation will be lower [18].

The initial research underlying the CNN discovery was first carried out by Hubel and Wiesel, who carried out visual cortex studies of the cat [19] visual sense. The visual cortex in animals is very powerful in existing visual processing systems until many studies have been inspired by how it works and has produced new models such as Neocognitron, HMAX [20], LeNet-5, and AlexNet.

The CNN method results from the development of the Multilayer Perceptron (MLP) method for two-dimensional processing data, for example, images or sounds [21].
The way CNN works is similar to MLP, but on CNN, each neuron present in two dimensions. Unlike MLP, where each neuron is only one dimension. Image processing can be started as a specific feature, such as brightness or edge increasing complexity on features that uniquely define objects according to layer thickness. In general, the existing layer types in Convolutional Neural Network divide into two, namely Extraction Layers Features (Feature Extraction Layer) and Classification Layer (Classification Layer).

The CNN layers have a 3-dimensional arrangement of neurons (width, height, depth). Width and height are the sizes of the layers, while depth refers to the number of layers. A CNN can have tens to hundreds of millions of layers, each of which learns to detect various images. Image processing applies to each training image at a different resolution, and the output of each image is processed and used as input to the next layer, as shown in figure 9.

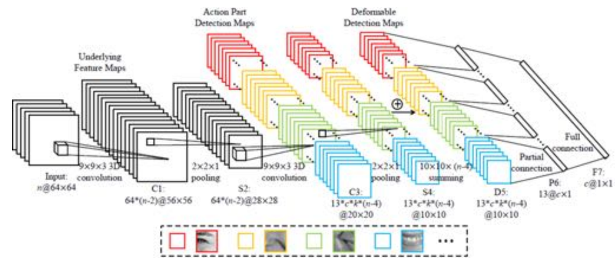


Fig. 9: Convolutional Neural Network Algoritm (source: [22])

F. CNN Model Training: Mini XCEPTION

The architecture used is designed based on the Xception model, which is one of the models developed from the general architecture of Convolution Neural Networks (CNN) which implements a convolution pattern and feature extraction by utilizing the additional layers used, namely a separate convolutional layer (Depthwise Separable Convolution) with Pointwise Convolution. Layer as an extractor with a mechanism to separate the feature extraction process and eliminate the fully-connected layer (Fully-Connected Layer). Some of the techniques used when training the CNN Mini XCEPTION model are Data Augmentation, Kernel Regularizer, Batch Normalization, Global Average Pooling, and Split Convolution.
Mini XCEPTION is one relatively small interesting architecture and achieves almost all of the facial expression classifier performance illustrated in the figure 10 [23].

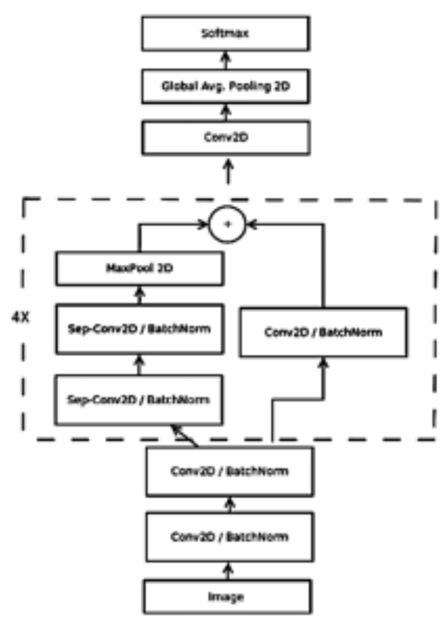


Fig. 10: Xception mini Architecture Model

The convolution layer design with a two-dimensional (2D) mechanism base on image input standards, wherein this layer there are dimensions of height (height) and width (width). However, in the convolution process, this layer stores parameters in the specified depth dimension (Depth) and produces the output tensor (dot-product) obtained from the kernel filter size of each dimension during the activation process. This establishment implements a regularization mechanism using the ReLU (Rectified Linear Unit) activation function, namely activation with a linear maximum value mechanism.

III. Design and Testing of Facial Micro Expression Systems

The first concept provides training data by evaluating the model to obtain an architectural model stored in a file with the extension .hdf5. The model evaluation process utilizes Facial Emotion Recognition (FER) from the FER-2013 dataset, operated on the Raspberry Pi mini-PC platform. Generally shown in the Figure 11 section to the left of the figure. After

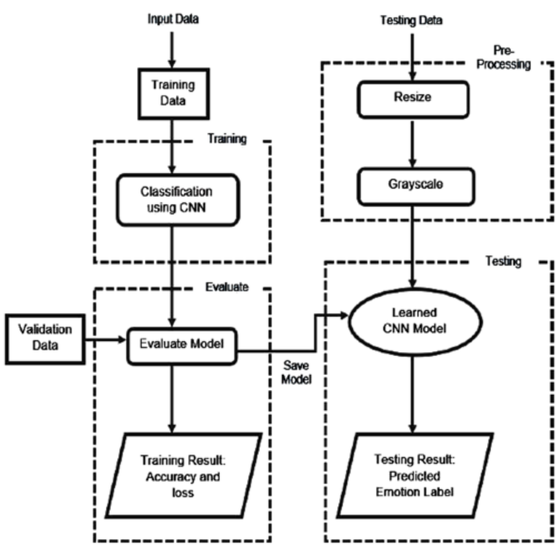


Fig. 11: Flowchart of Facial Micro Expression System

producing an architectural model in the .hdf5 file format, the second concept aims to increase low accuracy and reduce computational complexity. This study proposes learning by applying the Convolutional Neural Network (CNN) method using the Xception mini Architecture Model. The application of CNN carrier out with the intention is that the system carries out a classification learning process to predict 7 (seven) types of image results based on microexpressions on human faces. The second process is also operate using the Raspberry Pi as a processor to recognize facial expressions shown in Figure 11 to the right of the figure.

The research that the author has done to implement a microexpression system on human faces generally shown in Figure 11, where the design consists of 2 (two) stages, namely

1) Process Training
2) Testing Process

A. *Training process using dataset FER-2013*

The data training process in this study utilizing the FER-2013 dataset has been preprocessed and called the dataset with the file name fer2013.csv.
In the Figure 12, where the training process will be compiled and also perform data training that has been read. The training process is carried out with predetermined parameters (such as the value of epoch and batch_size) which will produce a trained model which is used as a prediction parameter and is stored in a file with the extension .hdf5. Based on the

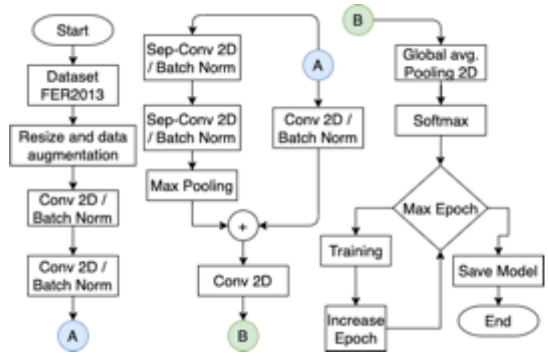


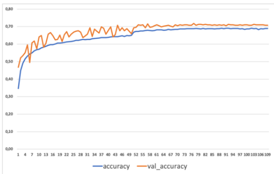
Fig. 12: FER-2013 Training dataset

picture 11 shown on the left of the Figure, the data training process algorithm is inputting training data and validation data. The training data is processed using the CNN algorithm to produce feature extraction that will be evaluated with data validation. The evaluation results will produce a trained model architecture to reach the maximum epoch value. The trained data model will then be used as a parameter for comparison with testing data (images obtained from the camera) to predict facial expressions for the real-time testing process. The CNN algorithm model in the training process, which combines the use of residual modules and combinations in a convolutional layer. The residual module modifies the selected mapping between the next two layers, so that features are extracted by means of a comparison mechanism between the original features and the selected features. As a result, the desired feature, namely H(x) is modified to make it easier to solve problems in the training process F(x) so that the equation is as follows:

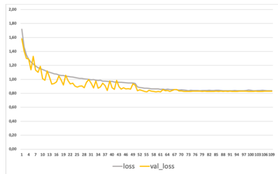
$$H(x) = F(x) + x \tag{1}$$

The architectural model (see Figure 10), has 4 (four) modules in a hidden layer, which contains a separate deep layer of convolution (Depthwise Separable Convolutions). Each convolution layer is followed by a batch normalization operation and a ReLU (Rectified Linear Unit) activation function. In the final layer, global average pooling will be carried out and activation of the Softmax function to generate predictions. This architectural model has about 60,000 parameters, with a reduction value of 10 times when compared to implementing CNN architectural models with fully connected layer removal functions, and a reduction value of 80 times when compared to standard CNN architectural models.

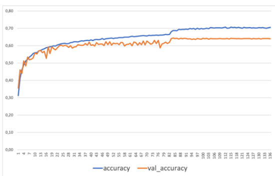
The data training process in this study was carried out with 2 training trials, where the first training process provided a test of 0.1 and the second training trial gave a testing value of 0.2. From the two experiments, the results of the accuracy and lost data are shown in Figure 13.



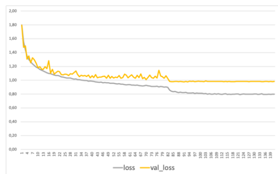
(a) Accuracy (d_test = 0.1)



(b) Lost (d_test = 0.1)



(c) Accuracy (d_test = 0.2)



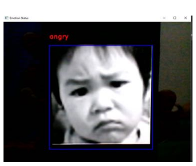
(d) Lost (d_test = 0.2)

Fig. 13: The result of Accuracy and Lost data

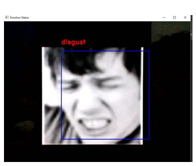
The stages of the training process shown in the Figure 13 show data accuracy and loss in the architectural model given a test value of 0.2. The use of test data with a value of 0.2 is optimal, where conceptually, in training an architectural model for lost data from the plot of the model training, when the value of the *train_lost* and *valid_loss* data in each number epoch decreases, the architectural model tries to understand the image data pattern. Furthermore, if the lost value is at peaks (increases), an epoch at the loss value-line indicates the pattern's difficulty in the training process.

While the training process for architectural models in finding the optimum of accuracy values (*train_accuracy* and *valid_accuracy*), when an accuracy value from the plot results increases for each epoch. If there is a line that results in valleys resulting in a decrease in the accuracy line's value, then the accuracy results indicate that the training process pattern is complicated.

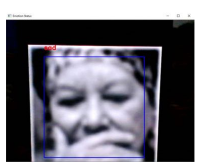
The first testing uses images from the FER-2013 dataset (some of the datasets in Figure 1). This experiment shows the results of the comparison and accuracy of the application in this research, where the results of detecting facial expressions are correct in the input image, as shown in Figure 14.



(a) an Angry



(b) a Disgust



(c) a Sad

Fig. 14: Test expressions to accuracy comparison using the dataset FER-20133

B. *Real-time testing to detect facial expressions*

The testing process is the process of detecting micro-expressions on faces that are carried out in real-time (see Figure 15). This process begins with preparing a tool for image input using a webcam and a feature module for detecting facial objects, namely the Haar Cascade Classifier method.

The testing process is shown in Figure 11 on the right of the image. The first step is to call and activate the webcam, which functions to capture the image. Each image frame will detect objects from the face. When the object from the face is recognized, the next step is to carry out pre-processing the image results from the face object. The pre-processing process continues with the image segmentation process using the CNN algorithm to predict microexpressions against facial input data. The trained model is defined as a facial expression prediction parameter on the system to carry out the labeling of the predicted image results and the display on the LCD.

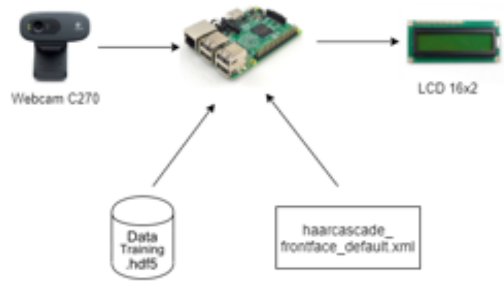


Fig. 15: Architecture of Micro Facial Expression System Design

The design shown in the system architecture in Figure 15 uses a Raspberry Pi. When the system detects a facial object as testing data, it will be used as input from the system for facial expression detection. The results of facial expression detection are shown in Figure 16.

Figure 16a represents an illustration of the experiment when the facial expression detection system is in the camera; the facial expression prediction results will display in an information module on the LCD. The system will also show the predicted results of images captured through the camera by labeling facial expressions found, namely happy expressions, as shown in Figure 16b.

(a) Experiment illustration

(b) Result of facial expression

Fig. 16: The process of predicting micro-expressions on faces

IV. Results Analysis

The design recognition of micro expressions on faces carried out in this study in several testing stages on the system.
The results showed that using the CNN architectural model in the facial expression detection system could be done optimally and in real-time. The evidence using a separate convolution layer when conducting data training to do training can carry out optimally, and the accuracy of the data in the trained model facial expression is an average of 65.97%. That is very necessary to analyze the results of the system that becomes implemented. The results of this test shown in the table II.

TABLE II: The Result of Facial Expression Testing

Testing in real-time	Corrected		Total	
	Yes	No	Test	(%)
Facial Expression	66	4	70	94.3
Face Position/Condition	159	86	245	64.9
Distance face versus Camera	109	66	175	62,3
From Image or Poster	104	141	245	42,4
Result Ekspremen (mean)				65.97

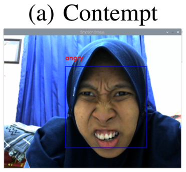
A. Facial Expression Prediction Test

This test aims to find out how successful the system is in recognizing facial expressions that have trained. The system's success rate in recognizing the user's face is 94.28%, so it can conclude that the system shows good results. In the Figure 17 is an example of the test results for facial expression recognition.
The experiment was carried out ten times for each expression based on the test results for sadness, surprise, contempt, and happy. The system successfully recognized all tests. While the results of the expression of anger had an error one time, the expression of disgust had an error two times, and the expression of fear of having an error one time.

(a) Contempt

(b) Happy

(c) Sad



(d) Angry

(e) Disgust

(f) Surprised

(g) Fear

Fig. 17: Test Results of Facial Expressions

The test results are ten analyses for each expression; then, the results are shown in Table III as the result of the confusion matrix. The table will show which expressions are easy to predict, often predictable, and hard to predict. The table III shows that the characteristics of the expression of disgust that get the wrongest results are two times out of 10 tries to predict facial expressions.

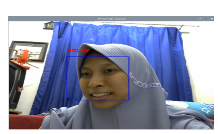
TABLE III: Confusion Matrix

		Number of Predictions						
		Contempt	Happy	Sad	Angry	Disgust	Surprise	Fear
micro-expression	Contempt	10	0	0	0	0	0	0
	Happy	0	10	0	0	0	0	0
	Sad	0	0	10	0	0	0	0
	Angry	0	0	0	9	1	0	0
	Disgust	0	0	0	2	8	0	0
	Surprise	0	0	0	0	0	10	0
	Fear	0	0	1	0	0	0	9

The factor that makes it very difficult to classify the prediction results for the expression disgust. Table II shows that the dataset FER-2013, especially the expression of disgust, has 547 data total. Other factors cause disgust to fail more because the facial expression is almost the same as the angry expression.

B. Facial Expression Recognition Test Based on the Condition or Face Tilt Position of the tester

In facial expression recognition testing based on conditions or the face's position tilted to the left or right, the face is looking up or looking down, and the face is turning left or right. The total number of tests based on the point of view of the camera is seven-position models. The purpose of the test is to pay attention to the results of the camera captured image. Facial expressions can recognize by the system based on the viewing angle of a camera.
The experiment to do five times in certain conditions and positions, and the classification of experimental expressions if there are seven classifications of expressions and 7 test positions so that the total test at this stage is 245 attempts. Based on the point of view captured by the camera, the correct prediction result was 159 times, while it failed 86 times. Then the success rate of the system in recognizing the overall facial expression is 64.89%. The effect of the number of unsuccessfulness is because the image contained in the FER-2013 dataset has very little variation in the viewpoint of the dataset. In both Figure 18 and Figure 19, the facial



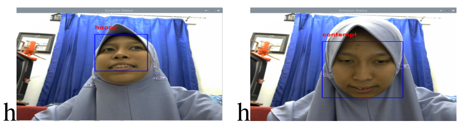
(a) Correct

(b) Incorrect

Fig. 18: Test facial expressions based on the image against the point of view from the camera

expression should be in a happy expression, which an example of a face test result based on the face position's condition. Figure 18a, the result can recognize the expression correctly when looking to the right, and the result when looking left at the Figure 18b, cannot recognize facial expressions because the result of prediction for facial expressions is contempt.

The other image results are shown in Figure 19a, where the position is looking up (can recognize expressions), and in Figure 19b, the tester is looking down cannot recognize facial expressions because the result of prediction for facial expressions is contempt too.



(a) Correct (b) Incorrect

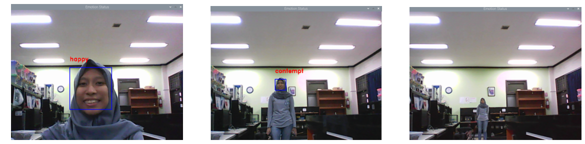
Fig. 19: Test facial expressions based on the image against the point of view from the camera

C. Facial Expression Recognition Test Based on the Distance of the tester's face from the camera

In testing facial expression detection based on the distance between the object and the camera, the user's face faces the camera perpendicularly with variations in the distance, namely 50 cm, 100 cm, 200, cm, 500 cm, and 1000 cm in the test. The purpose of testing is to determine the system's level of accuracy to recognize facial expressions displayed by a person from a certain distance.
The results of this test get an average presentation of 62,3%. The influencing factor is that the image results for 500 cm and more fail to detect faces and predict facial expressions.
The test is shown in Figure 20 is the test result of the happy expression. At 20a gets the correct result. Whereas in Figure 20b is an expression prediction error, and in Figure 20c does not get a prediction result because the system cannot detect the face object.

All tests are 175 attempts at this stage, where each facial expression (7 types of expression) tested five times.
Testing distance between object and camera from 50 cm to 200 cm, the system can detect most and recognize correct facial expressions 99 times out of 105 experiments. The testing distance of 500 cm allows the system to detect facial objects 22 times and detect and recognize facial expressions, and it only succeeded ten times out of 35 experiments. From the detected face detection process, 12 trials cannot recognize facial expressions.

However, testing at a distance of 1000 cm, even one expression by the system cannot detect the face, so the system cannot recognize facial expressions so that the total that can predict facial expressions correctly is 109 times out of 175 trials as a whole. In the Figure 20a, 20b, 20c are examples of the results of testing facial expressions based on the object's distance from the camera.



(a) Correct (b) Incorrect (c) Failed

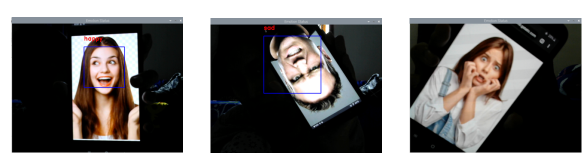
Fig. 20: Test facial expressions based on the distance between the object and the camera

D. Facial Expression Recognition Test Based on Photos or Face in Posters

In the last test, by looking for several photos of a person, the photo has a characteristic facial expression. The collection of photographs with the characteristic expression following this study through an image search engine, namely google search. Photos with facial expressions' characteristics will test by showing them as camera input with the camera's distance setting. The system will recognize faces based on camera distance in capturing facial objects, and the system will analyze facial objects using these photos to predict facial expressions or not.

Tests are performed 10 times, where the Figure 21a is corrected. The process for Figure 21b, an error occurs so that you can determine whether it is appropriate or not. Test process 3 shown in Figure 21c is an example of an unchecked result from a facial expression test result based on a photo or poster. Apart from being based on this distance, the photos will test by doing a specific angle rotation. This test aims to find out whether the system can succeed in recognizing facial expressions using photos by rotating the image with an angle of 0^o, 15^o, 45^o, 135^o, 225^o, 315^o, 345^o with the distance between the camera and the user's face that is 30 until 100 cm.

The system can detect and recognize seven facial expressions in the test with a photo at an angle of 0^o. The test with an angle rotation of 15^o and 345^o, the system can detect multiple photos with seven facial expressions. Test with the angular rotation of 45^o, 135^o, 225^o, and 315^o system cannot detect faces and also cannot recognize photos with seven facial expressions. In the Figure 21a, 21b and 21c are examples of the results of facial recognition testing based on photos or posters. The factors that affect the test results are because of



(a) Correct (b) Incorrect (c) Failed

Fig. 21: Facial expression recognition test based on photos or posters

the FER-2013 dataset. Generally, the image from the dataset rotated at typical degrees, around -15^o to +15^o. The influence of these factors results in this test to get an average of 42.4% presentation.

V. Conclusion

From the results of the trials that have carried out, overall, it has succeeded in designing a system with a general description of the research object using the Convolutional Neural Network (CNN) method for the prediction of 7 (Seven) Human Facial Expressions utilizing Facial Emotion Recognition (FER) from the Raspberry Pi-based FER-2013 dataset. The system design in this study uses the process described as follows:

1) The training process utilizes the Facial Expression Recognition 2013 Dataset or FER-2013 dataset and uses the Convolutional Neural Network (CNN) method as feature extraction and a suitable facial prediction process.
2) Direct facial expression recognition (real-time), facial objects found using the Haar Cascade method, and the Convolutional Neural Network or (CNN) method used to classify facial expressions.
3) When the facial expression recognition process is in progress will appear in the information viewer on the expression display board what the system detected.

Based on testing on the implementation of Deep Learning with the Convolutional Neural Network (CNN) method for the prediction of 7 (Seven) Human Facial Expressions utilizing Facial Emotion Recognition (FER) from the FER-2013 dataset based on Raspberry Pi, even though the system both hardware and software functions and works well. However, the test results have system weaknesses, including:

1) Uses a more convolutional layer and fully connected layer with the appropriate configuration, so get a deeper CNN architecture capable of better accuracy.
2) Use a new dataset or enter more training data so that when the data tested, it will produce higher accuracy, especially for tests with a distance of more than 5 meters, viewing angles and when photos rotated.
3) Choosing hardware such as a digital camera that has a higher resolution and has features such as autofocus to get a clear image even though the object is in motion to improve network performance before entering the model is good for detection as well as for recognition.

Acknowledgment

Gunadarma university fully supported this research collaborated with the Department of Informatics Engineering at Gorontalo Polytechnic. The authors gratefully acknowledge all participants of user testing during the experiment this research. Specially gratefully acknowledge to Universitas Gunadarma for permission to research activity to using the laboratory's research facilities at Gunadarma University, in Depok city, Indonesia.

References

[1] S. B. Musa and H. Tjandrasa, "Analisis fitur sinyal emosi eeg berdasarkan hybrid decompotion," *ENERGY*, vol. 7, no. 1, pp. 7–12, 2017.

[2] A. Saputra and T. A. B. Wirayuda, "Pengenalan ekspresi wajah menggunakan local binary pattern (lbp)," *eProceedings of Engineering*, vol. 2, no. 3, 2015.

[3] I. G. P. S. Wijaya, A. A. Firdaus, A. P. J. Dwitama, and Mustiari, "Pengenalan Ekspresi Wajah Menggunakan DCT dan LDA untuk Aplikasi Pemutar Musik (MOODSIC)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 2018.

[4] Z. Abidin, "Menggunakan Fisherface dan Jaringan Syaraf Tiruan Backpropagation," *Jurnal MIPA*, 2012.

[5] P. Musa, S. A. Sudiro, E. P. Wibowo, S. Harmanto, and M. Paindavoine, "Design and implementation of non-linear image processing functions for CMOS image sensor," *Optoelectronic Imaging and Multimedia Technology II*, vol. 8558, no. 0, p. 85580O, 2012.

[6] T. A. Wijaya and Y. Prayudi, "Implementasi Visi Komputer dan Segmetasi Citra," in *SNATI 2010*, 2010.

[7] P. Musa, N. Nuryuliani, and M. Lamsani, "Rancang bangun pengendali pintu automatis berdasarkan ciri wajah menggunakan metode euclidean distance dan fuzzy c-mean," *Jurnal Ilmiah Informatika dan Komputer*, vol. 13-1, 2008.

[8] A. L. Kurniawan, R. R. Isnanto, A. A. Zahra, K. Kunci, P. Wajah, V. Komputer, A. Wavelet, T. Gabor, and J. Euclidean, "Perancangansistem Pengenalan Wajah Menggunakan Metode Ekstraksi Ciri Susunan Tapis Wavelet Gabor 2D Dengan Jarak Euclidean," *Perancangansistem Pengenalan Wajah Menggunakan Metode Ekstraksi Ciri Susunan Tapis Wavelet Gabor 2D Dengan Jarak Euclidean*, 2015.

[9] A. L. Ramadhani, P. Musa, and E. P. Wibowo, "Human face recognition application using PCA and eigenface approach," *Proceedings of the 2nd International Conference on Informatics and Computing, ICIC 2017*, vol. 2018-Janua, pp. 1–5, 2018.

[10] J. Cristanto and K. R. R. Wardani, "Penerapan Metode Single-Layer Feed-Forward Neural Network Menggunakan Kernel Gabor untuk Pengenalan Ekspresi Wajah," *Jurnal Telematika,, Institut Teknologi Harapan Bangsa, Bandung*, 2017.

[11] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, and D.-H. &. others Lee, "Challenges in representation learning: A report on three machine learning contests," *International Conference on Neural Information Processing*, 2015.

[12] C. Pramerdorfer and M. Kampel, "Facial expression recognition using convolutional neural networks: state of the art," *arXiv preprint arXiv:1612.02903*, 2016.

[13] A. Lambert, F. Eadeh, S. Peak, L. Scherer, J. P. Schott, and J. Slochower, "Towards a greater understanding of the emotional dynamics of the mortality salience manipulation: Revisiting the "affect free" claim of terror management theory (in press)." *Journal of Personality and Social Psychology*, 2014.

[14] P. Ecman, "Micro Expressions," 2017. [Online]. Available: https://www.slideserve.com/tyrell/micro expressions

[15] J. Xia, F. Xie, Y. Zhang, and C. Caulfield, "Artificial intelligence and data mining: Algorithms and applications," 2013.

[16] Y. Zhang, S. Balochian, P. Agarwal, V. Bhatnagar, and O. J. Housheya, "Artificial intelligence and its applications," 2014.

[17] O. Arriaga, M. Valdenegro-Toro, and P. Plöger, "Real-time convolutional neural networks for emotion and gender classification," *arXiv preprint arXiv:1710.07557*, 2017.

[18] A. Ahmad, "Mengenal Articial Intelligence, MachineLearning, Neural Network, dan Deep Learning," *Yayasan Cahaya Islam, Jurnal Teknologi Indonesia*, 2017.

[19] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in Neural Information Processing Systems*, 2014.

[20] M. Paindavoine, J. Dubois, and P. Musa, "Neuro-inspired smart image sensor: analog Hmax implementation," *Image Sensors and Imaging Systems 2015*, vol. 9403, p. 94030J, 2015.

[21] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," in *3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings*, 2015.

[22] M. Liu, S. Li, S. Shan, R. Wang, and X. Chen, "Deeply learning deformable facial action parts model for dynamic expression analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9006, 2015, pp. 143–157.

[23] I. Putra, "Klasifikasi citra menggunakan convolutional neural network (cnn) pada caltech 101," Ph.D. dissertation, Institut Teknologi Sepuluh Nopember, 2016.