Facial Micro Expression Emotion Prediction System Using the FER-2013 Dataset and Convolutional Neural Network (CNN)

1st, Le Vu Hoang Tung,
Hanoi University of Science and Technology, Hanoi, Vietnam
Tung.lvh214137@sis.hust.edu.vn

2nd, Le Van Tung,
Hanoi University of Science and Technology, Hanoi, Vietnam
Tung.lv214136@sis.hust.edu.vn

Abstract - Humans **communicate with each other in many ways, including facial expressions. Research and development of technology in artificial intelligence using deep learning methods in human-computer interaction as an effective system application process. For example, a person shows and tries to recognize facial expressions when communicating. Sometimes predicting the expressions or emotions of the people who see them is not understood. In psychology, detecting emotions or facial expressions entails analyzing and evaluating decisions in anticipating the emotions of a person or a group of people with whom they are communicating. This study proposes to design a system that can predict and identify facial emotion classification based on feature extraction using the Convolutional Neural Network (CNN) algorithm in real-time with the OpenCV library, namely TensorFlow and Keras. The research design carried out on the Webcam consists of three main processes, namely: facial recognition, facial feature extraction, and facial emotion classification. The results of facial expression prediction in the study using the Convolutional Neural Network (CNN) method using the Facial Expression Recognition Dataset (FER-2013) reached 65.97%.**

Keywords — FER-2013 Dataset, CNN Model, Facial Emotion Recognition, Artificial Intelligence

I. INTRODUCE

Information technology is developing in the direction of creating applications to help people live and work more conveniently. The current development trend of modern information technology is artificial intelligence technology, also known as Artificial Intelligence (AI). The development of technology in today's modern era has changed the way humans interact with machines, from the use of buttons, screens, touchscreens to voice commands, to social work support, as well as in applications such as security control systems, interface between humans and computers. This is also the beginning of Human Face Recognition technology. The study [1] proposes a method of filtering EEG signals using experimental mode separation (EMD) and wavelet packet separation (WPD) to shape data on human emotional characteristics. One approach to understanding user feedback is facial expression recognition systems [2].

Facial expressions are the result of gestures or facial features that indicate the position of muscles on the human face, a form of non-verbal communication, and an important way to express one's emotions in the form of emotions, intentions, goals, and perspectives with others [3]. Currently, the process of predicting facial expression scores for services and utilities has been widely used by computers by manually selecting emotions for the level of satisfaction with the service displayed on the computer screen. Sometimes a person's emotional level by selecting on the screen is considered inaccurate in expressing customer satisfaction [4]. The use of real-time facial expression recognition systems that use artificial intelligence can increase accuracy. As a result, the system can directly recognize customers' facial expressions when they are served at the checkout counter.

We conducted research to test the accuracy of successfully predicting 7 (seven) facial expressions, namely joy, sadness, anger, fear, surprise, contempt, and disgust in real time using the Facial Expression Recognition (FER-2013) dataset.

II. RELATED WORK

A. Computer Vision

Computer vision is the process of converting data from the camera into specific decision results or new presentations to achieve specific goals. Images from the camera need to meet the standards of quality, speed and processing efficiency. In computer vision, image segmentation is used to classify objects, such as classifying the shape and size of chicken eggs [5]. Another application, Facial Recognition, uses algorithms to detect facial features and compare them to personal profiles, including name, gender, and age, for identity authentication and data security [6].

Facial expressions, from raised eyebrows to the direction of view, are a unique part of the human body that communicates intentions, ideas, and emotions. Facial expression reading methods are used in a variety of fields, from psychology to customer satisfaction assessments. The face detection in this study uses the Haar Waterfall Classification method [7]-[9]. Use the Haar Wavelet functions to process the image in a square shape and distinguish between dark and light areas, thereby detecting the position of the face in the image [10].

B. Facial Expression Recognition 2013 (FER-2013)

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

In this dataset, each face has been categorized based on

Figure 1: Image in the FER-2013 Dataset (source: [12])

emotion category, where the FER-2013 dataset is a grayscale image measuring 48pixel x 48pixel per image. The total FER-2013 dataset is 35,887 consisting of 7 (seven) different types of micro-expressions and marked with labels based on 7 (seven) different classifications starting from index labels 0 to 6, described in Table I.

Emotion	Training	Validation Data	Total
Angry	3993	960	4953
Disgust	436	111	547
Fear	4103	1018	5121
Happy	7164	1825	8989
Sadness	4938	1139	6077
Surprise	3205	797	4002
Neutral	4982	1216	6198

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

C. *Micro-expression of each emotion*

1. Happiness:

The micro-manifestations of happiness are usually associated with the activation of certain facial muscles. The corners of her mouth smirked, creating a smile, and her cheeks could be raised. The eyes also play an important role, with the outer corners wrinkled and sometimes accompanied by the formation of crow's feet. Additionally, the eyebrows may be slightly raised and the cheeks may push upwards, resulting in dimples.

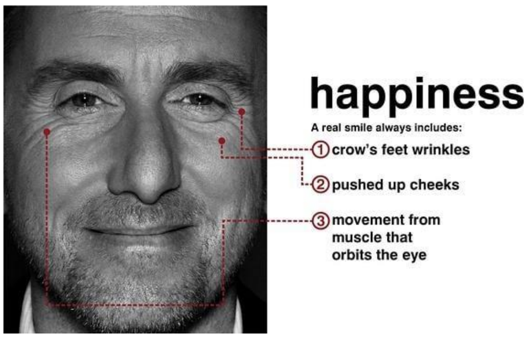


Figure 2: Characteristics of happiness manifestations (source: [14])

2. Sad:

The microscopic manifestations of sadness often involve turning the mouth down, creating a frown. The inner corners of the eyebrows can also move upwards, causing a slight wrinkle or forehead wrinkle. The eyes may appear watery or watery, and the eyelids may droop slightly. Overall, the face may appear gloomy and lacking in energy.

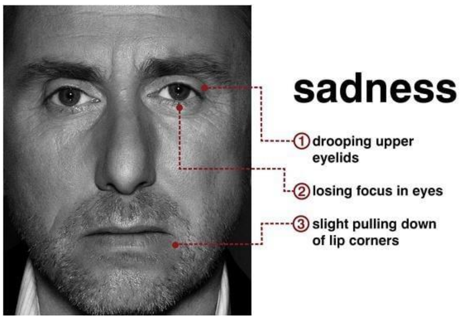


Figure 3: Characteristics of sadness (source: [14])

3. Disgusting:

Micro-manifestations of disgust are often associated with an upper lip lift and a wrinkled nose. The eyebrows can also come together and move downwards, causing a vertical wrinkle between the eyebrows. The eyes may be narrow or narrowed, and the lower eyelids may be tight or tight. These manifestations convey a feeling of aversion or disgust.

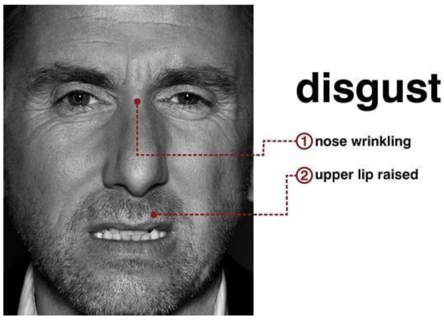


Figure 4: Manifestation of disgust (source: [14])

4. Fear:

Micro-manifestations of fear often involve eyes that are wide open, with eyebrows raised and pulled upwards. The forehead may wrinkle horizontally, and the mouth may open slightly. Lips can be stretched horizontally, making them look thinner. The overall expression is one of high apprehension and vigilance.

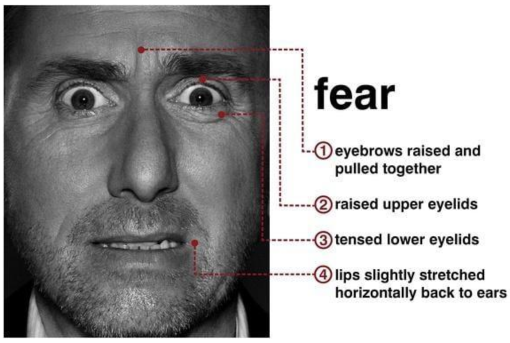


Figure 5: Typical manifestations of fear (source: [14])

5. Anger:

Micro-manifestations of anger are often associated with tightening and tightening the facial muscles. The eyebrows can be pulled down and inward, causing longitudinal grooves between them. The eyes may narrow, and the gaze may become more intense. The lips may press tightly together or curve down at the corners. The overall expression is one of intensity and aggression.

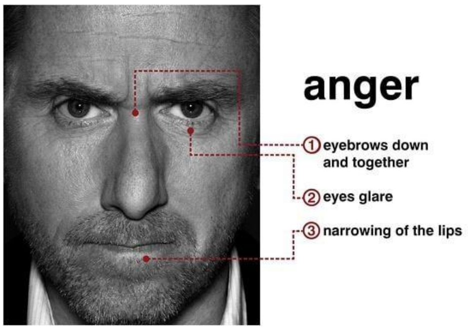


Figure 6: Characteristics of anger expression (source: [14])

6. Contempt:

The microscopic expression of contempt is characterized by a subtle facial movement. The face may not show strong emotions, but there is a bit of asymmetry, with a corner of the lip subtly raised. The mouth does not expand, but there is a distinct, albeit minimal, muscular movement. The eyebrows are still mostly relaxed and in their natural position, not raised or wrinkled significantly. The eyes may show a calm attitude, but with a look can be seen as disapproval or disapproval.

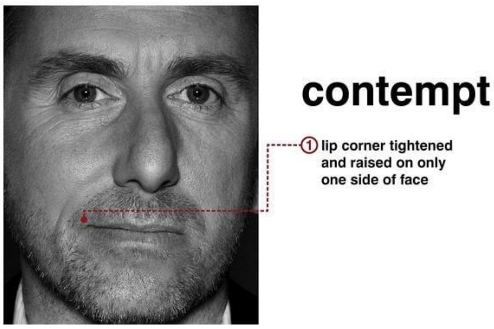


Figure 7: Characteristics of contempt (source: [14])

7. Surprise:

Microscopic manifestations of surprise involve eyes wide open, eyebrows raised, mouth open, and jaw dropping. The expression signifies surprise or sudden realization. It is usually short and intense, often accompanied by panting.

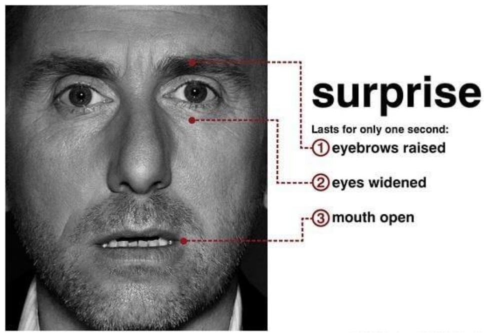


Figure 8: Characteristic expression of surprise (source: [14])

D. Condensed and deep learning neural networks

A convolutional neural network (CNN) is a deep layer of neural networks, notable for its deep architecture and significant superiority in image data processing. Deep learning (DL) is a neural network technique that uses specialized methods such as the Constrained Boltzmann Machine (RBM) to accelerate learning in neural networks, often using multiple layers, usually more than seven. With the advent of Deep Learning, the time required for training decreases due to the gradient problem disappearing lower during backpropagation [15].

The initial research that led to the discovery of CNNs was carried out by Hubel and Wiesel, who studied the visual cortex of cats. The visual cortex in animals is very powerful in current image processing systems until many studies inspired by its activity have created new models such as Neocognitron, HMAX, LeNet-5 and AlexNet [16].

The CNN method is the development of the multilayer Perceptron (MLP) method for processing two-dimensional data, such as images or audio. CNNs work similarly to MLPs, but in CNNs, each neuron exists in two dimensions, unlike in MLP where each neuron is one-dimensional. Image processing can start with specific features, such as brightness or edges, increasing complexity across layers to uniquely identify objects based on layer thickness as illustrated in figure 9.

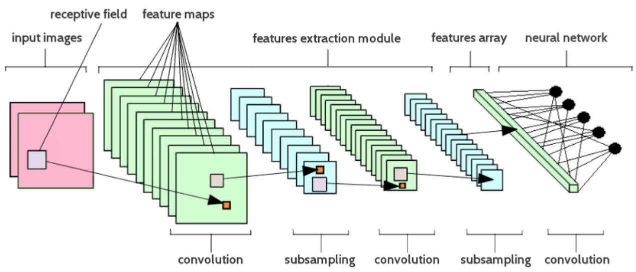


Figure 9: CNN Network

Generally, the class types in the Convolutional Neuron Network are divided into two categories: the Feature Extraction Layer and the Classification Class. The CNN layers have a three-dimensional structure of neurons (width, height, depth). Width and height refer to the size of the layers, while depth refers to the number of layers [16]. A CNN can have tens to hundreds of millions of layers, each learning how to detect different images. Image processing is applied to each training image at different resolutions, and the output of each processed image is used as input for the next layer [15].

E. *CNN Model Training: Mini XCEPTION*

In the field of computer vision, the Mini XCEPTION model is designed based on the advanced Xception structure, inherited and developed from the basic architecture of Convolutional Neural Networks (CNN) [17]. This model is characterized by the application of independent convolution models and intelligent feature extraction, through the use of Depthwise Separable Convolution in conjunction with Pointwise Convolution [18]. This approach optimizes processing and minimizes the need for a fully connected layer. Advanced methods such as Data Augmentation, Kernel Regularizer, Batch Normalization, Global Average Pooling, and Split Convolution are all applied to improve the efficiency of model training [19]. Mini XCEPTION, with its compact size, has proven high performance in facial expression classification [20].

The convolutional layer of the model is designed according to a two-dimensional (2D) mechanism, which conforms to image input standards and includes height and width dimensions. One difference in the condensation process is that this layer retains the parameters at a defined depth, thus generating an output tenx (dot product) based on the nuclear filter size during activation. The model also integrates an adjustment mechanism through the rectifier Linear Unit (ReLU) activation function, which ensures efficient activation with the maximum linear value [21], which is shown in detail in Figure 10.

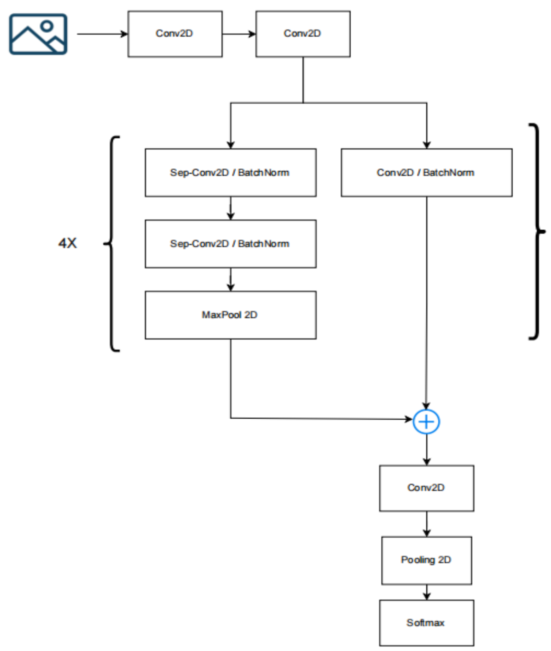


Figure 10: Xception Mini Architecture Model

III. Design and Testing of Facial Expression Recognition Systems

The first concept provides training data by evaluating the model to obtain an architectural model stored in a file with the .hdf5 extension. The model evaluation process uses Facial Expression Recognition (FER) from the FER-2013 dataset, which works on the Webcam. An overview is shown in Figure 12. After creating an architectural model in .hdf5 file format, the latter 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 is carried out with the purpose of the system performing a classification learning process to predict 7 (seven) types of image results based on microscopic facial expressions on human faces.

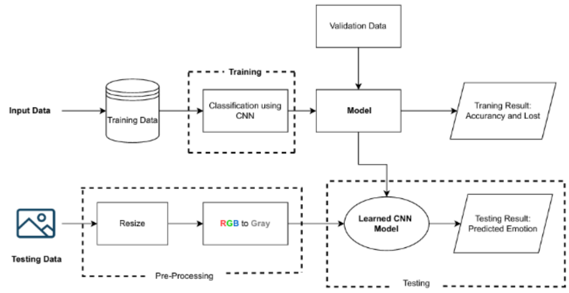


Figure 11: Flowchart of the facial micro-expression system

The research conducted by the authors to implement a facial micro-expression recognition system is shown in Figure 11. The design consists of two stages, namely:

1. Training Process
2. Test Procedure

A. *Training process with FER-2013*

The data training process in this study uses the FER-2013 dataset that has been published on Kaggle. Training performed with predefined parameters (such as era values and batch sizes) will generate a trained model that is used as a prediction parameter and stored in a file with the .hdf5 extension. Based on Figure 11, the data training process algorithm is to import training data and validation data. The training data is processed using the CNN algorithm to generate feature extracts that will be evaluated against the validation data. The results of the evaluation will create a model architecture that is trained to achieve maximum epoch value. The trained data model will then be used as a parameter to compare with the experimental data (images obtained from the camera) to predict facial expressions for real-time testing. Model the CNN algorithm during training, combined with the use of the remaining modules and combine in a convolutional layer. The other module modifies the selected map between the next two layers, so that the features are extracted by comparing the original features and the selected features. As a result, the desired feature, namely H(x) is modified to make it easier to solve the problem during F(x) training, with the following equation:

$$H(x) = F(x) + x \qquad (1)$$

The architectural model (see Figure 10) has 4 (four) modules in a hidden layer, containing a separate depth-separable convolutions (Depthwise Separable Convolutions). Each condensation layer is followed by a batch normalization operation and a ReLU (Rectifier Linear Unit) activation function. In the last layer, the global average synthesis and

Softmax trigger function is done to generate the prediction. This architectural model has about 60,000 parameters, which is a 10-fold reduction compared to the implementation of the CNN architecture model with the function of completely eliminating the connection layer, and an 80-fold reduction compared to the standard CNN architectural model.

This training for architectural modeling involves figuring out the optimal value of accuracy (training accuracy and validation accuracy), as the exact value from the graph results increases for each epoch. If there is a resulting line that creates a valley that results in a decrease in the value of the correct line, then the correct result indicates that the training process model is complex.

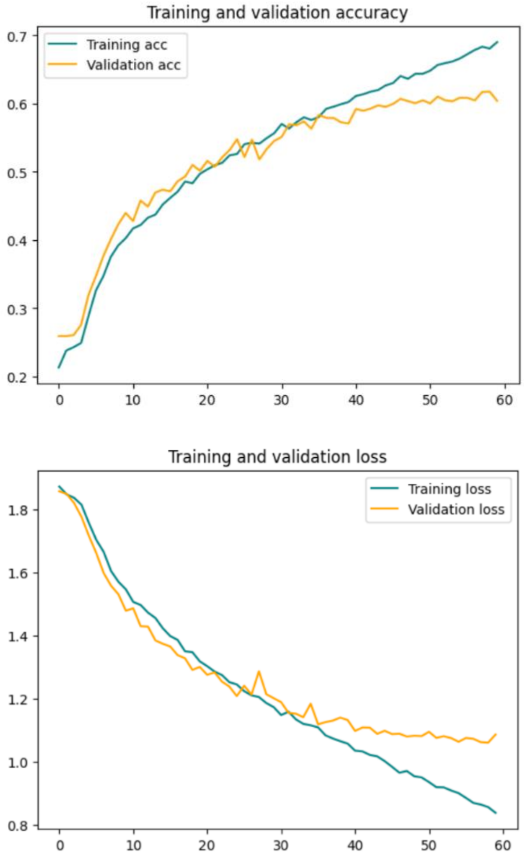


Figure 12: Results of Accuracy and Lost Data

B. *Test Procedure*

The first test used images from the FER-2013 dataset (some datasets are shown in Figure 1). This experiment demonstrates the comparative results and accuracy of the application in this study, where the results of detecting facial expressions are accurate in the input image, as shown in Figure 13.

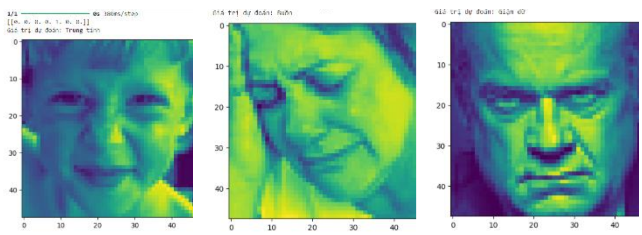


Figure 13: Testing with the FER-2013 dataset

The next test will use images from a laptop webcam with a real person's face. We will analyze this in detail in Part IV.

IV. Analyze the Results

Facial microfacial expression recognition in this study was carried out through the evaluation of classification reports. The results show that the use of the CNN architecture model in the facial expression recognition system can achieve optimal and real-time performance. The evidence is that the use of a separate convolutional layer when conducting data training can perform optimal training, and the accuracy of facial expression data in the trained model is on average 62%.

	precision	recall	f1-score	support
Tức giận	0.57	0.47	0.52	958
Ghê tởm	0.75	0.53	0.62	111
Sợ hãi	0.50	0.44	0.47	1024
Hạnh phúc	0.81	0.82	0.81	1774
Vô cảm	0.54	0.59	0.56	1233
Buồn	0.47	0.53	0.49	1247
Ngạc nhiên	0.76	0.77	0.77	831
accuracy			0.62	7178
macro avg	0.63	0.59	0.61	7178
weighted avg	0.62	0.62	0.62	7178

Figure 14: Classification report

Overall, the model seemed to work well on several types of emotions with an overall model accuracy of 62%. This shows that the model is capable of accurately predicting facial expressions at a relatively high rate. The F1 score is an important metric because it considers both accuracy and recall. Classes usually have relatively high F1 scores, indicating a model of balance of accuracy and good recall for grades. From the recall column data, it can be seen that the model can detect the test data layer at a fairly good rate. Averages such as macro averages and weighted averages show that the pattern works consistently across all layers, not just focusing on specific layers.

Analyzing the results of the system after implementation is essential. The results of this test are shown in Table II.

		Number of Predictions						
		Contempt	Happy	Sad	Angry	Disgust	Surprise	Surprise
Micro expresion	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
	Surprise	0	0	1	0	0	0	9

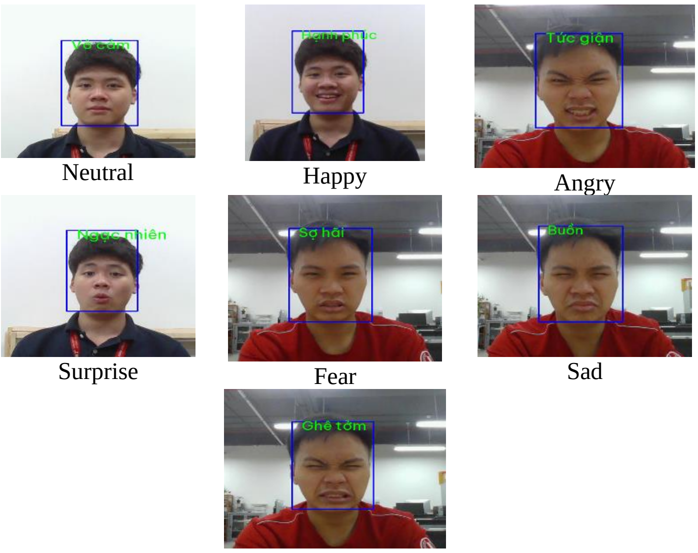
Table II: Confusion Matrix

Based on Table II, it seems that the model performs quite well at identifying most emotions, with high accuracy for 'Contempt', 'Happiness', 'Sadness', 'Surprise', and 'Fear'. The numbers on the diagonal, which represent accurate predictions, are mostly high, indicating good performance. However, there are a few instances of confusion between 'Surprise' and 'Fear', which may suggest that these two emotions are more difficult to distinguish for the model. This is not uncommon because these two emotions can share similar facial expressions. In general, the model has an accuracy on the actual test of 65.97%

A. *Facial Expression Prediction Test*

This test aims to determine whether the system is successful in recognizing the trained facial expressions. The success rate of the system in identifying the user's face is very high, so it can be concluded that the system is working well. Figure 15 is an example of a facial expression recognition test result.

The experiment was conducted ten times for each expression based on the results of tests of sadness, surprise, contempt, and happiness. The system has successfully identified all tests. While the result of the expression of anger has one error, the expression of disgust has two errors, and the expression of fear has one error.



Form. 15: Facial expression test results

Looking at the model's confusion matrix on the FER-2013 test kit, the model shows the best classification for "happy" and "unexpected" expressions. On the other hand, the model makes the most mistakes when it comes to categorizing between "sadness" and "fear." In addition, the low classification accuracy for the expression "sad" can be explained by the fact that they have a smaller number of samples in the initial training set. The misclassification between "fear" and "sadness" can be attributed to the similarity between the layers in the dataset, which is reflected in Figure 16.

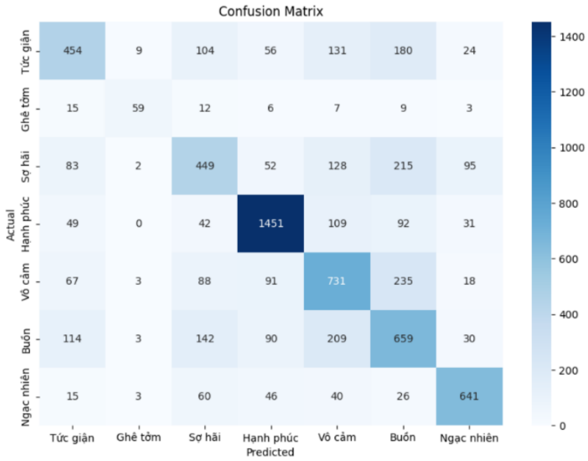


Figure 16: Confusion Matrix

B. *Recognize facial expressions based on facial condition or angle*

The facial expression recognition test is based on the condition or angle of the face including: face looking up, looking down, turning left or right. There are a total of seven test models based on the camera perspective. The purpose of this test is to analyze the system's facial expression recognition results based on the camera's shooting angle.

Experiments are carried out under certain conditions and locations. With seven expression types and seven test positions. Based on the camera's viewing angle, the system correctly predicted 141 times and failed 59 times. The overall success rate of the system in facial expression recognition is very high. This failure rate may be explained by the limited view of the data samples in the FER-2013 set. Specifically, in Figure 17 are examples of facial recognition tests based on facial position in the case of happy expressions.

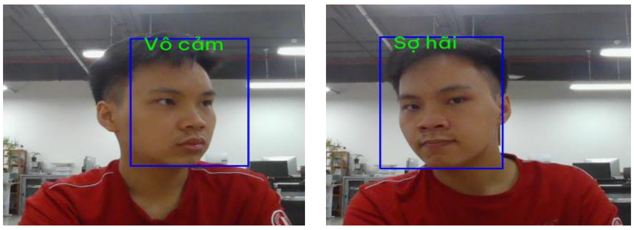


Figure 17a: Correct Figure 17b: Incorrect

Figure 17: Facial expression test based on the image compared to the viewfinder from the camera

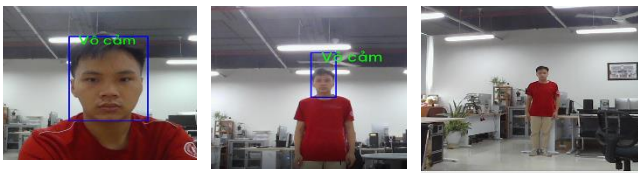
In Figure 17a, the system can correctly recognize the expression when the person looks to the right, but in Figure 17b (when looking to the left), the system cannot recognize the expression because it is misclassified as "Fear"

C. *Facial expression recognition based on the distance between the human face and the camera*

The facial expression recognition test based on the distance between the subject and the camera is performed by aligning the face perpendicular to the camera at different distances: 50 cm, 100 cm, 200 cm, 500 cm, and 1000 cm. The purpose of this test is to determine the accuracy of the system in recognizing facial expressions displayed from a specific distance. The average result of this test is 62.3%.

The factor affecting the result is that the image obtained at a distance of 500 cm or more cannot accurately detect faces and predict expressions. An illustrative example of this test is shown in Figure 18, which focuses on the expression of happiness. In Figure 18a, the system correctly recognizes the expression. In Figure 18b, the system makes an incorrect prediction about the expression. Figure 18c cannot make a prediction because the system did not detect the object being a human face.

A total of 35 experiments were carried out during this period, with each facial expression (7 types of expressions) tested five times. Within a distance of 50 cm to 300 cm, the system can detect most faces and accurately recognize expressions. At a distance of 500 cm, the system can detect objects that are human faces but can only predict expressions with reduced accuracy. In addition, the system cannot accurately recognize facial expressions. At a distance of 1000 cm, the system does not even detect a human face, so it cannot recognize any expressions. Summing up the results, the system accurately predicted facial expressions in 35 test cases. Figures 18a, 18b, and 18c are examples of facial expression recognition test results based on the distance between the subject and the webcam.



18a: Correct 18b: Incorrect 18c: Failed

Figure 18: Facial expression test based on the distance between the subject and the camera

V. CONCLUSION

Based on the experimental results, the study successfully designed a system that uses the Convolutional Neural Network (CNN) method to predict 7 human facial expressions through Facial Expression Recognition (FER) from the FER-2013 dataset. The system is designed according to the following process:

Training: Using the FER-2013 dataset and the CNN method to extract features and predict appropriate facial expressions. Real-time facial expression recognition: Detect facial objects using the Haar Cascade method and use CNNs to classify facial expressions. Show results: Recognized facial expressions will be displayed on the dashboard.

Although the software system works well, the test results still reveal some weaknesses:

Need for deeper CNN architecture: Use multiple condensation layers and fully connected layers with the right configuration to achieve greater accuracy.
Additional Training Data Required: Use new datasets or supplement additional training data to improve accuracy, especially for tests with distances of more than 5 meters, different viewing angles, and rotating images.

ACKNOWLEDGMENT

We would like to express our sincere thanks to Master Hoang Quang Huy for guiding us in this project, as well as some colleagues in the iBME laboratory, Hanoi University of Science and Technology, who have commented and supported us in our research process.

REFERENCES

[1] S. B. Musa and H. Tjandrasa, "Analysis of eeg emotion signal features based on hybrid decompotion," ENERGY, vol. 7, no. 1, pp. 7–12, 2017.

[2] A. Saputra and T. A. B. Wirayuda, "Facial Expression Recognition Using Local Binary Pattern (LBP)," eProceedings of Engineering, vol. 2, no. March 3, 2015.

[3] I. G. P. S. Wijaya, A. A. Firdaus, A. P. J. Dwitama, and Mustiari, "Introduction of Facial Expressions Using DCT and LDA for Music Player Applications (MOODSIC)," Journal of Information Technology and Computer Science (JTIIK), 2018.

[4] Z. Abidin, "Using Fisherface and Backpropagation Artificial Neural Network," Journal of Mathematics and Natural Sciences, 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, "Implementation of Computer Vision and Image Segregation," in SNATI 2010, 2010.

[7] P. Musa, N. Nuryuliani, and M. Lamsani, "Design and Design of Automatic Door Controllers Based on Facial Features Using the Euclidean Distance and Fuzzy C-mean Method," Scientific Journal of Informatics and Computers, vol. January 13, 2008.

[8] A. L. Kurniawan, R. R. Isnanto, A. Zahra, K. Key, P. Face, V. Computer, A. Wavelet, T. Gabor, and J. Euclidean, "Design of Facial Recognition System Using the Characteristic Extraction Method of 2D Gabor Wavelet Filter Arrangement with Euclidean Distance," Design of Facial Recognition System Using the Characteristic Extraction Method of 2D Gabor Wavelet Filter Arrangement with Euclidean Distance, 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-January, pp. 1–5, 2018.

[10] J. Cristanto and K. R. R. Wardani, "Application of Single-Layer Feed-Forward Neural Network Method Using Gabor Kernel for Facial Expression Recognition," Journal of Telematics, Harapan Bangsa Institute of Technology, 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 machines 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 "effect free" claim of terror management theory (in press)." Journal of Personality and Society Psychology, 2014.

[14] P. Ecman, "Micro Expressions," 2017. [Online]. Available: https://kknews.cc/psychology/gyb89y.html

[15] "The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi" vol.04 no. 3, 2021

[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. Ploger, "Real-time convolutional ¨ neural networks for emotion and gender classification," arXiv preprint arXiv:1710.07557, 2017.

[18] A. Ahmad, "Getting to Know Articial Intelligence, MachineLearning, Neural Network, and Deep Learning," Yayasan Cahaya Islam, Indonesian Journal of Technology, 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.