JSCE (Journal of System and Computer Engineering)
Vol.6, No.1, January 2025, pp. 35-44
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v6i1.1559

Facial Expression Recognition of Al-Qur'an Memorization Students Using Convolutional Neural Network

Ayu Lestari Perdana [1,a] , Suharni [2,b*]
[1] *Universitas Islam Makassar, Jl. Perintis Kemerdekaan KM.9, Makassar 90245, Indonesia*
[2] *Universitas Islam Makassar, Jl. Perintis Kemerdekaan KM.9, Makassar 90245, Indonesia*
[a] *ayulestariperdana.dty@uim-makassar.ac.id;* [b] *suharni.dty@uim-makassar.ac.id*
* *suharni.dty@uim-makassar.ac.id*

Abstract

Facial expression recognition technology has advanced significantly and has become an intriguing topic of study. This research focuses on the facial expressions of Al-Qur'an memorization students, which naturally reveal various aspects of their engagement, understanding, and emotional barriers about the verses being memorized. The issue is that facial expression recognition still lacks optimal accuracy, and the need for a better algorithmic model to improve accuracy is evident. Therefore, an intelligent computing system is required to address this problem. This study aims to enhance the accuracy of facial expression recognition in Al-Qur'an memorization students using the Convolutional Neural Network (CNN) method, classifying facial expressions such as happy, neutral, and tired based on collected facial image data, achieving improved accuracy. The first stage involves capturing image data via CCTV, followed by preprocessing, training the CNN model, result analysis, and model evaluation. By using the CNN method to recognize the facial expressions of Al-Qur'an memorization students, a high accuracy of 84% was achieved with a loss value of 14.9.

Keywords— *CNN, Facial Expression, Qur'an Memorizers, Students*

1. Introduction

Humans express various fundamental emotions through universal and basic facial expressions(Pise et al., 2022). Facial expressions represent a crucial aspect of non-verbal communication, allowing individuals to convey emotions without using words (Dalle Nogare et al., 2023), serving as a means to express individual emotions and sentiments(Revina & Emmanuel, 2021). In the realm of education, particularly among students dedicated to memorizing the Al-Qur'an, interpreting these expressions can provide valuable insights into their level of engagement, fatigue, or potential barriers in the memorization process. By identifying facial expressions that indicate confusion, anxiety, or enthusiasm, educators can promptly adjust their teaching methods and provide emotional support, thereby enhancing the effectiveness of the learning process and memorization targets. Understanding the emotions and psychological states of students becomes critical in the context of education, especially in the Al-Qur'an memorization process(Fata & Rosyadi, 2024). The intensive nature of memorization, which requires high levels of concentration, can elicit various emotional responses, ranging from joy and satisfaction to fatigue or frustration. However, facial expression recognition still lacks optimal accuracy, and the need for the best algorithmic model to improve accuracy remains(I Putu Agus Aryawan, I Nyoman Purnama, 2023). With facial expression recognition, schools can gain valuable insights to assist teachers in understanding students' enthusiasm for memorization, enabling teachers to be more empathetic and responsive.

Research on facial expression recognition has been widely conducted by several scholars such as Lestari Perdana et al., (2023)**,** Riadi & Sulaehan (2019)**,** Daffa Ulhaq et al., (2023) and Sidik et al.,(2021). Facial expression recognition has also been an active area of research over the past few decades. The application of deep learning has had the greatest impact in this field due to the significant advancements in neural networks(Pise et al., 2022), and it is one of the leading applications in the field of computer vision(Huang et al., 2023). In computer vision, Convolutional Neural Networks (CNN) are a very popular model used for emotion recognition. This model has been successfully applied to detect various objects in digital images with remarkable accuracy(Shahzad et al., 2023). CNN has become the core technology enabling high recognition accuracy(Adiatma et al., 2021)(Nurdiati et al., 2022). Essentially, CNN is a deep learning algorithm specifically designed to work with images and videos. It takes images as input, extracts and learns image features, and classifies them based on the learned features. CNN uses various filters to extract different information from images, such as edges and shapes, which are then combined to identify the image(Kaur et al., 2022). The development of this technology, including algorithm improvements, computational capabilities, and the availability of large datasets, has enabled the practical implementation of efficient and accurate facial expression recognition systems(Nour et al., 2020).

Alamsyah and Pratama demonstrated that CNN can be used for facial expression classification with satisfactory results, even in non-ideal data conditions(Alamsyah & Pratama, 2020) . This facilitates the development of applications that can automatically and in real-time (Bhagat et al., 2024) interpret students' facial expressions(Ihsan et al., 2021). Furthermore, they emphasized the importance of data preprocessing and model training in facial expression recognition, which can also be applied in the context of Al-Qur'an memorization(AL Sigit Guntoro et al., 2022).

The background of using CNN for facial expression recognition in Al-Qur'an memorization students reflects a combination of deep educational needs, technological advancements, and a vision for more adaptive and responsive learning. CNN is employed because it shows how its deep layers can automatically extract relevant facial features for expression recognition, from basic features to more abstract representations. Facial expression recognition using Convolutional Neural Networks (CNN) has significant implications for understanding the emotional state of students during Al-Qur'an memorization at Sekolah Dasar Hafidz Al-Qurba. This technology can improve educational strategies by providing insights into student engagement and emotional well-being. By utilizing this technology, it is expected to enhance the learning experience, making teaching approaches more personalized and effective, while also considering ethical and privacy aspects. The aim of this research is to improve the accuracy of facial expression recognition in Al-Qur'an memorization students using the CNN method by classifying facial expressions such as happy, neutral, and tired based on collected facial image data, ultimately enhancing memorization targets, engagement, and learning effectiveness.

2. Method

The method used in this research is the descriptive research method. The research steps to be conducted are as follows:

2.1 Dataset Collection

Data will be collected by capturing facial images sourced from CCTV installed in classrooms or memorization rooms. This data will then undergo preprocessing and be used as the dataset.

2.2 Eksperimen

The experiment will begin with inputting the dataset, followed by preprocessing, which includes resizing images and normalization. The next steps involve designing the CNN architecture, training the CNN model, evaluating and validating the research, testing, performance evaluation, and obtaining accurate results.

Figure 1. Experimental steps.

1. The collected data is processed and labeled according to the displayed expressions (Happy, Neutral, Tired).
2. The dataset is then divided into three subsets: training data, validation data, and testing data. The training data is used to train the model, the validation data is used to tune model parameters during training, and the testing data is used to evaluate the model's performance.
3. In the preprocessing stage, all images are resized uniformly and stored in designated folders.
4. After preprocessing is completed, the CNN model is trained. The CNN architecture is designed using deep learning libraries such as TensorFlow or Keras.
5. After model training is finished, validation data is used to test the model’s performance, followed by testing data to objectively measure the model's final performance.
6. After the model is trained, it needs to be evaluated to ensure its performance meets expectations. The model evaluation is conducted using several metrics, such as accuracy, confusion matrix, precision, recall, and F1 score.
7. Once the CNN model is trained and evaluated, it can be used to predict the class of unseen data.
8. Finally, a confusion matrix is used to analyze the experimental results, where higher precision and recall values indicate better performance of the classification model in predicting the correct class.

3. Results And Discussion

3.1. Data Collection

Data collection was conducted at SD Hafidz Al Qurbah Moncongloe. The data was gathered while students were engaged in memorizing the Al-Qur'an, from Monday to Friday, between 8:30 AM and 2:00 PM WITA. The data was recorded using CCTV installed in the classroom, ensuring clear capture of the students' faces. The CCTV footage was subsequently processed into images by selecting clear, non-blurry facial images. The processed images depicted students’ facial expressions captured naturally, without awareness of the camera, as they concentrated on memorizing and reciting the Al-Qur'an to their teachers (Ustadz/Ustadzah). This approach facilitated the classification of facial expressions, aiming to achieve high accuracy.

3.2 Preprocessing

The collected facial images were further preprocessed by cropping to focus on the facial area and resizing the images to optimize subsequent data processing.
Examples of the collected facial images are as follows:

Figure 2. Sample of Face Data of Students Memorizing the Qur'an.

The facial data samples were then preprocessed by resizing the images to 224 x 224 pixels, resulting in the following image output:

Figure 3. Resized Face Image.

The data consists of three types of expressions: Happy, Neutral, and Tired.

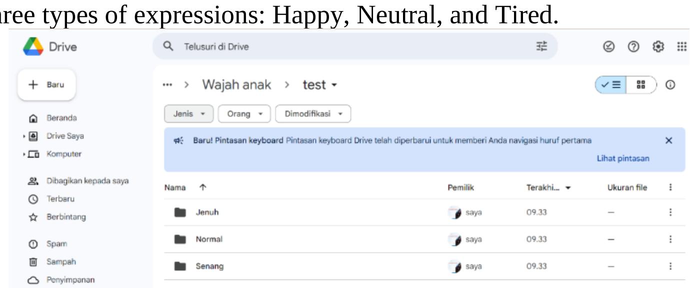

Figure 4. Children's Facial Expression Class.

The folder contains testing data (Test), training data (Train), and validation data (Valid).

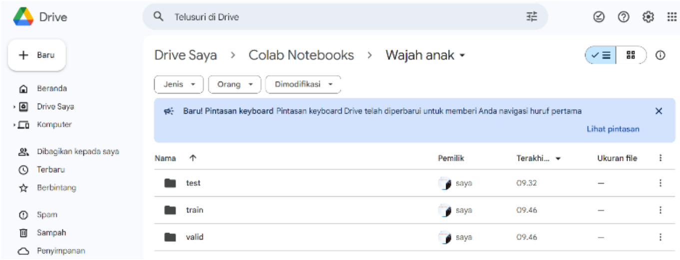

Figure 5. Data Testing, Training, dan Validasi.

3.3. Model Design CNN

The CNN model employed is a combination of EfficientNetB3 as the pre-trained base architecture and additional custom layers designed for classification tasks. This architecture leverages transfer learning by integrating custom layers on top of the base model. EfficientNet represents a CNN architecture that utilizes compound scaling techniques to efficiently balance depth, width, and input image resolution.

Custom Layers

1. Batch Normalization Normalizes the output of the base model to improve stability and performance during training.
2. Dense Layer A fully connected layer with 256 neurons, employing ReLU activation and L1 and L2 regularization to prevent overfitting
3. Dropout Applied with a dropout rate of 45% to enhance model generalization.
4. Output Layer A fully connected layer with the number of neurons corresponding to the number of classes, utilizing Softmax activation for multi-class classification.

Number of Layers

The base model, EfficientNetB3, consists of 66 layers. With the addition of custom layers such as Batch Normalization, Dense (256 neurons), Dropout, and Dense (output layer), the total architecture comprises approximately 70+ layers.

Filters and Kernel Sizes

In EfficientNetB3, the initial convolutional layer uses 32 filters, which exponentially increase to hundreds of filters in deeper layers. The kernel sizes are a combination of (3x3) and (5x5), depending on the convolutional block. The custom layers, being fully connected, do not incorporate filters or kernels.

Pooling Technique

Global Max Pooling is employed to reduce the output dimensions of the base model into a 1D vector. This is achieved using the parameter pooling='max'.

Optimization

The model optimization is carried out using the Adamax optimizer, with an initial learning rate set to 0.001.

3.4. Training Model

In this case, the CNN model has a total of 11,183,922 parameters. Of these, 11,093,547 parameters are trainable, while 90,375 parameters are non-trainable

Model: "sequential_7"

Layer (type)	Output Shape	Param #
efficientnetb3 (Functional)	(None, 1536)	10,783,535
batch_normalization_11 (BatchNormalization)	(None, 1536)	6,144
dense_14 (Dense)	(None, 256)	393,472
dropout_7 (Dropout)	(None, 256)	0
dense_15 (Dense)	(None, 3)	771

Total params: 11,183,922 (42.66 MB)
Trainable params: 11,093,547 (42.32 MB)
Non-trainable params: 90,375 (353.03 KB)

Figure 6. Preparation for Training the Model.

The training loss and validation loss graphs for the Convolutional Neural Network (CNN) are used to monitor model performance during training. The training and validation loss results can be seen in Figure 7 below:

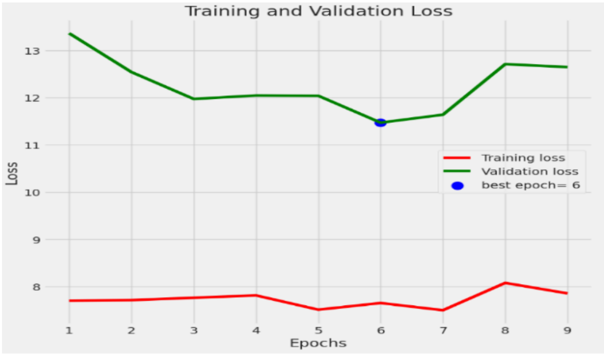

Figure 7. Training and Validation Loss Graphs.

Based on Figure 7, it was found that the training loss remained relatively stable, and the validation loss was higher than the training loss. This is a common occurrence and indicates that the model performs better on the training data than on unseen data (validation set). If the difference is not too large, the model still generalizes well. The validation loss decreased along with the training loss, and the best epoch was obtained at the 6th epoch.

The training accuracy and validation accuracy graphs for CNN provide information on how well the model learns to classify the training data and how well it generalizes to unseen data (validation set). The training and validation accuracy results can be seen in Figure 8 below:

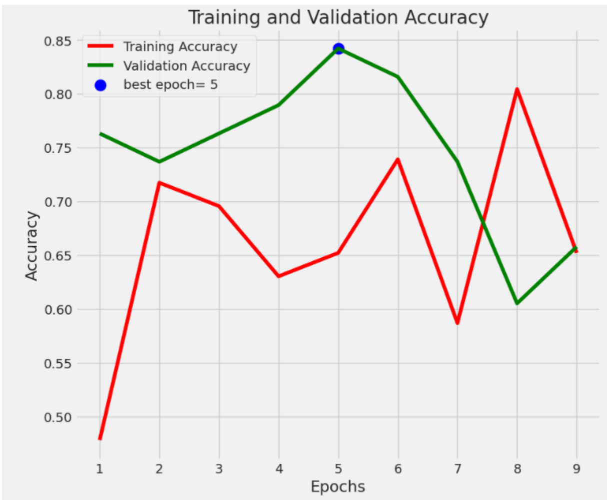

Figure 8. Training and Validation Accuracy Graph.

Training accuracy measures how well the model correctly predicts facial expressions on the training data, while validation accuracy measures how well the model predicts on unseen data during training, which is used to assess model generalization. Based on Figure 8, it is shown that training accuracy gradually increased with each epoch, indicating that the model was learning to recognize patterns in the training data more effectively and generalizing better. The highest validation accuracy was achieved at epoch 5, while the highest training accuracy occurred at epoch 8. However, the validation accuracy was low at epoch 8, indicating overfitting. This means that the model performed very well on the training data but failed to recognize patterns in the validation data. The model had learned too many specific details from the training data, which worsened its performance on new data. Nevertheless, the validation accuracy generally improved as the epochs progressed and approached the training accuracy, suggesting that the model was generalizing well.

3.5 Evaluation Model

The model evaluation results can be seen in the following figure:

```
[45] ts_length = len(test_df)
     test_batch_size = test_batch_size = max(sorted([ts_length // n for n in range(1, ts_length + 1) if ts_length%n
     test_steps = ts_length // test_batch_size
     train_score = model.evaluate(train_gen, steps= test_steps, verbose= 1)
     valid_score = model.evaluate(valid_gen, steps= test_steps, verbose= 1)
     test_score = model.evaluate(test_gen, steps= test_steps, verbose= 1)

     print("Train Loss: ", train_score[0])
     print("Train Accuracy: ", train_score[1])
     print('-' * 20)
     print("Validation Loss: ", valid_score[0])
     print("Validation Accuracy: ", valid_score[1])
     print('-' * 20)
     print("Test Loss: ", test_score[0])
     print("Test Accuracy: ", test_score[1])
```

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 22s 22s/step - accuracy: 0.8500 - loss: 12.7649
1/1 ━━━━━━━━━━━━━━━━━━━━ 8s 8s/step - accuracy: 0.8421 - loss: 14.9906
/usr/local/lib/python3.10/dist-packages/keras/src/trainers/data_adapters/py_dataset_adapter.py:121: UserWarning
  self._warn_if_super_not_called()
1/1 ━━━━━━━━━━━━━━━━━━━━ 5s 5s/step - accuracy: 0.8421 - loss: 14.9906
Train Loss:  12.764901161193848
Train Accuracy:  0.8500000238418579
--------------------
Validation Loss:  14.990635871887207
Validation Accuracy:  0.8421052694320679
--------------------
Test Loss:  14.990635871887207
Test Accuracy:  0.8421052694320679
```

Figure 9. Model Evaluation Results.

Table 1. The Performance of Model Evaluation Result.

Variable	Accuracy	Loss
Training	85%	12,76%
Validation	84%	14%
Testing	84%	14,99%

The evaluation of training, validation, and testing was relatively similar, indicating that the model had good generalization. This model demonstrated good performance on the training, validation, and testing data. Based on the testing results, an accuracy of 84% was obtained, with a loss value of 14.99. A slightly higher testing loss suggests room for improvement in making the model more adaptive to new data.

The model evaluation results using the confusion matrix can be seen in the following figure:

```
Confusion Matrix, Without Normalization
[[10  4  0]
 [ 1 16  0]
 [ 0  1  6]]
              precision    recall  f1-score   support

       Jenuh       0.91      0.71      0.80        14
      Normal       0.76      0.94      0.84        17
      Senang       1.00      0.86      0.92         7

    accuracy                           0.84        38
   macro avg       0.89      0.84      0.86        38
weighted avg       0.86      0.84      0.84        38
```

Figure 10. Confusion Matrix Results.

The confusion matrix provides deeper insights into the CNN model's performance for each class. Based on the model evaluation from the confusion matrix, several important evaluation metrics can be calculated. Accuracy measures how often the model makes correct predictions. Based on the results, an overall accuracy rate of 0.84 or 84% was achieved. The high accuracy indicates that the model performs well in predicting the correct class. Precision measures how accurately the model predicts a particular class. Precision for the "Tired" category was 91%, for the "Neutral" category 76%, and for the "Happy" category 100%. Recall (Sensitivity) measures how well the model detects all samples of a given class. The recall for the "Tired" category was 71%, for the "Neutral" category 94%, and for the "Happy" category 86%. High precision indicates that the model rarely misclassifies classes, while high recall indicates the model detects all instances of the respective class. A balance between the two is important, especially when dealing with imbalanced datasets [20]. The F1-score, which is the harmonic mean of precision and recall, is used to measure model balance. The F1-score for the "Tired (Jenuh)" category was 80%, for the "Neutral (Normal)" category 84%, and for the "Happy (Senang)" category 92%. Confusion Matrix can reveal some of the limitations of the model in recognizing facial expressions, especially in distinguishing between expressions that are almost similar to neutral and tired.

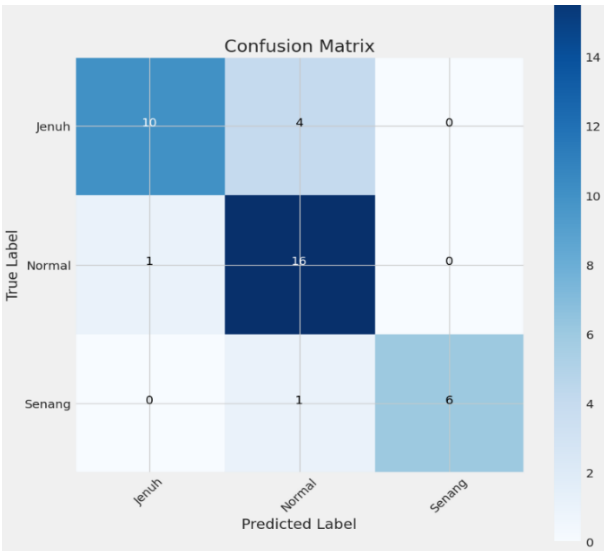

Figure 11. Prediction and Confusion Matrix.

The confusion matrix is critical for evaluating the performance of classification models. The prediction and confusion matrix in Figure 11 shows the comparison between the model's predictions and the actual labels of the test data, providing information on how well the model performed the classification.

Based on the CNN model performance, facial expressions were classified into three classes: Tired, Neutral, and Happy. After five epochs of progress, the model demonstrated good performance, as seen by the decreasing loss values and increasing accuracy with each epoch. The best results were achieved at epoch 5, with an accuracy of 84% and a loss of 14.99.

This study successfully developed a CNN model for recognizing the facial expressions of Al-Qur'an memorization students at SD Hafidz Al-Qurba, classifying them into three categories: Tired, Neutral, and Happy, with a high accuracy of 84% and a loss value of 14.99. This model can provide valuable information to teachers in understanding students' emotional states during the learning process. With further improvements, this system can become a more effective tool in supporting emotion-based education and responding to student needs. Enhancements could include the use of other datasets or different deep learning architectures, which might yield better results. Additionally, using data augmentation techniques to improve training efficiency and employing larger datasets could further improve accuracy. Image size, quality, and patterns significantly influence the facial recognition process(Sriyati et al., 2020).

4. Conclusions

His study successfully developed a CNN model to recognize facial expressions of Al-Qur'an memorizing students at SD Hafidz Al-Qurba, classified into three categories: Bored, Normal, and Happy, with a fairly high accuracy of 84% and a Loss value of 14.99. This model provides valuable information for teachers to understand students'emotional conditions during the learning process. With further improvements, this system has the potential to become a more effective tool in supporting emotion-based education and responding to students' needs.

Acknowledgements

Thanks to the Ministry of Education, Culture, Research, and Technology for the research funding support from the Beginner Lecturer Research (PDP) funding program and all related parties.

References

Adiatma, B. C. L., Utami, E., & Hartanto, A. D. (2021). Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network (CNN). *Explore*, *11*(2), 75. https://doi.org/10.35200/explore.v11i2.478

AL Sigit Guntoro, Edy Julianto, & Djoko Budiyanto. (2022). Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network. *Jurnal Informatika Atma Jogja*, *3*(2), 155–160. https://doi.org/10.24002/jiaj.v3i2.6790

Alamsyah, D., & Pratama, D. (2020). Implementasi Convolutional Neural Networks (CNN) untuk Klasifikasi Ekspresi Citra Wajah pada FER-2013 Dataset. *Jurnal Teknologi Informasi*, *4*(2), 350–355. https://doi.org/10.36294/jurti.v4i2.1714

Bhagat, D., Vakil, A., Gupta, R. K., & Kumar, A. (2024). Facial Emotion Recognition (FER) using Convolutional Neural Network (CNN). *Procedia Computer Science*, *235*(2023), 2079–2089. https://doi.org/10.1016/j.procs.2024.04.197

Daffa Ulhaq, M. R., Zaidan, M. A., & Firdaus, D. (2023). Pengenalan Ekspresi Wajah Secara Real-Time Menggunakan Metode SSD Mobilenet Berbasis Android. *Journal of Technology and Informatics (JoTI)*, *5*(1), 48–52. https://doi.org/10.37802/joti.v5i1.387

Dalle Nogare, L., Cerri, A., & Proverbio, A. M. (2023). Emojis Are Comprehended Better than Facial Expressions, by Male Participants. *Behavioral Sciences*, *13*(3). https://doi.org/10.3390/bs13030278

Fata, B. S., & Rosyadi, I. (2024). Kesejahteraan Emosional Santri Penghafal Al-Qur'an di

Pesantren an-Nuqthah, Tangerang. *Al- Fikrah*, *4*(2), 124–137.

Huang, Z. Y., Chiang, C. C., Chen, J. H., Chen, Y. C., Chung, H. L., Cai, Y. P., & Hsu, H. C. (2023). A study on computer vision for facial emotion recognition. *Scientific Reports*, *13*(1), 1–13. https://doi.org/10.1038/s41598-023-35446-4

I Putu Agus Aryawan, I Nyoman Purnama, K. Q. F. (2023). Analisis Perbandingan Algoritma Cnn Dan Svm Pada Klasifikasi Ekspresi Wajah. *Jurnal Teknologi Informasi Dan Komputer*, *9*(4), 399–408.

Ihsan, M., Niswatin, R. K., & Swanjaya, D. (2021). Deteksi Ekspresi Wajah Menggunakan Tensorflow. *Joutica*, *6*(1), 428. https://doi.org/10.30736/jti.v6i1.554

Kaur, N., Kaur, K., & Hafiz, A. (2022). Facial Expression Recognition Using Convolutional Network. *International Journal for Research in Applied Science & Engineering Technolog*, *10*(VI), 2779–2785. https://doi.org/https://doi.org/10.22214/ijraset.2022.44447

Lestari Perdana, A., Suharni, & Riadi, A. (2023). *Pengenalan Ekspresi Wajah Pengunjung Deal Coffee Menggunakan Metode Convolutional Neural Network (Cnn)*. *04*(November), 132–139. https://journal.unm.ac.id/index.php/JESSI/article/view/1024

Nour, N., Elhebir, M., & Viriri, S. (2020). Face Expression Recognition using Convolution Neural Network (CNN) Models. *International Journal of Grid Computing & Applications*, *11*(4), 1–11. https://doi.org/10.5121/ijgca.2020.11401

Nurdiati, S., Najib, M. K., Bukhari, F., Ardhana, M. R., Rahmah, S., & Blante, T. P. (2022). Perbandingan AlexNet dan VGG untuk Pengenalan Ekspresi Wajah pada Dataset Kelas Komputasi Lanjut. *Techno.Com*, *21*(3), 500–510. https://doi.org/10.33633/tc.v21i3.6373

Pise, A. A., Alqahtani, M. A., Verma, P., Purushothama, K., Karras, D. A., Prathibha, S., & Halifa, A. (2022). Methods for Facial Expression Recognition with Applications in Challenging Situations. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/9261438

Revina, I. M., & Emmanuel, W. R. S. (2021). A Survey on Human Face Expression Recognition Techniques. *Journal of King Saud University - Computer and Information Sciences*, *33*(6), 619–628. https://doi.org/10.1016/j.jksuci.2018.09.002

Riadi, A., & Sulaehani, R. (2019). Analisis Implementasi Preprocessing Dengan Otsu-Gaussian Pada Pengenalan Wajah. *ILKOM Jurnal Ilmiah*, *11*(3), 200–205. https://doi.org/10.33096/ilkom.v11i3.457.200-205

Shahzad, H. M., Bhatti, S. M., Jaffar, A., Akram, S., Alhajlah, M., & Mahmood, A. (2023). Hybrid Facial Emotion Recognition Using CNN-Based Features. *Applied Sciences (Switzerland)*, *13*(9), 1–14. https://doi.org/10.3390/app13095572

Sidik, A. D. W. M., Suryana, A., Edwinanto, Artiyasa, M., Pradiftha Junfithrana, A., Himawan Kusumah, I., & Imamulhak, Y. (2021). Pengenalan Ekspresi Wajah Menggunakan Teknik Filter Wavelet Gabor. *FIDELITY : Jurnal Teknik Elektro*, *3*(1), 1–4. https://doi.org/10.52005/fidelity.v3i1.84

Sriyati, S., Setyanto, A., & Luthfi, E. E. (2020). Literature Review: Pengenalan Wajah Menggunakan Algoritma Convolutional Neural Network. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, *8*(2). https://doi.org/10.30646/tikomsin.v8i2.463