

# Optimizing Facial Expression Recognition with Image Augmentation Techniques: VGG19 Approach on FERCC Dataset

Fahma Inti Ilmawati<sup>1)\*</sup>, Kusri<sup>2)</sup>, Tonny Hidayat<sup>3)</sup>

<sup>1,2,3)</sup> University Amikom Yogyakarta, Indonesia

<sup>1)</sup>[fahma.inti.ilmawati@students.amikom.ac.id](mailto:fahma.inti.ilmawati@students.amikom.ac.id), <sup>2)</sup>[kusri@amikom.ac.id](mailto:kusri@amikom.ac.id), <sup>3)</sup>[tonny@amikom.ac.id](mailto:tonny@amikom.ac.id)

**Submitted** : Feb 19, 2024 | **Accepted** : Feb 29, 2024 | **Published** : Apr 1, 2024

**Abstract:** In the field of facial expression recognition (FER), the availability of balanced and representative datasets is key to success in training accurate models. However, Facial Expression Recognition Challenge (FERC) datasets often face the challenge of class imbalance, where some facial expressions have a much smaller number of samples compared to others. This issue can result in biased and unsatisfactory model performance, especially in recognizing less common facial expressions. Data augmentation techniques are becoming an important strategy as they can expand the dataset by creating new variations of existing samples, thus increasing the variety and diversity of the data. Data augmentation can be used to increase the number of samples for less common facial expression classes, thus improving the model's ability to recognize and understand diverse facial expressions. The augmentation results are then combined with balancing techniques such as SMOTE coupled with undersampling to improve model performance. In this study, VGG19 is used to support better model performance. This will provide valuable guidelines for optimizing more advanced CNN models in the future and may encourage further research in creating more innovative augmentation techniques.

**Keywords:** Augmentation; Convolutional Neural Network; Image Generator; SMOTE; VGG19

## INTRODUCTION

Non-verbal communication, such as human expressions, plays a crucial role in everyday human interactions. Facial expressions are not just about expressing emotions such as anger, discomfort, or happiness, but can also be an important means to complement spoken language in conveying messages and understanding the feelings of others. However, human emotions can sometimes be very complex, and vary depending on the situation and surrounding environment (Pise et al., 2022). In the scope of computer vision and machine learning, facial expression recognition is an interesting research subject. Many facial expression recognition systems have been developed to interpret and recognize emotions from facial expressions represented in the form of visual data. However, understanding and interpreting facial expressions remains a challenge due to the complexity of human emotions.

Research by (Mehendale, 2020), there are seven basic emotional expressions in humans including anger, fear, happiness, sadness, disgust, surprise, and neutral. These expressions are considered a valuable communication tool across cultures, allowing humans to understand each other even without using words (Küntzler, Höfling, & Alpers, 2021). Although FER (Facial Expression Recognition) has great potential in improving human-computer interaction, there are still a number of challenges that need to be overcome. One of them is that the image quality in FERC datasets is often lower than the image quality obtained from digital cameras, which can affect the effectiveness of the algorithm in identifying faces (Febrian, Halim, Christina, Ramdhan, & Chowanda, 2022). The large number of FERC datasets also causes expensive computational costs and takes a lot of time (Prabaswera & Soeparno, 2023). The variability in faces such as differences in shape, size, pose, expression, lighting, and occlusion, also affects the ability of the model to generalize well (He, 2023). The model that cannot generalize well is also due to the imbalance of data in the class. Data imbalance in the class is also a big challenge (Li & Deng, 2022). This is because learning tends to focus on the majority class, resulting in a suboptimal and uneven model. Therefore, solving the data imbalance problem is an important priority in developing a reliable FERC model. The combination of the factors mentioned makes the Facial Expression Recognition Image Version of FERC dataset complex and challenging to analyze.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

In this research, the main focus is to optimize the VGG19 model by addressing the data imbalance problem encountered in the Facial Expression Recognition Image Version of (FERC) dataset. Data balancing is a major concern because the uneven distribution of various emotional expressions in the dataset may affect the performance of the facial expression recognition model. Therefore, this research will adopt an approach using augmentation techniques and the use of SMOTE (Synthetic Minority Over-sampling Technique) method, which is then followed by the data balancing process in the CNN architecture, to overcome the problem of data imbalance in the FERC dataset and produce a more reliable model in facial expression recognition.

## LITERATURE REVIEW

Before starting research on facial expressions, researchers conducted a literature review to review previous studies as a reference base in formulating the research conducted. In previous research conducted by (Pham et al., n.d.), proposed a New Masking Idea to improve CNN performance in facial expression recognition tasks using personal FER2013 and VEMO datasets. By combining the concept of Deep Residual Network with Unet-like architecture to form an advanced and effective residual Masking Network in accurately recognizing facial expressions. By adding data augmentation methods including left-right flipping and rotating to prevent overfitting (Białek, Matiolański, & Grega, 2023). The application of data augmentation is proven to increase data variation and reduce the occurrence of overfitting in the model (Tripathi, Khatri, & Greunen, 2022). Furthermore, in research (Isthigosah, Sunyoto, & Hidayat, 2023), applying data augmentation techniques such as image generator, SMOTE, and ADASYN. All three provided good accuracy values on the BreakHis breast cancer medical dataset. The use of SMOTE in the study (Kummer, Ruppert, Medvegy, & Abonyi, 2022), in addition to increasing the quantity of data, also increased model performance by 50%. The SMOTE algorithm utilizes the nearest sample data from the minority class to create new synthetic data, thus effectively improving the representation of the less dominant class (Balla, Habaebi, Elsheikh, Islam, & Suliman, 2023). Expanding the amount of data in the minority class, the SMOTE algorithm addresses the problem of class imbalance and improves the model's performance in detecting emotions in Alzheimer's patients (Dablain, Krawczyk, & Chawla, 2022). The application of SMOTE can be used to reduce bias towards the majority class (Chan, Kelly, & Schnabel, 2021), but it is important to be aware that the addition of synthetic samples can introduce noise to the minority class, potentially causing overfitting if the number of synthetic samples generated is too large (Chen, Chang, & Guo, 2021). The synthesized images may not fully interpret the original images and tend to be distributed as dots (Xu & Wali, 2023). This phenomenon may impact the quality of the synthetic data generated, ultimately affecting the model's ability to generalize well.

In this research, VGG19 will be the main focus. A number of studies have explored the application of deep learning models for facial expression recognition (FER), and VGG19 has proven to be a highly effective model for this task VGG19, an artificial neural network (CNN), has shown excellence in tackling the FER task thanks to its unique strategy in complex facial feature extraction (Pham et al., n.d.). These features made VGG19 chosen for the facial expression recognition task.

This research aims to utilize knowledge to deal with the constraints posed by imbalances and limitations, especially in the unevenly distributed FERC dataset. The VGG19 architecture is used as the main model. To overcome these challenges, a combination of image generator and SMOTE augmentation techniques will be implemented. The approach shows potential to improve model performance as well as reduce the impact of class imbalance. The evaluation will focus on analyzing the effectiveness of image generator and SMOTE augmentation techniques using VGG19 architecture to deal with FERC dataset imbalance.

## METHOD

In this research, there is a research framework structured in five main steps. The first step is to conduct a literature review, identify the FERC dataset, apply data preprocessing in the form of a combination of Image Generator and SMOTE augmentation techniques, training and validation of the model using the VGG19 algorithm, followed by evaluation of the model performance as shown in Figure 1. Dividing the dataset into three parts is very important in an effort to get an objective assessment of the model's performance. The FERC dataset is divided into 80:10:10, where 80% is used for model training, 10% is used for validation, and 10% for testing. The use of training data helps the model in acquiring knowledge and adjusting parameters. Meanwhile, validation data is used to optimize hyperparameters and prevent overfitting during the training process. On the other hand, testing data is used to measure the final performance of the trained model with data that has never been seen before, thus providing a more accurate picture of the model's ability to generalize new data. By dividing the FERC dataset according to a predefined ratio, this research seeks to provide more reliable and relevant results in evaluating the performance and effectiveness of augmentation techniques applied to the CNN model architecture. The research flow can be seen in Figure 1.

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

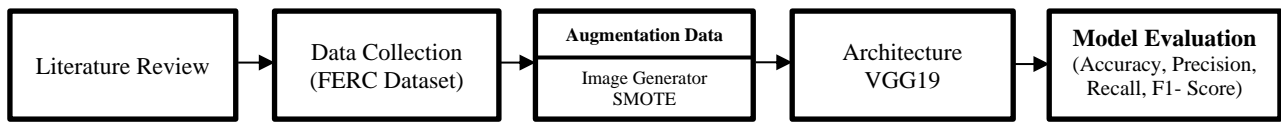


Fig. 1 Figure Dataset FERC

**Data Collection**

In this section, data collection is carried out with the aim of obtaining as much information as possible needed in order to carry out research activities. The dataset used in this research consists of images of human expression images that have been collected from the Facial Expression Recognition (FER) Challenge. The Facial Expression Recognition (FER) Challenge dataset is used to train and evaluate facial expression recognition models. This dataset has been used in the ICML competition and several research papers, making it one of the more challenging datasets with human-level accuracy. The following is an example dataset in FER in Figure 2.



Fig. 2 Figure Dataset FERC  
Source Image : Irmak et al., 2021

This dataset consists of 35,887 facial expression data labeled with seven emotion categories: Anger, disgust, fear, pleasure, sadness, surprise, and neutral. Previous research has used this dataset to develop and test deep learning models for recognizing facial expressions under various conditions, contributing to the advancement of facial expression recognition technology. The following is the distribution of FERC data in table 1.

Table 1. Class Distribution of FERC Dataset

Label	Total
Anger	4953
Disgust	547
Fear	5121

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Happiner	8989
Neutral	6198
Sadness	6077
Surprise	4002

### Data Augmentation

The main objective of this research is to evaluate the effectiveness of a combined augmentation technique using Image Generator and SMOTE in improving the performance of CNN models that use the VGG19 architecture. The main focus of this research is on the implementation of these augmentation techniques in the context of unbalanced and complex facial expression datasets. After the data is collected, the next step is to perform augmentation on the minority class data, especially the "disgust" class which has the least amount of data, namely 547 data. The augmentation techniques applied using Image Generator include rotation, range shift, zoom, and flip. Thus, this study aims to investigate whether the combination of these augmentation techniques can improve the CNN model's ability to recognize and learn complex patterns in facial expressions, especially on minority classes that are underrepresented in the dataset.

Table 2. Transformation Image Generator

Transformation	Specification
Rescale	1.0 / 255
Rotation range	30
Width shift range	0.2
Height shift range	0.2
Shear range	0.2
Zoom range	0.2
Horizontal flip	True
Vertical flip	True
Fill mode	Nearest

After the augmentation process with the image generator is complete, the next step is to apply the SMOTE (Synthetic Minority Oversampling Technique) method. SMOTE is used to overcome the imbalance of class distribution in a dataset by generating synthetic samples of minority classes. It has proven to be effective in dealing with class imbalance problems in classification tasks and is often used to improve the performance of machine learning models when faced with imbalanced data. The approach works by taking samples from the feature space, drawing a line between them, and creating synthetic samples at points along the line. Furthermore, to obtain optimal results, oversampling (SMOTE) is combined with under sampling of the majority class. After the application of SMOTE, the dataset is divided into training, validation, and testing sections to prepare the model for the learning process. Next, the image will be resized from the previous image to be sure it is 48x48 pixels. Then, followed by the split data process, the dataset will be randomly divided into 3 parts, namely 80% for training data, 10% for validation data, and 10% for testing.

### Convolutional Neural Networks

Convolutional Neural Networks (CNN) is one type of artificial neural network architecture that is often used in image processing tasks, such as image recognition or object detection. Feature Learning on the Convolutional Layer is the first step in CNN that aims to extract important features from the input image (Anantrasirichai & Bull, 2022). Convolution is done by applying various filters or kernels to the entire image to identify patterns such as edges, corners, and textures. The Activation Layer then applies an activation function, such as ReLU (Rectified Linear Unit), to the convolution results (Febrian et al., 2022). This helps in recognizing more complex features in the image. The Pooling Layer is then used to reduce the dimensionality of the data by taking the maximum or average value of certain regions, thus helping in reducing overfitting and saving computational resources.

Next, at the Structural Condition Classification stage, the process involves Flatten, which converts the feature extraction results from the convolution and activation layers into one-dimensional vectors so that they can be used by subsequent layers. Fully Connected Layer is the layer located after Flatten. This layer consists of neurons that are fully connected to the neurons in the previous layer, allowing the neural network to understand more complex relationships in the data. Finally, the Softmax layer is used in classification problems, generating a probability distribution for each possible class. The class with the highest probability is considered the prediction of the network.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



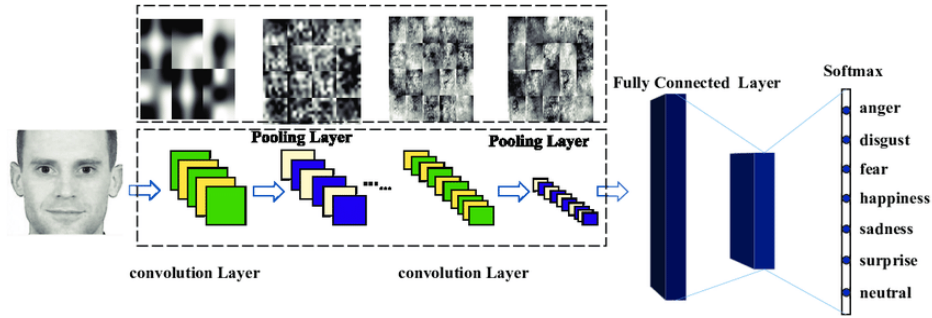


Fig. 4 Convolution Neural Network Process  
Source Image : Wang et al., 2019

**VGG19**

VGG19 (Visual Geometry Group 19) is one of the most well-known and commonly used types of Convolutional Neural Network (CNN) architecture in image processing tasks. The feature extraction process using VGG19 involves a series of steps (Gour, Jain, & Sunil Kumar, 2020). First of all, the network receives an input image of a specified size, e.g. 48x48 pixels, and before the image is fed into the VGG19 model, pre-processing is performed to normalize the pixel scale into an appropriate range. The VGG19 architecture consists of a series of convolution and pooling layers. Each convolution layer is followed by ReLU activation, which helps in extracting non-linear features. The feature extraction process takes place in the convolution layer, where convolution filters are applied to the input image to generate increasingly complex features. Each convolution layer is tasked with identifying a particular pattern or feature in the image.

After the convolution layer, a pooling layer is performed to reduce the dimensionality of the features and simplify the resulting representation. VGG19 uses a maximum pooling layer, where the maximum value is taken from each window shifted on the features generated by the previous convolution layer. After the convolution and pooling process, the extracted features are passed to the fully connected layer. This layer consists of several fully connected neurons, tasked with connecting the extracted features with their corresponding class labels. Finally, at the output layer, the results from the fully connected layer are connected to the softmax layer, which generates a prediction probability for each class. The class with the highest probability will be considered as the final prediction of the model. Briefly, VGG19 performs feature extraction by using a convolution layer to recognize important patterns in the image, followed by a pooling layer to reduce the dimensionality of the features, and a fully connected layer to associate the features with class labels. The end result is a predicted probability for each class, which can be used to perform image classification. The VGG19 architecture can be viewed in Fig 4.

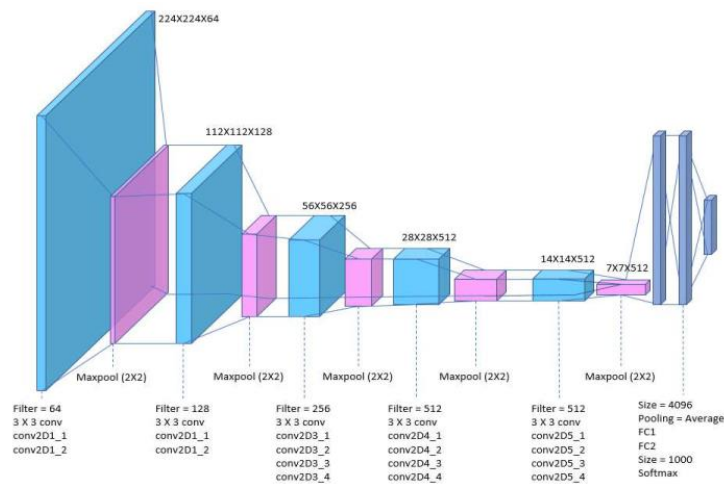


Fig. 4 VGG19 Architecture  
Source Image : Lagunas & Garces, 2017

In this study, the last three layers of VGG19 were converted into layers, namely one hidden layer (fully connected layer) with 1024 nodes, and used ReLU to activate the function, then added a dropout with a size of 0.5 to overcome overfitting. Furthermore, the number of outputs in the hidden layer (fully connected layer) which was originally 1000 classes was changed to 6 classes with a softmax of 0.5 to overcome overfitting.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

## Model Evaluation

The last step carried out in this research is the evaluation of the performance of the VGG19 model after applying the image generator and SMOTE augmentation techniques. The model was assessed using a series of tests generated by splitting the dataset before the model training process was performed. This research uses various metrics to compare the performance of the model including accuracy, precision, and recall.

Before being evaluated, the model undergoes a training process, some hyperparameter settings are adjusted first. The hyperparameters applied include the learning rate, optimizer, number of epochs, and three different batch sizes. Full details of the hyperparameter settings can be found in Table 3.

Table 3. Hyperparameter Details

Learning Rate	Optimizer	Epoch	Batch Size
Reduce Learning Rate (0.0001)	Adam	100	48

In addition to utilizing the hyperparameters as listed in Table 3. To improve accuracy, this study also implemented callback functions. The functions used in training include model checkpoints to save the model periodically during the training process, early stopping to stop the training early if there is no improvement, and decreased learning rate to control the learning rate to achieve more optimal convergence.

Furthermore, to evaluate the performance of the model, the values of accuracy, precision, and recall are used. Accuracy is a measure of the number of correct predictions given by the trained model. Precision is measured as the percentage of total cases correctly predicted by the model. Precision indicates the percentage of the model's predictions that would be correct if they were correct. Recall, also known as sensitivity, shows how often the model is able to correctly predict positive cases. Recall displays the percentage of positive cases out of all actual cases in the dataset.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \quad (1)$$

$$\text{Recall} = \frac{TP}{(FN+TP)} \times 100\% \quad (2)$$

$$\text{Precision} = \frac{TP}{(FP+TP)} \times 100\% \quad (3)$$

Based on equations (1), (2) and (3), TP represents the number of FER samples correctly classified by the system. TN represents the number of negative FER samples that were correctly classified, FP represents the number of negative samples that were misidentified as positive, and FN represents the number of positive samples that were misidentified as negative.

## RESULT

The research process involved the use of a dataset of 28,000 samples. The result of the data is the result of applying augmentation techniques as well as applying SMOTE oversampling followed by under sampling. Furthermore, the dataset was divided into three parts: 80% is used for the training process, 10% is used for validation, and 10% is used for testing. The selected architecture model is VGG19. In each experiment, we adopted a uniform approach with 100 epochs, optimization using Adam's algorithm, and applied three callback functions: checkpoint model check, learning speed reduction from 0.0001, and early termination. The results of these tests are presented in detail in Table 4.

Table 4. Accuracy and Loss

Augmentation	Accuracy Training	Accuracy Loss	Accuracy Validation	Loss Validation
Before Augmentation	0.7589	0.6593	0.6016	1.2108
After Augmentation	0.9939	0.0172	0.6114	1.1510

Before augmentation, the model had a training accuracy of 75.89%, with a loss rate of 65.93%. However, the validation accuracy was only 60.16%, with a validation loss rate of 1.2108. This suggests that the model may be

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

suffering from overfitting, where the model over-learns patterns specific to the training data and is less able to generalize the patterns found to data that has not been seen before.

After augmentation, there was a significant improvement in the performance of the model. The training accuracy increased dramatically to 99.39%, with a very low miss rate of 0.0172. Although the validation accuracy did not see a significant improvement (increasing only slightly to 61.14%), the validation miss rate managed to decrease to 1.1510. This shows that augmentation successfully improved the model's ability to generalize the patterns found in the training data to the validation data, thus reducing overfitting. However, it should be noted that very high training accuracy may show signs of overfitting in the training data, so further evaluation of the model is required.

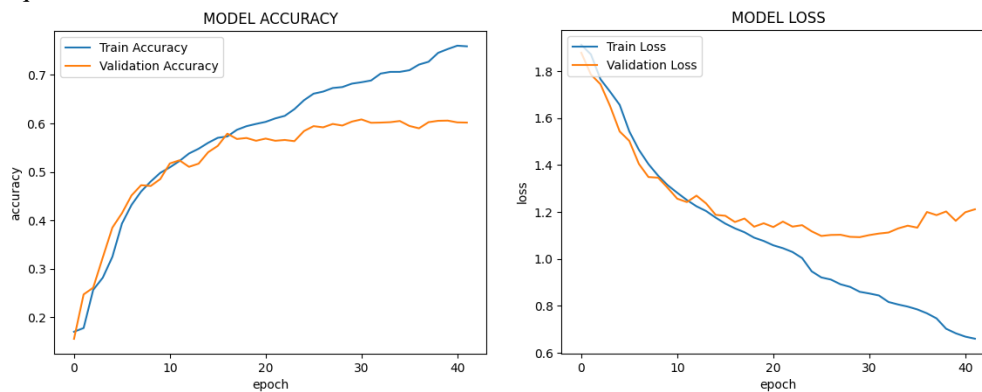


Fig.5 Model Accuracy and Loss before Augmentation

As seen in Figure 4, overfitting may occur because the model is not complex enough to handle the variation and complexity in the dataset. This could be due to the dataset size being too small, lack of variation in the data, or the model structure being too simple.

Further, we discuss the results of the VGG19 model architecture in the context of using augmentation techniques on our dataset. Before applying augmentation techniques, the training model faced challenges in achieving satisfactory accuracy and consistency in performance. Low accuracy and high loss values during the training process indicated that the model had difficulty in recognizing complex patterns in the data. However, the application of refinement techniques such as SMOTE and oversampling resulted in a noticeable, but not significant, improvement in model performance. During training, the accuracy continued to increase, and the loss rate decreased. This shows that the model can better understand the training data after augmentation. In addition, validation accuracy also improved although the decrease in validation loss was not significant. The results of these tests are presented in detail in Table 5.

Table 5. Accuracy, Precision, and Recall

Method	Accuracy	Precision	Recall
VGG19 (Before Augmentation)	62.73%	63.53%	63.65%
VGG19 (Without Augmentation)	73.93%	66.79%	66.38%

After looking at the performance evaluation results of the model using the VGG19 architecture before and after the application of augmentation techniques on Table 5, it can be concluded that augmentation has a significant impact on model performance. Before augmentation, the VGG19 model produced an accuracy of 62.73%, while after augmentation, the accuracy increased to 73.93%. This shows that augmentation successfully improves the model's ability to recognize complex patterns in the data. In addition, there was a significant increase in the precision of the model after augmentation, from 63.53% to 66.79%, which indicates that the model tends to give fewer false positives. However, there was a slight decrease in recall, from 63.65% to 66.38%, indicating that the model tended to miss more true positive cases after augmentation. Nonetheless, overall, the application of augmentation techniques to the VGG19 model has significantly improved the accuracy and precision of the model, which can be an indication of the improved quality and reliability of the model in performing data classification.

### DISCUSSIONS

Various studies have been conducted to date facial expression classification by applying augmentation techniques to facial expression data. Table 5 shows the evaluation results obtained through various architectures

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

and studies. However, it is difficult to compare these results fairly because the training and testing data used are different. Therefore, to compare with our approach, we selected studies that used data balancing and augmentation to report evaluations on the Facial Expression dataset. This approach allows the results to be compared with the performance evaluation of similar testing models. Our proposed model was able to outperform the results of previous studies with an accuracy rate of 96.09%. Overall, the use of image generators proved to be an effective method to improve data quality and diversity, resulting in improved performance of the VGG19 model on the FER dataset. In addition, it is important to choose a CNN model that matches the complexity of the image at hand. Therefore, choosing the right model is important when tackling complex tasks such as facial expression classification.

Table 4. Accuracy and Loss

Method	Accuracy
(Prabaswera & Soeparno, 2023)	71.80%
VGG19 (Without Augmentation)	62.73%
VGG19 (Augmentation)	73.93%

## CONCLUSION

The analysis results provide valuable insights into the application of augmentation techniques in facial expression datasets, enhancing model performance, particularly in addressing data imbalances. Augmentation enables training with additional varied data without the need for more physical data collection, saving time and cost. Careful selection of augmentation techniques is crucial, as it directly impacts the model's future performance. Model evaluation requires a consideration of classification task objectives, understanding the precision-recall trade-off, and analyzing relevant metrics and cases. Challenges, including potential noise from synthetic data, need addressing to ensure accurate evaluation. While optimization efforts have improved the VGG19 network for facial expression recognition, challenges persist due to complex and diverse human emotion expressions in datasets. Future research should focus on developing efficient techniques across diverse domains, considering real-world application impacts on model performance.

## REFERENCES

- Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: a review. *Artificial Intelligence Review*, 55(1), 589–656. <https://doi.org/10.1007/s10462-021-10039-7>
- Balla, A., Habaebi, M. H., Elsheikh, E. A. A., Islam, M. R., & Suliman, F. M. (2023). The Effect of Dataset Imbalance on the Performance of SCADA Intrusion Detection Systems. *Sensors*, 23(2). <https://doi.org/10.3390/s23020758>
- Białek, C., Matiolański, A., & Grega, M. (2023). An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks. *Electronics (Switzerland)*, 12(12). <https://doi.org/10.3390/electronics12122707>
- Chan, E., Kelly, M., & Schnabel, J. A. (2021). Comparison of classical machine learning deep learning to characterise fibrosis inflammation using quantitative MRI. *Proceedings - International Symposium on Biomedical Imaging, 2021-April*, 729–732. <https://doi.org/10.1109/ISBI48211.2021.9433962>
- Chen, Y., Chang, R., & Guo, J. (2021). Effects of Data Augmentation Method Borderline-SMOTE on Emotion Recognition of EEG Signals Based on Convolutional Neural Network. *IEEE Access*, 9, 47491–47502. <https://doi.org/10.1109/ACCESS.2021.3068316>
- Dablain, D., Krawczyk, B., & Chawla, N. V. (2022). DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data. *IEEE Transactions on Neural Networks and Learning Systems*, PP, 1–15. <https://doi.org/10.1109/TNNLS.2021.3136503>
- Febrian, R., Halim, B. M., Christina, M., Ramdhan, D., & Chowanda, A. (2022). Facial expression recognition using bidirectional LSTM-CNN. *Procedia Computer Science*, 216(2022), 39–47. <https://doi.org/10.1016/j.procs.2022.12.109>
- Gour, M., Jain, S., & Sunil Kumar, T. (2020). Residual learning based CNN for breast cancer histopathological image classification. *International Journal of Imaging Systems and Technology*, 30(3), 621–635. <https://doi.org/10.1002/ima.22403>
- He, Y. (2023). Facial Expression Recognition Using Multi-Branch Attention Convolutional Neural Network. *IEEE Access*, 11(December 2022), 1244–1253. <https://doi.org/10.1109/ACCESS.2022.3233362>
- Isthigosah, M., Sunyoto, A., & Hidayat, T. (2023). Image Augmentation for BreKHis Medical Data using Convolutional Neural Networks. *Sinkron*, 8(4), 2381–2392. <https://doi.org/10.33395/sinkron.v8i4.12878>
- Kummer, A., Ruppert, T., Medvegy, T., & Abonyi, J. (2022). Machine learning-based software sensors for machine state monitoring - The role of SMOTE-based data augmentation. *Results in Engineering*, 16(November). <https://doi.org/10.1016/j.rineng.2022.100778>
- Küntzler, T., Höfling, T. T. A., & Alpers, G. W. (2021). Automatic Facial Expression Recognition in Standardized

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



- and Non-standardized Emotional Expressions. *Frontiers in Psychology*, 12(May), 1–13. <https://doi.org/10.3389/fpsyg.2021.627561>
- Li, S., & Deng, W. (2022). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 13(3), 1195–1215. <https://doi.org/10.1109/TAFFC.2020.2981446>
- Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences*, 2(3). <https://doi.org/10.1007/s42452-020-2234-1>
- Pham, L., Huynh Vu, T., Anh Tran, T., Chi Minh City, H., Trung Ward, L., & Duc District, T. (n.d.). *Facial Expression Recognition Using Residual Masking Network*. Retrieved from <http://github.com/phamquilian/ResidualMaskingNetwork>
- 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>
- Prabaswera, D. R., & Soeparno, H. (2023). Facial Emotion Recognition Using Convolutional Neural Network Based on the Visual Geometry Group-19. *Jurnal TAM (Technology Acceptance Model)*, 14(1), 48. <https://doi.org/10.56327/jurnaltam.v14i1.1475>
- Tripathi, P., Khatri, S. K., & Greunen, D. Van. (2022). A transfer learning approach to implementation of pretrained CNN models for Breast cancer diagnosis. 6(4), 5816–5830.
- Xu, Y., & Wali, A. (2023). Handwritten Pattern Recognition using Birds-Flocking Inspired Data Augmentation Technique. *IEEE Access*, 11(May), 71426–71434. <https://doi.org/10.1109/ACCESS.2023.3294566>
- M. C. Irmak, M. B. H. Taş, S. Turan and A. Haşiloğlu, Emotion Analysis from Facial Expressions Using Convolutional Neural Networks, 2021 6th International Conference on Computer Science and Engineering (UBMK), Ankara, Turkey, 2021, pp. 570-574, doi: 10.1109/UBMK52708.2021.9558917.
- Wang, Y., Li, Y., Song, Y., & Rong, X. (2019). Facial expression recognition based on random forest and convolutional neural network. *Information (Switzerland)*, 10(12). <https://doi.org/10.3390/info10120375>
- Lagunas, M., & Garces, E. (2017). Transfer Learning for Illustration Classification. *27th Spanish Computer Graphics Conference, CEIG 2017*, 77–85. <https://doi.org/10.2312/ceig.20171213>

\*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.