IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 4, December 2024, pp. 4703~4710
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp4703-4710

Regularized Xception for facial expression recognition with extra training data and step decay learning rate

Elang Arkanaufa Azrien, Sri Hartati, Aufaclav Zatu Kusuma Frisky
Department of Computer Science and Electronics, Faculty of Science and Mathematics, Gadjah Mada University, Yogyakarta, Indonesia

Article Info

Article history:

Received Jan 6, 2024
Revised Apr 9, 2024
Accepted Apr 17, 2024

Keywords:

Augmentation data
Face expression recognition
Facial expression recognition2013
Regularized Xception
Step decay

ABSTRACT

Despite extensive research on facial expression recognition, achieving the highest level of accuracy remains challenging. The objective of this study is to enhance the accuracy of current models by adjusting the structure, the data used, and the training procedure. The incorporation of regularization into the Xception architecture, the augmentation of training data, and the utilization of step decay learning rate together address and surpass current constraints. A substantial improvement in accuracy is demonstrated by the assessment conducted on the facial expression recognition (FER2013) dataset, achieving a remarkable 94.34%. This study introduces potential avenues for enhancing facial expression recognition systems, specifically targeting the requirement for increased accuracy within this domain.

Corresponding Author:

Aufaclav Zatu Kusuma Frisky
Department of Computer Science and Electronics, Faculty of Science and Mathematics, Gadjah Mada Univesity
Sleman, Special Region of Yogyakarta, Indonesia
Email: aufaclav@ugm.ac.id

1. INTRODUCTION

Facial expression is a crucial kind of non-verbal communication that effectively conveys emotions and has a substantial impact on the emotions, ideas, and behaviors of others. They exert significant influence in determining how individuals see and engage with one another [1]. Mehrabian [2] indicates that non-verbal communication, including facial and bodily gestures, accounts for the highest percentage (55%) in conveying emotions, compared to vocal communication (38%), and verbal communication (7%) [2]. Facial expression recognition (FER) has been utilized in various domains of human-computer interaction (HCI), including autopilot systems, education, medical treatment, psychological treatment, and computer vision-based psychological analysis [3]. Recognition of facial expressions can be achieved through the utilization of advanced technologies such as machine learning and deep learning [4].

For its implementation, it requires the training of data to enable machines to comprehend and interpret facial emotions. Ekman [5] identifies happiness, anger, disgust, fear, sadness, and surprise as the fundamental human emotions. The FER2013 dataset, also known as the 2013 facial expression recognition dataset, was generated by Carrier and Courvill during the International Conference on Machine Learning in 2013 [6]. It is a dataset provided by Kaggle.

Various studies on deep learning-based FER have been conducted. Vignesh *et al*. [7] conducted a study that used a sophisticated deep learning framework termed segmentation visual geometry group (VGG)-19 to address the requirements of FER. The training model utilized the FER2013 dataset. The framework yielded an accuracy of 75.97% for the 2013 FER test data. Furthermore, it is worth noting that there are comparisons with other analogous studies that demonstrate the precision of the results mentioned below. An inherent challenge associated with the utilization of deep learning is the necessity for a substantial volume of data to achieve optimal model performance [8].

Research by Jaymon *et al.* [4], various architectures were employed for training, including the Xception model. The model achieved a training accuracy of 93.2% using the expanded FER2013 dataset. Nevertheless, the drawbacks lie in the validation accuracy and test accuracy sections, where only 64.4% and 65.2% were achieved. Research by Gunawan *et al.* [9], emotion recognition research was carried out using ConvNet on Google Colab. The training accuracy reached 97%, whereas the test accuracy was only 57.4%. It indicates that the model is overfitting. Overfitting is a common problem in supervised machine learning that hinders the ability to accurately generalize models to both training and testing data [10]. Regularization can be employed to mitigate overfitting [10]. Yao *et al.* [11] conducted a study to identify seven primary illnesses affecting peaches using a regularized Xception model. Prior to regularization, the validation accuracy achieved was merely 65.37%. However, following the implementation of L2-norm regularization, the validation accuracy significantly improved to 92.23%.

The learning rate is an essential hyperparameter that must be controlled and has a substantial impact on model training using gradient descent [12], [13]. Applying too high learning rates to complicated issues might adversely impact training and accuracy. Conversely, if the learning rate is set too low, the training progress will be hindered by minimal improvements to the work weight [14], resulting in a delayed process. Indeed, as time progresses, the rate at which validation errors reduce tends to slow down, resulting in a gradual fall in the variance of random noise. Inserting this component into the training procedure can improve its accuracy. An approach to accomplishing this is by utilizing the decay factor to simulate the decrease in variance of random noise [15].

This study will primarily focus on enhancing the precision of the FER2013 categorization. The given model can be utilized for the real-time classification of human faces. The paper presents the following contributions: i) the authors suggested incorporating regularized Xception models, increasing and augmenting the training data, and implementing step decay during the training process; ii) this study shows that using regularization techniques in the Xception model, along with increasing and augmenting the training dataset, can enhance the accuracy of training, validation, and testing. Additionally, it effectively mitigates the issue of overfitting that was observed in prior research; iii) furthermore, employing step decay can improve the model's convergence, allowing it to converge in 150 epochs; iv) the model's performance will be assessed using the FER2013 test data and is expected to demonstrate better results in comparison to earlier studies during the evaluation; and v) the study yielded a training accuracy of 96.51%, a validation accuracy of 94.73%, and a test accuracy of 94.34%.

The approach involves implementing the regularization technique proposed by Yao *et al.* [11] on the Xception architecture using the FER2013 dataset. To enhance precision, the training data will be supplemented with additional data from CK+ and extended & augmented Google FER. Further advancements in this research focus on the training process, specifically utilizing the step decay learning rate to enhance the precision of the model. This paper will be structured into distinct sections, including section 2 is method, section 3 is result and discussion, and section 4 is conclusion.

2. METHOD

The objective of this study is to enhance the precision of model on FER2013. The strategy involves augmenting the training dataset, applying regularization techniques to Xception architecture, and implementing a step decay mechanism throughout the training process. During the last phase, the model will undergo evaluation and prediction to assess its performance. The techniques we developed are illustrated in Figure 1.

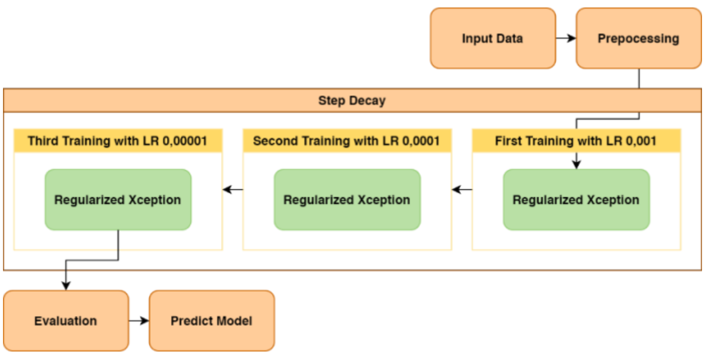


Figure 1. The stages of the proposed system for training method

At the beginning of the training, the program will receive datasets comprising training data, validation data, and test data. The data will undergo preprocessing, including normalization and data augmentation. The processed data is then utilized in the training process using the regularized Xception architecture with the step decay approach, which involves adjusting the learning rate at specific epochs. Following the completion of the training procedure, the finalized model will undergo evaluation to ascertain its performance. The technique will include detailed explanations for each procedure in discrete sections.

2.1. Dataset description

The primary dataset used in this research is FER2013 [16]. The dataset includes 35,887 grayscale photos with a resolution of 48×48 pixels. The dataset is structured as a.csv file, which has been split into three subsets: 28,709 training data, 3,594 validation data (publictest), and 3,589 test data (privatetest). Furthermore, the dataset has been categorized into seven distinct expressions: angry, disgust, fear, happy, sad, neutral, and surprise. To enhance the precision of the model, the training data is supplemented with additional data from CK+ [17] and extended and augmented Google FER (without the ‘contempt’ class) [18] from Kaggle. The data set of CK+ consists of 981 images, while the extended and augmented FER of Google is 38,191 images. The addition of this data brings the cumulative training data to a total of 67,881 photos. Figure 2 displays an example image from the FER2013 dataset.

Figure 2. Example images from FER2013 dataset

2.2. Prepocessing data

Preprocessing data involves transforming datasets from a csv file that contains the pixel values of an image into a numpy array with the data type float32. Next, we normalize the data by dividing each pixel value by 255 to bring it into the range of 0 and 1. Furthermore, we perform data augmentation on the training data using various criteria. Rotation with a maximum angle of 30 degrees chosen randomly. The width and height of the image might vary by up to 20%. Zoom in/out (zoom range) refers to the ability to increase or decrease the size of an image by a maximum factor of 20%. Horizontal flip refers to the reflection of an image along a vertical axis, resulting in a mirrored version of the original picture. Fill Mode fills empty areas that may emerge due to playback or shift. Fill Mode fills empty areas that may emerge due to playback or shift by filling the blank region with the pixel closest in proximity, using the "nearest" option.

2.3. Regularized Xception model

A widely used technique in deep learning is the convolutional neural network (CNN). This technique offers numerous key benefits such as image classification, segmentation, object detection, video processing, natural language processing, and audio recognition [19]. One of the most widely used CNN architectures is Xception.

Xception is a shortened form of "extreme inception" as it is grounded on the assumption that the process of mapping correlations between different channels and spatial correlations within CNN feature maps can be entirely disentangled. This hypothesis is an enhanced iteration of the hypotheses that form the foundation of the Inception architecture. The Xception design comprises 36 convolutional layers, which serve as the foundation for extracting network features. The architecture consists of 14 modules, each containing 36 convolution layers. All modules, save for the first and last ones, are connected by a residual linear link [20]. Fundamentally, the Xception architecture is a sequential arrangement of depthwise separable convolution layers with residual connections.

Xception is composed of three sections: the entry flow, the middle flow, and the exit flow. The entry flow component, which constitutes the initial segment of Xception, is tasked with the processing of the incoming image. The middle flow section serves as the central component of the Xception network and consists of a sequence of profound convolutional blocks. Nevertheless, the design in this study was altered by incorporating a layer dropout in the middle flow section, along with applying L2 regularization to the exit flow section. The purpose of this change was to mitigate overfitting in the model outcome [11]. Figure 3 displays the modified Xception architecture.

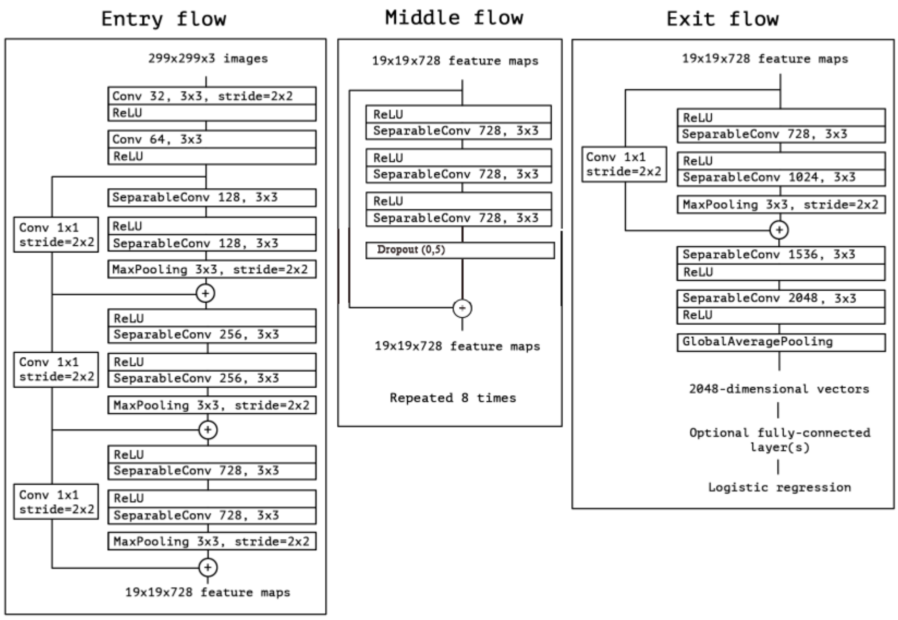


Figure 3. Modified xception architecture

2.4. Step decay method

Step Decay is a learning rate scheduling technique that entails decreasing the learning rate by a fixed factor at regular intervals, leading to a gradual decline in the learning rate over time. These schedules are frequently employed in the training of neural networks and have been demonstrated to enhance optimization and generalization [21]. This research used softmax activation and a step decay schedule for network training. The starting learning rate is set to 0.001 for a total of 100 epochs. Subsequently, the learning rate will be diminished to 0.0001 and reset for a total of 30 epochs. Ultimately, the learning rate will be reduced to 0.00001, and 20 epochs will be reloaded. The batch size will be configured as 64.

2.5. Evaluate and predict model

Model testing is conducted to assess the performance of the designed model and determine the success rate of classifying each facial expression. The model is evaluated using a data set from the FER2013 (private test) using the "model.evaluate" and "model.predict" modules [22]. The function model.evaluate will assess the model's suitability for the specified problem and data. This approach will provide a percentage score that quantifies the model's performance. The function Model.predict is utilized to obtain predictions from the model.

3. RESULTS AND DISCUSSION

3.1. First training

The initial phase of the training involves 100 epochs with a learning rate of 0.001, utilizing categorical crossentropy loss and a batch size of 64. Figure 4 displays the outcomes of the initial training session after the completion of the running process. The training data indicates a training accuracy of 90.67% and a loss of 0.271. The validation accuracy achieved the highest performance with a value of 87.94%, while the loss was measured at 0.364.

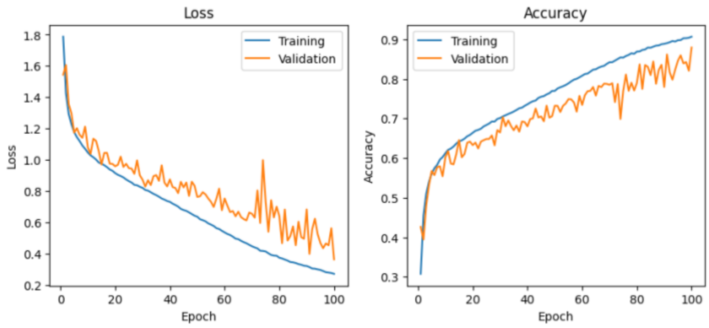


Figure 4. Accuracy and loss graph from first training

3.1. Second training

In the second part of the training, this is the advanced training of the first training. The learning rate on this training is changed to 0,0001 with an epoch of 30. Once run, the results of this second training are shown in Figure 5. It seems that the results of this training have improved quite significantly. This training gets the best training accuracy of 95.86% and a loss of 0.221. For the best validation accuracy, it achieves a score of 93.95% and a loss of 0.174.

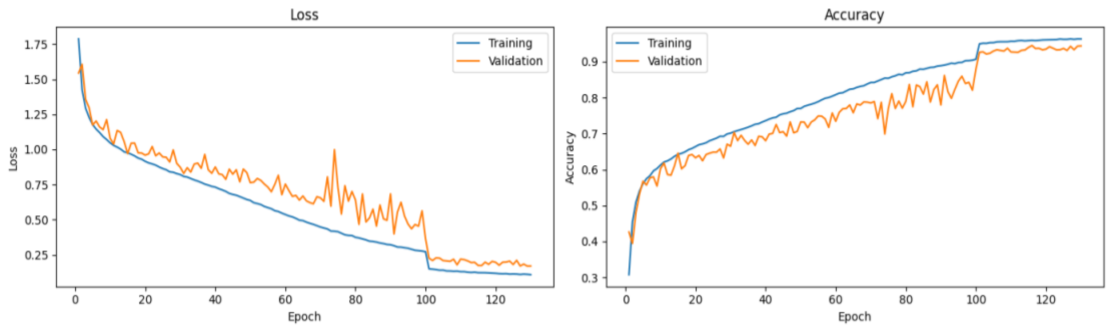


Figure 5. Accuracy and loss graph from second training

3.2. Third training

This is the last part of the model training, which is an advanced training session following the second training. For the third training session, the learning rate was adjusted to 0.00001 and the number of epochs was set to 20, resulting in a total of 150 epochs for the training process. Upon execution, the outcomes of this training are displayed in Figure 6. Based on the outcomes of this training, there has been a discernible enhancement in performance, although it is not very substantial. The model has achieved convergence. The third training session yielded a training accuracy of 96.51% and a loss of 0.1032. The validation accuracy was 94.73%, with a loss of 0.16.

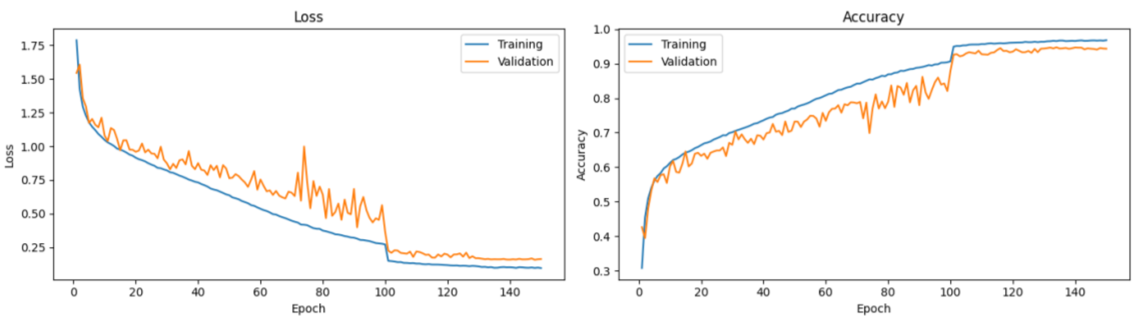


Figure 6. Accuracy and loss graph from third training

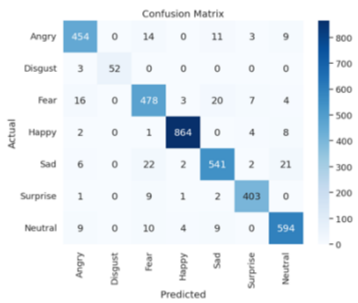
3.3. Confusion matrix and classification report

The FER model is evaluated using the FER2013 (privatetest) dataset, which consists of a total of 3,589 pictures. The model's performance will be evaluated by examining the outcomes of test accuracy, test loss, confusion matrix, and classification report, which include precision, recall, f1-score, support, macro average, and weighted average. Precision is the ratio of accurately anticipated positive observations to the total number of positive predictions. Recall refers to the proportion of accurately predicted positive observations out of all the observations in the actual class. The F1-score is calculated as the arithmetic mean of precision and recall. Support refers to the frequency of a class appearing in a given dataset.

Figure 7 displays the outcomes of the loss test and accuracy test conducted on the trained model. The exam has an accuracy rate of 94.34%. Figure 8 provides a more detailed explanation of the results through the confusion matrix and categorization report. The confusion matrix reveals that the happy class exhibits the highest count of true positives, with a total of 864. The other classes, namely neutral, sad, fearful, angry, surprised, and disgusted, follow in descending order. The quantity of photographs in each class influences this, with the happy classes having the highest count of 879 images.

```
113/113 [==============================] - 40s 348ms/step - loss: 0.1663 - accuracy: 0.9434
Test Loss: 0.1663
Test Accuracy: 94.34%
```

Figure 7. Test loss and test accuracy model




```
Classification Report:
              precision    recall  f1-score   support

       Angry       0.92      0.92      0.92       491
     Disgust       1.00      0.95      0.97        55
        Fear       0.90      0.91      0.90       528
       Happy       0.99      0.98      0.99       879
         Sad       0.93      0.91      0.92       594
    Surprise       0.96      0.97      0.97       416
     Neutral       0.93      0.95      0.94       626

    accuracy                           0.94      3589
   macro avg       0.95      0.94      0.94      3589
weighted avg       0.94      0.94      0.94      3589
```

Figure 8. Confusion matrix and classification report model

To ascertain the performance of each class, go to the classification report. The disgust class achieves the maximum level of accuracy, reaching a perfect score of 100%. This demonstrates that in a disgust test, when a non-disgust image is presented, the model consistently fails to predict that the image is not disgusting. The happy class achieves the highest recall and f1-score, with values of 98% and 99%, respectively. This suggests that the model demonstrates a favorable equilibrium between precision and recall only for the happy class.

3.4. Analysis of results

The research results show that the model achieved a training accuracy of 96.5%, a validation accuracy of 94.73%, and a test accuracy of 94.34%. In addition, each class achieves a f1-score value of > 90%, indicating a high level of balance between precision and recall for the findings in each class. This model exhibits superior performance in comparison to prior research. Table 1 presents a comparison of the model with previous research conducted on FER2013.

Table 1. The comparison of accuracy for each model

References	Method	Accuracy (%)
Jaymon *et al.* [4]	Simple CNN model	54
Gunawan *et al.* [9]	ConvNet model	57.4
Jaymon *et al.* [4]	Inception model	61.42
Jaymon *et al.* [4]	Xception model	65.2
Vignesh *et al.* [7]	Segmentation VGG-19	75.97
Zhang *et al.* [23]	CNN + image edge computing (FER2013 + LFW Dataset)	88.56
Debnath *et al.* [8]	Fussion features (CNN + LBP + ORB) + ConvNet	91.01
Proposed method	Regularized Xception + step decay learning rate (FER2013 + Extended & Augmented Google FER + CK+ dataset)	94.34

Although extra training data was added to enhance the accuracy of facial expression identification in the FER2013 dataset, this study still has limitations to address. Augmenting the training data can enhance the models' performance in FER, but the limited size of the FER2013 dataset is still a significant challenge. These constraints could impact models' capacity to extrapolate patterns seen in training datasets to more extensive real-world scenarios. Variable image quality, class imbalance, and inaccurate emotional class annotations can impact the model's accuracy in recognizing facial emotions. For instance, the emotion of happiness not only includes feelings of joy but also shares characteristics similar to those of peacefulness. The resemblance to peacefulness may lead to data being misclassified [24]. In addition, the images in the dataset have variations in illumination levels, which can affect the accuracy of extracting features from facial expressions [25]. These findings imply that while adding more training data improves accuracy, our model requires further attention to address restrictions and enhance performance in real-world applications.

4. CONCLUSION AND FUTURE RESEARCH

This document outlines strategies for enhancing the precision of FER2013 testing by utilizing a collection of Xception architectures. This approach entails augmenting the training data by incorporating CK+, extended and augmented Google FER datasets. Additionally, step-decay learning levels are employed during the training phase. The study demonstrates a substantial enhancement in the accuracy test outcomes, surpassing previous research findings with a score of 94.34%. Our recommendation for future research is to further investigate advanced methods in image processing and machine learning to improve models' capability in identifying facial expressions. Future research could concentrate on creating more sophisticated models using data from diverse sources and with greater variability in facial expressions. Furthermore, all data inside the FER2013 dataset can be verified to ensure alignment with the provided label, enhancing the accuracy of the generated model.

ACKNOWLEDGEMENT

This work was partially supported by the Department of Computer Science and Electronics, Universitas Gadjah Mada under the Publication Funding Year 2024.

REFERENCES

[1] G. A. V. Kleef and S. Côté, "The social effects of emotions," *Annual Review of Psychology*, vol. 73, pp. 629–658, 2022, doi: 10.1146/annurev-psych-020821-010855.

[2] A. Mehrabian, "Communication without words," in *Communication Theory: Second Edition*, Routledge, 2017, pp. 193–200, doi: 10.4324/9781315080918-15.

[3] Z. Y. Huang *et al.*, "A study on computer vision for facial emotion recognition," *Scientific Reports*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-35446-4.

[4] N. Jaymon, S. Nagdeote, A. Yadav, and R. Rodrigues, "Real time emotion detection using deep learning," in *Proceedings of the 2021 1st International Conference on Advances in Electrical, Computing, Communications and Sustainable Technologies, ICAECT 2021*, Feb. 2021, pp. 1–7, doi: 10.1109/ICAECT49130.2021.9392584.

[5] P. Ekman, "Are there basic emotions?," *Psychological Review*, vol. 99, no. 3, pp. 550–553, 1992, doi: 10.1037/0033-295x.99.3.550.

[6] I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," *Neural Networks*, vol. 64, pp. 59–63, Apr. 2015, doi: 10.1016/j.neunet.2014.09.005.

[7] S. Vignesh, M. Savithadevi, M. Sridevi, and R. Sridhar, "A novel facial emotion recognition model using segmentation VGG-19 architecture," *International Journal of Information Technology*, vol. 15, no. 4, pp. 1777–1787, Apr. 2023, doi: 10.1007/s41870-023-01184-z.

[8] T. Debnath, M. M. Reza, A. Rahman, A. Beheshti, S. S. Band, and H. A. -Rokny, "Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity," *Scientific Reports*, vol. 12, no. 1, Apr. 2022, doi: 10.1038/s41598-022-11173-0.

[9] T. S. Gunawan *et al.*, "Development of video-based emotion recognition using deep learning with Google Colab," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 5, pp. 2463–2471, Oct. 2020, doi: 10.12928/TELKOMNIKA.v18i5.16717.

[10] X. Ying, "An overview of overfitting and its solutions," *Journal of Physics: Conference Series*, vol. 1168, no. 2, Feb. 2019, doi: 10.1088/1742-6596/1168/2/022022.

[11] N. Yao *et al.*, "L2MXception: an improved Xception network for classification of peach diseases," *Plant Methods*, vol. 17, no. 1, Dec. 2021, doi: 10.1186/s13007-021-00736-3.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Massachusetts, USA: The MIT Press, 2016.

[13] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," *Neural Networks: Tricks of the Trade*, pp. 437–478, 2012, doi: 10.1007/978-3-642-35289-8_26.

[14] Z. Xu, A. M. Dai, J. Kemp, and L. Metz, "Learning an adaptive learning rate schedule," *arXiv-Computer Science*, pp. 1-6, 2019, doi: 10.48550/arXiv.1909.09712.

[15] Q. Dong and G. Luo, "Progress indication for deep learning model training: a feasibility demonstration," *IEEE Access*, vol. 8, pp. 79811–79843, 2020, doi: 10.1109/ACCESS.2020.2989684.

[16] D. Erhan, "Challenges in representation learning: facial expression recognition challenge," *Kaggle*, 2013. [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge

[17] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*, Jun. 2010, pp. 94–101, doi: 10.1109/CVPRW.2010.5543262.
[18] P. Sood, “Extended and augmented google FER,” *Kaggle*, 2021, [Online]. Available: https://www.kaggle.com/datasets/prajwalsood/google-fer-image-format/data
[19] Purwono, A. Ma’arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. U. Haq, “Understanding of convolutional neural network (CNN): a review,” *International Journal of Robotics and Control Systems*, vol. 2, no. 4, pp. 739–748, Jan. 2022, doi: 10.31763/ijrcs.v2i4.888.
[20] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Jul. 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.
[21] R. Ge, S. M. Kakade, R. Kidambi, and P. Netrapalli, “The step decay schedule: A near optimal, geometrically decaying learning rate procedure for least squares,” *Advances in Neural Information Processing Systems*, vol. 32, 2019, doi: 10.48550/arXiv.1909.09712.
[22] F. Chollet, “Training & evaluation with the built-in methods,” *Keras*, 2019, [Online]. Available: https://keras.io/guides/training_with_built_in_methods/
[23] H. Zhang, A. Jolfaei, and M. Alazab, “A face emotion recognition method using convolutional neural network and image edge computing,” *IEEE Access*, vol. 7, pp. 159081–159089, 2019, doi: 10.1109/ACCESS.2019.2949741.
[24] Y. H. Chin, S. H. Lin, C. H. Lin, E. Siahaan, A. Frisky, and J. C. Wang, “Emotion profile-based music recommendation,” in *Proceedings - 2014 7th International Conference on Ubi-Media Computing and Workshops, U-MEDIA 2014*, Jul. 2014, pp. 111–114, doi: 10.1109/U-MEDIA.2014.32.
[25] A. Z. K. Frisky, A. Harjoko, L. Awaludin, S. Zambanini, and R. Sablatnig, “Investigation of single image depth prediction under different lighting conditions,” *Journal on Computing and Cultural Heritage*, vol. 14, no. 4, 2021, doi: 10.1145/3465742.

BIOGRAPHIES OF AUTHORS

Elang Arkanaufa Azrien graduated with a Bachelor of Science degree from Gadjah Mada University, Indonesia, in 2024. Currently, he is one of the researchers in the Electronics and Instrumentation Studies Program at Gadjah Mada University. His main research focuses on machine learning and computer vision. With knowledge in both fields, he engages in research aimed at gaining in-depth understanding and developing practical solutions. He can be contacted at email: elang.arkanaufa.azrien@gmail.com or elang.arkanaufa.azrien@mail.ugm.ac.id.

Sri Hartati currently holds the position of Lecturer in the Department of Computer Science and Electronics at the Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada (UGM). She earned her Master's degree in Computer Science from the University of New Brunswick, Canada, during the period of September 1988 to October 1990. With a profound academic background. She has been actively contributing to the fields of computer science, computer reasoning, and artificial intelligence. Her dedication is reflected in her extensive publication record, boasting 114 articles in reputable journals and conferences. She can be contacted at email: at shartati@ugm.ac.id.

Aufaclav Zatu Kusuma Frisky was born in Surakarta, Indonesia, at 1990. He received the bachelor degree in Department of Computer Science and Electronics from Universitas Gadjah Mada, Indonesia at 2012 and M.Sc. degree in Department of Computer Science and Information Engineering from National Central University, Taiwan at 2015. He graduated as a Doctoral Student at Computer Vision Lab, Institute of Visual Computing and Human-Centered Technology, TU Wien, Austria, in 2023. His research interest includes computer vision, single image reconstruction, and its applications in cultural heritage preservation. He can be contacted at email: aufaclav@ugm.ac.id.