Jurnal TAM (Technology Acceptance Model)
Jurnal TAM, Volume 14, Number 1, July 2023
E ISSN: 2579-4221; P ISSN: 2339-1103, pp. 48-56
Accredited SINTA 4 Number 225/E/KPT/2022
https://jurnal.ftikomibn.ac.id/index.php/index

FACIAL EMOTION RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK BASED ON THE VISUAL GEOMETRY GROUP-19

Dwi Redjeki Prabaswera[1], Haryono Soeparno[2]
[1,2]Computer Science Graduate Program, BINUS University, Jakarta, DKI Jakarta
[1,2]Jl. Kebon Jeruk Raya No. 27, Kebon Jeruk, Jakarta Barat, Indonesia
E-mail: dwi.prabaswera@binus.ac.id[1*], haryono@binus.edu[2]

Article history:
Received: April 18, 2023
Revised: June 5, 2023
Accepted: June 12, 2023

Corresponding authors
***dwi.prabaswera@binus.ac.id**

Keywords:
Facial Emotion Recognition,
Convolutional Neural Network,
FER-2013,
VGG-16,
VGG-19,
Deep Learning

Abstract
Facial recognition is one of many popular and difficult tasks in computer vision. A variety of research have been conducted on this subject, each of which suggests a stand-alone approach. While many studies strive for more accuracy, this study research aims to increase the efficiency of human computers by classifying emotions based on human faces using a self-based neural network. The usage of a Convolutional Neural Network (CNN) based on the Visual Geometry Group - 19 (VGG-19) classification model, which has been employed in ImageNet data sets and improved for emotion classification, is proposed in this paper. The classification process was conducted using FER-2013 dataset, which consists of over 35,000 facial images captured in various settings and contains 7 different emotions. The dataset was divided into three subsets, with 80% allocated for training, 10% for validation, and 10% for testing. With an accuracy of 71.80%, the proposed technique surpasses most self-based models.

1. INTRODUCTION

Currently, facial expression recognition is a challenging and interesting task, as evidenced by numerous previous competitions by researchers [1–8], a group of available datasets [1,9–13], and research on the subject. Many publications [14–17] have described the progress made in computer vision in recognizing emotions with faces. The extensive paper by Shan Li [15] is of particular importance because it provides a thorough explanation and review of existing and commonly used data sets for facial emotion recognition, as well as state-of-the-art (SOTA) and their respective results.

Facial Emotion Recognition 2013 (FER-2013) [1], Static Facial Emotion in the Wild (SFEW) [9], Cohn-Kanade (CK) [10], Extended Cohn-Kanade (CK+) [11], Japanese Association of Female Facial Expression (JAFFE) [12] and Expression in the Wild (ExpW) [13] are among the many sets of facial emotion recognition datasets available [14–17]. These datasets differ in many ways, which are generally described by one or more of the following factors: the amount of data, the number of emotion classes, image-based or sequential, and in conditions such as laboratory or in the wild. Each dataset contains hundreds to tens of thousands of rows of data, with some predefined datasets containing training, validation, and/or testing data distribution. These data sets also differ in the number of emotional classes they contain, with most containing six to eight emotional classes that include anger, disgust, fear, happiness, sadness, surprise, humiliation, and are frequently supplemented with neutral emotions. Despite the fact that some datasets do not include neutral expressions and/or insults. Image-based or sequential datasets (video-based or image-based) also provide variation in the research conducted, as each type of dataset requires a different approach to processing. Another difference is the state of the datasets, with lab-like datasets differing from in-the-wild datasets. The former is captured in ideal conditions (proper lighting, proper facial alignment, and/or minimal or no use of facial accessories such as

glasses), whereas the latter is captured in non-ideal conditions from real-world scenarios.

This data set is used in many works related to the completion of facial emotion recognition tasks, some of which use conventional methods, deep learning, pre-trained models, ensemble neural networks, combinations of deep learning with handmade feature selection techniques, and other related works that will be described in the sections that follow. This method achieved a diverse set of results, which can be summarized by the test accuracy of most studies for laboratory-like data sets, which frequently exceeds 90%, while testing accuracy for data sets in the wild rarely exceeds 75% [15].

With previous studies low accuracy results for facial emotion recognition with in-the-wild datasets (as compared to laboratory-like datasets), this paper attempts to improve facial emotion recognition accuracy for FER-2013 image-based in-the-wild datasets. The proposed model achieves a test accuracy of 71.80%, which is greater than most existing studies using a self-based neural network architecture, while being simpler in terms of network depth and topology and having end-to-end training capabilities.

This introduction is followed by review a of the core concepts, a brief description of FER-2013 as the dataset used in this paper experiment, the existing approach to facial emotion recognition tasks using the FER-2013 dataset, and the inspiring works that serve as the basis the of the paper's model. Following that, the proposed model is thoroughly described. The proposed model's results are then discussed and compared to other related works to determine how well the model performs in comparison to other approaches taken. Hopefully, the paper concludes with final thoughts on the proposed model and future work to improve the task of facial emotion recognition.

Based on a review of previous research, this study will use deep learning to recognize facial expressions, because deep learning has been shown to provide better accuracy results in classifying images than other methods. The CNN method was used in this study, along with the Transfer Learning of the VGG architecture, which was modified by adding Global Average Pooling to the top layer.

II. LITERATURE

2.1. Related research

Various researchers have previously conducted research on the classification of facial expressions. Jung's [18] research used the CNN and DNN methods to study facial emotion recognition. They compared the performance of the two methods and found that CNN performed better than DNN, with an 86.54% recognition rate. They also used the Haar-like method to create systems for real-time face detection and feature extraction. The study suggests that DNN overfitting is a possibility. The CK+ dataset, which contains 327 image sequences covering 7 universal emotions, was used.

Pramerdorfer's [19] research used the FER2013 dataset to extract features for image-based facial expression recognition using CNN with VGG, RestNet, and Inception architectures. They reviewed existing literature on the subject, identified barriers, and proposed future research directions. They discovered that using a modern CNN ensemble resulted in significant performance gains without the need for additional training data or facial registration. The accuracy of the FER2013 test was 75.2%, outperforming previous work.

Yang's [20] research suggests changes to the CNN algorithm and SGD optimizer to improve feature recognition rates and reduce time costs. To avoid gradient problems and accelerate convergence, the MCNN-DS algorithm employs a quadratic CNN structure with fixed linear units as activation functions. To minimize cross-entropy, the SGD optimizer inserts dropout layers into the all-connected layer and output layers. The proposed algorithm performed well on benchmark data sets such as MNIST and HCL2000, but performed poorly on the EnglishHand test set. MCNN-DS required significantly less time than MLP-CNN and SVM-ELM.

Gunawan's [21] research implemented Google Colab and CNN to classify facial expressions and develop video-based emotion recognition through deep learning. The study examined the pre and post processes involved in the model methodology, and their work demonstrates a common architectural model for developing deep learning recognition systems. When the Haar cascade Technique was applied to the FER2013 dataset, it resulted in 97% accuracy on the training set and 57.4% accuracy on the testing device. Various performance parameters are used as benchmarks in various studies to demonstrate progress in this area, and the study includes a data set that researchers can use to contribute to this field.

Pranav's [22] research studied human facial expressions into seven emotions using a CNN. Before settling on the final CNN model, which has six convolutional layers, two max pooling layers, and two fully connected layers, various models were tested. The final accuracy obtained after tuning hyperparameters was 0.60. Because of their ability to capture spatial features, CNNs perform better in image recognition.

Cheng's [23] research used CNN and VGG methods to accurately classify expression images. They optimized the network structure and parameters of the VGG-19 model and used transfer learning to compensate for a lack of training data. They used different CNN models to train and analyze facial expression data and discovered that the improved VGG-19 model had the highest accuracy of 96%, outperforming other models.

Sajjanhar's [24] research tested a CNN model for recognizing facial expressions and used it as a baseline for evaluating other pre-trained CNN models. The researchers compared the performance of Inception and VGG, which had previously been trained for object recognition, to VGG-Face, which had previously been trained for facial recognition. The experiments were carried out with the help of publicly available face databases such as CK+, JAFFE, and FACES.

Most researchers require appropriate methods for measuring the rate of detection of facial emotions using image processing and CNN architecture. However, the use of various methods can have an impact on the detection of facial emotions. As a result, this study will compare the ability of VGG16 and VGG19 to detect facial emotions. It is hoped that by comparing these two methods, an appropriate method for detecting facial emotions will be discovered by modifying some feature learning from both methods.

2.2. Emotion Recognition

The process of recognizing human emotions is known as emotion recognition. Accuracy in reading others' emotions varies greatly among people. Research on using technology to aid in emotion recognition is still in its infancy. In general, technology functions best when it makes use of a variety of modalities. The majority of the work done so far has been focused on automating the recognition of facial expressions from video, spoken expressions from audio, written expressions from text, and wearable-measured physiology.

Decades of scientific research have gone into developing and testing methods for automatic emotion recognition. There is now a large body of literature proposing and evaluating hundreds of different types of methods that employ techniques from a variety of fields, including signal processing, machine learning, computer vision, and speech processing. To interpret emotions, various methodologies and techniques such as Bayesian networks [25], Gaussian Mixture models [26], Hidden Markov Models [27], and Deep Neural Networks [28] can be used.

2.3. VGG-NET

K. Simonyan and A. Zisserman proposed the VGG-NET model of convolutional neural networks in their paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" [29]. VGG-Net won the ILSVRC-2014 match positioning task and came in second in the classification task. His outstanding contribution was demonstrating that small convolution filters can express stronger features in input data while requiring fewer parameters. Increasing the network depth can significantly improve the model's effect, and VGG-Net has good generalization capabilities for image datasets. As a result, VGG-Net is frequently used to extract image features.

VGG-Net is a network developed from Alex-Net. Before that, LeNet5 networks used large convolution kernels to obtain similar image features, and Alex-Net networks used 11 x 11, 5 x 5 and 3 x 3 filters. A major advance of VGG-Net is to simulate the efect of a larger receptive field by using multiple 3 x 3 convolutions in turn. Assuming that all data have a C channel, one 7 x 7 convolution layer will contain C* (7 x 7 xC) = 49C x C parameter, while the combination of three 3 x 3 convolution layers has only 3 x (C x (3 x 3 xC)) = 27C x C parameter. Therefore, choosing a small-size convolution kernel is better than a large-size conv-lution kernel. The structure of the VGG-Net network is very consistent. From start to finish, it consists of 3 x 3 convolutions and 2 x 2 pooling. After extracting features from the feature extraction layer, three full connection layers are added to improve the network's nonlinear mapping capabilities, while limiting the network size.

III. RESEARCH METHODS

3.1. Research Model

Determining the type of facial expression can be done manually but requires a relatively long time and sometimes there are still errors in determining the type of facial expression, besides that special experience and knowledge about facial expressions are needed. From the results of literature studies conducted, research on the classification of facial expressions has been carried out. This study developed research [19] which conducted research on the classification of facial expressions using the Transfer Learning method with VGG (Visual Geometry Group) architecture. The stages of research are shown in Figure 1.

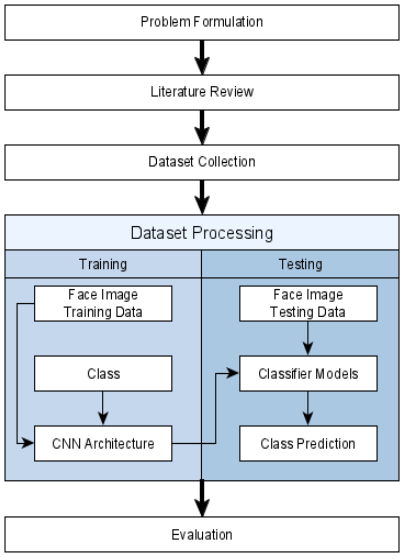


Figure 1. Research Stages

The research stage as shown in Figure 1 is problem formulation, literature review, data collection, training, testing, classification, evaluation in determining the best model of facial expression classification.

3.2. Dataset

The image dataset used in this study uses FER2013 data from www.kaggle.com, where this data consists of 35,887 images with an image size of 48x48 divided into 2 test sets: PublicTest and PrivateTest with 3,589 images each.

Figure 2. Facial emotion dataset (FER2013)

3.3. Data Acquisition & Preprocessing

The dataset that has been obtained, then processed first to facilitate the classification of facial expressions at the next stage, as well as to get more accurate results. In this study, data processing was carried out in several stages as follows:

1. *Filtering Image*

Images obtained from Kaggle are filtered back manually, there is no accurate filtering in the process.

2. *Labeling Image*

Images obtained from Kaggle and have gone through the filtering stage, are data that is still fairly raw, has not been labeled correctly, therefore it is necessary to do the image labeling process manually.

3. *Resize Image*

The labeled image is then resizing using the PIL library in Python to resize the image to 48*48, this is done to support the transfer learning process.

3.4. Classification Model

The Classification Model used in this study was obtained from the results of training using the proposed method. This research uses the VGG16 & VGG19 architecture as the basis for the proposed solution. Because based on the results shown in ILSVRC 2014 this architecture gets a top 5 accuracy of 90.10% which can be said to be almost similar to human capabilities.

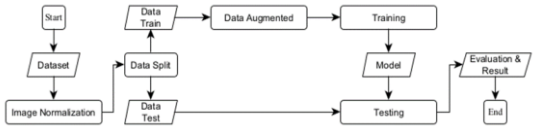


Figure 3. System Design Block Diagram

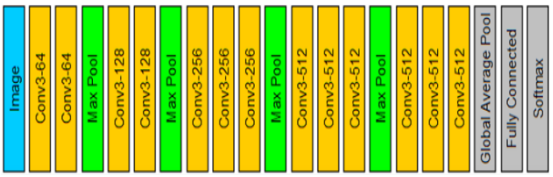


Figure 4. Model architecture VGG-16

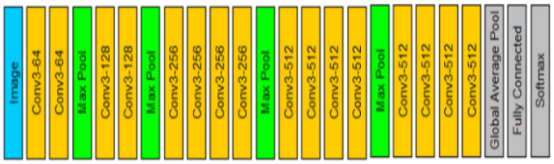


Figure 5. Model architecture VGG-19

3.5. Evaluation Methods & Experiment Design

Evaluation of the designed method is carried out by comparing the level of accuracy produced in this study with the level of accuracy of previous research. The parameters observed include: Accuracy of Validation, Accuracy of Testing, Loss of Validation, Loss of Training, Training Time, Testing Time.

The data used in this study amounted to 35,887 images, and were randomly divided into 28,709 images (80% of 35,887 images) as training data, 3,589 images (10% of 35,887 images) as validation data and 3,589 images (10% of 35,887 images) as test data. Evaluation of architectural modifications is carried out by observing testing accuracy and training time per epoch.

$$A = \frac{nc}{nt}\ x\ 100\%$$

Information[1]:

A = Accuracy results
nc = Number of correct data
nt = Total data

$$T = \frac{\varepsilon\tau}{\varepsilon p}$$

Information[2]:

T = Training time
$\varepsilon\tau$ = Total time of epoch
εp = Epoch parameter

Table 1. Analysis of Test Results

Architecture	Training Accuracy	Validation Accuracy	Testing Accuracy	Training Time	Testing Time
VGG16					
VGG16 + GAP					
VGG19					
VGG19 + GAP					
Prop. Method					

IV. RESULTS & DISCUSSION

4.1. Deep Learning Model Implementation

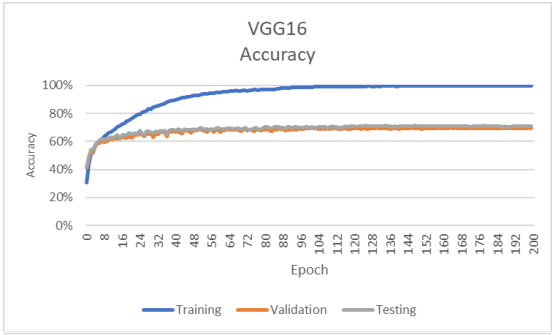
This study compared the results of 2 architectures between VGG16 and VGG19. The parameters that are considered are testing accuracy and training time per epoch. The entire experimental process in this study was implemented using PyTorch 1.4, Python 3.6.9 and GPU cloud platform with Google Colaboratory (based on Jupyter Notebook).

The entire training process in this experiment is carried out with the following conditions:

- Epoch: 200
- Optimizer: SGD
- Learning rate: 0.001
- Decay: 5e-4
- Using the FER2013 dataset containing 28,709 sample facial expression images.

4.2. Evaluation The Results of Existing Architectures.

The architecture tested in this study amounted to 4 models. Each of these architectures will be tested to determine the accuracy of the test results.



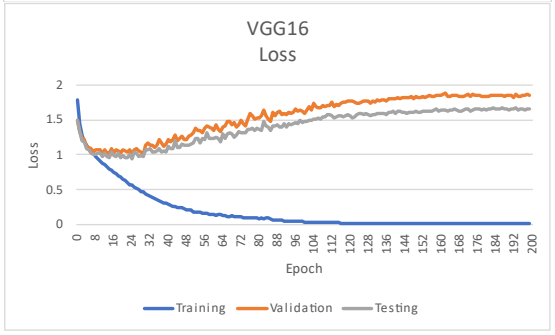
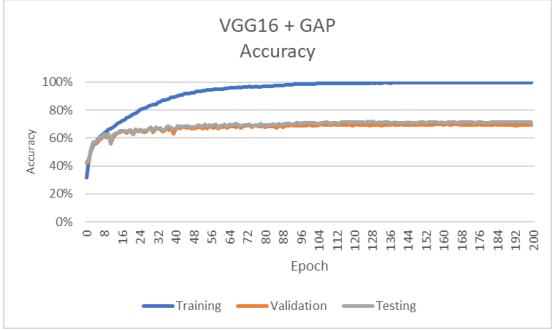


Figure 6. Experimental results using VGG16 architecture



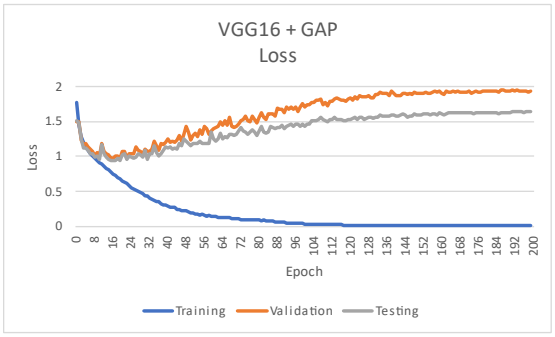
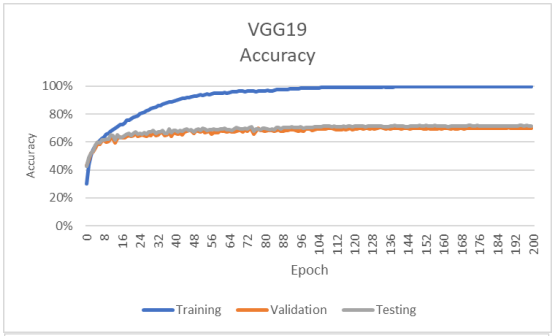


Figure 7. Experimental results using VGG16 + GAP architecture



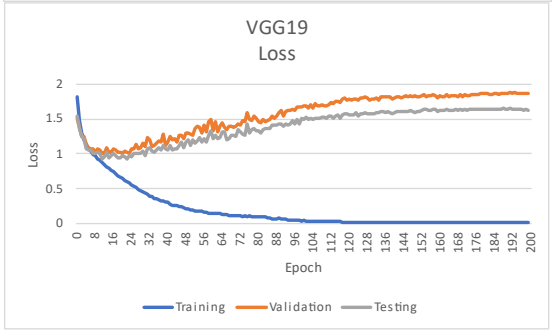
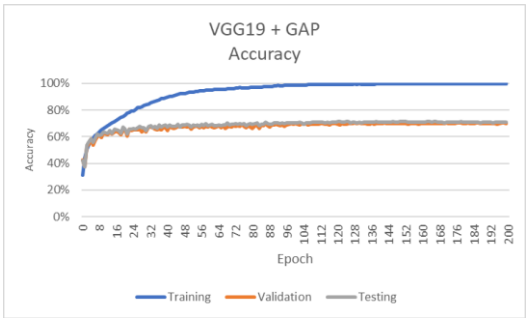


Figure 8. Experimental results using VGG19 architecture



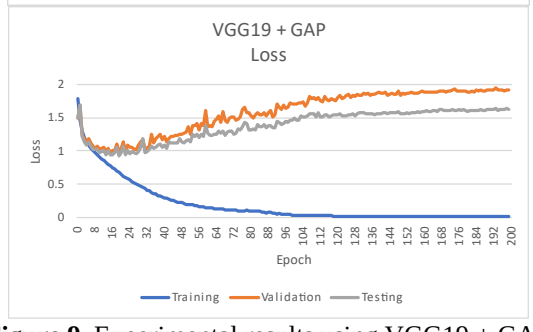


Figure 9. Experimental results using VGG19 + GAP architecture

Table 2. Comparison of Classification Existing Architectures Results

Architecture	Training Acc.	Validation Acc.	Testing Acc.
VGG-16	99.38%	69.32%	71.33%
VGG-16+SGD	99.34%	69.38%	71.80%
VGG19	99.47%	69.55%	71.80%
VGG19+SGD	99.19%	69.41%	71.55%

According to the graph, the experimental results in this study show that the method with the highest test accuracy value of 71.80% is VGG16 + SGD, and the method with the lowest accuracy value is VGG16 with an accuracy value of 71.33%.

Based on the results of the experiment, it appears that the more complex the architecture used, the lower the resulting accuracy value, and the higher the level of overfit. As a result, it is proposed to make changes to the architecture that has the best accuracy value between VGG16 + SGD and VGG19 by reducing several learning rates so that the parameters used become much less so that they can degrade the computational process.

4.3. Evaluation The Results of The Proposed Architectures.

In this study, there are two variations of the proposed method, the difference between each of these architectures is the number of convolutional layers owned. The first variation has 11 convolutional layers and the second variation has 8 convolutional layers.

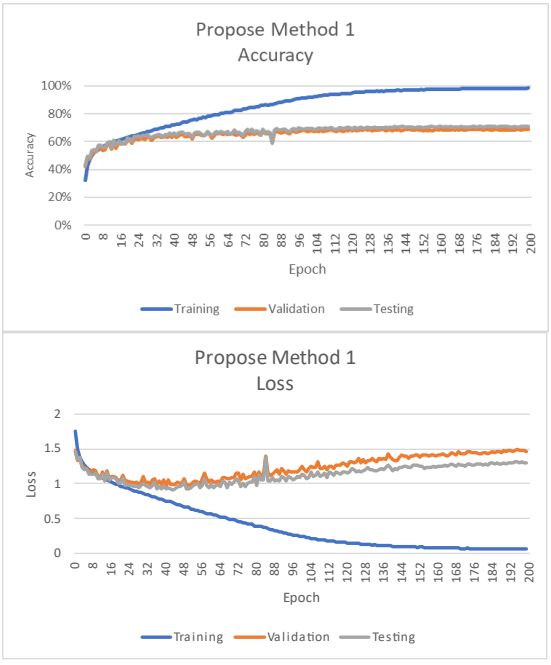


Figure 10. Experimental results using the Propose Method 1 architecture

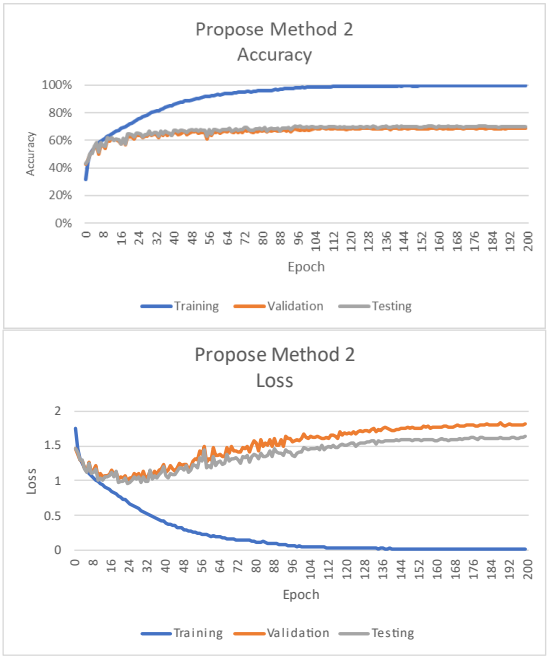


Figure 11. Experimental results using the Propose Method 2 architecture

Table 3. Comparison of Classification Propose Method Architectures Results

Architecture	Training Acc.	Validation Acc.	Testing Acc.	Training Time	Testing Time
VGG16 + GAP	99.34%	69.38%	71.80%	24.89s/ epoch	5s
8 Conv layer + 4 Maxpool + GAP	90.60%	66.87%	67.51%	24.19s/ epoch	4s
11 Conv layer + 4 Maxpool + GAP	99.51%	68.52%	70.27%	20.93s/ epoch	4s

Figure 12. Comparison of Proposed Architectures Results

According to the graph, the experimental results in this study show that VGG16 + Global Average Pooling is the method that produces the highest test accuracy value of 71.80%, while Proposed Method 2 produces the lowest accuracy value of 67.51%.

It can be seen that Proposed Method 1 can achieve an accuracies that are nearly identical to the best architecture from previous trials, namely 70.27%. Although the results are nearly identical, Proposed Method 1 has a few advantages, including the use of a more restrictive parameter, which reduces the time required for the training process and increases the likelihood of errors occurring during the training. Proposed Method 2 is being developed as a barometer to determine whether proposed method 1 is sufficiently complex or not. This architecture makes the accuracy decrease to 67.51% for accuracy test. This demonstrates that the Proposed Method 2 architecture is always more complex in terms of expressing facial expressions.

To achieve maximum accuracy, it is critical to use the appropriate architecture. A large architecture will have many weight parameters, resulting in heavy computational processes, longer training times, and overfitting. Large architectures, on the other hand, have the advantage of being able to study large and complex datasets. While a small architecture has fewer weight parameters, the computational process becomes lighter, and the training time becomes faster, making it suitable for training processes with few datasets and typical datasets that are not overly complex. However, a small architecture has the disadvantage of resulting in underfit.

V. CONCLUSION

This study describes the design and implementation of a face recognition system based on Convolutional Neural Networks. With the right architecture, the proposed method achieved good accuracy in face classification. Reduced learning rate or architecture simplification can improve accuracy while reducing computational burden and time. With 71.80% accuracy, the VGG16 + SGD usage scheme was found to be the most optimal.

To strengthen research results, the study suggests further development through the use of simpler datasets, transfer learning, designing a more optimal architecture, and validation by facial experts on the dataset used.

REFERENCES

[1] I.J. Goodfellow, D. Erhan, P. Luc Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, Y. Bengio, Challenges in representation learning: A report on three machine learning contests, Neural Netw. 64 (2015) 59–63. https://doi.org/10.1016/j.neunet.2014.09.005.

[2] A. Dhall, R. Goecke, J. Joshi, M. Wagner, T. Gedeon, Emotion recognition in the wild challenge 2013, in: Proc. 15th ACM Int. Conf. Multimodal Interact., ACM, Sydney Australia, 2013: pp. 509–516. https://doi.org/10.1145/2522848.2531739.

[3] A. Dhall, R. Goecke, J. Joshi, K. Sikka, T. Gedeon, Emotion Recognition In The Wild Challenge 2014: Baseline, Data and Protocol, in: Proc. 16th Int. Conf. Multimodal Interact., Association for Computing Machinery, New York, NY, USA, 2014: pp. 461–466. https://doi.org/10.1145/2663204.2666275.

[4] A. Dhall, O.V. Ramana Murthy, R. Goecke, J. Joshi, T. Gedeon, Video and Image based Emotion Recognition Challenges in the Wild: EmotiW 2015, in: Proc. 2015 ACM Int. Conf. Multimodal Interact., Association for Computing Machinery, New York, NY, USA, 2015: pp. 423–426. https://doi.org/10.1145/2818346.2829994.

[5] A. Dhall, R. Goecke, J. Joshi, T. Gedeon, Emotion recognition in the wild challenge 2016, in: Proc. 18th ACM Int. Conf. Multimodal Interact., Association for Computing Machinery, New York, NY, USA, 2016: pp. 587–588. https://doi.org/10.1145/2993148.3007626.

[6] A. Dhall, R. Goecke, S. Ghosh, J. Joshi, J. Hoey, T. Gedeon, From individual to group-level emotion recognition: EmotiW 5.0, in: Proc. 19th ACM Int. Conf. Multimodal Interact., Association for Computing Machinery, New York, NY, USA, 2017: pp. 524–528. https://doi.org/10.1145/3136755.3143004.

[7] A. Dhall, A. Kaur, R. Goecke, T. Gedeon, EmotiW 2018: Audio-Video, Student Engagement and Group-Level Affect Prediction, (2018). https://doi.org/10.48550/arXiv.1808.07773.

[8] A. Dhall, Goecke, Roland, Ghosh, Shreya, Gedeon, Tom, EmotiW 2019: Automatic Emotion, Engagement and Cohesion Prediction Tasks, in: 2019 Int. Conf. Multimodal Interact., Association for Computing Machinery, New York, NY, USA, 2019: pp. 546–550. https://doi.org/10.1145/3340555.3355710.

[9] A. Dhall, R. Goecke, S. Lucey, T. Gedeon, Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark, in: 2011 IEEE Int. Conf. Comput. Vis. Workshop ICCV Workshop, IEEE, Barcelona, Spain, 2011: pp. 2106–2112.

https://doi.org/10.1109/ICCVW.2011.6130508.

[10] T. Kanade, J.F. Cohn, Yingli Tian, Comprehensive database for facial expression analysis, in: Proc. Fourth IEEE Int. Conf. Autom. Face Gesture Recognit. Cat No PR00580, IEEE Comput. Soc, Grenoble, France, 2000: pp. 46–53. https://doi.org/10.1109/AFGR.2000.840611.

[11] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, in: 2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Workshop, IEEE, San Francisco, CA, USA, 2010: pp. 94–101. https://doi.org/10.1109/CVPRW.2010.5543262.

[12] M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with Gabor wavelets, in: Proc. Third IEEE Int. Conf. Autom. Face Gesture Recognit., IEEE Comput. Soc, Nara, Japan, 1998: pp. 200–205. https://doi.org/10.1109/AFGR.1998.670949.

[13] Z. Zhang, P. Luo, C.C. Loy, X. Tang, From Facial Expression Recognition to Interpersonal Relation Prediction, Int. J. Comput. Vis. 126 (2018) 550–569. https://doi.org/10.1007/s11263-017-1055-1.

[14] B. Ko, A Brief Review of Facial Emotion Recognition Based on Visual Information, Sensors. 18 (2018) 401. https://doi.org/10.3390/s18020401.

[15] S. Li, W. Deng, Deep Facial Expression Recognition: A Survey, IEEE Trans. Affect. Comput. 13 (2022) 1195–1215. https://doi.org/10.1109/TAFFC.2020.2981446.

[16] E. Owusu, E.K. Gavua, Z. Yong-Zhao, Facial Expression Recognition – A Comprehensive Review, Int. J. Technol. Manag. Res. 1 (2020) 29–46. https://doi.org/10.47127/ijtmr.v1i4.36.

[17] Samadiani, Huang, Cai, Luo, Chi, Xiang, He, A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data, Sensors. 19 (2019) 1863. https://doi.org/10.3390/s19081863.

[18] H. Jung, S. Lee, S. Park, B. Kim, J. Kim, I. Lee, C. Ahn, Development of deep learning-based facial expression recognition system, in: 2015 21st Korea-Jpn. Jt. Workshop Front. Comput. Vis. FCV, IEEE, Mokpo, South Korea, 2015: pp. 1–4. https://doi.org/10.1109/FCV.2015.7103729.

[19] C. Pramerdorfer, M. Kampel, Facial Expression Recognition using Convolutional Neural Networks: State of the Art, (2016). http://arxiv.org/abs/1612.02903 (accessed February 16, 2023).

[20] J. Yang, G. Yang, Modified Convolutional Neural Network Based on Dropout and the Stochastic Gradient Descent Optimizer, Algorithms. 11 (2018) 28. https://doi.org/10.3390/a11030028.

[21] T.S. Gunawan, A. Ashraf, B.S. Riza, E.V. Haryanto, R. Rosnelly, M. Kartiwi, Z. Janin, Development of video-based emotion recognition using deep learning with Google Colab, TELKOMNIKA Telecommun. Comput. Electron. Control. 18 (2020) 2463. https://doi.org/10.12928/telkomnika.v18i5.16717.

[22] E. Pranav, S. Kamal, C. Satheesh Chandran, M.H. Supriya, Facial Emotion Recognition Using Deep Convolutional Neural Network, in: 2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS, IEEE, Coimbatore, India, 2020: pp. 317–320. https://doi.org/10.1109/ICACCS48705.2020.9074302.

[23] S. Cheng, G. Zhou, Facial Expression Recognition Method Based on Improved VGG Convolutional Neural Network, Int. J. Pattern Recognit. Artif. Intell. 34 (2020) 2056003. https://doi.org/10.1142/S0218001420560030.

[24] A. Sajjanhar, Z. Wu, Q. Wen, Deep Learning Models for Facial Expression Recognition, in: 2018 Digit. Image Comput. Tech. Appl. DICTA, IEEE, Canberra, Australia, 2018: pp. 1–6. https://doi.org/10.1109/DICTA.2018.8615843.

[25] Y. Miyakoshi, S. Kato, Facial emotion detection considering partial occlusion of face using Bayesian network, in: 2011 IEEE Symp. Comput. Inform., IEEE, Kuala Lumpur, Malaysia, 2011: pp. 96–101. https://doi.org/10.1109/ISCI.2011.5958891.

[26] H.K. Vydana, P.P. Kumar, K.S.R. Krishna, A.K. Vuppala, Improved emotion recognition using GMM-UBMs, in: 2015 Int. Conf. Signal Process. Commun. Eng. Syst., IEEE, Guntur, India, 2015: pp. 53–57. https://doi.org/10.1109/SPACES.2015.7058214.

[27] B. Schuller, G. Rigoll, M. Lang, Hidden Markov model-based speech emotion recognition, in: 2003 IEEE Int. Conf. Acoust. Speech Signal Process. 2003 Proc. ICASSP 03, IEEE, Hong Kong, China, 2003: p. II-1–4. https://doi.org/10.1109/ICASSP.2003.1202279.

[28] P. Singh, G. Saha, M. Sahidullah, Non-linear frequency warping using constant-Q

transformation for speech emotion recognition, in: 2021 Int. Conf. Comput. Commun. Inform. ICCCI, IEEE, Coimbatore, India, 2021: pp. 1–6. https://doi.org/10.1109/ICCCI50826.2021.9402569.

[29] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, (2014). https://doi.org/10.48550/ARXIV.1409.1556.