InceptionV3, ResNet50, ResNet18 and MobileNetV2 Performance Comparison on Face Recognition Classification

Mohammad Rafka Mahendra Ariefwan [1*], I Gede Susrama Mas Diyasa [2], Kartika Maulidya Hindrayani [3]

[1*,3] Program Studi Sains Data, Fakultas Ilmu Komputer, UPN “Veteran” Jawa Timur, Surabaya, Indonesia

[2]Program Studi Magister Teknologi Informasi, Fakultas Ilmu Komputer, UPN “Veteran” Jawa Timur, Surabaya, Indonesia

* *20083010025@student.upnjatim.ac.id; igsusrama.if@upnjatim.ac.id; kartika.maulida.ds@upnjatim.ac.id*

DOI : *https://doi.org/10.56480/jln.v4i1.990*

Received: August 24, 2023
Revised: September 29, 2023
Accepted: Oktober 23, 2023

Abstract

This research aims to design a facial recognition system for faculty attendance in the Data Science Program at UPN "Veteran" East Java. The study proposes a solution by integrating CNN-based facial recognition technology with Android devices. The methodology involves training CNN models using a dataset of faculty faces and comparing various architectures such as ResNet, MobileNet, and InceptionV3. The research methodology encompasses data collection, data preprocessing, model creation, model comparison, performance evaluation, and implementation. Results from the study, utilizing Convolutional Neural Network models and testing various architectures, reveal the architecture with the best facial recognition performance achieving an average accuracy rate of 77%.

Keywords– *Face Recognition, Attendance, CNN, Android, Model Architecture*

1. Introduction

In various institutions, including schools, universities, and companies, attendance systems have become integral, providing accurate data for evaluations despite the prevalence of conventional methods. The Data Science Program at UPN "Veteran" East Java represents the forefront of technological utilization and development, reflecting the program's commitment to relevance and advancement. The current practice of utilizing fingerprint-based attendance systems is noted for inefficiencies and security concerns, prompting the exploration of alternative methods such as facial recognition (Chowdhury et al., 2020).

This research addresses the limitations and challenges associated with fingerprint-based systems, proposing a facial recognition system integrated with Android devices for efficient and accurate faculty attendance in the Data Science Program. Concerns such as time-consuming verification, the need for direct contact, and the complexity of biometric systems are discussed, emphasizing the importance of selecting an optimal and secure solution (Susrama et al., 2022). The security of faculty facial data is highlighted, underscoring the necessity of designing systems with robust privacy measures (Bahety et al., 2020).

In the context of rapid technological development, this study aims to contribute to the ongoing evolution of attendance systems, specifically tailored for the needs of the Data Science Program. By combining Convolutional Neural Network (CNN) technology with Android deployment, the research focuses on training the CNN model using a dataset of faculty faces, enabling efficient and secure facial recognition for attendance purposes (Riyantoko et al., 2021). The chosen CNN architecture is justified based on its widespread use in image processing, facial identification, and applications in diverse fields such as medical imaging and small and medium-sized enterprises (UMKM) analytics (Goel et al., 2023).

By bridging the gap between existing facial recognition research and the unique requirements of the Data Science Program, this study seeks to enhance the speed and accuracy of facial recognition for attendance tracking. Notable prior research on CNN and alternative models is cited, providing a foundation for the proposed approach. The introduction sets the stage for a comprehensive exploration of the methodology, results, and conclusions, promising valuable insights for the improvement of attendance systems in academic settings.

2. Method

The exploration of Convolutional Neural Network (CNN) architectures plays a crucial role in enhancing image recognition and classification tasks. Three notable architectures—ResNet, MobileNet, and InceptionV3—offer distinct features and advantages that make them suitable for various applications.

2.1 Residual Network

ResNet, short for Residual Network, introduces skip connections that facilitate training by learning residual functions. Particularly effective in image recognition, ResNet maintains superior performance even as its architecture deepens. The implementation involves skipping connections in two to three layers containing Rectified Linear Unit (ReLU) and batch normalization (Sandler et al., 2018).

2.2 Mobile Network

MobileNet prioritizes efficiency for mobile and embedded devices, utilizing depthwise separable convolutions to reduce computational requirements while preserving accuracy. Tailored for lightweight computing and low-resolution images, MobileNet employs depthwise separable convolution as its core component, significantly reducing parameters and computations while maintaining good accuracy (Dongmei et al., 2020).

2.3 InceptionV3

InceptionV3, designed for image classification and object detection, is part of the Inception family of CNNs developed by Google researchers. InceptionV3 enhances its predecessor, InceptionV2, by introducing improvements for increased performance and efficiency. Noteworthy is the Inception module, a layer block that enables the network to capture features at different scales by using parallel filters of various sizes (1x1, 3x3, and 5x5) (Srinivasu et al., 2021).

The Inception module employs factorized convolutions to reduce computations effectively, breaking down larger convolution kernels into smaller ones. Additionally, InceptionV3 addresses the vanishing gradient problem by including additional classifications in intermediate layers, aiding the training of deeper networks.

Comparing ResNet, MobileNet, and InceptionV3, ResNet uses residual blocks and shortcut connections to address gradient vanishing, while InceptionV3 employs the Inception module and additional classifications. MobileNetV2, designed for resource-limited mobile devices, utilizes depthwise separable convolutions for parameter reduction (Venkateswarlu et al., 2020).

2.4 Transfer Learning

Transfer learning, a technique involving the sharing of knowledge gained from solving one problem to solve a related problem, significantly reduces training time. Two common approaches are pre-trained models and custom output layers. Pre-trained models, such as ResNet and MobileNet trained on large benchmark datasets like ImageNet, provide a starting point for feature extraction. Custom output layers leverage pre-trained models for feature extraction, which can then be utilized for specific classification tasks. The integration of these CNN architectures and transfer learning techniques holds promise for advancing image recognition and classification in various domains (Venkateswarlu et al., 2020).

3. Result and Discussion

The results evaluation were conducted with various testing. The first is by observation from the model training epochs. The training process could be observed based on the model loss and accuracy and could be further plotted for better visualization. The second method of evaluation is using a confusion matrix. The confusion matrix was implemented using a new data that was not used in the training.

3.1. Model Training Evaluation

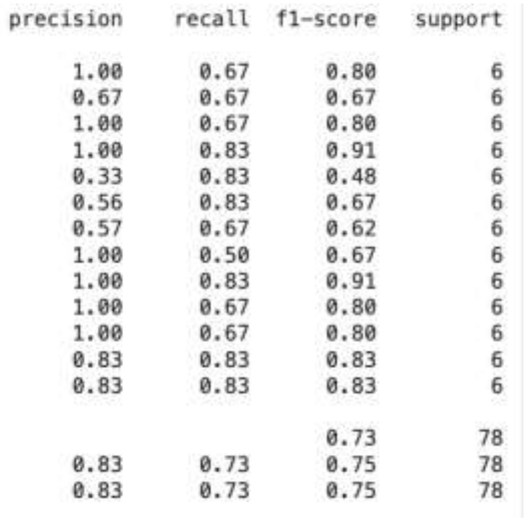
Table 1 Model and Optimizer Comparison

Model	**Optimizer**	**Training Accuracy *(%)***	**Validation Accuracy (%)**	**Training Loss**	**Validation Loss**
MobileNetV2 *Pre-trained*	Adam	97.11	98.72	0.62	0.46
	SGD	97.10	98.72	0.73	0.48
	RMSprop	97.54	97.44	0.46	0.40
MobileNetV2 Scratch	Adam	91.28	30.76	0.92	6.86
	SGD	88.74	87.61	0.26	0.30
	RMSprop	99.23	35.90	0.37	2.39
InceptionV3 *Pre-trained*	Adam	42.22	47.86	10.43	9.48
	SGD	33.27	36.32	10.22	9.48
	RMSprop	35.82	47.01	10.76	9.43
InceptionV3 Scratch	Adam	61.54	53.84	7.10	8.37
	SGD	11.43	11.54	10.07	9.65
	RMSprop	53.85	41.03	7.42	8.39
ResNet18 Scratch	Adam	94.01	94.87	0.46	0.19
	SGD	73.44	69.23	0.70	0.73
	RMSprop	87.54	67.45	0.42	0.75
ResNet50 *Pre-trained*	Adam	85.45	10.37	7.65	10.37
	SGD	84.30	8.34	6.21	8.34
	RMSprop	80.64	9.16	7.43	9.16

Based on Table 1, three models perform quite well, namely the three pre-trained MobileNetV2 models. These three models will then be tested using a confusion matrix with validation data that has never been used to assess whether the models can work with new data. The validation data used comes from video data that has not been cropped or converted to grayscale, or any other data alterations. Changes in accuracy levels occur because the validation data used has different accuracy and loss due to the fact that during training, validation data is taken from the training data, leading to data leakage. Therefore, a confusion matrix is conducted to revalidate the performance of the designed models.

3.2. Confusion Matrix Evaluation

The number of images included in the confusion matrix is adjusted based on the amount of validation data used during the model training. In this case, it is 6 images per class due to the data split during training being 90% and 10%. The confusion matrix is considered appropriate for retesting each model because it provides more metrics such as accuracy, precision, recall, and F1-score to measure the performance of a model.



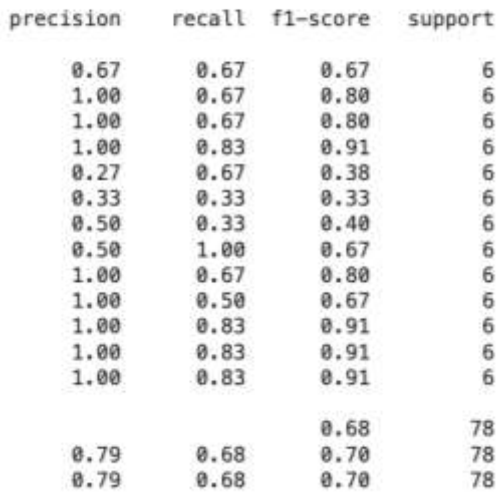
```
precision    recall  f1-score   support

     1.00      0.67      0.80         6
     0.67      0.67      0.67         6
     1.00      0.67      0.80         6
     1.00      0.83      0.91         6
     0.33      0.83      0.48         6
     0.56      0.83      0.67         6
     0.57      0.67      0.62         6
     1.00      0.50      0.67         6
     1.00      0.83      0.91         6
     1.00      0.67      0.80         6
     1.00      0.67      0.80         6
     0.83      0.83      0.83         6
     0.83      0.83      0.83         6

                         0.73        78
     0.83      0.73      0.75        78
     0.83      0.73      0.75        78
```

Figure 1 MobileNetV2 Pre-Trained Adam

In Figure 1, the confusion matrix results for the pre-trained Inception-V3 model with the Adam optimizer show an accuracy rate of approximately 73%. This model exhibits good average precision, including in recall and F1-score. Overall, the accuracy value of 73% indicates that the model performs well.



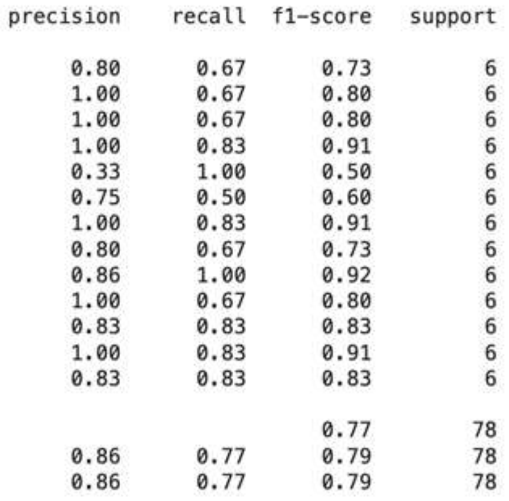
```
precision    recall  f1-score   support

     0.67      0.67      0.67         6
     1.00      0.67      0.80         6
     1.00      0.67      0.80         6
     1.00      0.83      0.91         6
     0.27      0.67      0.38         6
     0.33      0.33      0.33         6
     0.50      0.33      0.40         6
     0.50      1.00      0.67         6
     1.00      0.67      0.80         6
     1.00      0.50      0.67         6
     1.00      0.83      0.91         6
     1.00      0.83      0.91         6
     1.00      0.83      0.91         6

                         0.68        78
     0.79      0.68      0.70        78
     0.79      0.68      0.70        78
```

Figure 2 MobileNetV2 Pre-Trained SGD

In Figure 2, the confusion matrix results for the pre-trained Inception-V3 model with the SGD optimizer show an accuracy rate of approximately 68%. This model exhibits reasonably good average precision, including in recall and F1-score. Overall, the accuracy value of 68% indicates that the model performs well.



```
precision    recall  f1-score   support

     0.80      0.67      0.73         6
     1.00      0.67      0.80         6
     1.00      0.67      0.80         6
     1.00      0.83      0.91         6
     0.33      1.00      0.50         6
     0.75      0.50      0.60         6
     1.00      0.83      0.91         6
     0.80      0.67      0.73         6
     0.86      1.00      0.92         6
     1.00      0.67      0.80         6
     0.83      0.83      0.83         6
     1.00      0.83      0.91         6
     0.83      0.83      0.83         6

                         0.77        78
     0.86      0.77      0.79        78
     0.86      0.77      0.79        78
```

Figure 3 MobileNetV2 Pre-Trained RMSprop

In Figure 3, the confusion matrix results for the pre-trained Inception-V3 model with the RMSprop optimizer show an accuracy rate of approximately 77%. This model demonstrates good average precision, including in recall and F1-score. Overall, the accuracy value of 77% indicates that the model performs well.

The comprehensive analysis of confusion matrices suggests that the pre-trained Inception-V3 model with the RMSprop optimizer yields the most satisfactory results. Consequently, this model will be implemented on mobile devices, emphasizing its superior performance for the intended application.

4. Conclusion

Based on various evaluations, analyses, and experiments conducted, it can be concluded that the pre-trained MobileNetV2 model using the RMSprop optimizer exhibits reasonably good predictions, achieving a 77% accuracy in the confusion matrix and a training accuracy of 97%. It is important to acknowledge that the model may still have limitations in generalizing its predictions, but even with a limited dataset, the model demonstrates the ability to make predictions or recognize faces.

Although the MobileNetV2 model possesses a relatively high level of complexity, it is well-balanced, and with the application of regularization, it can effectively address the issue of overfitting. The architecture of this model consists of numerous layers, providing the capability to capture various features from images.

By leveraging the pre-trained model with embedded ImageNet weights, the model is equipped with strong prior knowledge of general visual features. This enables the model to inherit and apply this knowledge effectively in more specific classification tasks.

The RMSprop optimizer exhibits outstanding performance. RMSprop contributes to the stability of the training process, preventing the learning rate from oscillating excessively, particularly in situations where adaptive learning rates are crucial for various parameters.

Author Contributions: Conceptualization, Mohammad Rafka Mahendra Ariefwan and I Gede Susrama Mas Diyasa; methodology, Mohammad Rafka Mahendra Ariefwan, and I Gede Susrama Mas Diyasa; software, Mohammad Rafka Mahendra Ariefwan; validation, Mohammad Rafka Mahendra Ariefwan, I Gede Susrama Mas Diyasa, and Kartika Maulidya Hindrayani; formal analysis, Mohammad Rafka Mahendra Ariefwan; investigation, Mohammad Rafka Mahendra Ariefwan; resources, Mohammad Rafka Mahendra Ariefwan; data curation, Mohammad Rafka Mahendra Ariefwan; writing—original draft preparation, Mohammad Rafka Mahendra Ariefwan; writing—review and editing, I Gede Susrama Mas Diyasa, and Kartika Maulidya Hindrayani; visualization, Mohammad Rafka Mahendra Ariefwan; supervision, I Gede Susrama Mas Diyasa, and Kartika Maulidya Hindrayani; project administration, Mohammad Rafka Mahendra Ariefwan; funding acquisition, Mohammad Rafka Mahendra Ariefwan. All authors have read and agreed to the published version of the manuscript.", please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This research received no external funding

Acknowledgments: Data Science Study Program of UPN “Veteran” Jawa Timur

Declaration of Conflicting Interests: The authors declare no conflict of interest

References

Bahety, S. S., Kumar, K., Tejaswi, V., Balagar, S. R., & Anil, B. C. (2020). Implementation of Automated Attendance System using Facial Identification from Deep Learning Convolutional Neural Networks. International Journal of Engineering Research & Technology, 8(15), 170–174.

Chowdhury, S., Nath, S., Dey, A., & Das, A. (2020). Development of an Automatic Class Attendance System using CNN-based Face Recognition. ETCCE 2020 - International Conference on Emerging Technology in Computing, Communication and Electronics, 2–6. https://doi.org/10.1109/ETCCE51779.2020.9350904

Dongmei, Z., Ke, W., Hongbo, G., Peng, W., Chao, W., & Shaofeng, P. (2020). Classification and identification of citrus pests based on InceptionV3 convolutional neural network and migration learning. 2020 International Conference on Internet of Things and Intelligent Applications, ITIA 2020. https://doi.org/10.1109/ITIA50152.2020.9312359

Goel, A., Goel, A. K., & Kumar, A. (2023). The role of artificial neural network and machine learning in utilizing spatial information. Spatial Information Research, 31(3), 275–285. https://doi.org/10.1007/s41324-022-00494-x

Riyantoko, P. A., Sugiarto, & Hindrayani, K. M. (2021). Facial Emotion Detection Using Haar-Cascade Classifier and Convolutional Neural Networks. Journal of Physics: Conference Series, 1844(1). https://doi.org/10.1088/1742-6596/1844/1/012004

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

Srinivasu, P. N., Sivasai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J. (2021). Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM. Sensors, 21, 1–27.

Susrama, I. G., Putra, A. H., & Ariefwan, M. (2022). Feature Extraction for Face Recognition Using Haar Cascade Classifier. International Seminar of Research Month 2021, 197–206. https://doi.org/10.11594/nstp.2022.2432

Venkateswarlu, I. B., Kakarla, J., & Prakash, S. (2020). Face mask detection using MobileNet and Global Pooling Block. 4th IEEE Conference on Information and Communication Technology, CICT 2020, 20, 0–4. https://doi.org/10.1109/CICT51604.2020.9312083*r*.