Journal of Optimization and Decision Science (JODS)

Website: https://ioinformatic.org/

April 2026. Vol. 1. No. 1

Identification of Batak Ethnic Groups Based on Facial Features Using Convolutional Neural Network Algorithm

Muhammad Abid Syuja[1]*, Said Iskandar Al Idrus[2], Hermawan Syahputra[3], Debi Yandra Niska[4]

[1.2.3.4] Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, Indonesia
mhdabidsyuja@gmail.com [1], saidiskandar@unimed.ac.id [2], hsyahputra@unimed.ac.id [3], debiyandraniska@gmail.com [4]*

Abstract

Facial-based ethnic identification is a challenging task in computer vision due to subtle inter-class variations and high intra-class similarities. The Batak ethnic group in Indonesia exhibits distinctive facial characteristics that can be analyzed using deep learning approaches. This study aims to identify Batak and Non-Batak ethnic groups based on facial features using a Convolutional Neural Network (CNN) and a pretrained ResNet50 model. The research process includes image preprocessing, feature extraction, model training, and performance evaluation. Model performance is assessed using accuracy and confusion matrix analysis. Experimental results show that the CNN model achieved an accuracy of 67%, outperforming the ResNet50 model which obtained an accuracy of 50%. These findings indicate that simpler CNN architectures can be more effective than deep pretrained models when applied to limited and locally collected facial datasets.

Keywords: *Batak ethnicity; convolutional neural network; deep learning; face classification; ResNet50*

1. Introduction

Face-based ethnic classification has attracted significant attention in the field of computer vision due to its potential applications in security systems, human–computer interaction, and computational anthropology. Facial images contain rich visual information that can represent identity, gender, age, and ethnic characteristics. Advances in artificial intelligence, particularly deep learning, have enabled automatic extraction of discriminative facial features without relying on handcrafted descriptors.

Convolutional Neural Networks (CNNs) have become the dominant approach for facial image analysis due to their ability to learn spatial hierarchies of features through convolution and pooling operations. CNN-based models have demonstrated strong performance in face recognition, facial expression analysis, and biometric identification. However, the effectiveness of CNN models is highly influenced by architecture design and dataset size.

Deep pretrained architectures such as ResNet50 introduce residual connections that allow very deep networks to be trained effectively by addressing the vanishing gradient problem. ResNet50 has achieved state-of-the-art results in many image classification tasks. Nevertheless, its performance may decrease when applied to limited datasets without extensive fine-tuning.

In Indonesia, research focusing on facial-based ethnic classification remains limited, particularly for local ethnic groups such as Batak. The Batak ethnic group possesses distinctive facial morphological characteristics, making it a relevant subject for facial-based ethnic analysis. Therefore, this study aims to compare the performance of a custom CNN model and a pretrained ResNet50 model in identifying Batak and Non-Batak ethnic groups based on facial images.

2. Methodology

2.1. Dataset Collection

The dataset used in this study consists of facial images categorized into two classes: Batak and Non-Batak. Images were collected under controlled conditions with three facial orientations: front, left, and right views. This variation aims to improve model robustness in recognizing facial features from different angles.

2.2. Image Preprocessing

All facial images are resized to a uniform resolution and normalized to enhance training stability. Face detection and cropping are performed to focus on facial regions and reduce background influence. Data augmentation techniques such as rotation and horizontal flipping are applied to increase data diversity and reduce overfitting.

2.3. Convolutional Neural Network

The CNN model extracts spatial features from facial images through convolution operations defined as:

$$Y(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X\,(i+m, j+n).K(m,n) \qquad (1)$$

where X represents the input image, K denotes the convolution kernel, and Y is the resulting feature map.

Pooling layers are applied to reduce spatial dimensions and computational complexity.

2.4. ResNet50 Architecture

ResNet50 is a deep residual network consisting of 50 layers with residual connections that allow the network to learn identity mappings. In this study, ResNet50 is used as a pretrained model with weights learned from large-scale image datasets. Fine-tuning is applied to adapt the model to the facial ethnic classification task.

2.5. Classification and Evaluation

The output layer uses the Softmax function to classify facial images into Batak and Non-Batak classes:

$$P(c_i) = \frac{e^{zi}}{\sum_{j=1}^{C} e^{zj}} \qquad (2)$$

Model performance is evaluated using accuracy calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3)$$

A confusion matrix is also used to analyze classification performance in detail.

3. Results and Discussion

3.1. Performance Comparison Between CNN and ResNet50

Based on the experimental evaluation, the custom Convolutional Neural Network (CNN) achieved an accuracy of 67%, whereas the pretrained ResNet50 model achieved 50% accuracy in classifying Batak and Non-Batak facial images. This result indicates a noticeable performance gap between the two models, despite ResNet50 having a significantly deeper architecture.

From an interpretative perspective, these findings suggest that higher model complexity does not automatically lead to better performance, particularly when applied to limited datasets. Although ResNet50 is designed to extract highly abstract features through deep residual layers, such complexity requires sufficient training data to adapt effectively to a new domain.

The implication of this result is important for researchers working with local or small-scale datasets. It demonstrates that a well-designed CNN architecture tailored to the dataset characteristics can outperform deep pretrained models, making CNN a more practical and efficient choice in similar ethnic facial classification tasks.

3.2. Analysis of CNN Performance

The CNN model demonstrates superior performance due to its ability to learn dataset-specific facial features effectively. The relatively shallow architecture allows the network to focus on essential facial characteristics such as facial contours, eye spacing, nose structure, and overall face shape, which are relevant for ethnic differentiation.

From a theoretical standpoint, CNNs are known to perform well when the learned features closely match the data distribution. In this study, the CNN model was trained entirely on the target dataset, enabling it to adapt directly to the unique facial patterns of Batak and Non-Batak individuals without relying on external feature representations.

The practical implication of this result is that CNN-based models are more suitable for scenarios where data availability is limited. Additionally, the use of data augmentation techniques such as rotation and horizontal flipping further improves generalization and reduces overfitting, contributing to the overall performance of the CNN model.

3.3. Analysis of ResNet50 Performance

The ResNet50 model achieved lower accuracy compared to the CNN model, which can be attributed to several factors. One key issue is the domain mismatch between the pretrained ImageNet dataset and the facial ethnic dataset used in this study. ImageNet primarily contains object-centric images, which differ significantly from facial image characteristics.

From the perspective of transfer learning theory, pretrained models require sufficient fine-tuning data to adapt high-level features to new tasks. In this research, the limited dataset size restricts the ability of ResNet50 to effectively adjust its deep residual layers, resulting in suboptimal feature representation for ethnic classification.

These findings imply that while ResNet50 is powerful, it is not always the optimal choice for small datasets. Without sufficient data or extensive fine-tuning, deep pretrained architectures may underperform compared to simpler models trained specifically on the target domain.

3.4 Effect of Dataset Size and Characteristics

The dataset used in this study consists of locally collected facial images with limited size and controlled acquisition conditions. Such dataset characteristics play a crucial role in determining model performance. Smaller datasets often favor simpler architectures that require fewer parameters and are less prone to overfitting.

Theoretically, deep learning models with a large number of parameters require extensive data to achieve stable convergence. When training data is insufficient, models like ResNet50 may fail to generalize effectively, whereas CNN models with fewer layers can still capture meaningful patterns.

This result reinforces previous research emphasizing the importance of aligning model complexity with dataset scale. For local ethnic studies with limited data, lightweight CNN architectures provide a more reliable and efficient solution.

3.5. Comparison with Previous Studies

Previous studies in facial image classification have reported strong performance using CNN-based approaches, particularly when models are tailored to specific datasets. Research by LeCun et al. highlights that CNNs excel in learning spatial hierarchies of visual features when trained directly on relevant data distributions.

The findings of this study are consistent with earlier research demonstrating that simpler CNN architectures can outperform deeper networks under data constraints. Similar results have been reported in ethnic and demographic facial analysis, where dataset-specific training yields better performance than generic pretrained models.

This study extends previous work by providing empirical evidence in the context of Indonesian ethnic facial classification, specifically focusing on the Batak ethnic group, which has been underexplored in existing literature.

3.6. Scientific and Practical Implications

From a scientific perspective, this research contributes to the understanding of model selection for ethnic facial classification. The results emphasize that architectural simplicity combined with appropriate preprocessing can yield better performance than complex pretrained models in small-scale studies.

Practically, the findings suggest that CNN-based systems are more feasible for real-world applications such as local biometric systems, academic research, and cultural documentation projects that rely on limited datasets. CNN models require lower computational resources and are easier to deploy compared to deep pretrained architectures.

Furthermore, this study highlights the importance of dataset quality and relevance in facial analysis research. Future studies may explore hybrid approaches, larger datasets, or advanced fine-tuning strategies to further improve classification accuracy.

Acknowledgement

The authors would like to express their sincere gratitude to the Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Negeri Medan, for providing academic support and facilities during the completion of this research. Special appreciation is also extended to the supervisors and lecturers who provided valuable guidance, suggestions, and motivation throughout the research process. Furthermore, the authors would like to thank all parties who contributed directly or indirectly to the successful completion of this study.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[4] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4–20, 2004.
[5] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," ACM Computing Surveys, vol. 35, no. 4, pp. 399–458, 2003.
[6] S. Z. Li and A. K. Jain, Handbook of Face Recognition. New York, NY, USA: Springer, 2011.
[7] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in Proceedings of the British Machine Vision Conference (BMVC), 2015.
[8] H. Wang and B. Raj, "On the origin of deep learning," IEEE Access, vol. 5, pp. 1351–1368, 2017.
[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, 2012.
[10] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, 1991.