APPLIED ARTIFICIAL INTELLIGENCE AND ENGINEERING JOURNAL
E-ISSN: 3124-1212, DOI: 00.000/XXXX.YYYY.II.NNN
Vol. 2 No. 1 2026 pp. 01-09

Deep Learning-Based Celebrity Face Recognition Using Convolutional Neural Networks with Image Augmentation

Aidilfansury[1], Aulia Syahrul Mubaraq[2], Jordi Rusli[3*]

[1,2] Jurusan Teknologi Informasi dan Komputer, Politeknik Negeri Lhokseumawe, Lhokseumawe, Indonesia
[3] Fakultas Rekayasa Industri, Telkom University, Bandung, Indonesia

*Corresponding Author: jordirusli@student.telkomuniversity.ac.id

Article info: Received 03/03/2026, Revised 15/03/2026, Accepted 02/04/2026

Abstract

Celebrity face recognition has become an important research area in computer vision due to its wide range of applications in biometric authentication, security systems, and digital media analysis. Advances in deep learning, particularly Convolutional Neural Networks (CNNs), have enabled automated face recognition by learning highly discriminative visual features directly from image data. This study aims to develop a CNN-based celebrity face recognition model using a custom dataset consisting of 1,200 facial images categorized into four classes: Blackpink, Ariana Grande, BTS, and Taylor Swift. Following image validation and selection, the dataset was divided into training and testing subsets with an 80:20 ratio, resulting in 70 images for model evaluation. The preprocessing pipeline included image resizing to 128 × 128 pixels, pixel value normalization, contrast enhancement, and data augmentation techniques, including random rotations of up to ±20°, 20% translation, 20% zoom, horizontal flipping, and brightness variation. The CNN model was trained for 30 epochs using the Adam optimizer and categorical cross-entropy loss function. Experimental results showed that the proposed model achieved an overall classification accuracy of 63%. Among the four classes, the Blackpink class obtained the highest F1-score of 0.79, whereas the Taylor Swift class recorded the lowest F1-score of 0.29. These findings indicate that the CNN model is capable of learning fundamental facial characteristics for celebrity recognition; however, further improvements in model architecture, dataset diversity, and optimization strategies are required to enhance recognition performance and classification robustness.

Keywords: celebrity face recognition, Convolutional Neural Network (CNN), deep learning, computer vision; image augmentation

1. Introduction

Face recognition has become one of the most important research topics in the field of computer vision and has experienced rapid development over the past decade. It has been widely adopted in numerous applications, including security systems, biometric authentication, identity verification, attendance management, and digital media content analysis [1], [2]. The rapid growth of digital devices and social media platforms has generated an enormous volume of visual data, thereby increasing the demand for face recognition systems that are not only highly accurate but also computationally efficient and robust under diverse real-world conditions. However, face recognition systems continue to face several challenges, including variations in illumination, viewing angles, facial expressions, and image quality, all of which can significantly affect classification performance. Consequently, there is a need for approaches capable of extracting highly discriminative facial features while maintaining robustness against these variations. Recent advances in deep learning, particularly through the application of Convolutional Neural Networks (CNNs), have provided an effective solution to these challenges. CNNs are capable of automatically learning hierarchical feature representations directly from image data without requiring manual feature engineering, which has long been a major limitation of conventional face recognition methods [3]–[5]. Numerous CNN architectures have been developed to improve recognition performance, ranging from very deep

networks to lightweight and computationally efficient models. Nevertheless, the effectiveness of CNNs strongly depends on the quantity and diversity of the training data, as well as the computational resources available during the training process [6]. When only limited training data are available, CNN models are more susceptible to overfitting, resulting in reduced generalization capability on unseen samples.

Previous studies have consistently demonstrated the effectiveness of CNN-based approaches for face recognition tasks. Deep architectures such as ResNet enhance training stability and feature representation through residual learning mechanisms, enabling improved recognition performance [7]. Similarly, other deep CNN architectures, including VGG and Inception, have achieved remarkable performance in image classification and complex visual pattern recognition tasks [8], [9]. In addition, data augmentation techniques have been shown to significantly improve the generalization capability of CNN models by introducing variations in facial pose, illumination, and expression during training [10]. Although embedding-based face recognition methods, such as FaceNet, DeepFace, and ArcFace, have achieved state-of-the-art recognition accuracy, these approaches generally require large-scale datasets and substantial computational resources, making them less suitable for applications with limited data availability and constrained hardware environments [11]–[13].

Based on these observations, there remains a research gap in developing face recognition systems that maintain satisfactory performance under limited-data conditions and modest computational resources. Therefore, this study focuses on developing a CNN-based celebrity face recognition model using a subset of the CelebA dataset consisting of 1,200 facial images categorized into four classes: Blackpink, Ariana Grande, BTS, and Taylor Swift [14]. The proposed framework incorporates image preprocessing and data augmentation techniques to improve data quality and increase training data diversity, thereby enabling the CNN model to learn more representative facial characteristics. Model performance is evaluated using a confusion matrix and a classification report to assess prediction accuracy for each class [15]. The novelty of this study lies in its empirical analysis of CNN performance under limited dataset conditions, providing insights into the development of efficient, lightweight, and practical face recognition systems suitable for resource-constrained environments.

2. Methods

This study develops a celebrity face recognition model through two main phases: training and testing. The training phase begins with an image preprocessing stage designed to improve the quality and consistency of facial images, enabling the Convolutional Neural Network (CNN) to extract discriminative visual features more effectively [5]. Subsequently, data augmentation techniques are applied to increase the diversity of the training dataset and enhance the model's generalization capability.

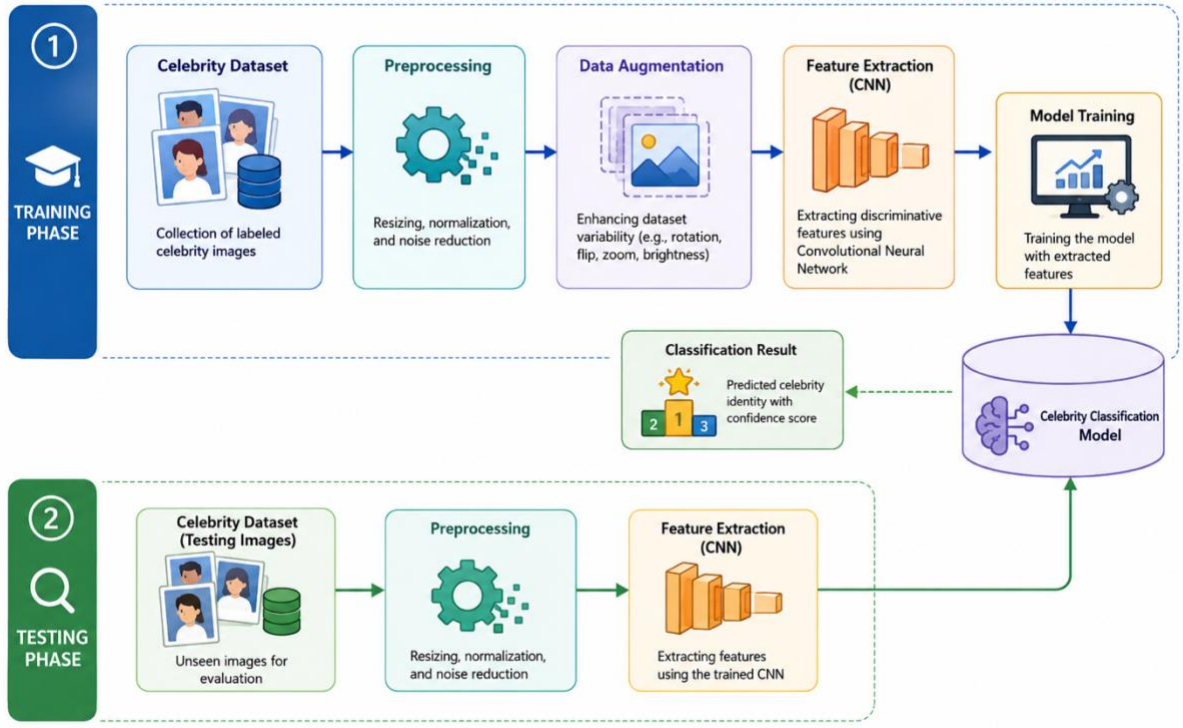


Figure 1. Research Framework

The next stage involves hierarchical facial feature extraction using the CNN architecture, followed by model training to learn representative facial patterns and generate the final classification model [10]. During the testing phase, each input facial image undergoes the same preprocessing and feature extraction procedures as those used

during training to ensure consistency. The extracted features are then classified using the trained CNN model to predict the corresponding celebrity class. This two-stage framework enables the proposed model to learn robust facial representations during training and accurately recognize unseen facial images during the evaluation process. Fig 1 depicts the complete research framework, outlining the main stages of the proposed methodology.

2.1. Dataset

This study employs two celebrity datasets: a filtered subset of the CelebA dataset containing selected artists and an additional dataset of artist images collected manually from the internet.

A. CelebA Subset Dataset

The CelebA subset dataset was derived from the original CelebA dataset by selecting facial images belonging to the four artist classes considered in this study: Blackpink, Ariana Grande, BTS, and Taylor Swift [12]. The resulting dataset comprises 1,200 facial images, with 300 images allocated to each class. The facial images vary in resolution and include a wide range of poses, facial expressions, and illumination conditions. Such diversity is essential for improving the model's generalization capability and enhancing its robustness in real-world applications [3]. For model development and evaluation, the dataset was divided into two subsets: 80% for training and 20% for testing. This split resulted in 960 images for training and 240 images for testing. The distribution of images across each class is summarized in Table 1.

B. Additional Celebrity Images Dataset

The additional celebrity images dataset consists of facial photographs collected from publicly available online sources. This dataset was incorporated to increase visual diversity and improve the model's ability to recognize faces under various pose, illumination, and background conditions. The dataset contains 600 additional images with varying resolutions, distributed across the same four artist classes used in this study [11].

Table 1. CelebA Subset Dataset

No	Image	Class
1		Blackpink
2		Ariana Grande
3		BTS
4		Taylor Swift

Prior to model training, all images were standardized to a uniform size and manually verified to ensure the correctness of their class labels. Similar to the CelebA subset, the additional dataset was divided into 80% training data and 20% testing data, resulting in 480 images for training and 120 images for testing. The distribution of

images for each class is presented in Table 2.

2.2. Preprocessing

Before being used for model training, all facial images underwent a preprocessing stage aimed at improving image quality and ensuring consistency across the dataset. This stage is essential for enabling the model to learn representative visual features effectively, particularly when the dataset contains significant variations in facial pose, illumination, and image quality [5].

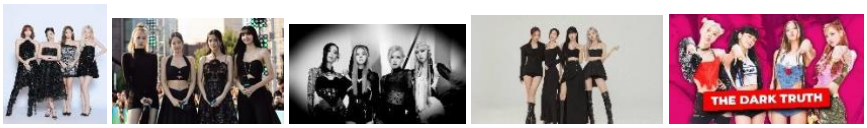
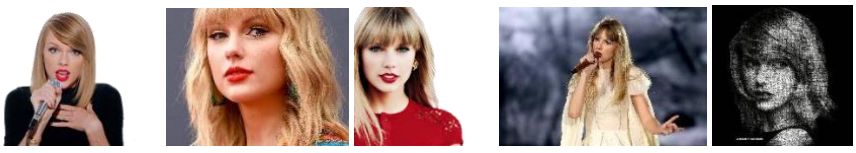
a. Before Face Alignment b. After Face Detection & Alignment

Figure 2. Example of facial images before and after face detection and alignment.

A. Face Detection and Alignment

Each facial image was first processed using the Multi-task Cascaded Convolutional Networks (MTCNN) algorithm to detect the face region, eyes, and other facial landmarks. Based on the detected eye landmarks, a face alignment procedure was then applied to normalize the orientation of each face. This process ensures that all facial images are consistently aligned, thereby reducing geometric variations and improving the robustness of the feature extraction process.

Table 2. Additional Celebrity Images Dataset

No	Image	Class
1		Blackpink
2		Ariana Grande
3		BTS
4		Taylor Swift

B. Image Resizing

After face detection and alignment, each facial image was resized to a resolution of 128 × 128 pixels. This resizing step standardizes the input dimensions required by the CNN model while reducing computational complexity, thereby improving training efficiency. An example of a facial image before and after the resizing

process is shown in Fig. 3.

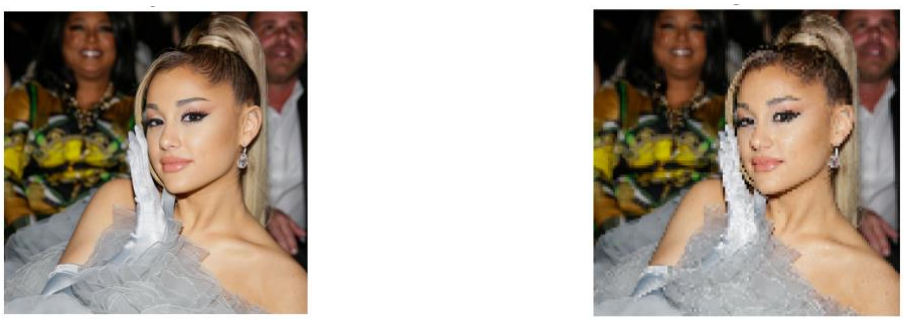
(a) Before Resize (b) After Resize

Figure 3. Facial images before and after resizing

After the resizing stage, all facial images were converted from their original, varying dimensions to a fixed resolution of 128 × 128 pixels. As a result, the images became more uniform while preserving their essential visual content. This process standardizes the input dimensions across the dataset, ensuring compatibility with the CNN architecture and facilitating subsequent processing stages.

C. Rescaling (Pixel Value Normalization)

To improve training stability and accelerate model convergence, the pixel values of all images were normalized from the original range of 0–255 to a floating-point range of 0–1. This normalization was performed by dividing each pixel value by 255, resulting in a standardized input scale that facilitates more efficient optimization during the training process. Fig. 4 shown an example image before and after rescaling process.



a. Before Rescaling b. After Rescaling

Figure 4. Facial images before and after Rescaling

After the rescaling stage, the resized facial images were normalized by converting their pixel values to the 0–1 range. Although this transformation produces little noticeable visual difference to the human eye, as the images appear nearly identical, it provides a more stable numerical representation that is better suited for machine learning algorithms. The normalized pixel values contribute to improved training stability and more efficient model optimization.

D. Contrast Adjustment

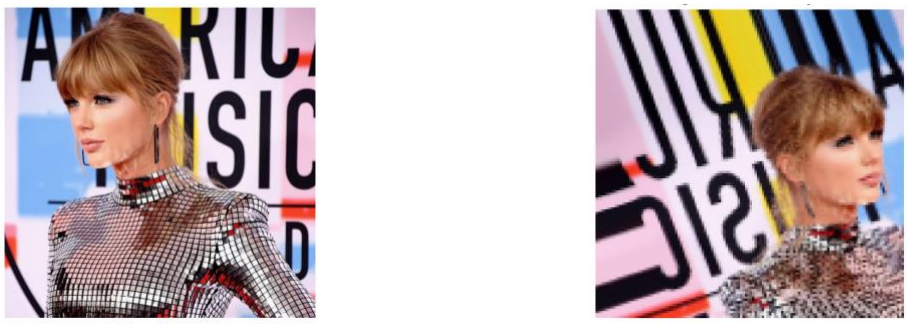
Contrast adjustment was performed by applying brightness variations within the range of 0.8–1.2. This augmentation step was introduced to mitigate the effects of varying illumination conditions commonly found in celebrity facial images collected from different sources. By exposing the model to images with different brightness levels, the model becomes more robust to lighting variations encountered in real-world scenarios. Fig. 5 shown an example image before and after contrast adjustment process. After the contrast adjustment stage, the differences between bright and dark regions of the face became more distinguishable compared to the previous images. Facial details, such as the eyes, nose, and facial contours, appeared more prominent. These results indicate that contrast enhancement can effectively strengthen important facial features within the images.

a. Before Contrast Adjustment b. After Contrast Adjustment

Figure 5. Example Image Before and After Contrast Adjustment

E. Image Augmentation

To increase the amount and diversity of training data, an image augmentation process was applied. Several augmentation techniques were employed, including: Rotation, Translation (horizontal and vertical shifts), Zooming, Horizontal flipping, and Brightness variation. The augmentation process enables the model to learn more robust facial representations by exposing it to various image conditions, poses, and viewing angles. This approach helps improve the model's generalization capability when recognizing faces under diverse real-world conditions. The augmentation process applied in this study does not rely on a single technique; instead, it combines multiple image augmentation methods. In the implemented configuration, rotation, translation (horizontal and vertical shifts), zooming, horizontal flipping, and brightness variation were applied simultaneously using the ImageDataGenerator framework.



a. Before Augmentation b. After Augmentation

Figure 6. Example Image Before and After Augmentation

Consequently, the augmented Taylor Swift facial images may undergo transformations such as rotation up to ±25 degrees, positional shifts of up to 20%, image scaling through zooming, horizontal mirroring, and variations in brightness levels. Therefore, the AUGMENTATION (Taylor Swift) category represents a combined image augmentation approach based on the predefined augmentation parameters. This strategy increases data diversity and enables the model to learn more robust facial features under various appearance conditions.

2.3. Convolutional Neural Network (CNN)

After undergoing the preprocessing stage, the facial images were processed using a Convolutional Neural Network (CNN) architecture to extract facial features and subsequently classify them into four celebrity classes [5].

A. Transfer Learning

The CNN model employed a transfer learning approach by utilizing the **NASNetMobile** architecture, which was previously trained on the **ImageNet** dataset. All base layers of NASNetMobile were frozen and set as non-trainable layers, allowing the pretrained network to function as a feature extractor for facial image representations.

B. CNN Model Architecture

The proposed CNN architecture consists of several main components, including:

- NASNetMobile base model with `include_top=False`
- Conv2D layer with 32 filters, a 3×3 kernel size, *same* padding, and ReLU activation function
- MaxPooling2D layer with a 2×2 pool size
- Dropout layer with a rate of 0.5 to reduce the risk of overfitting
- Flatten layer to convert extracted feature maps into a one-dimensional feature vector

- Dense output layer with 4 neurons and Softmax activation, corresponding to the number of target classes [10]

C. Model Compilation

The model was compiled using the following configurations:

- Loss function: Categorical cross-entropy
- Optimizer: Adam with a learning rate of 0.001
- Evaluation metric: Accuracy

D. Training Process

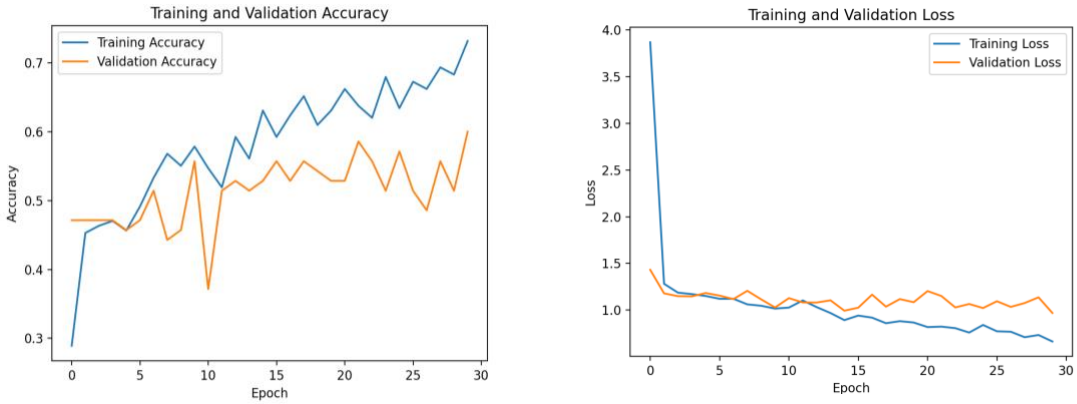
The model was trained for 30 epochs with a batch size of 64. The training dataset was augmented to improve model generalization, while the validation and testing datasets were processed only through rescaling without augmentation. The model evaluation was performed using the `flow_from_dataframe` method to manage image input and assess classification performance.

3. Results and Discussion

3.1. CNN Model Training Results

During the training phase of the Convolutional Neural Network (CNN) model for celebrity face recognition, the accuracy and loss values of both the training and validation datasets were monitored to evaluate the model's ability to learn facial visual patterns from each celebrity class. This monitoring process aims to assess the classification performance, measure prediction errors, and determine the model's generalization capability on unseen data. Furthermore, comparing training and validation performance is essential for identifying potential overfitting, ensuring that the obtained training results remain reliable and effective [6], [15].

Based on the accuracy curve, the training accuracy gradually increased and reached approximately 73% in the final epoch. Meanwhile, the validation accuracy fluctuated within the range of 50–60%, indicating that the model was able to learn meaningful patterns from the training data but remained less stable when applied to validation data. Although the difference between training and validation accuracy is still within an acceptable range, this performance gap suggests a tendency toward mild overfitting. The loss curve shows that the training loss consistently decreased throughout the training process, indicating that the model optimization proceeded effectively. In contrast, the validation loss remained relatively stable at approximately 1.0, suggesting that additional learning from the training data did not lead to significant improvement on the validation dataset. Therefore, the model's generalization capability remained limited, as illustrated in Figure 7.



a. Training and Validation Accuracy b. Training and Validation Loss

Figure 7. Training and validation accuracy and loss curves during CNN model training.

3.2. Model Performance Evaluation

To evaluate the capability of the Convolutional Neural Network (CNN) model in classifying facial images into four celebrity classes, namely Blackpink, Ariana Grande, BTS, and Taylor Swift, a confusion matrix and classification report were employed. The confusion matrix provides information regarding the number of correct and incorrect predictions for each class, allowing the identification of classes with the highest recognition accuracy and those with

the most frequent misclassifications. Meanwhile, the classification report presents evaluation metrics, including precision, recall, and F1-score, which are used to measure the accuracy and balance of the model's predictions for each class. This evaluation provides a comprehensive analysis of model performance and helps identify classification error patterns that can be considered for future model improvements [13], [9].

Based on the confusion matrix presented in Figure 9, the Blackpink class achieved the best performance, with 28 correct predictions out of 33 samples, indicating that this class was the easiest for the model to recognize. The BTS class was correctly classified in 7 out of 13 samples, while the remaining 6 samples were misclassified as Blackpink. The Ariana Grande class achieved 6 correct predictions out of 11 samples**, with the remaining samples being incorrectly classified into other classes. The Taylor Swift class obtained the lowest performance, with only 3 correct predictions out of 13 samples, while the remaining 10 samples were misclassified, primarily as Blackpink and Ariana Grande. These differences indicate that the model was more effective in identifying facial characteristics associated with the Blackpink class, whereas other classes exhibited higher feature similarity or were affected by a smaller number of training samples.

In addition to the confusion matrix, model performance was further evaluated using a classification report, which includes precision, recall, and F1-score values for each celebrity class. The classification report shows that the model achieved an overall accuracy of 63%. The Blackpink class achieved the highest performance, with a precision of 0.74, recall of 0.85, and an F1-score of 0.79, indicating relatively accurate recognition performance. The BTS and Ariana Grande classes demonstrated moderate performance, while the Taylor Swift class obtained the lowest performance with an F1-score of 0.29, indicating that the model still faced difficulties in distinguishing this class from other categories.

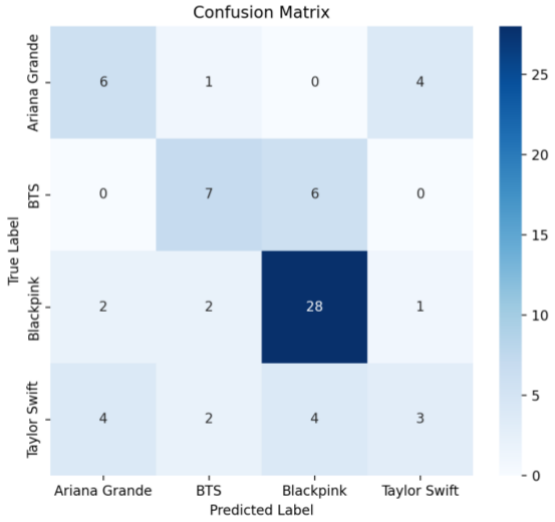


Figure 9. Confusion matrix illustrating the classification performance of the celebrity face recognition model

As presented in Table 3, the model achieved satisfactory overall performance; however, the average F1-score of 0.61 indicates that improvements are still required, particularly for classes with lower recognition performance. These results suggest that increasing data diversity and improving feature representation may help enhance the model's classification capability.

Table 3. Classification Report

	Precision	Recall	F1-score	support
Ariana Grande	0.50	0.55	0.52	11
BTS	0.58	0.54	0.56	13
Blackpink	0.74	0.85	0.79	33
Taylor Swift	0.38	0.23	0.29	13
Accuracy			0.63	70
Macro avg	0.55	0.54	0.54	70
Weighted avg	0.60	0.63	0.61	70

4. Conclusions

The results of this study demonstrate that the developed Convolutional Neural Network (CNN) model is capable of classifying and recognizing celebrity faces across four classes: Blackpink, Ariana Grande, BTS, and Taylor Swift. The initial dataset was collected using a web scraping approach through Bing Image Downloader, resulting in a total of 1,200 images, with 300 images allocated for each class. However, not all downloaded images were suitable for use due to corrupted files, incompatible formats, duplicate images, or images that could not be properly

detected by the system. After automated filtering and preprocessing, only valid images were utilized for the model training and testing processes. The valid image data were divided using the ImageDataGenerator framework with an 80% training and 20% testing split. As a result, 70 images were used as testing data. The CNN model achieved approximately 73% training accuracy and 63% testing accuracy after 30 training epochs. These results indicate that the model was able to learn fundamental facial feature patterns from each celebrity class; however, its generalization capability to unseen data remains limited.

Performance evaluation using the confusion matrix and classification report revealed variations among the four classes. The Blackpink class achieved the best performance, with 28 correct predictions out of 33 testing samples, along with a precision of 0.74, recall of 0.85, and F1-score of 0.79. These results indicate that the model was more effective in capturing the visual characteristics of this class. The BTS and Ariana Grande classes achieved moderate performance, while the Taylor Swift class obtained the lowest performance, with only 3 correct predictions out of 13 testing samples and an F1-score of 0.29.The differences in classification performance were influenced by variations in facial features, image acquisition angles, illumination conditions, and differences in the number of valid samples among classes. The implementation of preprocessing techniques, including resizing images to 128×128 pixels, pixel value normalization, and data augmentation strategies such as rotation up to $\pm 20^{\circ}$, translation of 20%, zooming of 20%, horizontal flipping, and brightness variation, contributed to improving the model's learning capability and reducing the risk of overfitting. Nevertheless, the findings indicate that the proposed system still has potential for further improvement. Future studies may enhance performance by increasing the quantity and quality of facial image data, balancing the distribution of samples across classes, and applying deeper CNN architectures or more advanced transfer learning-based models.

REFERENCES

[1] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," 2007.
[2] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition BT - British Machine Vision Conference," 2015.
[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks BT - Advances in Neural Information Processing Systems," 2012.
[5] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a Convolutional Neural Network BT - International Conference on Engineering and Technology," 2017.
[6] F. Chollet, Deep Learning with Python. Manning Publications, 2017.
[7] K. He, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition BT - Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition," 2016.
[8] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition BT - International Conference on Learning Representations," 2015.
[9] C. Szegedy, W. Liu, and Y. Jia, "Going Deeper with Convolutions BT - Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition," 2015.
[10] C. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," J. Big Data, vol. 6, no. 1, 2019.
[11] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering BT - Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition," 2015.
[12] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification BT - Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition," 2014.
[13] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition BT - Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition," 2019.
[14] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild BT - Proceedings of the IEEE International Conference on Computer Vision," 2015.
[15] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," Inf. Process. Manag., vol. 45, no. 4, 2009.