APPLIED ARTIFICIAL INTELLIGENCE AND ENGINEERING JOURNAL
E-ISSN: 0000-0000, DOI: 00.000/XXXX.YYYY.II.NNN
Vol. I No. 1 2026 pp. 71-79

A NASNetMobile-Based Deep Learning Framework for Cat and Dog Recognition

M.Ijlal Ziyad Akbar[1], Muhammad Syahrul[2], Ahmad Imran Ramadhan[3*]

[1,2] Program Studi Teknik Informatika, Jurusan Teknologi Informasi dan Komputer, Politeknik Negeri Lhokseumawe, Indonesia
[3] Program Studi Teknik Informatika , Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Syiah Kuala, Indonesia

*Corresponding Author: imranramadhan2005@gmail.com

Article info: Received month dd, yyyy, Revised month dd, yyyy, Accepted month dd, yyyy

Abstract

Animal recognition has emerged as a non-invasive approach that is increasingly adopted in various applications, including smart farming and image-based animal identification. This study proposes a Convolutional Neural Network (CNN)-based animal recognition system for classifying two companion animal species, namely cats and dogs. The experiments were conducted using the Kashtanka Pets (Animal Recognition) dataset, which consists of 400 images, including 200 cat images and 200 dog images. Prior to model training, the dataset underwent several preprocessing steps, including image resizing, pixel value normalization, contrast enhancement, and image augmentation to improve data diversity and reduce the risk of overfitting. The proposed model employs a transfer learning approach using the NASNetMobile architecture as the feature extractor, followed by a customized classification head for binary classification. The network was trained using the Adam optimizer with the categorical cross-entropy loss function and evaluated using accuracy, confusion matrix, precision, recall, and F1-score. Experimental results demonstrate that the proposed model achieved an overall test accuracy of 70%, while maintaining relatively balanced precision, recall, and F1-score across both classes. The confusion matrix analysis further indicates that misclassification errors were distributed relatively evenly between the cat and dog classes.The findings suggest that the transfer learning-based CNN is capable of effectively extracting discriminative visual features for animal recognition, although there remains considerable room for improving classification performance. Therefore, the proposed approach provides a promising foundation for the development of more robust image-based multi-species animal recognition systems.

Keywords: Animal Recognition, Deep Learning, Convolutional Neural Network (CNN), Transfer Learning, NASNetMobile

1. Introduction

Recent advances in computer vision and deep learning have significantly improved image analysis capabilities, enabling the development of automated animal recognition systems. These technologies have been widely adopted in applications such as smart farming, wildlife conservation, and non-invasive animal monitoring, where accurate identification is essential for effective population management and health assessment [1]. Compared with conventional identification techniques, image-based animal recognition offers a more efficient, scalable, and sustainable solution by eliminating the need for direct physical interaction with animals. Traditional animal identification methods, including ear tags, tattoos, and microchips, present several limitations. These approaches are susceptible to marker loss, require relatively high installation and maintenance costs, and may cause stress or discomfort to animals during the identification process [2]. Consequently, image-based animal recognition has emerged as a promising non-invasive alternative that provides safer, more efficient, and cost-effective identification for various livestock management and wildlife monitoring applications [3].

Despite these advantages, automated animal recognition remains more challenging than human recognition because of the substantial morphological variations among animal species. Differences in body shape, fur texture, coat color, and anatomical structures, together with uncontrolled environmental conditions such as illumination changes, animal movement, background complexity, and viewpoint variations, considerably increase the difficulty of extracting discriminative visual features [4]. These challenges limit the effectiveness of traditional handcrafted feature extraction methods, which often fail to capture the complex and non-linear visual patterns present in animal images [5]. The rapid advancement of deep learning has established Convolutional Neural Networks (CNNs) as one of the most effective approaches for animal recognition tasks. CNNs automatically learn hierarchical feature representations directly from raw images, enabling the extraction of low-level textures as well as high-level semantic features without manual feature engineering [6]. Numerous studies have demonstrated that CNN-based models outperform conventional machine learning approaches by providing greater robustness to variations in illumination, pose, and background while achieving superior classification performance across diverse image recognition tasks [7].

CNN-based animal recognition systems have recently been incorporated into numerous practical applications, including animal health monitoring, behavioral analysis, individual tracking, and camera-based population management. Modern smart farming and wildlife conservation systems increasingly rely on deep learning models to perform real-time animal recognition without requiring additional sensors or physical identification devices [8]. This capability improves operational efficiency while reducing maintenance costs and minimizing animal disturbance, making image-based recognition an attractive solution for intelligent livestock and conservation management [9]. Although CNN-based animal recognition has achieved encouraging performance, existing studies are predominantly designed for specific animal species or application domains, limiting the generalizability of the developed models to other species [10]. To address this limitation, the present study employs the Kashtanka Pets dataset, which contains images of two companion animal species—cats and dogs. The substantial visual differences between these species provide an appropriate benchmark for evaluating the effectiveness of CNN-based feature learning in inter-species recognition tasks.

This study proposes an animal recognition framework based on a Convolutional Neural Network using transfer learning with the NASNetMobile architecture. The proposed model is designed to accurately classify cat and dog images while maintaining robustness against visual variations. The findings are expected to contribute to the advancement of computer vision technologies for intelligent animal recognition, with potential applications in smart farming, wildlife conservation, and image-based animal monitoring systems [11].

2. Methods

This study develops an animal recognition model through two primary stages: the training phase and the testing phase. The proposed methodology follows a standard deep learning pipeline for image classification using a Convolutional Neural Network (CNN), consisting of image preprocessing, data augmentation, feature extraction, model training, and performance evaluation [10].

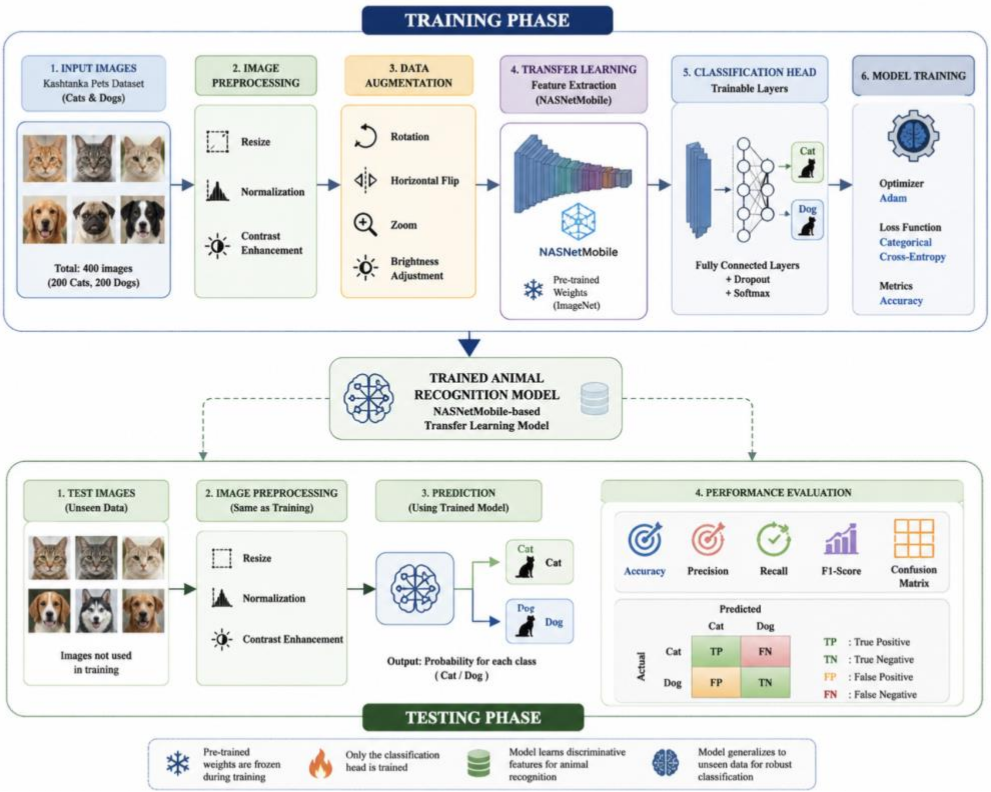


Figure 1. Workflow of the proposed animal classification model based on transfer learning using the NASNetMobile architecture

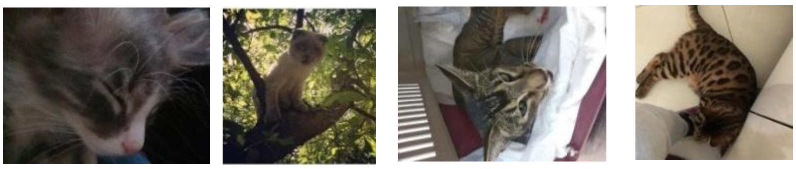
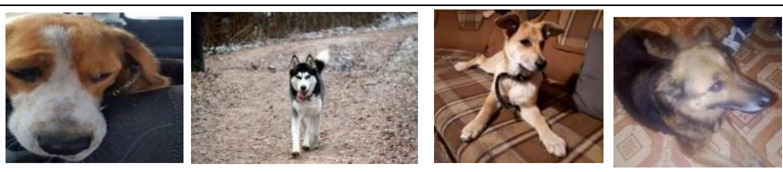
During the training phase, the animal images undergo a series of preprocessing operations to ensure consistent image quality and standardized input dimensions. Data augmentation techniques are subsequently applied to increase dataset diversity, thereby improving the model's robustness and reducing the risk of overfitting by exposing the network to a wider range of visual variations [3]. Feature extraction is then performed using a transfer learning approach based on the NASNetMobile architecture, which automatically learns hierarchical visual representations, including texture, shape, and other discriminative characteristics relevant to animal recognition [12]. The testing phase employs previously unseen images that are excluded from the training process to evaluate the model's generalization capability. This evaluation assesses the robustness of the proposed model under varying visual conditions and determines its effectiveness in recognizing different animal categories, consistent with established evaluation protocols for image-based animal recognition systems [15].

Fig. 1. Overall workflow of the proposed animal recognition system based on transfer learning using the NASNetMobile architecture. The framework consists of two main stages: (a) the training phase, including image preprocessing, data augmentation, feature extraction, and model training, and (b) the testing phase, where unseen images are classified and evaluated using accuracy, precision, recall, F1-score, and the confusion matrix.

2.1 Dataset

The dataset used in this study is the Kashtanka Pets (Animal Recognition) dataset, a publicly available benchmark containing images of companion animals, specifically cats and dogs. This dataset was selected because it provides substantial variations in pose, illumination, viewing angle, and fur texture, making it suitable for evaluating the robustness and generalization capability of convolutional neural network (CNN)-based animal recognition models [1]. A subset consisting of 400 images was employed in this study, including 200 cat images and 200 dog images. To ensure input consistency, all images were resized to 256 × 256 pixels, following a standard preprocessing procedure widely adopted in contemporary deep learning studies [13]. The dataset was then divided into training, validation, and testing sets with a ratio of 240:64:80, corresponding to 60%, 16%, and 20% of the total dataset, respectively. The training set was used to optimize the network parameters, the validation set was utilized to monitor the training process and mitigate overfitting, and the testing set was reserved for evaluating the model's generalization performance on previously unseen data [7]. Representative examples of the animal images used in this study are presented in Table 1.

Table 1. Examples of the images in the Kashtanka Pets Dataset

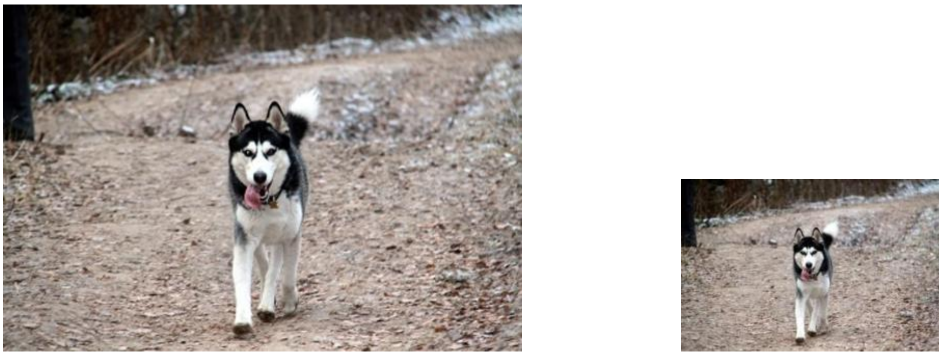
No.	Images	Class
1		Cat
2		Dog

2.2 Image Preprocessing

The image preprocessing stage was performed to improve image quality and ensure that all input data were standardized before being used for model training. This step is essential because variations in image resolution, illumination, and object pose can significantly affect the feature learning capability and classification performance of Convolutional Neural Networks (CNNs).

2.2.1 Image Resizing

Image resizing was performed to standardize the dimensions of all input images prior to model training. In this study, every image was resized to 256 × 256 pixels, ensuring a consistent input size compatible with the NASNetMobile architecture. Standardizing image dimensions not only reduces computational complexity but also facilitates efficient batch processing and stable feature extraction throughout the training process. Furthermore, uniform image resolution minimizes variations caused by different image sizes, allowing the network to focus on learning discriminative visual features rather than scale-related differences [13]. Examples of the original and resized images are presented in Fig. 2.



(a) Original (256 × 256 pixel) (b) After Resizing (128 × 128 pixel)

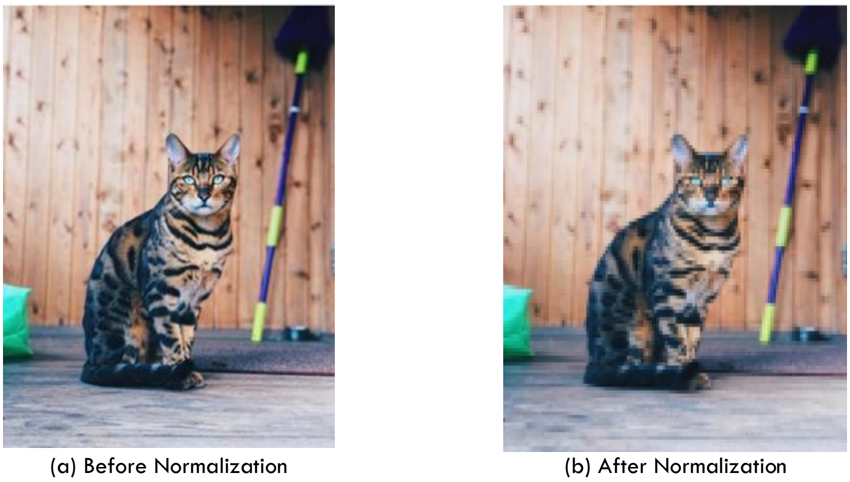
Figure 2. Dog image before and after Resizing

2.2.2 Pixel Normalization

Following the resizing process, pixel values were normalized to the range of 0–1 by dividing each pixel intensity by 255. Pixel value normalization is a fundamental preprocessing technique that reduces the influence of variations in image intensity and illumination while improving numerical stability during model training. This operation also accelerates the convergence of the optimization process and prevents large pixel values from dominating the learning process, thereby enabling the network to learn more discriminative visual representations [8]. For an input image $I(x,y)$, the normalized pixel value $I_{\text{norm}}(x,y)$ is computed as follows:

$$I_{\text{norm}}(x,y) = \frac{I(x,y)}{255} \quad (1)$$

where $I(x,y)$ denotes the original pixel intensity in the range [0, 255], and $I_{\text{norm}}(x,y)$ represents the normalized pixel value within the interval [0, 1]. This normalization ensures a consistent input scale for the NASNetMobile model, contributing to more stable training and improved classification performance. Examples of the original and normalized images are presented in Fig. 3.

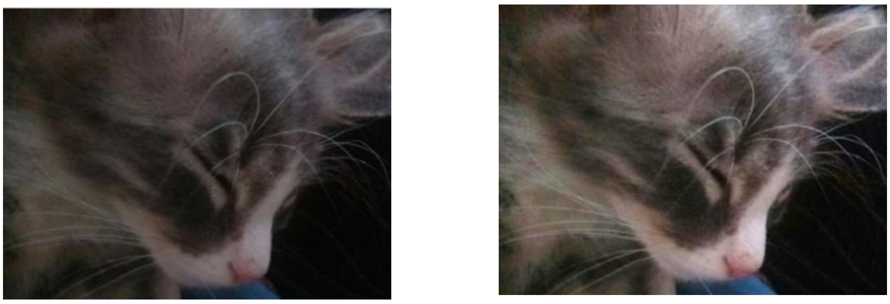


(a) Before Normalization (b) After Normalization

Figure 3. Example image before and after Normalization

2.2.3 Contrast Enhancement

Contrast enhancement was applied to improve the visibility of discriminative visual features in animal images, particularly those captured under suboptimal lighting conditions. Enhancing image contrast increases the distinction between foreground and background regions, enabling the model to better capture fine-grained characteristics such as fur texture, color distribution, and structural patterns that are essential for accurate animal recognition [2]. Examples of the original and contrast-enhanced images are presented in Fig. 4.

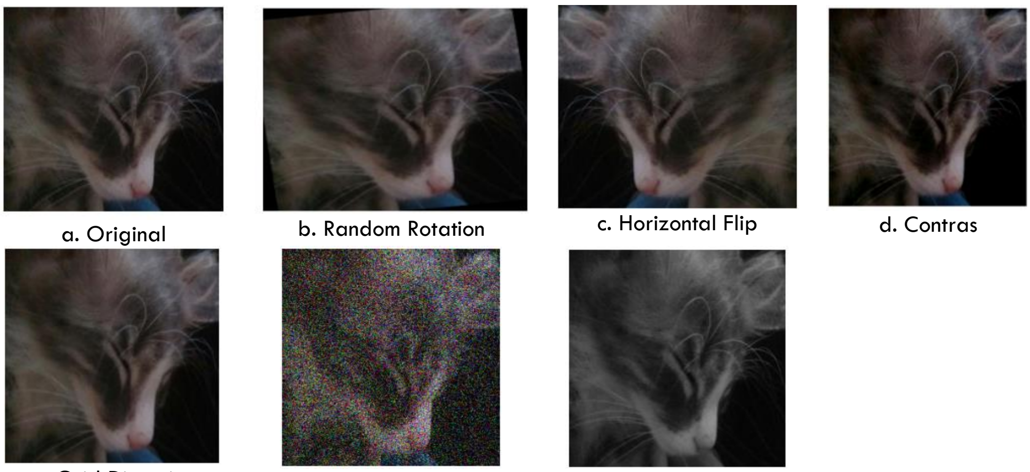


(a) Original Image (b) After Applying Contrast Enhancement

Figure 4. Example of Image before anda after Contrast Enhancement

2.2.4 Image Augmentation

Data augmentation was applied to increase the diversity of the training dataset and improve the generalization capability of the proposed model while reducing the risk of overfitting [3]. By generating multiple transformed versions of the original images, the model becomes more robust to variations that may occur in real-world scenarios, such as changes in orientation, position, and scale. In this study, the augmentation techniques included random rotation, horizontal flipping, zooming, and width/height shifting. These transformations preserve the semantic information of the images while introducing sufficient variability to enable the NASNetMobile-based transfer learning model to learn more discriminative and invariant feature representations. Representative examples of the augmented images are presented in Fig. 5.



a. Original b. Random Rotation c. Horizontal Flip d. Contras

e. Grid Distortion f. Noise Gaussian g. Grayscale

Figure 5. Image Augmentation Results

2.3 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) was employed as the primary classification model in this study due to its ability to automatically learn hierarchical visual features from images. Unlike conventional machine learning methods that rely on manually engineered features, CNN extracts discriminative representations directly from raw image data through convolutional operations. This capability enables the model to effectively capture complex visual patterns, including fur texture, facial structure, ear shape, and other distinctive characteristics that differentiate cats and dogs. Consequently, CNN has become one of the most effective deep learning architectures for image classification tasks because of its high feature extraction capability and robust classification performance [10].

- Transfer Learning
  Transfer learning was adopted to improve classification performance while reducing training time and computational cost. This study employed the NASNetMobile architecture as the backbone feature extractor due to its lightweight design and strong capability in learning discriminative visual representations from natural images. NASNetMobile was pre-trained on the ImageNet dataset, allowing the model to leverage previously learned low-level and high-level features, such as edges, textures, shapes, and object patterns. During training, the extracted features were fine-tuned using the cat and dog image dataset to adapt the model to the target classification task. This approach has been widely reported to achieve competitive performance, particularly when the available training data are relatively limited [12].

- CNN Model Architecture
  The proposed classification model combines the pre-trained NASNetMobile network with several additional neural network layers to improve classification capability. The feature extraction stage utilizes the convolutional layers of NASNetMobile, followed by a custom classification head consisting of a Global Average Pooling layer, Dropout, and fully connected (Dense) layers. The dropout layer was incorporated to reduce overfitting by randomly deactivating a proportion of neurons during training, thereby improving the model's generalization performance. Finally, a Dense output layer with a Softmax activation function was employed to generate the probability distribution for the two target classes, namely cat and dog. This architecture enables the model to efficiently learn discriminative features while maintaining computational efficiency suitable for practical image classification applications [10].

- Model Compilation
  Before training, the model was compiled using the Adam optimizer, which adaptively updates the learning rate for each parameter and has demonstrated excellent convergence performance in deep learning applications. The categorical cross-entropy loss function was selected because the task involves multi-class probability prediction using one-hot encoded labels. Model performance was evaluated using classification accuracy as the primary evaluation metric throughout the training and validation processes. This compilation configuration has been widely adopted in image classification studies due to its stability and effectiveness in optimizing deep neural networks [14].

- Training Process
  The proposed model was trained for 40 epochs with a batch size of 64. Prior to training, the dataset was divided into training, validation, and testing subsets. Data augmentation techniques, including random rotation, horizontal flipping, zooming, and shifting, were applied exclusively to the training set to increase data diversity and reduce overfitting. In contrast, the validation and testing datasets underwent only pixel normalization to ensure that the evaluation accurately reflected the model's generalization capability without introducing artificial variations. During training, both training and validation accuracy and loss were monitored to assess convergence and detect potential overfitting. The model with the best validation performance was subsequently used for the final evaluation on the testing dataset [3].

3. Result and Discussions

The NASNetMobile-based CNN model was trained using 240 training images, validated with 64 validation images, and evaluated on 80 testing images. Model performance was assessed by analyzing the training and validation accuracy and loss curves throughout the training process, followed by quantitative evaluation using a confusion matrix and classification report. This evaluation strategy provides a comprehensive assessment of both the learning behavior and the classification capability of the proposed model.During training, the model was optimized for 40 epochs, while monitoring the accuracy and loss on both the training and validation datasets. These learning curves are commonly used to evaluate the convergence behavior of deep learning models and to identify potential overfitting or underfitting.

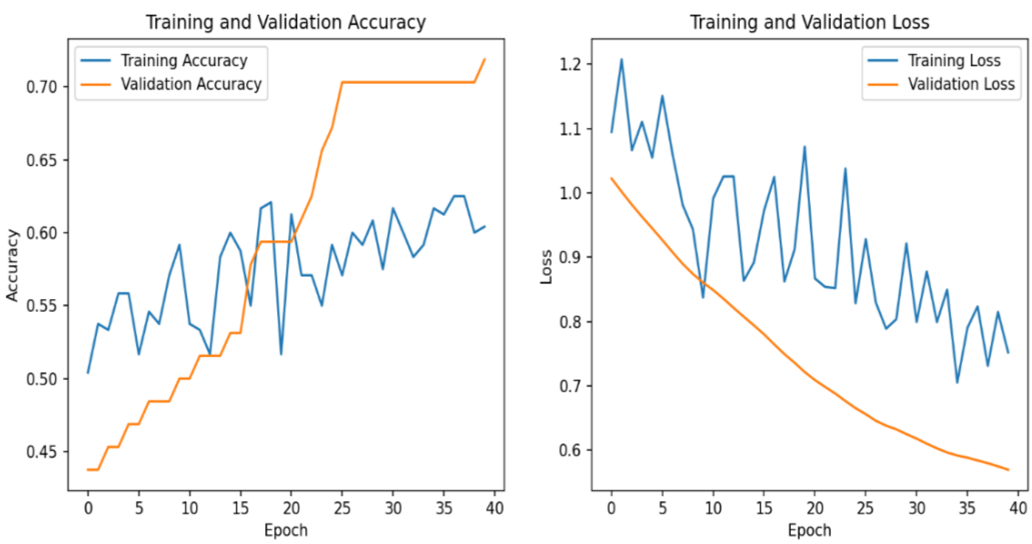


Figure 6. Training performance of the NASNetMobile-based CNN model in terms of accuracy and loss over 40 epochs

As illustrated in Fig. 6, the training accuracy exhibited moderate fluctuations but showed a gradual upward trend, reaching approximately 0.60 at the final epoch. In comparison, the validation accuracy increased more consistently and achieved approximately 0.72, indicating that the model successfully learned discriminative visual features and generalized reasonably well to unseen images.

The corresponding loss curves further demonstrate the effectiveness of the training process. Both training and validation loss gradually decreased throughout the 40 epochs, although minor fluctuations were observed in the training loss. The validation loss exhibited a smoother and more consistent decline, suggesting continuous improvement in the model's predictive capability as the network parameters were updated. A decreasing loss trend generally indicates successful optimization and effective feature learning in CNN-based image classification models. Furthermore, the gap between the training and validation curves remained relatively small, suggesting that the model maintained a good balance between fitting the training data and generalizing to new samples. No significant signs of overfitting were observed during training. This behavior can be attributed, in part, to the application of data augmentation techniques, which increased the diversity of the training samples and improved the model's robustness despite the relatively limited size of the Kashtanka Pets dataset. Overall, the experimental results demonstrate that the proposed NASNetMobile-based CNN successfully learned representative visual characteristics of cat and dog facial images, achieving stable convergence and satisfactory generalization performance. These findings support previous studies reporting that transfer learning with lightweight CNN architectures can provide reliable and efficient performance for animal image classification tasks, particularly when only a limited amount of labeled training data is available.

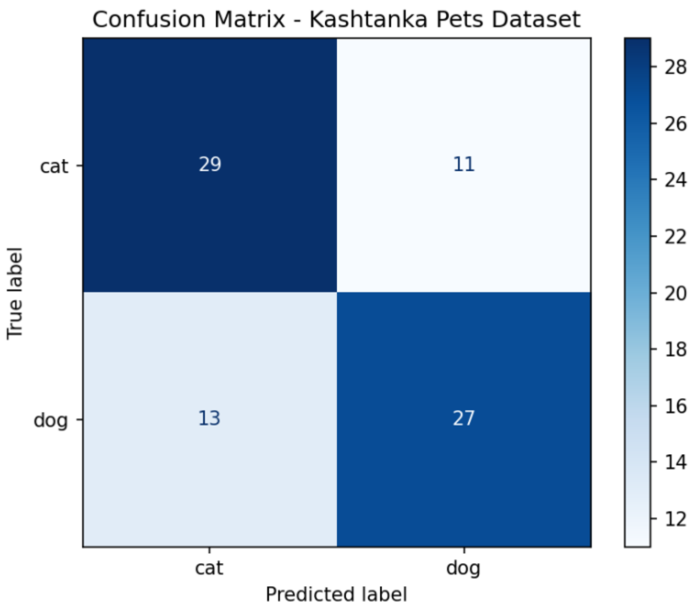


Figure 7. Confusion matrix of the proposed NASNetMobile-based CNN model

The confusion matrix obtained from the Kashtanka Pets Dataset is presented in Fig. 7. Of the 40 cat facial images included in the testing set, 29 images were correctly classified as cats, while the remaining 11 images were incorrectly classified as dogs. Similarly, for the 40 dog facial images, the proposed model correctly identified 27 images as dogs, whereas 13 images were misclassified as cats. Overall, the confusion matrix indicates a relatively balanced distribution of misclassification errors between the two classes, suggesting that the proposed model does not exhibit a significant prediction bias toward either cats or dogs. Nevertheless, the observed classification errors indicate that distinguishing between the two animal classes remains a challenging task under certain conditions. Misclassifications are likely associated with the high visual similarity between cat and dog facial characteristics, particularly when images are captured under varying viewpoints, illumination conditions, facial expressions, or partial occlusions. Such variations can reduce the discriminative capability of the extracted feature representations, leading to incorrect predictions.

Despite these challenges, the confusion matrix demonstrates that the NASNetMobile-based CNN successfully captured the essential visual characteristics required for binary animal classification. The remaining errors highlight opportunities for further performance improvement through the use of larger and more diverse datasets, advanced data augmentation strategies, or fine-tuning of deeper network layers. Similar observations have been reported in previous studies, where variations in pose, texture, and lighting conditions were identified as major factors affecting the classification performance.

The detailed classification performance of the proposed model on the Animal Faces Dataset is summarized in Table 3. The classification report indicates that the model achieved relatively balanced performance across both target classes. For the Cat class, the model obtained a precision of 0.69 and a recall of 0.72, indicating that most cat images were correctly identified, although a number of samples were still misclassified as dogs. Conversely, for the Dog class, the model achieved a precision of 0.71 and a recall of 0.68, demonstrating comparable classification capability while also revealing some difficulty in correctly recognizing certain dog images. The F1-score, which provides a balanced measure of precision and recall, reached 0.71 for the Cat class and 0.69 for the Dog class. These comparable values suggest that the model maintains consistent predictive performance across both categories without exhibiting a substantial bias toward either class. The relatively balanced precision, recall, and F1-score further confirm that the NASNetMobile-based CNN effectively learned representative facial features from both animal categories.

Overall, the proposed model achieved an overall classification accuracy of 70% on the testing dataset. Although this result demonstrates the capability of the NASNetMobile-based CNN to perform binary animal face classification with satisfactory performance, there remains considerable potential for improvement. Future work may focus on increasing the size and diversity of the training dataset, optimizing the fine-tuning strategy of the pre-trained network, and investigating more advanced deep learning architectures or attention mechanisms to enhance feature discrimination and classification accuracy. These improvements are expected to increase the model's robustness, particularly when dealing with challenging images exhibiting variations in pose, illumination, facial expression, and background complexity.

Table 2. Classification Reports

Class	Precision	Recall	F1-score	Support
Cat	0.69	0.72	0.71	40
Dog	0.71	0.68	0.69	40
Accuracy			0.70	80
Macro avg	0.70	0.70	0.70	80
Weighted avg	0.70	0.70	0.70	80

4. Conclusion

This study successfully developed a Convolutional Neural Network (CNN)-based animal recognition system for binary classification of cats and dogs using the Kashtanka Pets Dataset. By employing a transfer learning approach with the NASNetMobile architecture, the proposed model achieved an overall testing accuracy of 70%, accompanied by balanced precision, recall, and F1-score values across both classes. These results demonstrate that the model is capable of effectively learning representative facial features while maintaining comparable classification performance for cats and dogs. The training process exhibited stable convergence, as reflected by the gradual improvement in validation accuracy and the consistent reduction in validation loss over 40 training epochs. Furthermore, the confusion matrix analysis revealed a relatively balanced distribution of misclassification errors between the two classes, indicating that the model did not exhibit a significant prediction bias toward either category. These findings suggest that the combination of transfer learning and data augmentation contributed to improved model generalization despite the limited size of the training dataset.Although the achieved accuracy indicates satisfactory performance, there remains room for improvement. Classification errors were primarily associated with the visual similarity between cat and dog facial characteristics, as well as variations in pose, illumination, and image quality. Future research should investigate larger and more diverse datasets, more extensive fine-tuning strategies, and advanced deep learning architectures, such as attention-based networks or transformer-based models, to further enhance feature representation and classification accuracy.Overall, this study demonstrates that the NASNetMobile-based CNN provides an effective and computationally efficient approach for animal recognition, highlighting its potential for practical applications in intelligent pet identification, animal monitoring systems, and other computer vision-based animal recognition tasks.

REFERENCES

[1]. H. K. Kuehl and T. Burghardt, "Animal biometrics: Quantifying and detecting phenotypic appearance," Trends in Ecology & Evolution, vol. 28, no. 7, pp. 432–441, 2013. DOI: 10.1016/j.tree.2013.02.013.

[2]. S. Kumar and S. K. Singh, "Visual animal biometrics: Survey," IET Biometrics, vol. 6, no. 3, pp. 139–156, 2017. DOI: 10.1049/iet-bmt.2016.0017.

[3]. C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, art. no. 60, 2019. DOI: 10.1186/s40537-019-0197-0.

[4]. R. Chen, R. Little, L. Mihaylova, and R. Delahay, "Wildlife surveillance using deep learning methods," Ecology and Evolution, vol. 9, no. 17, pp. 9453–9466, 2019. DOI: 10.1002/ece3.5410.

[5]. A. Hitelman, Y. Edan, A. Godo, R. Berenstein, J. Lepar, and I. Halachmi, "Biometric identification of sheep via a machine-vision system," Computers and Electronics in Agriculture, vol. 194, art. no. 106713, 2022. DOI: 10.1016/j.compag.2022.106713.

[6]. H. Nguyen, S. J. Maclagan, T. D. Nguyen, T. Nguyen, D. S. Phung, and R. C. Stone, "Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring," in Proc. IEEE Int. Conf. Data Science and Advanced Analytics (DSAA), 2017. DOI: 10.1109/DSAA.2017.31.

[7]. F. Chollet, Deep Learning with Python, 2nd ed. Shelter Island, NY, USA: Manning Publications, 2021. ISBN: 978-1-61729-686-4.

[8]. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015. DOI: 10.1038/nature14539.

[9]. B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 8697–8710. DOI: 10.1109/CVPR.2018.00907.

[10]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594.

[11]. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. International Conference on Learning Representations (ICLR), 2015. DOI: 10.48550/arXiv.1412.6980.

[12]. J. Kim, J. Ko, and C. Kwon, "Comparison of fine grained classification of pet images using image processing and CNN," Journal of Broadcast Engineering, vol. 26, no. 2, pp. 175–183, 2021. DOI: 10.5909/JBE.2021.26.2.175.

[13]. E. Okafor, "Deep learning for animal recognition," Ph.D. dissertation, University of Groningen, Groningen, The Netherlands, 2019. ISBN: 978-94-034-1459-1.

[14]. [14] P. Sykora, P. Kamencay, R. Hlavata, and R. Hudec, "Overview and comparison of deep neural networks for wildlife recognition using infrared images," AI, vol. 5, no. 4, pp. 2801–2828, 2024. DOI: 10.3390/ai5040135.

[15]. D. Schneider et al., "Recognition of European mammals and birds in camera trap images using deep neural networks," IET Computer Vision, 2024. DOI: 10.1049/cvi2.12294.