Vol 43, No 1 (2026): April 2026

Jurnal Penelitian Pendidikan
https://journal.unnes.ac.id/journals/JPP

Application of Fine-Tuning on Convolutional Neural Networks to Improve Classification Accuracy of Feline Skin Diseases

Syafa Nabilah Syahputra*, Riza Arifudin

Universitas Negeri Semarang, Indonesia

*Corresponding Author: syafa.nabilah@students.unnes.ac.id

Abstract
Skin conditions are among the most frequent reasons for veterinary visits, yet they remain notoriously difficult to distinguish by eye alone. For the average pet owner or general practitioner, overlapping visual symptoms between diseases like ringworm and scabies often lead to diagnostic uncertainty. This study addresses this challenge by developing an automated classification system based on the MobileNetV2 architecture. By employing a two-stage transfer learning strategy, where initial feature extraction is followed by targeted fine-tuning of layers from index 100 onwards, we adapted a general-purpose model to the specific nuances of veterinary dermatology. Our results indicate a significant performance leap: while standard training struggled with the complexities of skin textures, the fine-tuned model achieved a validation accuracy of 92%. These findings suggest that fine-tuning is not just a technical optimization, but a necessary step in making deep learning a viable, accessible tool for real-world veterinary diagnostics.

Keywords: MobileNetV2, fine-tuning, feline skin diseases, computer vision

INTRODUCTION

For most pet owners, a cat's skin health is a primary indicator of their overall well-being. However, early signs of dermatological issues are often subtle and easily misread. Traditional diagnosis relies heavily on the subjective eye of a veterinarian or costly lab tests, which are not always accessible to every pet owner. This gap in care often results in delayed treatment or, worse, the accidental spread of zoonotic diseases to humans.

While Deep Learning and Convolutional Neural Networks (CNNs) have shown incredible promise in human medicine, their application to feline health is hampered by a lack of massive, high-quality datasets. Without a large volume of data, deep models tend to "overfit," memorizing specific images rather than learning the actual characteristics of a disease. This study explores the role of fine-tuning as a bridge across this gap. By taking a model already "educated" on millions of general objects and carefully refining its upper layers to recognize the specific patterns of feline fur and skin lesions, we aim to provide a diagnostic accuracy that rivals professional observation.

The primary challenge in veterinary computer vision is "data scarcity," where the volume of annotated medical images for felines is significantly lower than that of human datasets. Without sufficient data, traditional CNNs risk high variance and poor generalization. This research addresses this by implementing transfer learning with MobileNetV2 architecture. By utilizing a model pre-trained on the ImageNet dataset, we can extract sophisticated low-level features such as edges and textures and repurpose them for the specific domain of veterinary dermatology. This approach significantly reduces the training time and data requirements while maintaining a diagnostic accuracy that bridges the gap between digital accessibility and professional veterinary consultation [1].

In the Indonesian veterinary landscape, the ratio of veterinarians to pet owners remains significantly unbalanced, especially in suburban areas. Many pet owners resort to "self-diagnosis" via unverified internet sources, which often leads to the misuse of medications or the worsening of zoonotic conditions that could spread to humans. The development of a MobileNetV2-based diagnostic system offers a scalable solution that can be integrated into mobile applications. By providing a preliminary screening tool that is both fast and computationally efficient, this technology empowers owners to seek professional help sooner, thereby improving animal welfare standards and public health safety across the region [2].

The journey toward automated veterinary diagnostics has evolved significantly in recent years. Early efforts, such as the work by Putri et al. [3], utilized the Modified K-Nearest Neighbor (MKNN) method. While groundbreaking, these traditional algorithms often struggled with the high degree of visual similarity between different feline skin conditions. More recently, Pangestu et al. [4] shifted the focus toward accessibility by developing an Android-based detection system, noting that over 65% of cat owners find it difficult to identify skin issues manually.

Despite these advancements, the challenge of balancing model "weight" with accuracy remains. High-performance models often require immense computing power, making them impractical for mobile use. MobileNetV2 [5] solved the efficiency problem through its inverted residual structure, but its general-purpose training is not enough for the complexities of dermatology. This research builds upon the work of Pangestu and Putri by specifically investigating how surgical-level fine-tuning can improve a model's ability to distinguish between nearly identical clinical symptoms.

METHOD

This research adopts a systematic scientific framework to ensure methodological rigor in developing a deep learning–based classification model for feline skin diseases. The research workflow follows established best practices in machine learning predictive modeling. Such a structured approach is essential to reduce bias, improve reproducibility, and ensure reliable performance assessment, particularly in medical image classification tasks with limited datasets. A clearly defined methodological pipeline encompassing data preparation, model development, and evaluation is critical for building robust and generalizable predictive models [6].

Dataset dan Preprocessing

The dataset used in this study was obtained from the public repository Roboflow Universe, *Feline Skin Disease* dataset, that available at https://universe.roboflow.com/bios-mz7bh/penyakit-kulit-pada-kucing-obdto-fby1x. The dataset consists of 1,543 labeled images representing six common feline skin disease categories, namely *demodicosis*, *dermatitis*, *flea allergy*, *fungus*, *ringworm*, and *scabies*. The dataset was divided into training (1,235 images), validation (155 images), and testing (153 images) subsets using a class-specific directory structure. Although the overall dataset size is adequate for model development and evaluation, variations in class distribution may influence the model's generalization performance if not properly addressed.

To prepare the data for model training, all images were resized to 224×224 pixels and normalized through rescaling operations to align with the specifications of the MobileNetV2 architecture. To enhance dataset diversity and mitigate the risk of overfitting, data augmentation techniques were employed, specifically *Random Flip* (horizontal and vertical), *Random Rotation* (0.2), *Random Zoom* (0.2), and *Random Contrast* (0.2). This approach aligns with the study by Shorten and Khoshgoftaar, which demonstrated that variations in geometric and photometric transformations can improve the generalization of deep learning models across various image domains [7].

Convolutional Neural Networks

The core methodology of this research relies on the hierarchical feature extraction capabilities of Convolutional Neural Networks (CNNs). Unlike traditional image processing, CNNs utilize convolutional kernels to automatically learn spatial hierarchies of features, ranging from simple edges in the initial layers to complex pathological patterns in the deeper layers. In the context of feline dermatology, these layers identify the distinct textures of lesions and inflammatory responses against the background of varied fur patterns [8]. This automated extraction eliminates the need for manual feature engineering, which is often prone to human error in medical diagnostics.

To optimize the model for mobile accessibility, this study utilizes the MobileNetV2 architecture, which introduces the concept of Inverted Residuals and Linear Bottlenecks. The primary methodology involves depth-wise Separable Convolutions, which split a standard convolution into two separate layers: a depth-wise convolution for filtering and a point-wise convolution for combining. This significantly reduces the number of parameters and computational cost (floating-point operations) without a substantial loss in accuracy [5]. For feline skin disease classification, this efficiency allows the model to perform high-speed inference on devices with limited hardware resources, such as smartphones used by pet owners in the field [9].

Model Architecture

MobileNetV2 pre-trained on ImageNet served as the backbone, with its original top classification layers removed [10]. The architecture was extended with a custom head consisting of *Global Average Pooling 2D* for spatial dimension reduction, a Dropout layer (0.3) for regularization, and a Dense output layer using SoftMax activation, with the number of neurons matching the specific target classes.

The selection of the MobileNetV2 architecture for feline skin disease classification is fundamentally driven by its optimized balance between computational efficiency and representational power [1]. Unlike traditional deep learning models that use standard convolutional layers, MobileNetV2 utilizes depth-wise Separable Convolutions [5]. This mechanism factorizes the convolution process into two distinct steps: a depth-wise Convolution for spatial filtering and a Pointwise Convolution (1 x 1) for feature combination. This approach reduces computational costs by nearly 8 to 9 times compared to standard CNNs, facilitating real-time diagnostic performance on mobile devices via TensorFlow Lite [11].

Furthermore, architecture introduces Inverted Residual Blocks and Linear Bottlenecks to enhance feature propagation while maintaining a low memory footprint. While standard residual blocks connect high-dimensional layers, the inverted residual structure in MobileNetV2 connects thin "bottleneck" layers, which ensures that large intermediate tensors never fully materialize in the main memory [9]. To prevent significant information loss in these low-dimensional spaces, the model employs Linear Bottlenecks, removing the non-linear activation functions in the narrow layers to preserve the representational capacity required to detect subtle textures in feline skin lesions [12].

Finally, the custom head of the model replaces traditional "Flatten" layers with Global Average Pooling (GAP) 2D, which significantly reduces the total number of trainable parameters and provides implicit regularization. This is particularly vital for specialized veterinary datasets to minimize the risk of overfitting. The integration of a Dropout layer (0.3) and a SoftMax activation layer ensures that the model can generalize effectively across various target classes of dermatological conditions [13].

Training Protocol

The evaluation was conducted using two experimental scenarios to compare the efficacy of transfer learning for cat skin disease classification. The first scenario (Feature Extraction Only) involved freezing all pre-trained layers of the MobileNetV2 architecture, allowing only the custom-added classifier to be trained using the Adam optimizer with a learning rate of 1e-4 for 50 epochs.

The second scenario (Direct Fine-Tuning) involved unfreezing the MobileNetV2 layers from index 100 onwards from the onset of training. This model was trained using the Adam optimizer with a reduced learning rate of 1e-5 to facilitate delicate weight adjustments in the high-level feature maps. This comparative strategy aimed to assess how adapting the base model's upper layers to the specific visual characteristics of cat skin lesions enhances performance compared to relying solely on generic pre-trained features.

RESULTS AND DISCUSSION

This section presents and discusses the experimental results obtained from the fine-tuned

MobileNetV2 model for feline skin disease classification, analyzing model performance before and after fine-tuning using accuracy and confusion matrix–based metrics to evaluate improvements in classification capability.

Evaluation Matrix

To obtain a complete picture of performance, analysis went beyond simple accuracy. A Confusion Matrix was utilized to identify exactly where the model was getting "confused," measuring Precision, Recall, and the F1-score for each disease. Based on the parameters above, the following mathematical formulas are used to calculate the model's efficacy:

- Accuracy: Represents the overall percentage of correct predictions (both positive and negative) out of the total dataset.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

- Precision: Measures the model's exactness by calculating the ratio of correctly predicted positive observations to the total predicted positives, which is crucial for avoiding misdiagnosis.

$$Precision = \frac{TP}{TP+FP}$$

- Recall (Sensitivity): Measures the model's completeness by calculating the ratio of correctly predicted positive observations to all actual observations in that class [14].

$$Recall = \frac{TP}{TP+FN}$$

- F1-Score: The harmonic mean of Precision and Recall, providing a balanced metric that is particularly useful when dealing with imbalanced veterinary datasets [15].

$$F1\text{-}Score = 2 \times \frac{Precission \times Recall}{Precission+Recall}$$

Beyond the basic formulas, integrating a Confusion Matrix allows for Error Analysis. In the context of cat skin diseases, many conditions present with similar visual symptoms (e.g., redness or alopecia) [16]. The Confusion Matrix reveals "inter-class similarity," showing if the model is specifically struggling to distinguish between, for example, *Feline Acne* and *Allergic Dermatitis* [17].

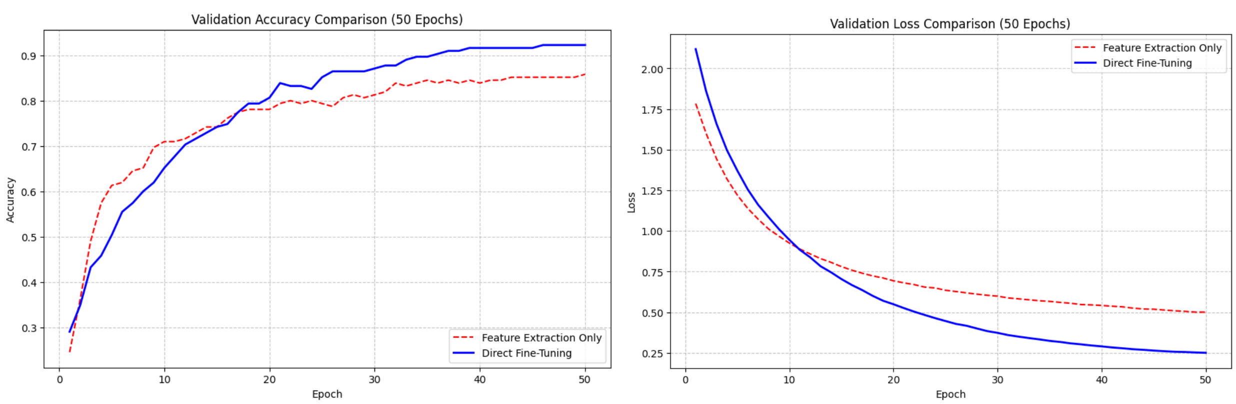

Figure 1. Accuracy vs Loss

Experimental Results

The impact of fine-tuning was immediate and clearly observable in both the validation accuracy and validation loss curves. Without fine-tuning, the model exhibited a rapid increase in validation accuracy during the initial epochs but soon reached a plateau at approximately 83%, indicating limited adaptation to the domain-specific characteristics of feline skin disease images. In this configuration, the model frequently misclassified scabies as dermatitis, reflecting difficulties in distinguishing visually similar inflammatory conditions.

Conversely, implementing fine-tuning caused a notable improvement in how the model performed. The fine-tuned model demonstrated a steady and sustained increase in validation accuracy, ultimately reaching approximately 92% by the final epoch. This improvement suggests that the model successfully adapted to the specific textures and visual patterns present in the dataset. Additionally, the validation loss decreased more sharply and converged to a lower value compared to the model without fine-tuning, indicating improved optimization stability and reduced generalization error.

Discussion

The comparison of the confusion matrices obtained before and after applying fine-tuning to the MobileNetV2 architecture highlights significant differences in feline skin disease classification performance. In the scenario without fine-tuning, the model exhibits a noticeable level of misclassification, particularly among disease classes with visually similar characteristics. For instance, fungus is frequently misclassified as scabies, while scabies is often confused with fungus and dermatitis. These errors indicate that the baseline model has limited capability in capturing subtle texture differences between closely related skin conditions.

After fine-tuning is applied, a substantial improvement in classification performance is observed across nearly all classes. The number of correct predictions along the main diagonal of the confusion matrix increases, while inter-class misclassifications are significantly reduced. Notably, fungus and scabies, which previously exhibited the highest confusion rates, demonstrate marked gains in classification accuracy. Furthermore, ringworm and demodicosis are classified with near-perfect accuracy, suggesting that fine-tuning enables the model to learn more discriminative and domain-specific visual features.

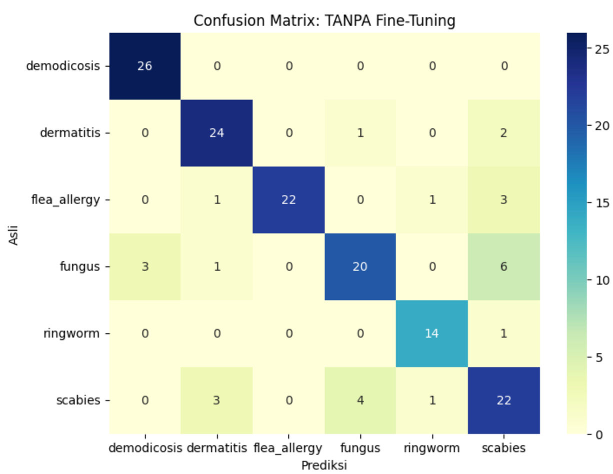

Figure 2. Confusion Matrix without Fine-Tuning

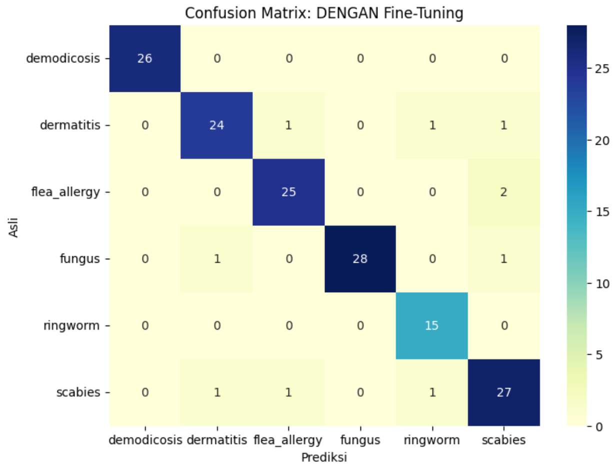

Figure 3. Confusion Matrix with Fine-Tuning

The experimental results highlight an important clinical insight: certain feline skin diseases are inherently easier to identify than others based on visual characteristics. Demodicosis demonstrated the strongest performance, achieving a near-perfect F1-score of 0.96 and a recall of 1.00, indicating that all demodicosis cases were correctly identified by the model. This high performance can be attributed to the distinct visual markers commonly associated with demodicosis lesions, which make them more separable from other disease categories.

A more detailed class-level analysis using the classification report further reveals variations in model performance across disease types. Dermatitis achieved high precision (0.96) and recall (0.93), reflecting reliable classification despite its known visual similarity to other skin conditions. However, dermatitis remained one of the most challenging classes, as it was frequently confused with fungal infections. This observation is consistent with the findings of Putri et al. [3], who described dermatitis as a “visual chameleon” in veterinary dermatology due to its overlapping symptoms with other skin diseases.

The flea allergy and fungus classes showed slightly lower recall values, indicating the presence of false negatives that may stem from subtle differences in texture and lesion patterns. Ringworm achieved perfect recall, suggesting that the model successfully detected all true cases; however, its lower precision indicates occasional false positives. Scabies recorded the lowest F1-score among all classes, which may be caused by subtle and less distinctive visual features that are more difficult for the model to differentiate from other inflammatory skin conditions.

Table 1. Classification Result

Diseases	**Precision**	**Recall**	**F1-score**
Demodicosis	0.93	1	0.96
Dermatitis	0.93	0.93	0.93
Flea Allergy	1	0.89	0.94
Fungus	0.93	0.83	0.88
Ringworm	0.79	1	0.88
Scabies	0.93	0.93	0.93
Accuracy	0.92	0.92	0.92

CONCLUSION

This study confirms that fine-tuning is the critical "missing link" in adapting general AI models for veterinary medicine. The proposed approach achieved 92% classification accuracy by refining MobileNetV2 with structured preprocessing and optimization. The results indicate that deep learning models, when properly adapted, can effectively distinguish between multiple categories of feline skin diseases with a high level of reliability and computational efficiency.

The findings confirm that fine-tuning plays a critical role in bridging the gap between general-purpose pretrained models and domain-specific veterinary applications. Fine-tuning enabled the model to refine previously learned features and adapt them to the visual characteristics of feline skin conditions, resulting in a significant improvement in classification performance compared to models without fine-tuning. This process effectively reduced overfitting and enhanced the model's ability to generalize, particularly when working with limited and imbalanced medical image datasets.

The outcomes of this research highlight the potential of fine-tuned deep learning models as supportive diagnostic tools in veterinary medicine. The achieved accuracy and computational efficiency suggest that such models can be integrated into lightweight and accessible systems, including mobile-based applications, to assist pet owners and veterinary professionals in early disease detection. By facilitating faster and more consistent preliminary assessments, this approach has the potential to encourage earlier consultation with veterinary experts, ultimately contributing to improved treatment outcomes and quality of life for feline companions.

REFERENCES

[1] Aji, D. S., Ariyus, D., & Saputra, W. M. (2025). Classification of Cat Skin Diseases Using MobileNetV2 Architecture with Transfer Learning. *Journal of Applied Informatics and Computing (JAIC)*, 9(6).

[2] Wicaksono, A. S., & Ariswati, N. K. (2022). Klasifikasi Penyakit Kulit pada Hewan Menggunakan Metode Convolutional Neural Network dengan Arsitektur MobileNetV2. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, 11(3), 184-191.

[3] M. B. P. Putri, E. Santoso, and Marji, "Diagnosis Penyakit Kulit Pada Kucing Menggunakan Metode Modified K-Nearest Neighbor," J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 1, no. 12, pp. 1797-1803, 2017.

[4] I. Y. Pangestu and S. R. Ramadhani, "Perancangan Sistem Deteksi Penyakit Kulit Pada Kucing Menggunakan Deep Learning Berbasis Android," Politeknik Caltex Riau, 2023.

[5] M. Sandler, et al., "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 4510-4520, 2018.

[6] Moons, K. G. M., Royston, P., Vergouwe, Y., Grobbee, D. E., & Altman, D. G., "Systematic review identifies the design and methodological conduct of studies on machine learning-based prediction models," Journal of Clinical Epidemiology, 154, 8–22, 2022.

[7] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," J. Big Data, vol. 6, no. 60, 2019.

[8] Faris, M., Rustam, R., & Arisanti, D. (2023). Analisis Performa Algoritma Convolutional Neural Network dalam Identifikasi Penyakit Hewan. *Jurnal Ilmiah Teknologi Sistem Informasi*, 4(1), 15-22

[9] Putra, R. A., & Wijayanto, A. W. (2022). Implementasi Deep Learning untuk Klasifikasi Penyakit Kulit Menggunakan Arsitektur MobileNetV2. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 9(4), 725-732.

[10] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). "ImageNet: A Large-Scale Hierarchical Image Database," IEEE Conference on Computer Vision and Pattern Recognition, 248-255.

[11] Furqon, A., Malik, K., & Fajri, F. N. (2024). Detection of Eight Skin Diseases Using Convolutional Neural Network with MobileNetV2 Architecture for Identification and Treatment Recommendation on Android Application. Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI), 10(2), 373-384.

[12] Tarisafitri, N., & Aljabar, A. (2025). Classification of Skin Diseases Using Digital Image Processing with MobileNetV2 Architecture. Nusantara Journal of Artificial Intelligence and Information Systems, 1(1), 35-44

[13] Hermawan, R. A., Taufik, I., & Gerhana, Y. A. (2025). Klasifikasi Citra Ras Kucing Berbasis CNN dengan Metode MobileNet-V2. INTERNAL (Information System Journal), 8(1), 70-84.

[14] Liu, J., et al. (2021). Deep Learning in Veterinary Medicine: Current Applications and Future Perspectives. Veterinary Sciences. (Discusses the application of CNNs in animal diagnostics and the

importance of high Sensitivity/Recall).
[15] Taufiqoh, S. D., & Purnamasari, P. D. (2025). Development of Detection Model for Skin Diseases in Pets Using Image Processing and Deep Learning Techniques. International Journal of Electrical, Computer, and Biomedical Engineering, 3(2).
[16] Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process, 5(2), 1. (Classic reference for why F1-Score is superior to Accuracy in imbalanced data).
[17] Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for Multi-class Classification: an Overview. *arXiv preprint arXiv:2008.05756. (Provides the mathematical justification for Macro vs. Micro averaging in multi-disease classification).*