Veterinary Dermatology

ESVD European Society of Veterinary Dermatology | ACVD American College of Veterinary Dermatology

ORIGINAL ARTICLE OPEN ACCESS

Artificial Intelligence-Based Identification of Common Canine Skin Lesions From Clinical Images

Soh-Yoon Kang[1] | Yeong-Hun Kang[1,2] | Hyun-Jung Kim[3] | Eun-A. Huh[3] | Jong-Soo Jeon[3] | Hyun-Su Kim[3] | Min-Sun Kim[1] | Anisha Tiwari[4] | Cheol-Yong Hwang[1]

[1]Laboratory of Veterinary Dermatology and the Research Institute for Veterinary Science, College of Veterinary Medicine, Seoul National University, Seoul, Korea | [2]Western Animal Medical Center, Seoul, Korea | [3]AIFORPET Corporation, Seoul, Korea | [4]Suvidha Pet Care and Derma Centre, Jammu, Jammu and Kashmir, India

Correspondence: Cheol-Yong Hwang (cyhwang@snu.ac.kr)

Received: 1 September 2025 | **Revised:** 31 March 2026 | **Accepted:** 7 April 2026

Keywords: convolutional neural networks | deep learning | dog | EfficientNet | teledermatology

ABSTRACT

Background: Accurate evaluation of skin lesions is an essential component of dermatological examination, yet it can be time-consuming and subject to interobserver variability. While artificial intelligence (AI) models have shown reliability in diagnosing specific skin diseases, lesion-level identification remains underexplored in veterinary dermatology.

Objectives: To develop and evaluate a convolutional neural network (CNN)-based AI model for the automated identification of four skin lesion types in dogs: erythema, lichenification, alopecia and erosion/ulcer.

Animals: Clinical skin images were collected from dogs presented to a veterinary medical teaching hospital.

Materials and Methods: Four EfficientNet models were independently trained, one for each lesion type. Model performance was evaluated by comparing the prediction results with veterinary surgeon-labelled data, using six metrics: accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and F1 score.

Results: All four CNN models achieved an accuracy of >90%, indicating a reliable performance across all lesion types. The alopecia model yielded the highest accuracy (98.12%) and F1 score (98.18%). The erythema and erosion/ulcer models exhibited balanced performance across all metrics. For the lichenification model, the sensitivity (87.02%) and F1 score (89.88%) were the lowest among the four lesion types.

Conclusions and Clinical Relevance: The CNN-based AI models developed in this study demonstrated validity in identifying common canine skin lesions from clinical images. These models may facilitate rapid and objective dermatological evaluation, supporting clinical diagnosis and lesion monitoring throughout treatment.

1 | Introduction

Artificial intelligence (AI) is a technology that enables computers to simulate human intelligence, including learning, perception, reasoning and decision-making. AI has been increasingly developed and applied in various fields of medicine and healthcare [1]. Deep learning is a subset of AI that autonomously learns patterns from large datasets. Convolutional neural networks (CNNs) are one of the most widely used deep learning models, particularly effective in image analysis. CNNs automatically detect and extract visual features from images through multiple layers of processing [2–4].

This study was presented at the Asian Meeting of Animal Medicine Specialties (AMAMS) 2025 in Thailand.

EfficientNet, an advanced type of CNN model, was developed to achieve better accuracy and efficiency compared to traditional CNNs. By using a method called compound scaling, it improved performance while reducing computational requirements [5]. EfficientNet has been widely applied in the medical field, where minimising the risk of misdiagnosis is critical. It is suitable for processing high-resolution images such as skin lesion identification, and computed tomography (CT) or magnetic resonance imaging (MRI) image analysis [6, 7].

Skin diseases are one of the most common disorders in dogs visiting primary-care veterinary clinics, which impact the quality of life of patients and their owners [8, 9]. For accurate diagnosis and management of skin diseases, identifying abnormalities through dermatological examination is a primary step. Because the type, severity and distribution of skin lesions vary depending on pathogenesis, precise observation and palpation of the skin are essential for developing a differential diagnosis list [10].

Among various skin lesions, erythema, lichenification, alopecia and erosion/ulcer are commonly observed in clinical practice and serve as key indicators of various skin diseases [11–13]. These lesions also are components of the Canine Atopic Dermatitis Extent and Severity Index, 4th iteration (CADESI-04) and the Canine Atopic Dermatitis Lesion Index (CADLI), which are widely used scoring systems in veterinary dermatology [14, 15].

Although lesion evaluation is an essential process, it can be challenging owing to its time-consuming nature and interobserver variability. Given these challenges, AI has been explored as a supportive tool for efficient and consistent dermatological examination. However, most previous studies in AI dermatology have focused on classifying and diagnosing skin diseases, rather than identifying individual lesions. For example, in humans, AI has demonstrated success in diagnosing various skin diseases such as melanoma, non-melanoma skin cancer, acne and psoriasis [2, 16–19]. Likewise, in veterinary dermatology, AI models have been developed to detect canine skin diseases, including infectious dermatitis, allergies, pododermatitis and otitis externa [20–23]. While these studies have successfully developed models for diagnosing specific diseases, they may have limited clinical applicability across a broader range of dermatological conditions.

Automated lesion identification using AI could enable rapid and objective skin evaluation. Furthermore, such models could assist in diagnostic approaches and lesion monitoring throughout treatment. Therefore, this study aimed to develop and validate a CNN-based AI model for the automated identification of four lesion types—erythema, lichenification, alopecia and erosion/ulcer—based on clinical images.

2 | Materials and Methods

2.1 | Image Collection

In this study, clinical skin images were retrospectively collected from canine patients that visited the Department of Dermatology, Seoul National University Veterinary Medical Teaching Hospital (SNU-VMTH), between January 2023 and February 2024. Owners provided informed consent for the collection and use of clinical data for research purposes.

In order to reflect the natural diversity of the clinical population encountered in daily practice, images were collected consecutively from all available patients without breed-specific or phenotype-specific (e.g., coat colour or hair length) exclusion criteria. Images were acquired from various body regions without restrictions. These clinical photographs were captured using various smartphones during routine practice. Various lesion states were included, such as pre-and post-crust removal, as well as clipped areas. The captured images were standardised in JPEG format with file sizes ranging from approximately 0.1–2.0 MB and an average of 0.3 MB. Subsequently, all images were cropped to centre on the lesion area. Specifically, when a single source image contained multiple target lesions, distinct cropped images were generated, with each crop centred on the respective target lesion.

Dataset cleaning and quality control were conducted by a machine learning engineer. Images with improper brightness or poor focus, and other low-quality images, were excluded from the dataset. During this preliminary screening, a veterinary surgeon validated the clinical evaluability of the images to ensure that they were suitable for model training. The primary criterion for excluding these 'low-quality' images was the inability to visually identify the lesion. Specifically, this included images with severe motion blur, images focused on the hair rather than the skin, or images taken in significantly inadequate lighting. Lesions only partially covered by hair were retained to reflect real-world clinical cases.

After image collection, separate training datasets were created for the four types of lesions: erythema, lichenification, alopecia and erosion/ulcer. Erosion and ulcer were grouped into a single class as a cosnequence of the difficulty in visually determining the depth of tissue damage in images. A single image could be included in multiple datasets, as distinct cropped images were generated from source images with multiple lesions. Consequently, the final datasets used for CNN model training consisted of: 3968 images for erythema; 2512 images for lichenification; 2763 images for alopecia; and 1985 images for erosion/ulcer.

2.2 | Image Labelling

Four datasets dedicated to erythema, lichenification, alopecia and erosion/ulcer were independently labelled as either positive (lesion present) or negative (lesion absent). Negative images were defined as images in which the target lesion type was not present. Explicitly, images of completely healthy skin as well as images exhibiting other nontarget lesions (e.g., an erosion/ulcer without erythema) were labelled as negative. For example, an erythematous erosion/ulcer was labelled as positive in both the erythema and erosion/ulcer datasets, while simple erythema was labelled as negative in the erosion/ulcer dataset. Lesion labelling was based on visually recognisable clinical features without applying a strict quantitative threshold, reflecting routine dermatological assessment. Representative examples of positive and negative images for each lesion type are shown in Figure 1.

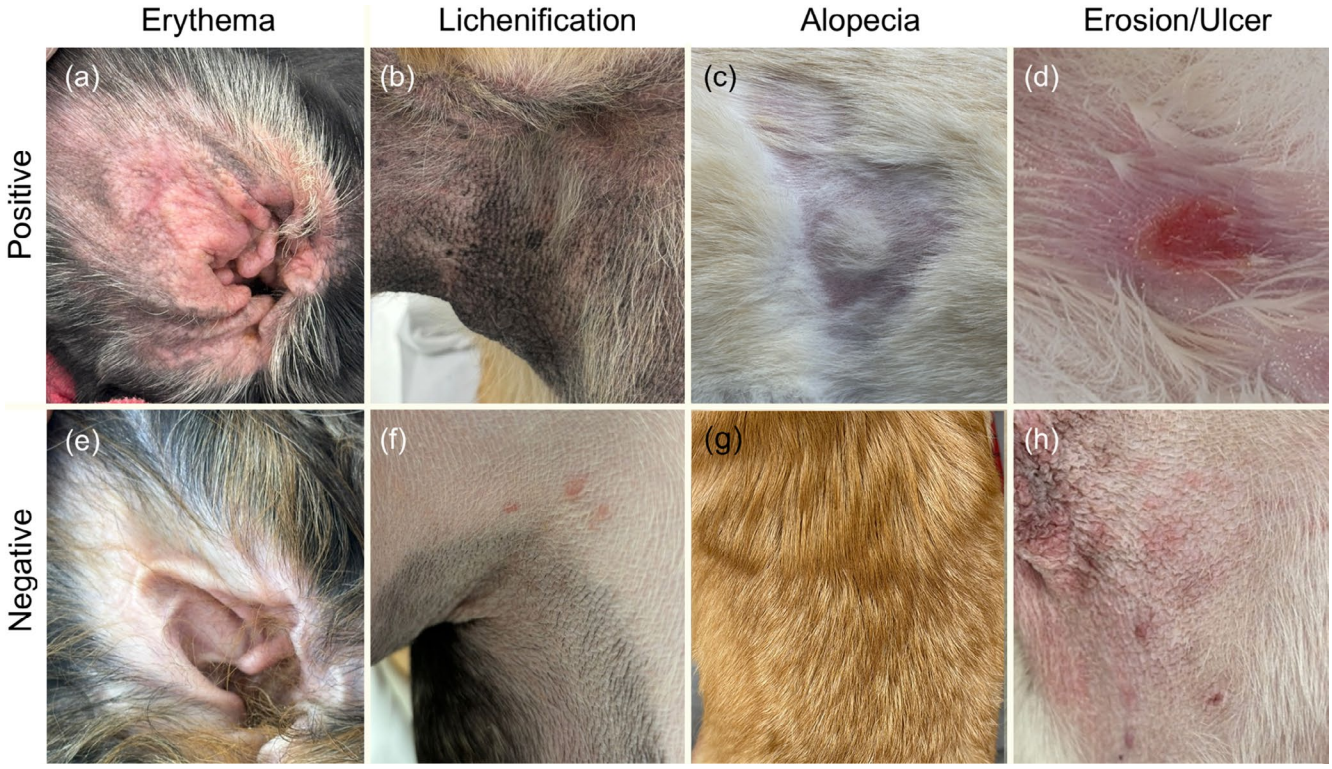

FIGURE 1 | Representative images of each target lesion (upper row) and their corresponding negative images (lower row). (a–d) Positive images showing the presence of erythema, lichenification, alopecia and erosion/ulcer, respectively. Negative images: (e) normal skin without erythema. (f) hyperpigmentation and erythematous macules without lichenification. (g) normal skin without alopecia. (h) erythematous macules and alopecia without erosion/ulcer.

Labelling was performed by veterinary surgeons, using software designed for data labelling and annotation (Lime Research; Lime Solution Corp). The labelling veterinary surgeons comprised both those who examined the patients and independent evaluators blinded to the clinical history. After three veterinary surgeons had independently performed initial labelling, cross-validation was conducted. Any disagreements in labelling were reviewed, and the evaluators reached a consensus based strictly on visual evidence. Final confirmation was achieved through a review and correction by a board-certified referral clinician (DipAiCVD), ensuring consistency of labelling results.

2.3 | Training the CNNs

EfficientNet, a type of CNN model, was employed for this study. Four separate models were independently trained for each lesion.

The images were resized to 600×600 pixels for training. The models were pretrained on ImageNet, a large-scale benchmark image dataset widely used in deep learning research [24]. After pretraining, the models were fine-tuned using the labelled images from this study. Training was conducted using TensorFlow (v2.11.0, CUDA v11.4), an open-source deep learning library developed by Google. The process was run on an NVIDIA A100 GPU machine with 80 GB of VRAM.

2.4 | Performance Evaluation of CNNs

In order to evaluate the performance of the CNNs after training, independent test datasets were prepared. While collected from the same clinical settings as the training dataset, the test images were prospectively acquired. The test datasets were deliberately sampled to maintain a balanced number of images across the four lesion types, ensuring a comparable evaluation of each model's performance.

The test dataset sizes of each CNNs were as follows: 970 images for erythema, 880 images for lichenification, 743 images for alopecia and 792 images for erosion/ulcer. All test images were evaluated and labelled by a board-certified referral clinician to ensure reliability and consistency.

The performance of the CNN models was evaluated by comparing their identification results with a veterinary surgeon's assessment. Six metrics were used to evaluate the CNN performance: accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and F1 score. These metrics were calculated based on the values of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

Accuracy is defined as the proportion of correctly identified images among all images, reflecting the overall performance of the model. However, in imbalanced datasets where one class (either positive or negative) dominates, accuracy alone may

not provide a reliable assessment. Therefore, additional metrics also were used for better evaluation. The F1 score, which is the harmonic mean of PPV and sensitivity, helps evaluate how well the model detects positive cases while balancing FP and FN [25]. A high F1 score indicates that the model achieves both high PPV (low FP rate) and high sensitivity (low FN rate).

Accuracy and F1 score are calculated by the following equations:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{F1 Score} = 2 \times \frac{PPV \times \text{Sensitivity}}{PPV + \text{Sensitivity}}$$

3 | Results

3.1 | Training Dataset and Labelling Results

Training dataset sizes varied across the four lesion types. Erythema had the largest dataset (3968 images), while erosion/ulcer had the smallest (1985 images).

Among the total 11,228 training images, breed information was available for 3049 images. The most common breeds were miniature poodle (428 images, 14.04%), Maltese (425 images, 13.94%) and shih tzu (342 images, 11.22%). Other commonly represented breeds included mixed-breed dogs (285 images, 9.35%), Pomeranian (214 images, 7.02%) and Cocker spaniel (198 images, 6.49%). In total, 42 different breeds were included in the training dataset, as detailed in Table S1.

TABLE 1 | Image labelling results for the training datasets.

Lesion type	Positive (*n*)	Negative (*n*)	Total (*n*)	P:N ratio
Erythema	2451	1517	3968	1.6:1
Lichenification	825	1687	2512	1:2.0
Alopecia	1353	1410	2763	1:1.0
Erosion/ulcer	339	1646	1985	1:4.9
	4968	6260	11,228	1:1.3

Note: Each image was labelled as either positive (lesion present) or negative (lesion absent) for a single lesion type—erythema, lichenification, alopecia or erosion/ulcer.
Abbreviation: P:N ratio, positive-to-negative image ratio.

The labelling results of training datasets for the four lesion types are summarised in Table 1. The proportion of positive and negative images varied among lesion types. The alopecia dataset exhibited a balanced distribution, with positive and negative images in an approximate 1:1 ratio. By contrast, the erosion/ulcer dataset revealed an imbalance, with 339 positive images and 1646 negative images, resulting in a ratio of 1:4.9. The erythema and lichenification datasets showed mild imbalance, with ratios of 1.6:1 and 1:2, respectively.

3.2 | CNN Performance

Each CNN model was independently evaluated using a separate test dataset. In total, the test datasets consisted of 3385 clinical images from 40 breeds, including shih tzu, Maltese, mixed-breed dogs, Pomeranian, miniature poodle and French bulldog. A detailed breed composition is provided in Table S2.

The performance evaluation results of four CNN models are presented in Table 2. Metrics were calculated from confusion matrix values—TP, TN, FP and FN (Table S3). Heatmaps illustrating the regions that contributed to model predictions are shown in Figure 2. Warmer colours represent areas with greater influence on the model's decisions.

All four CNN models achieved an accuracy of over 90%, indicating a reliable performance across all lesion types. The alopecia model attained the highest accuracy of 98.12%. The F1 score of 98.18% reflected a balanced performance between sensitivity (98.69%) and PPV (97.67%).

For the lichenification model, the sensitivity (87.02%) and F1 score (89.88%) were the lowest among the four lesion types, while specificity (94.66%) remained high. The lower sensitivity indicates a relatively higher false negative rate in detecting lichenification.

The erythema and erosion/ulcer models demonstrated balanced performance across all metrics. The erythema model showed an accuracy of 94.74% and an F1 score of 94.68%, while the erosion/ulcer model exhibited an accuracy of 93.81% and an F1 score of 93.17%.

4 | Discussion

This study developed CNN-based AI models for identifying common skin lesions in dogs. The lesion-specific CNN

TABLE 2 | Performance of convolutional neural network (CNN) models on the test dataset.

Lesion type	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F1 score (%)
Erythema	94.74	93.22	96.27	96.19	93.37	94.68
Lichenification	91.25	87.02	94.66	92.93	90.04	89.88
Alopecia	98.12	98.69	97.51	97.67	98.60	98.18
Erosion/ulcer	93.81	93.30	94.24	93.04	94.46	93.17

Note: Four CNN models were independently trained for each lesion type and assessed using six metrics: accuracy, sensitivity, specificity, positive predictive value, negative predictive value and F1 score.
Abbreviations: NPV, negative predictive value; PPV, positive predictive value.

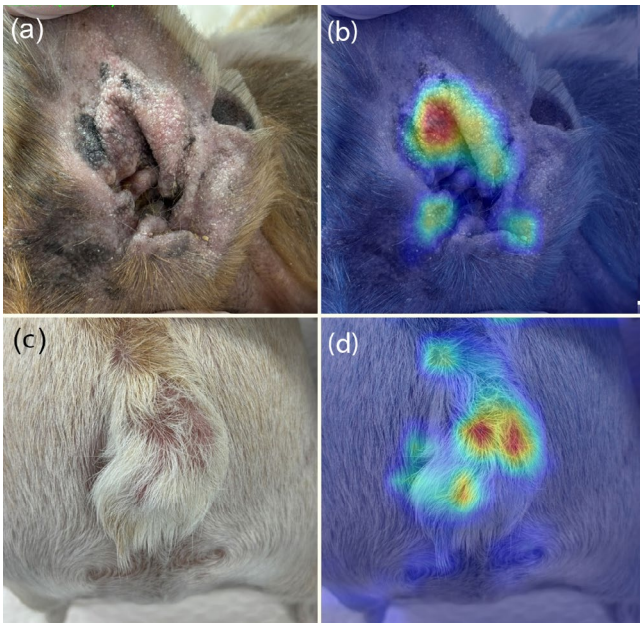

FIGURE 2 | Heatmaps generated from convolutional neural network (CNN) model predictions for lichenification and alopecia. (a) Original image of lichenification and (b) corresponding heatmap for lichenification. (c) Original image of alopecia and (d) corresponding heatmap for alopecia. Warmer colours (e.g., red, yellow) indicate regions with greater influence on the model's decision.

models, trained for erythema, lichenification, alopecia and erosion/ulcer, achieved reliable performance in identifying each lesion type.

All four models provided overall reliability, yet variations in their performance were observed. These differences may have been influenced by factors such as dataset size and class imbalance [26]. The favourable performance of the erythema model may be attributed to the relatively large size of its training dataset, which was approximately twice that of the smallest dataset. By contrast, the erosion/ulcer training dataset had the smallest size and a relatively pronounced class imbalance. This imbalance between positive and negative images in the erosion/ulcer dataset may reflect the naturally low prevalence of these lesions in clinical practice. Despite these limitations, the erosion/ulcer model achieved performance metrics comparable to the erythema model. This finding suggests that intrinsic lesion characteristics may have compensated for the quantitative limitations of the dataset. To rigorously assess the model's reliability given the class imbalance, the F1 score was evaluated. While the F1 score is typically more sensitive to imbalances than accuracy, in this study, it did not differ substantially from the accuracy. This similarity reflects the consistently low FP and FN rates, confirming that the class imbalance did not lead to biased results.

The morphological complexity of lesions significantly affected the performance of CNN models. A previous study on an AI model for human acne detection reported the influence of morphological clarity on model performance. The study explained the model's highest performance in detecting nodular/cystic lesions to their distinct colour and larger size [19]. In the present study, the impact of morphology is evident when comparing alopecia and lichenification. Although both models were trained on datasets of similar size, their performance outcomes diverged significantly. The relatively clear differentiation between alopecic and non-alopecic areas is likely to have contributed to the alopecia model's robust performance. By contrast, the lichenification model exhibited relatively lower sensitivity and F1 score, which may be a consequence of the visual ambiguity of the lesion. Lichenification can be confused with natural skin folds or skin atrophy, depending on the degree of skin thickening, leading to an increased false negative rate. Also, lichenification often presents in combination with hyperpigmentation, which can obscure lesion boundaries and negatively affect model performance. Furthermore, the stable performance of the erosion/ulcer model—comparable to that of the erythema model despite having a much smaller and more imbalanced dataset—highlights the impact of lesion clarity. The clear boundaries of the erosion/ulcer lesion may have enabled more effective feature extraction, compensating for the data shortage. These variations in performance, which correspond to morphological complexity, support the view that the models are identifying specific lesion features rather than merely detecting background artefacts.

Regarding the model selection, the use of EfficientNet was prioritised over classical machine learning approaches that require manual outlining. The implementation of transfer learning using a model pretrained on the large-scale ImageNet dataset mitigated the risks associated with small datasets. By utilising image-level classification rather than relying on manual segmentation, we minimised subjective human bias inherent in drawing artificial lesion borders. This enabled the model to autonomously learn complex morphological features directly from the raw pixel data, contributing to the robust performance.

The models developed in this study focused on lesion-level identification of individual skin lesions and are intended to serve as complementary tools to disease-specific diagnostic models. By identifying lesions that are common across various dermatological conditions, the lesion-specific models may achieve greater generalisability. These models show the potential to be utilised in daily clinical practice as supportive diagnostic tools, particularly in primary-care clinics.

The selection of lesion types in this study was based on their clinical significance and applicability. Erythema, lichenification, alopecia and excoriation are prevalent lesions in various dermatological conditions. These lesions are also included in CADESI-04 and CADLI, which serve as core instruments in standardised outcome sets for clinical trials in dogs [27]. However, excoriation—defined as an erosion or ulcer caused by self-trauma—often requires behavioural context for accurate identification [10]. Therefore, erosion/ulcer was selected instead, as it offers a clearer definition based solely on images.

Constructing datasets with diverse breeds is essential for ensuring the reliability of AI models in real-world clinical settings owing to breed-specific characteristics such as coat colour, skin colour and hair length. In human dermatology, several studies have reported differences in AI performance across skin tones [28, 29]. They revealed the significantly lower diagnostic accuracy on darker skin compared to lighter skin. These differences have been attributed to the predominance of lighter skin images

and lack of darker skin images in datasets, resulting in limited generalisability [30, 31]. Likewise, in canine dermatology, a lack of breed diversity in datasets may lead to biased models that perform well only within certain breeds. In this study, although complete breed information was not available for all training images, the training dataset included ≥42 distinct breeds. The models exhibited stable performance when evaluated on a test dataset composed of images from 40 different breeds. These findings suggest the potential applicability of the models across diverse breeds in clinical practice. However, it should be noted that all images were collected in Korea, where small-breed dogs are predominant. Consequently, application of the models to regions with different breed distributions, such as countries where large-breed dogs are more prevalent, may potentially result in underperformance.

Labelling of the training dataset and the evaluation of the test dataset in this study were exclusively conducted by veterinary surgeons specialised in dermatology, ensuring high reliability. In previous veterinary studies on AI-based lesion detection, the labelling process was often insufficiently described, with unclear information on whether trained experts, such as veterinary surgeons, were involved [20, 21, 32]. By ensuring expert involvement throughout both the training and test phases, this study minimised potential labelling errors and improved dataset consistency. Furthermore, to minimise classification bias, the labelling panel included both veterinary surgeons who examined the patients and independent evaluators. Regarding dataset characteristics, images collected at a veterinary teaching hospital could feature characteristic lesions, which may differ from cases encountered in primary care. However, establishing a robust baseline with clear images is optimal for CNNs to accurately learn the morphological features of each lesion.

The application of AI in veterinary medicine has been rapidly expanding. A recent study using a large language model (ChatGPT-4.5) demonstrated diagnostic accuracy comparable to veterinary ophthalmologists in evaluating feline ocular diseases [33]. While its approach differs from our CNN-based model in terms of processing speed, reproducibility and consistency, this finding highlights the growing importance of AI in veterinary clinical practice.

In our study, the models were developed to classify the presence or absence of lesions only, without the ability to assess their severity. Accordingly, this study should be regarded as an initial step toward automated lesion assessment rather than a comprehensive clinical tool. Future studies may extend this framework to incorporate evaluation of lesion severity, which could enable the assessment of lesion progression and ultimately support automated scoring systems such as CADESI-04 and CADLI. Additionally, developing models for other lesion types—such as papule, pustule, crust, epidermal collarette and scale—could further support diagnostic approaches for a broader range of skin diseases.

This study has some limitations. First, subanalysis of model performance based on specific phenotypic variables, such as coat colour and hair type, was not conducted. Although the models were developed using datasets with breed diversity, further studies are needed to determine whether model performance differs across these specific characteristics as well as different breeds. Second, owing to the retrospective data, pseudo-alopecic conditions (e.g., skin folds, clipped areas) and environmental variables (e.g., fluffy hair, crusts) were included. Manual review of the error cases confirmed these as primary causes of model errors, yet their exact proportions were not quantified. Additionally, imaging conditions—such as lighting, focus and camera quality—were not standardised, which may introduce bias in real-world practice. These failures highlight the inherent 'black-box' nature of deep learning, which makes it difficult to explain the model's exact diagnostic logic. Future research should explore providing exact confidence scores to quantify the model's reliability for each prediction and identifying the specific features that contribute to prediction uncertainty. Another limitation is the lack of cross-testing across different lesion types. Consequently, it remains unclear whether the models truly identify specific features or merely recognise a general difference from normal skin. Finally, further validation on mixed-lesion datasets is required to ensure that each model correctly identifies its target without interference from other co-existing lesions.

5 | Conclusions

This study confirmed the validity of CNN-based AI models for the automated identification of common canine skin lesion types: erythema, lichenification, alopecia and erosion/ulcer. Four AI models were independently trained and validated, exhibiting reliable performance across multiple evaluation metrics. The application of the models could enable rapid and accurate dermatologic examination, serving as supportive tools in clinical diagnosis and lesion monitoring throughout treatment.

Author Contributions

Soh-Yoon Kang: conceptualization, data curation, investigation, methodology, writing – original draft, writing – review and editing. **Yeong-Hun Kang:** data curation, investigation, writing – review and editing. **Hyun-Jung Kim:** conceptualization, methodology, project administration. **Eun-A. Huh:** conceptualization. **Jong-Soo Jeon:** formal analysis, software. **Hyun-Su Kim:** software. **Min-Sun Kim:** investigation. **Anisha Tiwari:** investigation. **Cheol-Yong Hwang:** conceptualization, project administration, supervision, writing – review and editing.

Acknowledgements

The authors thank the owners of the dogs that participated in this study and the clinical staff at Seoul National University Veterinary Medical Teaching Hospital for their assistance with image collection.

Funding

This study was supported by AIFORPET Corp.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Processed data or an evaluation script may be available from the corresponding author upon reasonable request and with company approval. In addition, processed datasets can be accessed within a secure data

enclave under controlled conditions, subject to privacy, ethical and proprietary restrictions.

References

1. M. Dave and N. Patel, "Artificial Intelligence in Healthcare and Education," *British Dental Journal* 234 (2023): 761–764.

2. V. V. Pai and R. B. Pai, "Artificial Intelligence in Dermatology and Healthcare: An Overview," *Indian Journal of Dermatology, Venereology and Leprology* 87 (2021): 457–467.

3. J. Y. Kim, H. E. Lee, Y. H. Choi, S. J. Lee, and J. S. Jeon, "CNN-Based Diagnosis Models for Canine Ulcerative Keratitis," *Scientific Reports* 9 (2019): 14209.

4. P. S. Basran and R. B. Appleby, "The Unmet Potential of Artificial Intelligence in Veterinary Medicine," *American Journal of Veterinary Research* 83 (2022): 385–392.

5. M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, CA, USA: 2019; 6105–6114.

6. K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass Skin Cancer Classification Using EfficientNets–A First Step Towards Preventing Skin Cancer," *Neuroscience Informatics* 2 (2022): 100034.

7. H. A. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul, and J.-M. Kang, "A Robust Approach for Brain Tumor Detection in Magnetic Resonance Images Using Finetuned Efficientnet," *IEEE Access* 10 (2022): 65426–65438.

8. E. Kim, C. Choe, J. G. Yoo, et al., "Major Medical Causes by Breed and Life Stage for Dogs Presented at Veterinary Clinics in the Republic of Korea: A Survey of Electronic Medical Records," *PeerJ* 6 (2018): e5161.

9. D. G. O'Neill, H. James, D. C. Brodbelt, D. B. Church, and C. Pegram, "Prevalence of Commonly Diagnosed Disorders in UK Dogs Under Primary Veterinary Care: Results and Applications," *BMC Veterinary Research* 17 (2021): 69.

10. W. H. Miller, C. E. Griffin, and K. L. Campbell, *Muller and Kirk's Small Animal Dermatology*, 7th ed. (Elsevier Mosby, 2013), 66–78.

11. P. B. Hill, A. Lo, C. A. N. Eden, et al., "Survey of the Prevalence, Diagnosis and Treatment of Dermatological Conditions in Small Animals in General Practice," *Veterinary Record* 158 (2006): 533–539.

12. K. Ural, H. Voyvoda, B. Ulutaş, S. Paşa, N. Aysul, and M. Gültekin, "Understanding Primary and Secondary Skin Lesions Among Infectious Dermatoses in Dogs: Lessons We Learned From Cases," *Animal Health and Production Hygiene* 1 (2012): 86–99.

13. M. M. Welle, "Canine Noninflammatory Alopecia: An Approach to Its Classification and a Diagnostic Aid," *Veterinary Pathology* 60 (2023): 748–769.

14. T. Olivry, M. Saridomichelakis, T. Nuttall, E. Bensignor, C. E. Griffin, and P. B. Hill, "Validation of the Canine Atopic Dermatitis Extent and Severity Index (CADESI)-4, a Simplified Severity Scale for Assessing Skin Lesions of Atopic Dermatitis in Dogs," *Veterinary Dermatology* 25 (2014): 77–85, e25.

15. J. D. Plant, K. Gortel, M. Kovalik, N. L. Polissar, and M. B. Neradilek, "Development and Validation of the Canine Atopic Dermatitis Lesion Index, a Scale for the Rapid Scoring of Lesion Severity in Canine Atopic Dermatitis," *Veterinary Dermatology* 23 (2012): 515-e103.

16. T.-C. Pham, C.-M. Luong, V.-D. Hoang, and A. Doucet, "AI Outperformed Every Dermatologist in Dermoscopic Melanoma Diagnosis, Using an Optimized Deep-CNN Architecture With Custom Mini-Batch Logic and Loss Function," *Scientific Reports* 11 (2021): 17485.

17. V. Venugopal, N. I. Raj, M. K. Nath, and N. Stephen, "A Deep Neural Network Using Modified EfficientNet for Skin Cancer Detection in Dermoscopic Images," *Decision Analytics Journal* 8 (2023): 100278.

18. S. Zhao, B. Xie, Y. Li, et al., "Smart Identification of Psoriasis by Images Using Convolutional Neural Networks: A Case Study in China," *Journal of the European Academy of Dermatology and Venereology* 34 (2020): 518–524.

19. Q. T. Huynh, P. H. Nguyen, H. X. Le, et al., "Automatic Acne Object Detection and Acne Severity Grading Using Smartphone Images and Artificial Intelligence," *Diagnostics* 12 (2022): 1879.

20. A. Upadhyay, G. Singh, S. Mhatre, and P. Nadar, "Dog Skin Diseases Detection and Identification Using Convolutional Neural Networks," *SN Computer Science* 4 (2023): 250.

21. A. Smith, P. W. Carroll, S. Aravamuthan, et al., "Computer Vision Model for the Detection of Canine Pododermatitis and Neoplasia of the Paw," *Veterinary Dermatology* 35 (2024): 138–147.

22. N. Apostolopoulos, S. Murray, S. Aravamuthan, and D. Döpfer, "Detection of Canine External Ear Canal Lesions Using Artificial Intelligence," *Veterinary Dermatology* 36 (2025): 424–432.

23. S. Hwang, H. K. Shin, J. M. Park, B. Kwon, and M.-G. Kang, "Classification of Dog Skin Diseases Using Deep Learning With Images Captured From Multispectral Imaging Device," *Molecular & Cellular Toxicology* 18 (2022): 299–309.

24. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification With Deep Convolutional Neural Networks," *Communications of the ACM* 60 (2017): 84–90.

25. J. Zuluaga-Gomez, Z. Al Masry, K. Benaggoune, S. Meraghni, and N. Zerhouni, "A CNN-Based Methodology for Breast Cancer Diagnosis Using Thermal Images," *Computational Methods in Biomechanics and Biomedical Engineering. Imaging & Visualization* 9 (2021): 131–145.

26. W. Zheng and M. Jin, "The Effects of Class Imbalance and Training Data Size on Classifier Learning: An Empirical Study," *SN Computer Science* 1 (2020): 71.

27. T. Olivry, E. Bensignor, C. Favrot, et al., "Development of a Core Outcome Set for Therapeutic Clinical Trials Enrolling Dogs With Atopic Dermatitis (COSCAD'18)," *BMC Veterinary Research* 14 (2018): 238.

28. R. Daneshjou, K. Vodrahalli, R. A. Novoa, et al., "Disparities in Dermatology AI Performance on a Diverse, Curated Clinical Image Set," *Science Advances* 8 (2022): eabq6147.

29. M. Groh, O. Badri, R. Daneshjou, et al., "Deep Learning-Aided Decision Support for Diagnosis of Skin Disease Across Skin Tones," *Nature Medicine* 30 (2024): 573–583.

30. H. K. Jeong, C. Park, R. Henao, and M. Kheterpal, "Deep Learning in Dermatology: A Systematic Review of Current Approaches, Outcomes, and Limitations," *JID Innovations* 3 (2023): 100150.

31. K. P. Venkatesh, M. M. Raza, G. Nickel, S. Wang, and J. C. Kvedar, "Deep Learning Models Across the Range of Skin Disease," *NPJ Digital Medicine* 7 (2024): 32.

32. Y. Yashu, V. Kukreja, P. Srivastava, A. Garg, and S. Hariharan, "Paws and Pixels: Enhancing Canine Dermatology Through Image Analysis," In *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*, Mysore, India: 2024; 1–6.

33. S. Okur, E. Modoğlu, B. Baykal, et al., "Comparison of Diagnostic Performance Between Large Language Models and Veterinary Evaluators in Feline Ocular Diseases Based on Clinical Summaries and Anterior Segment Photographs," *Veterinary Ophthalmology* 29 (2026): e70052.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Breed composition of the training dataset. **Table S2:** Breed composition of the test dataset. **Table S3:** Confusion matrix values for each CNN model on the test dataset.