Advanced Skin Disease Classification in Dogs and Cats Using an Integrated CNN and Vision Transformer Model

Park John
Handong Global University (Pohang, South Korea)

Abstract:

Accurate classification of skin diseases in pets, particularly dogs and cats, is crucial for early detection and effective treatment. This paper introduces a hybrid CNNViT model designed to improve the classification of skin diseases in dogs and cats. Conventional Convolutional Neural Networks (CNNs) are excellent at capturing local texture features, while Vision Transformers are effective at extracting high-level global dependencies. To leverage the strengths of both architectures, the proposed model integrates CNN for feature extraction and Linformer-ViT for efficient attention mechanisms, reducing computational complexity compared to standard ViT. The proposed model was trained and evaluated on a comprehensive dataset of 200,000 skin disease images covering 13 different types, achieving an accuracy of 95.43%, precision of 95.00%, recall of 95.14%, and F1-score of 94.86%. The integration of Convolutional Neural Networks and Linformer-based Vision Transformer architectures, combined with data augmentation techniques, proved highly effective in improving the model's ability to accurately classify skin diseases in dogs and cats, making it a promising tool for veterinary applications.