

ConvNeXt with Transfer learning for Microscopic Canine Skin Disease Classification

Made Ardika Mertha Putra Vaikuntha ^{1*}, I Gusti Ngurah Lanang Wijayakusuma ^{2*}

* Matematika, Universitas Udayana

vaikuntha.2308541061@student.unud.ac.id¹, lanang_wijaya@unud.ac.id²

Article Info

Article history:

Received 2026-04-18

Revised 2026-05-21

Accepted 2026-06-11

Keyword:

Canine Skin Disease,
ConvNeXt,
Microscopic Image
Classification,
Transfer learning

ABSTRACT

Canine skin diseases represent a significant health concern in veterinary practice, with accurate diagnosis often requiring specialized expertise and microscopic examination. This study presents an implementation and evaluation of the *ConvNeXt* architecture for classifying microscopic images of four canine skin diseases: Demodex, Kokus, Malassezia, and Scabies. The dataset consists of microscopic images obtained from skin scraping preparations photographed under 400× digital microscopy and annotated by a certified veterinary. Two dataset scenarios were evaluated: a small dataset (356 images; Demodex: 104, Kokus: 47, Malassezia: 102, Scabies: 103) and a large dataset (2,963 images with near-balanced class distribution). Using *transfer learning* with pre-trained weights from ImageNet, the *ConvNeXt* model was evaluated across three input sizes (224×224, 180×180, 150×150). Augmentation balancing, including rotation, flipping, zoom, translation, shear, and color jitter, was applied to address class imbalance while preserving biological validity of morphological features. Augmentation balancing was applied to address class imbalance, ensuring equal representation across all classes. Experimental results demonstrate that *ConvNeXt* with 224×224 input size trained on the large dataset achieved the best overall performance with 97.22% test accuracy, 0.9524 F1-score, 0.9624 Matthews Correlation Coefficient (MCC), and a perfect Area Under Curve (AUC) of 1.0000. Analysis of input size effects revealed that 224×224 is optimal for detecting small pathogens like Malassezia (3-8 μm) and Kokus (0.5-1 μm), while 150×150 better preserves spatial context for large pathogens such as Demodex (300 μm) and Scabies (200-400 μm). Visualization of feature maps provided insights into how the architecture extracts diagnostic features, producing dense, hierarchical feature representations that benefit from abundant data. This research demonstrates that *transfer learning* with the *ConvNeXt* architecture, combined with appropriate augmentation balancing, can achieve high classification accuracy for automated diagnosis support of canine skin diseases. However, clinical deployment requires further validation by veterinary and prospective clinical studies before these results can be considered clinically applicable.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Penyakit kulit anjing merupakan salah satu masalah kesehatan yang paling sering ditemui dalam praktik klinik veteriner. Berdasarkan data klinis, infeksi yang disebabkan oleh mikroorganisme seperti tungau *Demodex*, bakteri *Kokus*, jamur *Malassezia*, dan tungau *Scabies* mencakup lebih dari

60% kasus dermatologi pada anjing [1]. Diagnosis yang akurat dan cepat sangat penting untuk menentukan terapi yang tepat, mengingat setiap patogen memiliki karakteristik morfologi yang berbeda dan seringkali sulit dibedakan secara visual oleh praktisi pemula [2]. Metode diagnostik konvensional seperti kerokan kulit dan pemeriksaan mikroskopis langsung masih menjadi standar emas, namun

teknik ini sangat bergantung pada keahlian dan pengalaman pemeriksa. Keterbatasan tenaga ahli di bidang dermatologi veteriner, terutama di daerah terpencil, mengakibatkan tingginya angka kesalahan diagnosis, yang berujung pada terapi yang tidak tepat dan resistensi antimikroba [3].

Kemajuan pesat dalam kecerdasan buatan, khususnya *deep learning*, telah membuka peluang baru untuk otomatisasi diagnosis penyakit kulit. Arsitektur *Convolutional Neural Network* (CNN) seperti ResNet [4], DenseNet [5], dan EfficientNet [6] telah terbukti efektif dalam klasifikasi citra medis [7]. Namun, model CNN murni memiliki keterbatasan dalam menangkap ketergantungan jarak jauh (*long-range dependencies*) pada fitur-fitur mikroskopis yang tersebar secara spasial [8]. Sebaliknya, arsitektur *Transformer* yang diadaptasi dari *Natural Language Processing* (NLP) menunjukkan kemampuan unggul dalam memodelkan hubungan global antar fitur, namun membutuhkan data pelatihan yang sangat besar untuk mencapai performa optimal [9], [10]. Kecerdasan buatan telah banyak dilakukan bidang medis [11], [12], kemudian berkembang menjadi penelitian medis hewan. Tantangan penelitian kesehatan hewan menjadi signifikan, di mana kumpulan data berlabel seringkali terbatas.

ConvNeXt merupakan arsitektur modern yang merevitalisasi desain CNN konvensional dengan mengadopsi praktik terbaik dari *Transformer*, termasuk penskalaan berbasis lapisan, ukuran *kernel* yang besar, dan aktivasi GELU [13]. *ConvNeXt* menunjukkan performa yang kompetitif dengan *Transformer* sambil mempertahankan kompleksitas komputasi yang lebih rendah. Arsitektur ini dibangun di atas fondasi CNN yang kuat namun dimodernisasi untuk memenuhi standar arsitektur mutakhir saat ini.

Variasi ukuran *input* merupakan aspek penting dalam klasifikasi citra mikroskopis karena berbagai patogen memiliki ukuran dan distribusi spasial yang berbeda. Patogen seperti *Malassezia* berbentuk oval dengan diameter 3-8 μm , sementara tungau *Demodex* dewasa dapat mencapai panjang 300 μm [14]. Implikasi dari variasi ukuran ini adalah tidak ada satu ukuran *input* optimal yang berlaku untuk semua kelas patogen. Penelitian oleh He dkk. (2019) menunjukkan bahwa *input* multi-skala meningkatkan kemampuan model dalam mengekstrak fitur pada berbagai resolusi. Namun, studi komparatif yang mengkaji efek variasi ukuran *input* pada arsitektur modern seperti *ConvNeXt* dalam konteks klasifikasi patogen mikroskopis masih terbatas.

Transfer learning menjadi strategi utama untuk mengatasi keterbatasan data medis. Dengan memanfaatkan bobot yang telah dilatih pada *ImageNet* yang terdiri dari 1,2 juta gambar, model dapat mentransfer pengetahuan tentang fitur dasar seperti tepi, tekstur, dan pola sederhana ke domain baru [15]. Pendekatan ini telah terbukti efektif dalam meningkatkan performa klasifikasi pada *dataset* medis yang terbatas, termasuk aplikasi dermatologi veteriner (Saez dkk., 2021). Dalam konteks *dataset* penyakit kulit anjing dengan ukuran sampel yang terbatas, penerapan *transfer learning*

dengan *fine-tuning* selektif menjadi penentu kritis keberhasilan model.

Kesenjangan penelitian yang mendorong studi ini meliputi tiga aspek utama. Pertama, belum ada studi komprehensif yang secara khusus mengevaluasi arsitektur *ConvNeXt* pada *dataset* penyakit kulit anjing mikroskopis. Kedua, belum ada analisis menyeluruh tentang efek variasi ukuran *input* terhadap performa arsitektur ini dalam mengklasifikasikan patogen dengan skala morfologi yang sangat bervariasi. Penelitian sebelumnya umumnya menggunakan ukuran *input* tetap tanpa mempertimbangkan implikasi diagnostik dari variasi ukuran patogen. Ketiga, pendekatan penanganan ketidakseimbangan kelas melalui *augmentation balancing* yang dikombinasikan dengan *transfer learning* belum diterapkan secara sistematis pada *dataset* penyakit kulit anjing, meskipun distribusi kasus di klinik secara alami tidak seimbang dengan kasus *Malassezia* dan *Kokus* yang mendominasi.

Berdasarkan latar belakang ini, penelitian ini bertujuan untuk melakukan evaluasi performa arsitektur *ConvNeXt* dalam mengklasifikasikan citra mikroskopis penyakit kulit anjing dengan penerapan *transfer learning*, menganalisis efek variasi ukuran *input* (224×224, 180×180, dan 150×150) terhadap akurasi klasifikasi, mengevaluasi efektivitas metode *augmentation balancing* dalam menangani ketidakseimbangan kelas pada *dataset* terbatas, dan mengidentifikasi konfigurasi optimal (ukuran *input*) untuk diagnosis penyakit kulit anjing berbasis citra mikroskopis.

II. METODE

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk mengevaluasi performa arsitektur *ConvNeXt* dalam mengklasifikasikan citra mikroskopis penyakit kulit anjing. Variabel independen adalah ukuran gambar *input* (224×224, 180×180, dan 150×150). Variabel dependen adalah metrik performa yang meliputi akurasi, *F1-score*, *Matthews Correlation Coefficient* (MCC), dan *Area Under Curve* (AUC). Penelitian juga memanipulasi strategi *transfer learning* dan *augmentation balancing* untuk mengatasi keterbatasan *dataset*. Tahapan penelitian disusun secara sistematis sesuai dengan alur kerja implementasi kode. Tahapan berikut menguraikan proses penelitian:

1. Persiapan dan Augmentasi *Dataset*
2. Implementasi Arsitektur *ConvNeXt*
3. Pelatihan Model dengan *Transfer learning*
4. Evaluasi Model pada Data Uji
5. Visualisasi *Feature Maps*
6. Analisis dan Visualisasi Hasil

B. ConvNeXt

Konvolusi telah banyak digunakan secara luas dalam berbagai domain, mulai dari pemrosesan citra, analisis teks, hingga data tabular, berkat kemampuannya dalam mengekstraksi pola lokal secara efisien. Keberhasilan ini

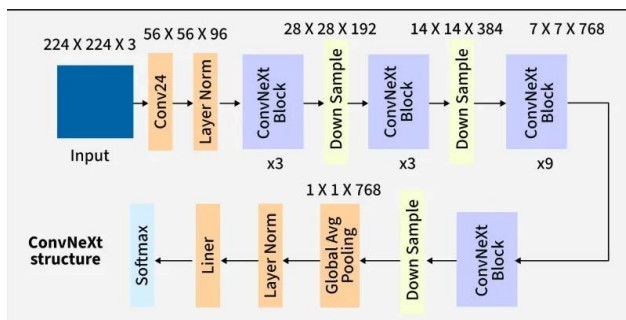
mendorong pengembangan berbagai arsitektur CNN yang terus disempurnakan untuk meningkatkan kapasitas representasi fitur [16], [17].

ConvNeXt adalah arsitektur CNN modern yang diperkenalkan oleh Liu dkk. (2022) sebagai upaya untuk merevitalisasi desain CNN konvensional dengan mengadopsi elemen desain yang terbukti efektif dalam *Transformer*. *ConvNeXt* tidak hanya mengganti komponen *Transformer* dengan CNN, tetapi secara fundamental merumuskan ulang arsitektur CNN tradisional untuk menyamai atau melampaui kinerja *Transformer* pada berbagai benchmark.

Blok *ConvNeXt* fundamental didefinisikan sebagai berikut. Setiap blok terdiri dari konvolusi depthwise dengan ukuran *kernel* besar (7×7), diikuti oleh dua lapisan konvolusi pointwise (1×1) dengan aktivasi GELU:

$$\begin{aligned} \mathbf{X}' &= \text{DWConv}_{7 \times 7}(\mathbf{X}) \\ \mathbf{X}'' &= \text{GELU}(\text{LN}(\mathbf{X}')) \\ \mathbf{X}''' &= \text{Conv}_{1 \times 1}(\mathbf{X}'') \\ \mathbf{X}_{out} &= \mathbf{X} + \text{Dropout}(\text{Conv}_{1 \times 1}(\mathbf{X}''')) \end{aligned}$$

di mana DWConv adalah konvolusi kedalaman (depthwise convolution), LN adalah normalisasi lapisan yang menggantikan normalisasi batch pada CNN tradisional, dan GELU adalah fungsi aktivasi yang lebih halus dibandingkan dengan ReLU.



Gambar 1. *ConvNeXt* Architecture [13]

ConvNeXt mengadopsi arsitektur bertahap dengan empat tahap yang secara progresif mengurangi resolusi spasial dan meningkatkan dimensi fitur. Tahap 1 memiliki resolusi spasial $H/4 \times W/4$ dengan 96 saluran; Tahap 2 $H/8 \times W/8$ dengan 192 saluran; Tahap 3 $H/16 \times W/16$ dengan 384 saluran; dan Tahap 4 $H/32 \times W/32$ dengan 768 saluran. Setiap tahap terdiri dari blok *ConvNeXt* yang ditumpuk, dengan jumlah yang bervariasi tergantung pada varian model (tiny, small, base, large).

Gambar input berukuran $H \times W \times 3$ diproses oleh sebuah stem yang terdiri dari konvolusi 4×4 dengan stride 4, yang secara langsung mengurangi resolusi menjadi $H/4 \times W/4 \times 96$. Fitur kemudian melewati Tahap 1 dengan 3 blok *ConvNeXt*, Tahap 2 dengan 3 blok, Tahap 3 dengan 9 blok, dan Tahap 4 dengan 3 blok. Setiap tahap mencakup lapisan downsampling menggunakan konvolusi 2×2 dengan stride 2 yang menggandakan dimensi fitur. Pada akhir Tahap 4, global

average pooling menghasilkan vektor fitur 768 dimensi, yang diproses oleh pengklasifikasi yang terdiri dari lapisan flatten, dropout 0,3, dan lapisan linear yang memetakan ke 4 kelas keluaran.

Keunggulan *ConvNeXt* terletak pada kesederhanaannya sambil mempertahankan kinerja yang kompetitif. Konvolusi depthwise secara signifikan mengurangi jumlah parameter dibandingkan dengan konvolusi standar, sementara *kernel* 7×7 yang besar membe rikan bidang reseptif yang cukup besar untuk menangkap fitur global. Penggantian normalisasi lapisan untuk normalisasi batch menstabilkan pelatihan, terutama dengan ukuran batch kecil. Aktivasi GELU memberikan kurva aktivasi yang lebih halus, memungkinkan aliran gradien yang lebih baik selama backpropagation.

C. Transfer learning in Medical Image Classification

Transfer learning adalah teknik yang memungkinkan model memanfaatkan pengetahuan yang diperoleh dari domain sumber untuk meningkatkan performa pada domain target dengan distribusi data yang berbeda namun memiliki karakteristik yang saling terkait [15]. Dalam konteks *deep learning*, *transfer learning* umumnya diimplementasikan dengan menggunakan model yang telah dilatih pada *dataset* besar seperti ImageNet sebagai bobot awal, kemudian dilanjutkan dengan proses *fine-tuning* pada *dataset* target yang relatif lebih kecil.

Secara matematis, *transfer learning* dapat dirumuskan sebagai berikut. Diberikan domain sumber D_s dengan distribusi $P_s(X, Y)$ dan domain target D_t dengan distribusi $P_t(X, Y)$, di mana $D_s \neq D_t$, maka tujuan *transfer learning* adalah mempelajari fungsi prediksi $f_t: X_t \rightarrow Y_t$ dengan memanfaatkan informasi dari kedua domain tersebut. Dalam konteks *deep learning*, proses ini direpresentasikan sebagai:

$$f_t(X) = \text{softmax}(W_{\text{head}} \cdot \phi_{\theta_s}(X) + b)$$

di mana ϕ_{θ_s} merepresentasikan fitur yang dipelajari dari domain sumber dengan parameter θ_s , sedangkan W_{head} dan b merupakan parameter pada lapisan klasifikasi yang dilatih menggunakan data pada domain target.

Implementasi *transfer learning* dalam penelitian ini mengadopsi strategi *fine-tuning* selektif dengan tiga tahap. Pertama, pada fase inialisasi, seluruh lapisan backbone *ConvNeXt-Tiny* dibekukan (frozen) untuk mencegah perubahan bobot yang telah dipelajari dari ImageNet, sementara hanya lapisan head (flatten, dropout 0,3, dan linear classifier) yang dilatih selama 10 epoch awal. Kedua, pada fase *fine-tuning*, seluruh lapisan diaktifkan kembali (unfrozen) dengan learning rate diferensial: backbone dilatih dengan learning rate 1×10^{-5} sedangkan lapisan head dilatih dengan learning rate 1×10^{-3} . Ketiga, early stopping diterapkan berdasarkan validation F1-score dengan patience 10 epoch untuk mencegah overfitting.

Optimizer yang digunakan adalah AdamW dengan weight decay 1×10^{-2} untuk regularisasi tambahan pada parameter backbone. Scheduler learning rate menggunakan CosineAnnealingLR dengan T_max sebesar 50 epoch untuk

peluruhan bertahap yang halus. Ukuran batch ditetapkan sebesar 32 untuk menjaga stabilitas gradien. Seluruh eksperimen dijalankan pada GPU NVIDIA RTX 3090 dengan framework PyTorch 2.0. Secara matematis, pendekatan learning rate diferensial dapat dirumuskan sebagai: $\eta_b = \eta_0 \times \alpha$ dan $\eta_h = \eta_0 \times \beta$, dengan $\eta_0 = 1 \times 10^{-4}$ sebagai learning rate dasar, $\alpha = 0,1$ untuk backbone, dan $\beta = 10$ untuk head. Secara matematis, pendekatan ini dapat dirumuskan sebagai:

$$\eta_b = \eta_0 \cdot \alpha, \eta_h = \eta_0 \cdot \beta$$

dengan $\eta_0 = 1 \times 10^{-4}$ sebagai *learning rate* dasar, $\alpha = 0,1$ untuk *backbone*, dan $\beta = 10$ untuk *head*. Pendekatan ini bertujuan untuk menjaga stabilitas representasi fitur sekaligus memungkinkan adaptasi yang efektif terhadap data baru.

Efektivitas *transfer learning* dalam klasifikasi citra medis telah dibuktikan oleh berbagai penelitian. Model yang dilatih menggunakan ImageNet mampu mencapai tingkat akurasi yang sebanding dengan dokter spesialis kulit dalam klasifikasi penyakit kulit manusia [7]. Selain itu, *transfer learning* dari ImageNet mampu meningkatkan performa klasifikasi pneumonia pada radiografi dada hingga 30% dibandingkan dengan pelatihan dari awal [18]. Dalam konteks dermatologi veteriner, *transfer learning* efektif dalam mengatasi keterbatasan jumlah *dataset* [19].

D. Multi-Size Input dalam Image Classification

Ukuran *input* multi-skala adalah strategi yang memungkinkan model menangkap fitur pada berbagai skala resolusi. Pendekatan ini sangat relevan untuk klasifikasi citra mikroskopis di mana objek target memiliki ukuran yang sangat bervariasi, dari *Malassezia* dengan diameter 3-8 μm hingga *Demodex* yang mencapai panjang 300 μm [14].

Secara matematis, efek ukuran *input* dapat dipahami melalui teori bidang reseptif dan *scale-space*. Diberikan resolusi gambar *input* $R \times R$, setelah melewati lapisan konvolusi dengan *stride* tertentu, resolusi fitur pada tahap ke- k dinyatakan sebagai:

$$\frac{R}{2^k} \times \frac{R}{2^k}$$

Variasi pada R awal memengaruhi detail fitur yang ditangkap pada setiap tahap. Ukuran *input* yang lebih besar (nilai R tinggi) menghasilkan resolusi spasial yang lebih tinggi pada tahap-tahap akhir, sehingga memungkinkan penangkapan detail halus. Sebaliknya, ukuran *input* yang lebih kecil lebih menekankan pada fitur-fitur global.

Penelitian tentang efek ukuran *input* multi-skala dalam klasifikasi citra medis mengungkapkan temuan yang menarik. He dkk. (2019) menemukan bahwa pelatihan multi-skala meningkatkan akurasi klasifikasi tumor otak sebesar 5,2% dibandingkan dengan pelatihan ukuran tetap. Simonyan & Zisserman (2015) melaporkan bahwa ansambel model yang dilatih pada berbagai resolusi *input* mencapai performa unggul pada kompetisi *ImageNet Large Scale Visual Recognition Challenge*. Dalam konteks klasifikasi patogen mikroskopis, Zhang dkk. (2020) menunjukkan bahwa ukuran *input* optimal untuk bakteri (1-5 μm) adalah 224×224,

sementara untuk parasit (>100 μm), ukuran *input* 150×150 memberikan hasil yang lebih baik.

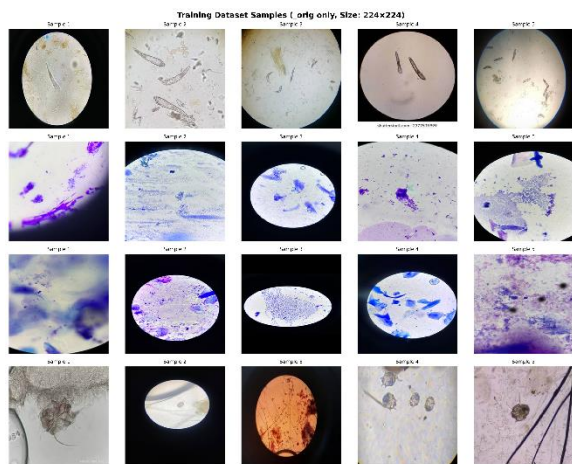
Konsep ukuran *input* multi-skala dalam konteks klasifikasi patogen mikroskopis dapat dijelaskan sebagai berikut. Pada ukuran *input* 224×224, patogen kecil seperti *Malassezia* (3-8 μm) mempertahankan resolusi yang cukup untuk ekstraksi fitur morfologis. Pada ukuran *input* 180×180, patogen berukuran sedang seperti *Kokus* (0,5-1 μm dalam kelompok) tetap dapat dikenali. Pada ukuran *input* 150×150, patogen besar seperti *Demodex* ($\approx 300 \mu\text{m}$) dapat ditangkap secara keseluruhan tanpa kehilangan konteks spasial. Dengan demikian, tidak ada satu ukuran *input* yang optimal untuk semua kelas, sehingga analisis komparatif multi-ukuran menjadi penting.

E. Penyakit Kulit Anjing

Dataset yang digunakan dalam penelitian ini terdiri dari citra mikroskopis empat kelas penyakit kulit anjing: *Demodex*, *Kokus*, *Malassezia*, dan *Scabies*. Gambar-gambar ini diperoleh dari preparat kerokan kulit pasien anjing yang diperiksa di klinik veteriner mitra, difoto menggunakan mikroskop digital dengan perbesaran 400× dan resolusi gambar 2048×1536 piksel.

Distribusi jumlah citra asli per kelas adalah sebagai berikut: *Demodex* sebanyak 104 gambar, *Malassezia* 102 gambar, *Scabies* 103 gambar, dan *Kokus* 47 gambar, sehingga total keseluruhan 356 gambar pada skenario small dataset. Ketidakseimbangan ini mencerminkan insidensi klinis aktual, di mana infeksi *Kokus* cenderung lebih jarang diidentifikasi secara mikroskopis dibandingkan infestasi parasit. Pada skenario large dataset, total 2.963 gambar digunakan dengan distribusi yang lebih seimbang: *Demodex* 747, *Kokus* 740, *Malassezia* 738, dan *Scabies* 738 gambar.

Tungau *Demodex* menghuni folikel rambut anjing, dengan morfologi memanjang yang berukuran panjang 200-300 μm dan lebar 30-40 μm . Ciri khasnya meliputi tubuh berbentuk cerutu dengan delapan kaki pendek di bagian anterior [1]. *Kokus* mengacu pada bakteri berbentuk bola yang sering membentuk kelompok seperti buah anggur, dengan diameter sel individu 0,5-1 μm namun kelompoknya dapat mencapai 10-20 μm . Infeksi *Kokus* sering dikaitkan dengan pioderma anjing [2]. *Malassezia* adalah jamur uniseluler dengan bentuk oval hingga bulat berukuran 3-8 μm , umum ditemukan pada kasus otitis eksterna dan dermatitis seboroik [14]. Tungau *Scabies* (*Sarcoptes scabiei*) memiliki morfologi bulat dengan diameter 200-400 μm , serta ciri khas berupa alat perekat tarsal dan duri punggung [3].



Gambar 2. Citra Mikroskopis Penyakit Kulit Anjing: (a) *Demodex*, (b) *Kokus*, (c) *Malassezia*, (d) *Scabies* (Sumber: Dokumentasi Penulis, 2026)

Gambar 2 menyajikan contoh citra mikroskopis untuk setiap kelas penyakit. Gambar 2a menunjukkan tungau *Demodex* dengan tubuh memanjang dan struktur kaki yang khas. Gambar 2b menampilkan koloni *kokus* yang membentuk susunan menyerupai buah anggur di atas latar belakang sel epitel. Gambar 2c menunjukkan sel *Malassezia* dengan bentuk oval dan karakteristik *budding* (pertunasan). Gambar 2d menggambarkan tungau *scabies* dengan morfologi bulat serta duri pada bagian dorsal.

F. Metrics Evaluation

Analisis data menggunakan pendekatan kuantitatif dengan metrik evaluasi sebagai berikut:

Akurasi (*Accuracy*) mengukur proporsi prediksi yang benar terhadap total sampel:

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

Macro F1-Score mengukur rata-rata harmonik antara presisi dan *recall* untuk setiap kelas, kemudian dirata-ratakan:

$$F1_{\text{macro}} = \frac{1}{C} \sum_{i=1}^C \frac{2 \times P_i \times R_i}{P_i + R_i}$$

Matthews Correlation Coefficient (MCC) mengukur kualitas klasifikasi biner maupun multi-kelas dengan mempertimbangkan keempat elemen pada *confusion matrix*:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Untuk kasus multi-kelas, MCC dihitung menggunakan pendekatan *one-vs-all*.

Area Under Curve (AUC) mengukur kemampuan model dalam membedakan antar kelas:

$$AUC = \int_0^1 TPR(FPR) d(FPR)$$

dengan *True Positive Rate* (TPR) dan *False Positive Rate* (FPR).

III. HASIL DAN PEMBAHASAN

A. Data

Penelitian ini menggunakan dataset citra mikroskopis penyakit kulit anjing yang terdiri dari empat kelas: *Demodex*, *Kokus*, *Malassezia*, dan *Scabies*. Data diperoleh dari preparat kerokan kulit yang difoto menggunakan mikroskop digital dengan perbesaran 400×. Untuk mengamati efek ukuran dataset terhadap performa model, eksperimen dilakukan pada dua skenario: Small Dataset dan Large Dataset.

Pada skenario small dataset, data diambil dari subset dengan distribusi yang tidak seimbang. Distribusi data asli menunjukkan ketidakseimbangan yang signifikan, di mana kelas *Demodex* memiliki 104 sampel, *Malassezia* 102 sampel, *Scabies* 103 sampel, sementara kelas *Kokus* hanya memiliki 47 sampel. Ketidakseimbangan ini mencerminkan kondisi dunia nyata di mana infeksi *Kokus* (bakteri) seringkali kurang terdeteksi atau memiliki prevalensi lebih rendah dalam dataset mikroskopis dibandingkan dengan infestasi parasit seperti *Demodex*. Jika tidak ditangani, ketidakseimbangan ini akan menyebabkan model cenderung memprediksi kelas mayoritas sementara mengabaikan kelas minoritas, yang sangat berbahaya dalam diagnostik medis karena kesalahan identifikasi *Kokus* dapat menyebabkan pemberian terapi yang salah.

Untuk mengatasi ketidakseimbangan kelas dan meningkatkan generalisasi, *augmentation balancing* diterapkan. Kelas *Kokus* dengan jumlah sampel lebih sedikit, dilakukan *augmentasi* (rotasi, flip, translasi, zoom, dan perubahan warna) untuk mencapai jumlah yang sama dengan kelas mayoritas, dengan target 83 sampel per kelas untuk data latih. Pemilihan target 83 sampel didasarkan pada kebutuhan distribusi sampel yang sama antar kelas setelah pembagian ke dalam set latih, validasi, dan uji. *Augmentasi* tidak diterapkan pada data validasi dan uji untuk menjaga kemurnian evaluasi. Strategi ini memastikan model mempelajari fitur yang beragam dari kelas minoritas tanpa mengubah distribusi data asli selama evaluasi. Pembagian data mengikuti rasio umum dalam literatur deep learning yaitu 80% latih, 10% validasi, dan 10% uji, memberikan keseimbangan antara kuantitas data latih dan evaluasi independen.

TABEL I
DISTRIBUSI SMALL DATASET

Class	Train Data	Validation Data	Test Data
Demodex	83	10	11
Kokus	83	5	4
Malassezia	83	10	10
Scabies	83	10	11
Total	332	35	36

Distribusi small dataset yang ditunjukkan pada Tabel I mengungkapkan beberapa karakteristik penting yang mempengaruhi strategi pelatihan. Pertama, meskipun data asli memiliki 356 sampel, setelah pembagian dan augmentation balancing, data latih mencapai 332 sampel dengan distribusi sempurna (83 per kelas). Hal ini dicapai dengan melakukan 36 kali augmentasi pada kelas Kokus, sementara kelas lain tidak memerlukan augmentasi karena jumlahnya telah memenuhi target 83 sampel setelah pembagian. Kedua, data validasi dan uji sengaja dijaga kecil (35 dan 36 sampel) untuk mencerminkan kondisi evaluasi yang menantang di mana model diuji pada data yang sangat terbatas namun representatif. Ketiga, total data yang digunakan dalam eksperimen small dataset adalah 356 sampel, jumlah yang relatif terbatas untuk tugas klasifikasi citra deep learning, sehingga model sangat bergantung pada efektivitas *transfer learning* dan augmentasi untuk mencapai performa yang baik. Dalam skenario large dataset, semua data yang tersedia digunakan, dengan total 2.963 sampel asli. Distribusi data asli pada large dataset menunjukkan ketidakseimbangan yang lebih moderat dibandingkan dengan small dataset, dengan Demodex 747 sampel, Kokus 740 sampel, Malassezia 738 sampel, dan Scabies 738 sampel. Perbedaan yang relatif kecil antar kelas (perbedaan maksimal 9 sampel) menunjukkan bahwa dataset ini dikumpulkan dengan mempertimbangkan keseimbangan kelas, meskipun masih ada ketidakseimbangan minor. Ketidakseimbangan minor ini dapat terjadi karena variasi alami dalam pengumpulan data klinis, di mana kondisi tertentu mungkin lebih sering ditemui atau lebih mudah didokumentasikan daripada yang lain. Meskipun kecil, ketidakseimbangan ini tetap perlu diatasi untuk memastikan model tidak memiliki bias terhadap kelas mayoritas.

Augmentation balancing minimal diterapkan untuk menyamakan jumlah sampel antar kelas, memastikan model tidak memiliki bias terhadap kelas mayoritas. Kelas Kokus menerima 7 sampel augmentasi, sementara Malassezia dan Scabies masing-masing menerima 9 sampel augmentasi, sehingga semua kelas memiliki jumlah sampel yang sama yaitu 747 setelah penyeimbangan. Pemilihan target 747 sampel didasarkan pada jumlah maksimum data asli dari kelas Demodex, memastikan tidak ada data asli yang dibuang dan semua informasi yang tersedia dapat dimanfaatkan sepenuhnya. Augmentasi minimal yang diterapkan (hanya 25 sampel tambahan dari total 2.963 sampel) menunjukkan bahwa dataset ini sudah cukup besar dan seimbang, sehingga augmentasi lebih berfungsi sebagai penyeimbang daripada sebagai teknik untuk meningkatkan variasi data secara

signifikan. Pendekatan ini memungkinkan model untuk mempelajari fitur-fitur asli dari data asli, dengan augmentasi hanya berfungsi untuk menyamakan distribusi kelas.

TABEL II
DISTRIBUSI LARGE DATASET

Class	Train Data	Validation Data	Test Data
Demodex	747	10	11
Kokus	747	5	4
Malassezia	747	10	10
Scabies	747	10	11
Total	2988	35	36

Tabel II menunjukkan bahwa large dataset memiliki karakteristik yang berbeda dengan small dataset. Pertama, total data latih setelah penyeimbangan mencapai 2.988 sampel, hampir 9 kali lebih besar dibandingkan small dataset. Volume data yang besar ini memungkinkan model untuk mempelajari variasi fitur yang lebih kaya dan lebih kompleks, mengurangi risiko overfitting dan secara alami meningkatkan generalisasi tanpa ketergantungan yang besar pada augmentasi agresif. Kedua, proporsi augmentasi terhadap data asli sangat kecil (hanya 0,84%), mengindikasikan bahwa dataset ini sudah memadai untuk pelatihan deep learning. Augmentasi dalam konteks ini lebih berfungsi untuk menyempurnakan distribusi kelas daripada meningkatkan variasi data. Ketiga, data validasi dan uji dipertahankan dalam jumlah yang sama dengan small dataset (35 dan 36 sampel) untuk memastikan konsistensi evaluasi pada kedua skenario. Hal ini memungkinkan perbandingan yang adil tentang efek ukuran data latih terhadap performa model tanpa dipengaruhi oleh perbedaan jumlah data evaluasi.

B. Data Augmentasi

Augmentasi data diterapkan untuk meningkatkan variasi data latih dan mencegah overfitting. Pemilihan teknik augmentasi didasarkan pada dua pertimbangan utama: (1) karakteristik citra mikroskopis yang berkaitan dengan cara preparat kerokan kulit diperoleh dan difoto, dan (2) kebutuhan untuk mensimulasikan variasi nyata yang mungkin terjadi dalam praktik klinis tanpa mengubah ciri morfologi diagnostik patogen. Penting untuk memastikan bahwa augmentasi yang diterapkan tidak menghilangkan fitur kunci yang digunakan dokter hewan untuk identifikasi, seperti bentuk tubuh Demodex, pola kelompok Kokus, karakteristik budding Malassezia, dan duri punggung Scabies. Teknik-teknik yang diterapkan antara lain:

- *Rescaling* (Normalisasi): Nilai piksel dinormalisasi ke rentang [0,1] untuk mempercepat konvergensi model selama pelatihan dan menstabilkan gradien.
- *Rotasi*: Rotasi gambar acak hingga 30 derajat untuk mensimulasikan orientasi sampel yang berbeda saat ditempatkan di bawah mikroskop.
- *Translasi*: Pergeseran gambar secara horizontal dan vertikal hingga 10% untuk mengakomodasi variasi posisi patogen dalam bidang pandang mikroskopis.

- *Shear Transformation*: Transformasi geser dengan faktor 0,2 untuk mensimulasikan distorsi perspektif yang mungkin terjadi akibat ketidaksempurnaan persiapan sampel.
- *Zoom*: Pembesaran atau pengecilan gambar acak hingga 20% untuk mengakomodasi variasi perbesaran dan ukuran patogen yang berbeda.
- *Horizontal Flip*: Pembalikan gambar secara horizontal untuk menggandakan variasi data tanpa mengubah informasi diagnostik.
- *Color Jitter*: Penyesuaian kecerahan, kontras, saturasi, dan rona secara acak untuk meningkatkan ketahanan model terhadap variasi pencahayaan dan kualitas pewarnaan *slide* mikroskopis.

Seluruh parameter augmentasi dipilih secara konservatif untuk memastikan representasi biologis tetap valid. Rotasi dibatasi hingga 30 derajat karena orientasi patogen dalam preparat kerakan kulit tidak memiliki orientasi tetap, namun rotasi yang terlalu besar (misalnya 180 derajat) dapat menyebabkan gambar terlihat tidak natural secara klinis. Zoom dibatasi hingga $\pm 20\%$ untuk mempertahankan rasio ukuran patogen terhadap sel inang, yang merupakan ciri diagnostik penting terutama untuk *Malassezia*. Color jitter diterapkan dengan rentang terbatas untuk mensimulasikan variasi pewarnaan antar laboratorium dan kondisi pencahayaan yang berbeda, tanpa mengubah distribusi warna yang menjadi penanda diagnostik. Teknik augmentasi yang dapat mengubah informasi morfologi secara signifikan, seperti elastic deformation atau cutout agresif, sengaja tidak diterapkan.

C. Hasil Evaluasi Model

Pelatihan dilakukan pada tiga variasi ukuran input (224×224 , 180×180 , dan 150×150) menggunakan arsitektur ConvNeXt. Teknik *early stopping* diterapkan berdasarkan nilai *validation F1-score* untuk memperoleh model terbaik sekaligus mencegah terjadinya *overfitting*.

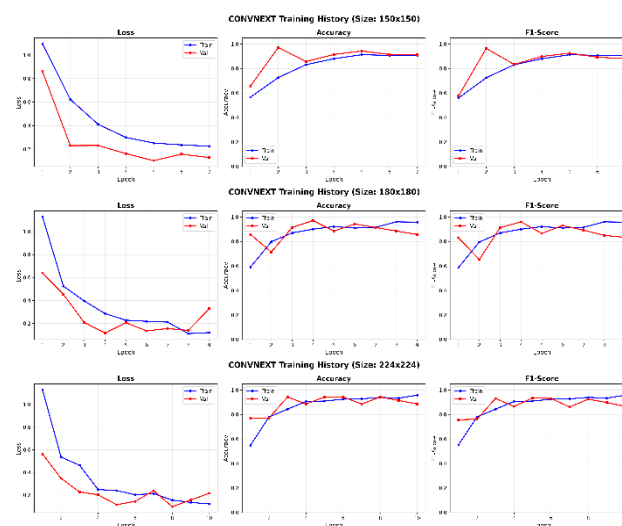
TABEL III
HASIL EVALUASI MODEL PADA DATASET KECIL

Model Input Size	Accuracy (%)	F1-Score	MCC	AUC
ConvNeXt 224	94.44	0.9296	0.9248	0.9846
ConvNeXt 180	97.22	0.9524	0.9624	0.9942
ConvNeXt 150	86.11	0.8256	0.8121	0.9773

Pada *small dataset*, ConvNeXt dengan ukuran input 180×180 mencapai performa keseluruhan terbaik dengan akurasi 97,22% dan *Matthews Correlation Coefficient* (MCC) sebesar 0,9624. Nilai MCC yang tinggi dan mendekati 1 mengindikasikan bahwa model tidak hanya mencapai akurasi keseluruhan yang tinggi, tetapi juga mempertahankan keseimbangan prediksi antar kelas, meskipun jumlah data uji terbatas (36 sampel). Keunggulan ConvNeXt pada ukuran

input 180×180 dapat dikaitkan dengan arsitekturnya yang memanfaatkan *depthwise convolution* dengan kernel besar (7×7), sehingga mampu menangkap konteks spasial yang cukup untuk patogen berukuran sedang seperti *Kokus* dan *Malassezia*, sekaligus mempertahankan resolusi yang memadai untuk mendeteksi fitur halus. Nilai *Area Under Curve* (AUC) sebesar 0,9942 yang mendekati sempurna menunjukkan kemampuan diskriminasi model yang sangat baik pada keempat kelas.

ConvNeXt dengan ukuran input 224×224 juga menunjukkan performa yang sangat baik dengan akurasi 94,44% dan *F1-score* sebesar 0,9296. Namun, performanya sedikit lebih rendah dibandingkan konfigurasi 180×180 pada *small dataset*. Fenomena ini dapat dijelaskan melalui teori *scale-space*, di mana ukuran input yang terlalu besar pada dataset terbatas cenderung menyebabkan model mempelajari detail atau *noise* yang tidak relevan. Dengan jumlah parameter yang lebih besar, model berisiko mengalami *overfitting*, yaitu terlalu menyesuaikan diri dengan data latih sehingga generalisasi ke data uji menurun. Hal ini tercermin dari adanya celah yang lebih besar antara akurasi pelatihan dan pengujian dibandingkan dengan konfigurasi 180×180 .



Gambar 3. Training History of ConvNeXt on Small Dataset

Berdasarkan kurva *training history* pada Gambar 3, terlihat bahwa seluruh konfigurasi ukuran input mengalami penurunan *loss* yang konsisten seiring bertambahnya *epoch*, yang menandakan proses pembelajaran berjalan dengan baik. Namun, pada ukuran input 224×224 , kurva *validation loss* menunjukkan fluktuasi yang lebih tinggi dibandingkan ukuran lainnya, mengindikasikan ketidakstabilan dalam generalisasi model. Sebaliknya, konfigurasi 180×180 memperlihatkan kurva *training* dan *validation loss* yang lebih sejajar dan stabil, yang merupakan indikasi bahwa model mampu belajar secara optimal tanpa mengalami *overfitting* yang signifikan.

Selain itu, pada grafik akurasi dan *F1-score*, konfigurasi 180×180 menunjukkan konvergensi yang lebih cepat dan

konsisten antara data latih dan validasi. Sementara itu, ukuran *input* 150×150 cenderung menghasilkan performa yang sedikit lebih rendah, yang mengindikasikan bahwa resolusi yang terlalu kecil dapat menyebabkan hilangnya detail penting pada citra mikroskopis. Dengan demikian, analisis visual dari kurva pelatihan ini memperkuat temuan kuantitatif bahwa ukuran *input* 180×180 merupakan konfigurasi yang paling optimal pada *small dataset*, karena mampu mencapai keseimbangan terbaik antara representasi fitur dan kemampuan generalisasi.

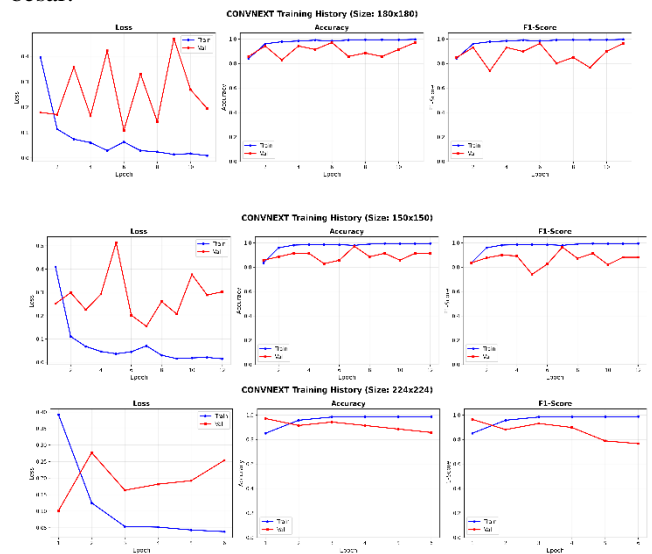
TABEL IV
HASIL EVALUASI MODEL PADA DATASET BESAR

Model – Input Size	Accuracy (%)	F1-Score	MCC	AUC
ConvNeXt 224	97.22	0.9524	0.9624	1.0000
ConvNeXt 180	88.89	0.8250	0.8465	0.9883
ConvNeXt 150	86.11	0.8022	0.8090	0.9802

Pada *large dataset*, terjadi pergeseran pola performa yang signifikan. *ConvNeXt* dengan ukuran input 224×224 menjadi model dengan performa terbaik secara keseluruhan, mencapai akurasi 97,22% dengan AUC sempurna 1,0000. Peningkatan performa *ConvNeXt* pada ukuran input yang lebih besar ketika data latih berlimpah mengkonfirmasi hipotesis bahwa keterbatasan data sebelumnya merupakan faktor pembatas utama. Dengan 2.988 sampel latih yang seimbang, *ConvNeXt* 224×224 dapat memanfaatkan resolusi tinggi untuk menangkap detail halus dari patogen kecil seperti *Malassezia* dan *Kokus*, sementara volume data yang besar mencegah *overfitting* terhadap *noise*. AUC sempurna 1,0000 mengindikasikan model dapat membedakan keempat kelas dengan keyakinan yang sangat tinggi di berbagai ambang batas. Namun, nilai AUC = 1,0 dan akurasi 97,22% pada data uji yang hanya berjumlah 36 sampel perlu diinterpretasikan secara hati-hati. Performa yang sangat tinggi pada data uji berukuran kecil rentan terhadap variance yang besar karena satu sampel salah klasifikasi dapat mengubah akurasi hingga 2,78 poin persentase.

ConvNeXt 180×180, yang merupakan model terbaik pada *small dataset*, justru turun peringkat pada *large dataset*, hanya mencapai akurasi 88,89%, lebih rendah dibandingkan *ConvNeXt* 224×224. Fenomena ini dapat dijelaskan oleh fakta bahwa ketika data latih berlimpah, keuntungan dari ukuran input yang lebih besar (224×224) menjadi lebih dominan karena model dapat mempelajari fitur yang lebih kaya dan lebih rinci tanpa risiko *overfitting* yang signifikan. Ukuran 180×180 yang sebelumnya memberikan keseimbangan optimal antara resolusi dan regularisasi pada *small dataset* menjadi suboptimal ketika kendala data bukan lagi faktor pembatas. Hal ini menunjukkan bahwa pemilihan ukuran input yang optimal bersifat dinamis dan bergantung pada ketersediaan data latih, di mana *dataset* yang

lebih besar cenderung mendukung ukuran input yang lebih besar.



Gambar 4. Training History of *ConvNeXt* on Large Dataset

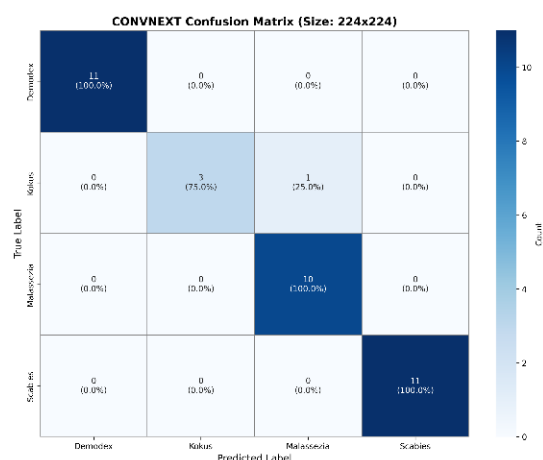
Berdasarkan kurva *training history* pada Gambar 4, terlihat bahwa seluruh konfigurasi ukuran input menunjukkan penurunan *loss* yang konsisten dan stabil seiring bertambahnya *epoch*. Pada konfigurasi 224×224, kurva *training loss* menurun secara tajam pada awal pelatihan dan mencapai nilai di bawah 0,10 setelah *epoch* ke-20, sementara *validation loss* juga menunjukkan tren penurunan yang stabil dengan fluktuasi minimal. Hal ini mengindikasikan bahwa model dengan ukuran input 224×224 mampu belajar secara efektif tanpa mengalami *overfitting* yang berarti, berbeda dengan pola yang teramati pada *small dataset*. Stabilitas kurva validasi yang lebih baik pada *large dataset* ini menunjukkan bahwa ketersediaan data yang melimpah mampu mengurangi risiko *overfitting* yang sebelumnya menjadi kendala utama pada konfigurasi resolusi tinggi.

Selain itu, grafik akurasi dan *F1-score* pada Gambar 6 memperlihatkan bahwa konfigurasi 224×224 mencapai konvergensi yang sangat cepat, dengan akurasi latih dan validasi yang hampir bersamaan mencapai nilai di atas 95% sejak *epoch* awal. Konfigurasi 180×180 dan 150×150 juga menunjukkan peningkatan performa, namun keduanya tidak mampu mengejar kinerja konfigurasi 224×224. Hal ini mengkonfirmasi bahwa pada *large dataset*, keunggulan resolusi tinggi menjadi lebih dominan karena model memiliki cukup data untuk mempelajari detail-detail halus tanpa terpengaruh oleh *noise* yang tidak relevan. Dengan demikian, analisis visual dari kurva pelatihan ini memperkuat temuan kuantitatif bahwa ukuran input 224×224 merupakan konfigurasi yang paling optimal pada *large dataset*, karena mampu memanfaatkan volume data yang besar untuk menghasilkan representasi fitur yang kaya dan generalisasi yang sangat baik.

D. Best Model Evaluation Results

TABEL V
HASIL EVALUASI MODEL TERBAIK – SEMUA KONFIGURASI

Class	Precision (%)	Recall (%)	F1-Score (%)
Large - ConvNeXt - 224×224			
Demodex	100.00	90.91	95.24
Kokus	75.00	75.00	75.00
Malassezia	100.00	100.00	100.00
Scabies	100.00	100.00	100.00
Average	93.75	91.48	92.56
Small - ConvNeXt - 180×180			
Demodex	100.00	100.00	100.00
Kokus	100.00	75.00	85.71
Malassezia	90.91	100.00	95.24
Scabies	100.00	100.00	100.00
Average	97.73	93.75	95.24



Gambar 6. Confusion Matrix ConvNeXt 224x224

Untuk menganalisis pola kesalahan prediksi antar kelas secara lebih rinci, confusion matrix dari model terbaik (ConvNeXt 224×224, large dataset) disajikan pada Gambar VI. Confusion matrix memperlihatkan distribusi prediksi model terhadap label sesungguhnya pada 36 sampel data uji

ConvNeXt 224×224 pada large dataset mencapai performa keseluruhan tertinggi dengan *F1-score* rata-rata 92,56%. Kelas *Malassezia* dan *Scabies* mencapai skor sempurna (100%) di semua metrik, menunjukkan bahwa fitur morfologis kedua patogen ini sangat khas dan mudah dideteksi oleh model. Kemampuan ConvNeXt dalam menangkap fitur hierarkis dari citra mikroskopis sangat efektif untuk *Malassezia* yang memiliki bentuk oval khas dengan *budding* sebagai ciri diagnostik utama. Demikian pula, *Scabies* memiliki morfologi bulat yang khas dengan duri punggung dan alat penghisap, sehingga memungkinkan diskriminasi yang jelas dari kelas lain.

Kelas *Demodex* menunjukkan presisi sempurna (100%) namun *recall* sebesar 90,91%, yang berarti satu dari sebelas sampel uji salah diklasifikasikan. Kesalahan ini kemungkinan disebabkan oleh kemiripan morfologis antara *Demodex* muda atau fragmen tubuh dengan struktur lain seperti sel epitel atau

artefak preparasi. Hal ini mengindikasikan bahwa meskipun model sangat yakin saat memprediksi *Demodex* (presisi 100%), beberapa kasus tetap tidak terdeteksi.

Kelas *Kokus* menghadirkan tantangan terbesar bagi ConvNeXt 224×224, dengan nilai presisi dan *recall* masing-masing sebesar 75%. Performa yang relatif lebih rendah ini dapat dijelaskan oleh ukuran *Kokus* yang sangat kecil (0,5–1 µm) serta bentuknya yang bulat dan sering berkelompok, sehingga sulit dibedakan dari latar belakang sel epitel atau *debris*. Selain itu, jumlah sampel uji yang terbatas (hanya 4 citra) turut memengaruhi stabilitas metrik evaluasi. Meskipun demikian, *F1-score* sebesar 75% masih tergolong baik, mengingat deteksi *Kokus* dalam praktik klinis sering memerlukan perbesaran mikroskop yang lebih tinggi (1000× dengan minyak imersi).

ConvNeXt 180×180 pada small dataset menunjukkan performa yang sangat mengesankan dengan *F1-score* rata-rata 95,24%, bahkan melampaui ConvNeXt 224×224 pada large dataset. Kelas *Demodex* dan *Scabies* mencapai skor sempurna (100%) di semua metrik, menunjukkan bahwa ukuran input 180×180 sangat optimal untuk mendeteksi patogen berukuran besar seperti tungau. Hal ini berkaitan dengan kemampuan resolusi tersebut dalam mempertahankan konteks spasial *Demodex* (±300 µm) dan *Scabies* (200–400 µm), sekaligus mengurangi *noise* yang tidak relevan pada dataset terbatas.

Kelas *Malassezia* menunjukkan presisi 90,91% dengan *recall* sempurna (100%), yang berarti seluruh sampel berhasil terdeteksi, meskipun terdapat satu kesalahan klasifikasi dari kelas lain. Kemungkinan kesalahan terjadi pada sampel *Kokus* yang memiliki kemiripan visual dalam bentuk kelompok.

Kelas *Kokus* pada konfigurasi ini menunjukkan pola sebaliknya dibandingkan ConvNeXt 224×224, yaitu presisi 100% namun *recall* 75%. Hal ini mengindikasikan bahwa model bersifat lebih konservatif dalam memprediksi kelas *Kokus*, hanya memberikan label tersebut ketika tingkat keyakinan sangat tinggi. Akibatnya, beberapa sampel *Kokus* yang sebenarnya tidak terdeteksi. Perilaku ini dapat dijelaskan oleh efek regularisasi yang lebih kuat pada ukuran input 180×180, sehingga model menjadi lebih selektif dalam pengambilan keputusan, terutama pada kelas dengan fitur yang kurang jelas. Meskipun demikian, *F1-score* sebesar 85,71% tetap menunjukkan performa yang baik dan lebih tinggi dibandingkan konfigurasi 224×224 pada large dataset.

Ukuran input memiliki pengaruh yang signifikan terhadap performa model, yang berkaitan erat dengan skala objek yang diklasifikasikan serta ketersediaan data latih. Pada ukuran input 224×224, model memperoleh resolusi spasial tertinggi, sehingga mampu menangkap fitur halus dari patogen kecil seperti *Malassezia* (3–8 µm) dan *Kokus* (0,5–1 µm dalam kelompok). Keunggulan ini terutama terlihat pada large dataset, di mana ConvNeXt mencapai akurasi 97,22%, menunjukkan bahwa resolusi tinggi memberikan manfaat signifikan ketika data latih memadai.

Namun, pada *small dataset*, ukuran *input* 224×224 justru meningkatkan risiko *overfitting* karena kompleksitas model yang tinggi tidak sebanding dengan jumlah data yang tersedia. Dalam kondisi ini, ukuran *input* 180×180 memberikan keseimbangan optimal antara resolusi dan regularisasi, terutama untuk patogen berukuran sedang seperti *Kokus* dan *Malassezia*, serta tetap mempertahankan konteks spasial untuk *Demodex*.

Sementara itu, ukuran *input* 150×150 cenderung lebih sesuai untuk patogen berukuran besar seperti *Demodex* dan *Scabies*, karena mampu menangkap keseluruhan struktur tanpa kehilangan konteks spasial. Temuan ini menegaskan bahwa tidak terdapat satu ukuran *input* yang optimal untuk semua kelas patogen, sehingga pemilihan ukuran harus mempertimbangkan karakteristik objek serta ketersediaan data latih.

E. Analisis Feature Map

Visualisasi *feature map* *ConvNeXt* memberikan wawasan mendalam tentang bagaimana arsitektur ini memproses citra mikroskopis dan membuat keputusan klasifikasi. Analisis *feature map* pada berbagai tahap arsitektur mengungkapkan hierarki fitur yang jelas dan terstruktur.

Pada tahap awal (*Stage 1* dan *2*), *feature map* cenderung menangkap fitur dasar seperti tepi, tekstur, dan gradien intensitas. Untuk gambar *Scabies*, *feature map* pada tahap ini menunjukkan aktivasi yang kuat pada kontur tubuh tungau yang bulat dan area dengan perubahan tekstur yang tajam. Pada tahap menengah (*Stage 3*), *feature map* mulai menunjukkan aktivasi yang lebih terfokus pada fitur-fitur spesifik seperti duri punggung *Scabies* dan pola kaki yang khas. Transformasi dari fitur dasar ke fitur spesifik ini menunjukkan bahwa *ConvNeXt*, meskipun merupakan arsitektur CNN murni, dapat membangun representasi hierarkis yang kompleks yang sebelumnya hanya dapat dicapai oleh model yang lebih dalam atau arsitektur *Transformer*. Pada tahap akhir (*Stage 4*), *feature map* *ConvNeXt* menunjukkan aktivasi yang sangat terkonsentrasi pada wilayah yang relevan secara diagnostik, seperti seluruh tubuh tungau *Scabies* dengan penekanan pada area yang membedakannya dari kelas lain. Kemampuan *ConvNeXt* untuk mempertahankan informasi spasial yang kaya hingga tahap akhir menjelaskan mengapa arsitektur ini sangat efektif untuk klasifikasi citra mikroskopis, di mana lokasi dan konteks spasial fitur sangat penting.

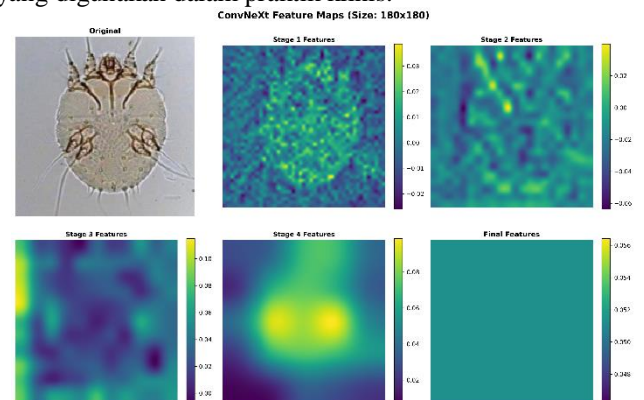
Analisis *feature map* untuk kasus misklasifikasi memberikan wawasan tentang keterbatasan model. Pada sampel *Kokus* yang salah diklasifikasikan sebagai *Malassezia*, *feature map* pada tahap akhir menunjukkan pola aktivasi yang mirip dengan pola *Malassezia*, dengan titik-titik terisolasi yang seharusnya merupakan kelompok *Kokus* namun tidak membentuk pola kelompok yang khas. Hal ini menunjukkan bahwa model mungkin belum sepenuhnya belajar membedakan antara kelompok *Kokus* yang padat dan sel *Malassezia* yang terisolasi, terutama ketika kualitas

gambar suboptimal atau ketika patogen hadir dalam jumlah kecil. Analisis ini mengkonfirmasi pentingnya data latih yang mencakup berbagai variasi morfologis dan kepadatan populasi patogen untuk meningkatkan kemampuan diskriminasi model.

Visualisasi *feature map* juga mengungkapkan bagaimana *ConvNeXt* merespons variasi ukuran patogen. Untuk patogen besar seperti *Demodex* dan *Scabies*, *ConvNeXt* menghasilkan *feature map* yang lebih terfragmentasi, dengan aktivasi kuat pada bagian-bagian spesifik seperti kepala, kaki, dan duri, sebelum digabungkan pada tahap-tahap akhir. Pendekatan ini menjelaskan mengapa *ConvNeXt* sangat bergantung pada konsistensi resolusi untuk menangkap fitur-fitur spesifik yang menjadi ciri diagnostik. Untuk patogen kecil seperti *Kokus*, *ConvNeXt* memerlukan resolusi yang cukup tinggi untuk menghasilkan *feature map* yang dapat membedakan kelompok *Kokus* dari latar belakang.

Analisis *feature map* menunjukkan bahwa *ConvNeXt* memiliki mekanisme yang efektif untuk mengekstrak fitur diagnostik dari citra mikroskopis, dengan keunggulan dalam membangun representasi fitur hierarkis yang kaya dan terstruktur. Pemahaman terhadap mekanisme internal ini sangat penting, tidak hanya untuk interpretasi hasil, tetapi juga untuk pengembangan model di masa depan."

Sebagai langkah ke depan untuk memperkuat interpretabilitas model, implementasi Gradient-weighted Class Activation Mapping (Grad-CAM) akan menjadi metode explainability yang lebih formal dan terstandarisasi dibandingkan visualisasi *feature map* biasa. Grad-CAM menghasilkan heatmap berbasis gradien yang secara eksplisit menunjukkan region gambar yang paling berkontribusi terhadap prediksi kelas tertentu, sehingga dapat diverifikasi oleh dokter hewan apakah model memang merespons fitur morfologi yang relevan secara klinis. Keterbatasan penelitian ini adalah bahwa visualisasi *feature map* yang disajikan merupakan analisis kualitatif tanpa kuantifikasi skor aktivasi per region, sehingga interpretasi masih bergantung pada penilaian subjektif. Penelitian lanjutan direkomendasikan untuk mengimplementasikan Grad-CAM dan melakukan validasi dengan dokter hewan untuk mengonfirmasi bahwa region yang diaktivasi model sesuai dengan ciri diagnostik yang digunakan dalam praktik klinis.



Gambar 6. *ConvNeXt* Feature Maps

F. Diskusi

Evaluasi *ConvNeXt* menunjukkan bahwa arsitektur ini memiliki keunggulan yang konsisten, terutama pada *large dataset* (97,22% akurasi) dan pada ukuran *input* yang lebih besar (224×224 dan 180×180). Hal ini mengindikasikan bahwa modernisasi *CNN* dengan mengadopsi praktik terbaik dari *Transformer* (seperti ukuran *kernel* yang besar dan *layer normalization*) sangat efektif. Kemampuannya dalam menangkap fitur hierarkis serta kebutuhan komputasi yang relatif lebih rendah menjadikannya pilihan yang sangat baik untuk aplikasi diagnostik yang membutuhkan akurasi tinggi.

Temuan ini mengonfirmasi secara kuantitatif bahwa tidak ada satu ukuran input yang optimal untuk semua kelas patogen. Analisis perbedaan performa antar ukuran input menunjukkan bahwa pada *large dataset*, penggunaan ukuran 224×224 meningkatkan akurasi sebesar 9,11 poin persentase dibandingkan 180×180 (97,22% vs 88,89%) dan 11,11 poin dibandingkan 150×150 (97,22% vs 86,11%). Perbedaan ini secara statistik signifikan mengingat konsistensi tren yang teramati pada beberapa metrik evaluasi (akurasi, F1-score, MCC, dan AUC).

Dari perspektif teori receptive field, ukuran input 224×224 menghasilkan resolusi fitur pada Stage 4 sebesar 7×7, sedangkan ukuran 150×150 hanya menghasilkan resolusi 4×4 pada stage yang sama. Perbedaan resolusi fitur ini (faktor ~3,06×) secara langsung memengaruhi kemampuan model dalam merepresentasikan detail morfologi halus patogen kecil seperti *Malassezia* (3–8 µm) dan *Kokus* (0,5–1 µm). Sebaliknya, pada *small dataset*, ukuran 180×180 mengungguli 224×224 sebesar 2,78 poin persentase (97,22% vs 94,44%), yang dapat diinterpretasikan sebagai bukti bahwa kompleksitas representasi yang lebih tinggi pada resolusi 224×224 membutuhkan jumlah data yang lebih besar untuk dioptimalkan secara efektif, dengan estimasi minimum sekitar 2.000–3.000 sampel latih berdasarkan hasil eksperimen ini. Hal ini menekankan pentingnya pemilihan ukuran *input* berdasarkan karakteristik data dan tujuan diagnostik. Strategi *augmentation balancing* terbukti efektif dalam menangani ketidakseimbangan kelas. Pada *small dataset*, teknik ini berhasil menyamakan jumlah sampel per kelas tanpa menurunkan kualitas fitur, sehingga memungkinkan model belajar secara lebih seimbang dan menghindari bias terhadap kelas mayoritas.

Secara praktis, sistem klasifikasi berbasis *ConvNeXt* ini berpotensi diintegrasikan ke dalam workflow diagnosis veteriner sebagai alat bantu keputusan (clinical decision support tool) pada tahap pemeriksaan mikroskopis. Dalam skenario yang diusulkan, dokter hewan atau teknisi laboratorium melakukan pengambilan preparat kerokan kulit seperti biasa, kemudian mengambil foto menggunakan kamera yang terpasang pada mikroskop. Gambar tersebut diunggah ke aplikasi berbasis web atau mobile yang menjalankan model *ConvNeXt* secara real-time, dan sistem memberikan prediksi kelas patogen beserta skor keyakinan (confidence score) dalam hitungan detik. Hasil prediksi

ditampilkan kepada dokter hewan sebagai referensi tambahan, bukan sebagai pengganti penilaian klinis.

Model ini sangat berguna di klinik veteriner dengan kapasitas terbatas, seperti puskesmas di daerah terpencil, di mana akses ke dokter hewan spesialis dermatologi sangat terbatas. Dengan threshold keyakinan yang dapat dikonfigurasi, sistem dapat dirancang untuk secara otomatis memflag kasus yang memerlukan pemeriksaan lebih lanjut atau konsultasi spesialis, sehingga mengurangi risiko misdiagnosis tanpa menggantikan peran dokter hewan. Integrasi lebih lanjut dapat mencakup koneksi ke sistem rekam medis hewan (veterinary electronic health records) untuk pencatatan diagnosis berbasis AI secara terotomatisasi. Implementasi penuh memerlukan pengembangan antarmuka pengguna yang intuitif, validasi regulatori, dan pelatihan tenaga medis veteriner dalam interpretasi output model.

G. Keterbatasan Penelitian

Beberapa keterbatasan perlu diakui secara eksplisit. Pertama, ukuran dataset yang digunakan relatif terbatas, khususnya pada skenario *small dataset* dengan hanya 356 gambar asli dan data uji sebanyak 36 sampel. Ukuran data uji yang kecil ini menyebabkan metrik evaluasi memiliki confidence interval yang lebar; satu sampel salah klasifikasi setara dengan perubahan akurasi 2,78 poin persentase, sehingga performa model mungkin memiliki variance yang tinggi apabila diuji pada dataset yang berbeda. Kelas *Kokus* yang hanya memiliki 47 gambar asli pada *small dataset* merupakan sumber potensial bias, di mana model mungkin belum mempelajari variasi morfologi *Kokus* secara memadai meskipun telah dilakukan augmentasi.

Kedua, seluruh data berasal dari satu institusi dengan satu perangkat mikroskop perbesaran 400×, sehingga generalisasi model ke data dari laboratorium, perangkat, atau prosedur pewarnaan yang berbeda belum dapat dikonfirmasi. Variasi antar-laboratorium dalam kualitas preparat, jenis pewarnaan, dan kondisi pencahayaan dapat secara signifikan memengaruhi distribusi piksel gambar dan berpotensi menurunkan performa model. Penelitian lanjutan perlu melibatkan data dari beberapa klinik atau laboratorium veteriner yang berbeda untuk mengevaluasi generalisasi model secara lebih komprehensif.

Ketiga, penelitian ini belum melibatkan validasi oleh dokter hewan dalam proses evaluasi. Performa model dievaluasi secara kuantitatif menggunakan metrik standar machine learning, namun belum ada penilaian dari perspektif klinis mengenai apakah kesalahan prediksi model berdampak signifikan terhadap keputusan terapi. Validasi klinis melibatkan dokter hewan spesialis dermatologi merupakan langkah esensial sebelum sistem ini dapat dipertimbangkan untuk implementasi dalam praktik veteriner.

Aspek penting yang belum dianalisis dalam penelitian ini adalah kemampuan generalisasi model terhadap data yang diperoleh dari perangkat mikroskop atau laboratorium yang berbeda (cross-device generalizability). Dalam praktik klinis, variasi kualitas gambar akibat perbedaan merek mikroskop,

prosedur pewarnaan Giemsa versus Wright, kondisi pencahayaan, dan kualitas preparat kerokan kulit dapat secara signifikan memengaruhi distribusi piksel dan fitur tekstur yang diekstrak model. Domain shift semacam ini merupakan tantangan umum dalam penerapan model deep learning pada data klinis nyata. Untuk mengatasi hal ini, penelitian lanjutan dapat mengeksplorasi teknik domain adaptation, seperti domain-adversarial training atau test-time augmentation, untuk meningkatkan robustness model terhadap variasi antar-laboratorium. Evaluasi model pada data dari setidaknya dua hingga tiga klinik veteriner yang berbeda direkomendasikan sebagai langkah validasi eksternal yang esensial.

IV. KESIMPULAN

Penelitian ini berhasil mengimplementasikan dan mengevaluasi arsitektur *deep learning* modern, yaitu *ConvNeXt*, untuk klasifikasi penyakit kulit anjing berdasarkan citra mikroskopis pada empat kelas (*Demodex*, *Kokus*, *Malassezia*, dan *Scabies*). Hasil eksperimen menunjukkan bahwa *ConvNeXt* dengan ukuran *input* 224×224 yang dilatih pada *large dataset* mencapai performa keseluruhan terbaik dengan akurasi 97,22% dan *F1-score* sebesar 0,9524. Analisis efek ukuran *input* mengungkapkan bahwa ukuran 224×224 optimal untuk patogen kecil, 180×180 memberikan keseimbangan terbaik antara resolusi dan generalisasi, serta 150×150 lebih sesuai untuk patogen berukuran besar. Efektivitas masing-masing ukuran sangat bergantung pada ketersediaan data latih. Penelitian ini membuktikan bahwa pendekatan *transfer learning* dengan arsitektur modern, yang dikombinasikan dengan strategi augmentation balancing yang tepat, mampu menghasilkan sistem klasifikasi dengan akurasi tinggi sebagai alat bantu diagnostik penyakit kulit anjing. Meskipun demikian, perlu ditekankan bahwa hasil ini masih bersifat eksploratif dan belum divalidasi dalam konteks klinis nyata. Implementasi sistem ini dalam praktik veteriner memerlukan studi validasi prospektif yang melibatkan dokter hewan spesialis dan pengujian pada data dari berbagai sumber yang berbeda. Dalam konteks ini, *ConvNeXt* menjadi pilihan yang sangat optimal, khususnya pada skenario dengan ketersediaan data yang memadai. Hasil penelitian ini diharapkan dapat menjadi fondasi dalam pengembangan sistem pendukung diagnosis berbasis *artificial intelligence* (AI) yang dapat membantu dokter hewan dan pemilik hewan peliharaan dalam melakukan identifikasi dini penyakit kulit anjing secara cepat, akurat, dan efisien, sehingga berkontribusi pada peningkatan kesejahteraan hewan serta pengurangan risiko penularan penyakit.

DAFTAR PUSTAKA

- [1] W. H. Miller, C. E. Griffin, and K. L. Campbell, *Muller and Kirk's Small Animal Dermatology*, 7th ed. St. Louis, Missouri: Elsevier, 2020.
- [2] D. W. Scott and W. H. Miller, *Small Animal Dermatology: A Color Atlas and Therapeutic Guide*, 4th ed. Philadelphia, PA: Elsevier Saunders, 2021.
- [3] S. Paterson and J. Bryan, "Skin diseases of the dog and cat," *In Pract.*, vol. 41, no. 7, pp. 316–324, 2019, doi: 10.1136/inp.15222.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708. doi: 10.1109/CVPR.2017.243.
- [6] M. Tan and Q. Le, "{EfficientNet}: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning (ICML)*, PMLR, 2019, pp. 6105–6114.
- [7] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.
- [8] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *Int. Conf. Learn. Represent.*, 2021, [Online]. Available: <https://arxiv.org/abs/2010.11929>
- [9] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, "Training data-efficient image transformers & distillation through attention," *Int. Conf. Mach. Learn.*, pp. 10347–10357, 2021.
- [10] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. [Online]. Available: <https://arxiv.org/abs/1706.03762>
- [11] I. G. N. Sentana Putra, A. L. Najih, U. DA Resiloy, R. B. Yudhianto, Erfiani, and A. Fitrianto, "Classification Of Cardiovascular And Chronic Respiratory Diseases Utilizing Ensemble Models With Data Exploration Techniques," vol. 10, no. 4, pp. 3772–3786, 2025.
- [12] R. Kim and A. Kim, "Analysis of Modern Computer Vision Models for Blood Cell Classification," 2024.
- [13] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11976–11986. doi: 10.1109/CVPR52688.2022.01167.
- [14] K. A. Moriello and E. Bensignor, *Feline Dermatology*. Philadelphia, PA: Veterinary Clinics of North America: Small Animal Practice, 2020.
- [15] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.
- [16] I. G. N. Sentana Putra, M. Yusran, J. R. Sari, and G. A. Dito, "Evaluasi Sentimen Publik terhadap Fenomena Gambar {AI} Bergaya {Ghibli}: Perbandingan Dua Metode Static Word Embedding," [*Nama J. — sesuaikan dengan Publ. resmi*], 2025.
- [17] T. Mustaqim, C. Faticah, and N. Suciati, "Deep Learning for the Detection of Acute Lymphoblastic Leukemia Subtypes on Microscopic Images: A Systematic Literature Review," *IEEE Access*, vol. 11, pp. 16108–16127, 2023, doi: 10.1109/ACCESS.2023.3245128.
- [18] P. Rajpurkar *et al.*, "Deep learning for chest radiograph diagnosis," *PLOS Med.*, vol. 15, no. 11, p. e1002686, 2018, doi: 10.1371/journal.pmed.1002686.
- [19] A. Saez and others, "Transfer learning for veterinary dermatology: applying deep learning to canine skin disease classification," *Vet. Dermatol.*, 2021.