

PERBANDINGAN KINERJA DAN EFISIENSI ARSITEKTUR CNN VGG-16 DAN RESNET50 DALAM KANKER KULIT DATASET HAM10000

Richard Leonardo^{1*}, Ryan Putra Laksana², Habibullah Akbar³, Imam Sutanto⁴

^{1,2,3,4}Fakultas Ilmu Komputer, Universitas Esa Unggul; Jl. Arjuna Utara No.9, Duri Keba, Kec. Kb. Jeruk, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta 11510; (021) 39529950

Keywords:

CNN,
Dataset,
HAM10000, Resnet50,
VGG16

Correspondent Email:

richardleo155@student.esa
unggul.ac.id

Abstrak. Penelitian ini bertujuan untuk membandingkan performa dan efisiensi dua arsitektur Convolutional Neural Network (CNN) populer, yaitu ResNet50 dan VGG16, dalam klasifikasi citra lesi kulit menggunakan dataset HAM10000. Dataset ini berisi lebih dari 10.000 gambar dermatologis yang merepresentasikan berbagai jenis lesi, baik jinak maupun ganas, termasuk melanoma. Kedua model dilatih menggunakan pendekatan transfer learning dengan bobot pretrained dari ImageNet, serta diuji dengan parameter pelatihan yang konsisten, termasuk penerapan fine-tuning, data augmentation, dan dropout. Hasil eksperimen menunjukkan bahwa ResNet50 lebih konsisten dalam menghasilkan performa yang lebih unggul dibandingkan VGG16, baik dari segi akurasi validasi maupun akurasi pengujian. ResNet50 mencapai akurasi tertinggi sebesar 82.14% pada data pengujian, sedangkan VGG16 hanya mencapai 71.32%. Selain itu, ResNet50 juga menunjukkan keunggulan dalam precision dan f1-score, bahkan melampaui beberapa pendekatan klasikal seperti Weighted K-NN, M-SVM, dan Naïve Bayes dari penelitian sebelumnya. Di sisi lain, meskipun VGG16 lebih ringan secara komputasi dan mampu mengenali kelas mayoritas dengan baik, model ini mengalami penurunan performa signifikan dalam mendeteksi kelas minoritas seperti melanoma dan dermatofibroma.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. The primary focus of this research is to compare the capabilities and efficiency of two Convolutional Neural Network (CNN) architectures: ResNet50 and VGG16. This comparison is conducted for the task of classifying skin lesion images using the HAM10000 dataset. This dataset comprises dermatoscopic images representing various types of skin lesions, including melanoma, nevus, and other benign and malignant tumors. Both models were optimized using transfer learning with pretrained weights from ImageNet and trained with identical parameters. The research findings indicate that ResNet50 outperformed VGG16, achieving an accuracy of 80.96% on the testing data, whereas VGG16 only reached 74.85%. While ResNet50 demonstrated superior results in terms of validation and testing accuracy, as well as generalization capability, VGG16 performed better on majority classes. Both models encountered difficulties in recognizing minority classes, such as melanoma and dermatofibroma. This challenge is likely attributable to the imbalance in the number of samples across different classes. Although ResNet50 showed overall higher accuracy, this study also highlights the necessity for further approaches, such as data balancing and augmentation, to enhance performance on minority classes. This system holds significant potential to serve as a foundation for developing AI-based skin cancer detection applications, which could assist medical professionals in accelerating diagnoses and improving detection accuracy.

1. PENDAHULUAN

Kanker kulit adalah suatu kondisi di mana terjadi pertumbuhan sel-sel kulit yang tidak terkendali, umumnya karena paparan sinar *ultraviolet* (UV) dari matahari atau sumber buatan seperti tanning bed. Kanker ini dapat berkembang di berbagai bagian tubuh, terutama area yang sering terpapar langsung oleh sinar matahari seperti leher, wajah, tangan, dan lengan [1]. Kanker kulit merupakan satu dari beberapa jenis kanker yang paling banyak kasusnya pada seluruh dunia dan menjadi permasalahan kesehatan masyarakat yang makin meningkat setiap tahunnya. Di Indonesia, kasus kanker kulit termasuk dalam sepuluh besar kanker terbanyak dengan tingkat prevalensi yang terus naik.

Berdasarkan laporan dari *Global Cancer Observatory* dan *World Health Organization* (WHO) pada tahun 2024, tercatat sekitar 408.661 kasus kanker baru di Indonesia, di mana kanker kulit *non-melanoma* (NMSC) adalah satu dari beberapa jenis yang sering ditemukan pada saat ini. WHO juga melaporkan pada tahun 2020 sebanyak 3.756 kematian akibat kanker kulit di Indonesia, berkontribusi sebesar 0,22% dari total kematian nasional [2]. Fakta ini menunjukkan pentingnya deteksi dini sebagai upaya kunci untuk meningkatkan peluang kesembuhan pasien kanker kulit.

Dalam bidang kesehatan, machine learning sangat berguna untuk mengklasifikasikan dan memprediksi data, bahkan bisa mengidentifikasi golongan atau tipe penyakit baru [3]. Metode deteksi konvensional yang mengandalkan observasi visual oleh dokter kulit sangat bergantung pada pengalaman dan keterampilan klinis, sehingga berpotensi menimbulkan ketidakkonsistenan diagnosis. Tantangan lainnya adalah keterbatasan akses terhadap tenaga medis di wilayah terpencil, yang memperburuk keterlambatan diagnosis [4]. Oleh karena itu, dibutuhkan sistem berbasis teknologi yang mampu mendukung tenaga medis, terutama dalam proses skrining dan diagnosis awal.

Seiring dengan majunya teknologi *artificial intelligence* yang dikenal juga sebagai kecerdasan buatan, pendekatan berbasis deep learning telah digunakan secara luas dalam

bidang medis, termasuk untuk klasifikasi dan deteksi citra dermatologis. Di antara berbagai metode deep learning, *Convolutional Neural Networks* (CNN) terbukti sangat ampuh untuk melakukan klasifikasi citra, terutama dalam deteksi pola visual kompleks pada citra kulit [5]. CNN bisa secara otomatis melakukan ekstraksi fitur dari gambar dermatologis walaupun tidak menggunakan hand-crafted features, dan memberikan hasil klasifikasi yang kompetitif [6]. Salah satu dataset benchmark yang sering digunakan dalam penelitian deteksi kanker kulit adalah HAM10000, yang berisi lebih dari 10.000 gambar dari berbagai jenis lesi kulit, termasuk melanoma, nevi jinak, dan keratosis.

Penelitian ini secara khusus membandingkan dua arsitektur CNN populer, yaitu VGG16 dan ResNet50. VGG16 telah digunakan secara luas dalam klasifikasi citra medis karena konsistensi strukturnya dalam menggunakan filter 3x3 dan kemampuan untuk menangkap fitur spasial secara efektif, meskipun memiliki jumlah parameter yang besar [7]. Sebaliknya, ResNet50 memanfaatkan residual connections yang memungkinkan pelatihan jaringan yang lebih dalam secara stabil. Studi oleh Xuhui Li menunjukkan bahwa baik VGG16 maupun ResNet50 mampu menghasilkan performa tinggi dalam klasifikasi kanker kulit, bahkan setara dengan ahli dermatologi dalam beberapa kasus [8].

Beberapa penelitian sebelumnya telah membahas performa CNN dalam klasifikasi kanker kulit. Meedeniya menekankan pentingnya preprocessing dan augmentasi data untuk meningkatkan performa model CNN [9]. Fatima menunjukkan bahwa model pretrained seperti ResNet50 yang di fine-tune dengan dataset khusus dapat mengatasi permasalahan overfitting dan mencapai akurasi tinggi pada klasifikasi melanoma. Di sisi lain, Mahbod menunjukkan bahwa pemilihan arsitektur CNN seperti VGG16 dan ResNet50 sangat mempengaruhi akurasi serta efisiensi komputasi dalam klasifikasi lesi kulit [7]. Di sisi lain, juga membuktikan bahwa kombinasi *fine-tuning* dan ensemble CNN mampu meningkatkan performa klasifikasi melanoma secara signifikan [10].

Penelitian ini dilakukan untuk membandingkan kinerja kedua arsitektur tersebut dalam mendeteksi kanker kulit berdasarkan dataset HAM10000, yang merupakan benchmark global dalam studi dermatologi berbasis citra. Hasil penelitian menunjukkan bahwa ResNet50 secara konsisten memberikan performa yang lebih baik dibandingkan VGG-16, baik dari sisi akurasi, *precision*, maupun *f1-score*. ResNet50 juga terbukti mengungguli metode klasikal seperti Weighted K-NN dalam hal akurasi dan kemampuan generalisasi, serta menunjukkan hasil yang kompetitif dibandingkan pendekatan yang dilaporkan dalam literatur sebelumnya [11].

Kesimpulan penelitian ini adalah evaluasi terhadap waktu pelatihan menunjukkan bahwa meskipun ResNet50 memerlukan durasi pelatihan yang lebih lama dibandingkan VGG-16, peningkatan performa yang dihasilkan sebanding dengan tambahan waktu komputasi tersebut. Berdasarkan hasil ini, pendekatan berbasis ResNet50 dengan strategi *fine-tuning* dan *preprocessing* yang tepat terbukti efektif dalam mendeteksi lesi kulit secara otomatis dan akurat. Penelitian ini memberikan kontribusi nyata dalam pengembangan sistem deteksi kanker kulit berbasis AI yang tidak hanya akurat dan efisien, tetapi juga potensial untuk diterapkan dalam sistem klinis modern guna mendukung pemerataan layanan kesehatan di masa depan.

2. TINJAUAN PUSTAKA

2.1. Kanker Kulit

Kanker kulit merupakan pertumbuhan sel kulit yang abnormal akibat kerusakan DNA, terutama karena paparan sinar UV, faktor genetik, dan gangguan sistem imun (Shraim et al., 2023).

2.2. Machine Learning

Machine learning adalah cabang kecerdasan buatan yang memungkinkan sistem belajar dari data untuk mengenali pola dan membuat prediksi tanpa diprogram secara eksplisit. Secara umum terbagi menjadi:

- *Supervised learning*: menggunakan data berlabel
- *Unsupervised learning*: tanpa label

- *Reinforcement learning*: berbasis interaksi dan reward

2.3. Deep Learning

Deep learning merupakan bagian dari machine learning yang menggunakan jaringan saraf berlapis untuk mempelajari representasi data secara otomatis [12]. Keunggulannya adalah mampu melakukan ekstraksi fitur tanpa rekayasa manual, dan banyak digunakan pada citra, teks, dan suara.

2.4. TensorFlow

TensorFlow adalah pustaka open-source untuk machine learning yang dikembangkan oleh Google, berbasis *computational graph* dan mendukung CPU, GPU, serta TPU [13]. Framework ini banyak digunakan untuk membangun dan melatih model deep learning.

2.5. Python

Python merupakan bahasa utama dalam machine learning karena sintaks sederhana dan dukungan pustaka seperti NumPy, Pandas, scikit-learn, TensorFlow, dan PyTorch [5].

2.6. Epoch

Epoch adalah satu siklus pelatihan di mana seluruh dataset diproses oleh model. Jumlah epoch memengaruhi kemampuan model belajar, namun terlalu banyak dapat menyebabkan overfitting [14].

2.7. Perbandingan Kinerja

Perbandingan kinerja digunakan untuk menilai efektivitas suatu metode berdasarkan aspek seperti akurasi, kecepatan, efisiensi, dan konsistensi (Bartz-Beielstein et al., 2020).

2.8. Efisiensi

Efisiensi mengukur kemampuan model dalam menggunakan sumber daya (waktu dan memori) secara optimal tanpa mengurangi performa [15].

2.9. Arsitektur

Arsitektur adalah struktur model yang menentukan susunan layer, neuron, dan alur data dalam memproses informasi[4].

2.10. ResNet50

ResNet50 adalah arsitektur CNN dengan 50 layer yang menggunakan *residual connection*

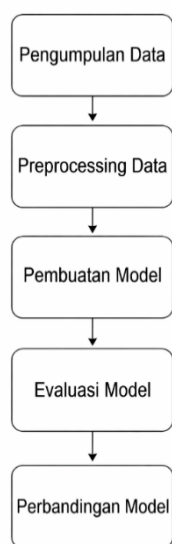
untuk mengatasi masalah degradasi pada jaringan dalam. Model ini terdiri dari blok bottleneck (1x1, 3x3, 1x1) dan banyak digunakan dalam computer vision [16].

2.11. VGG16

VGG16 adalah arsitektur CNN dengan 16 layer yang menggunakan filter kecil 3x3 secara berulang untuk meningkatkan kedalaman dan kualitas ekstraksi fitur [17].

3. METODE PENELITIAN

Tahap-tahap metodologi penelitian yang akan diterapkan dalam riset ini dapat dilihat pada Gambar 1.



Gambar 1. Rangkaian Proses Penelitian

Penelitian yang dilakukan seperti pada Gambar 1 ini menggunakan dataset HAM10000 (*Human Against Machine*) dengan 10.000 gambar untuk pelatihan dan dapat diunduh dari Kaggle ([Skin Cancer MNIST: HAM10000](#)). Dataset ini terdiri dari gambar lesi kulit yang telah dibagi kedalam tujuh kelas utama. Ketujuh kelas utama ini adalah: *Actinic keratoses (ak)*, *Basal cell carcinoma (bcc)*, *Benign keratosis-like lesions (bkl)*, *Dermatofibroma (df)*, *Melanocytic nevi (nv)*, *Melanoma (mel)*, *Vascular lesions (vasc)*.

Proses ini dilakukan dengan mengunduh citra dan metadata, kemudian mencocokkan *image_id* dengan file gambar menggunakan Python dan Pandas . Tujuan dari tahap ini adalah mempersiapkan dataset HAM10000

agar sesuai dengan kebutuhan model klasifikasi berbasis CNN. Tahapan preprocessing meliputi beberapa langkah utama sebagai berikut:

1. Mapping Image File Locations

Tahap awal dilakukan dengan mencocokkan setiap *image_id* pada metadata dengan lokasi file gambar di direktori dataset. Proses ini dilakukan secara otomatis menggunakan Python agar seluruh gambar dapat dikenali dan dimuat dengan benar.

2. Image Filtering

Setelah pemetaan, dilakukan penyaringan untuk memastikan hanya file gambar yang tersedia dan valid yang digunakan. Hal ini penting untuk menghindari error akibat file yang hilang atau rusak selama proses pelatihan.[18]. Meski berbeda tujuan, keduanya menekankan pentingnya kualitas input untuk proses pemrosesan citra yang optimal..

3. Split Data

Dataset dibagi menjadi tiga bagian menggunakan *stratified splitting* agar distribusi kelas tetap seimbang, yaitu 75% data latih, 15% data validasi, dan 15% data uji. Pembagian ini bertujuan agar model dapat belajar dengan optimal serta dievaluasi secara objektif. Penggunaan stratifikasi dalam split data sangat dianjurkan dalam praktik deep learning modern, khususnya pada data medis, untuk menjaga integritas distribusi label di seluruh subset data [19].

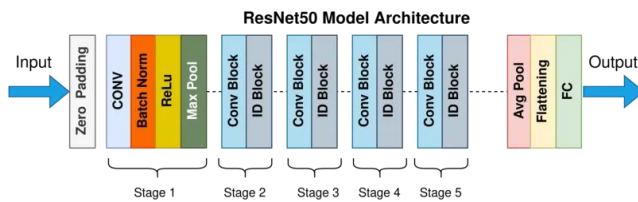
4. Image Augmentation

Augmentasi dilakukan pada data latih untuk meningkatkan kemampuan generalisasi model. Teknik yang digunakan meliputi rotasi, pergeseran, zoom, perubahan pencahayaan, dan pembalikan horizontal, sehingga model lebih robust terhadap variasi data. Strategi augmentasi seperti ini telah diakui efektif dalam mengatasi keterbatasan dataset medis, termasuk ketidakseimbangan kelas dan ukuran data yang terbatas [20]. Sementara itu, data validasi dan pengujian tidak mengalami augmentasi agar evaluasi tetap objektif..

5. Data Normalization

Tahap akhir adalah normalisasi nilai piksel agar sesuai dengan kebutuhan model CNN. Pada ResNet50 digunakan fungsi *preprocess_input()* dari Keras untuk menyesuaikan dengan standar ImageNet, sehingga proses pelatihan menjadi lebih stabil dan efisien. Normalisasi ini penting untuk

memastikan input data memiliki distribusi yang sesuai dengan model pretrained dan dapat mempercepat serta menstabilkan proses pelatihan model CNN [21]. Dengan rangkaian tahapan *preprocessing* ini, citra dari dataset HAM10000 disiapkan secara optimal untuk proses klasifikasi oleh model *deep learning* yang dibangun dalam penelitian ini.



Gambar 2 ResNet50

Model ResNet50 (Residual Network) merupakan arsitektur *deep convolutional neural network* yang dirancang untuk mengatasi masalah penurunan performa (*degradation problem*) dan *vanishing gradient* pada jaringan yang sangat dalam. Permasalahan ini menyebabkan model sulit belajar karena nilai gradien semakin kecil selama proses *backpropagation*.

Model ResNet50 atau Residual Network 50-layer merupakan arsitektur CNN yang dirancang secara khusus untuk mereduksi *vanishing gradient* yang terjadi pada jaringan saraf tiruan yang dalam. ResNet50 memperkenalkan konsep residual learning yang memungkinkan jaringan untuk tidak hanya mempelajari representasi fitur dari data secara langsung, tetapi juga mempelajari residual atau selisih antara input dan output suatu blok.

Mekanisme residual yang digunakan dalam arsitektur ResNet dapat dilihat pada Gambar 2. Dalam blok residual tersebut, input awal yang dilambangkan sebagai x diproses melalui dua buah lapisan berparameter (*weight layer*) yang masing-masing diikuti oleh fungsi aktivasi ReLU. Hasil dari dua lapisan tersebut dinotasikan sebagai $\mathcal{F}(x)$, yang merepresentasikan transformasi non-linear dari input awal. Selanjutnya, output dari transformasi ini tidak langsung menjadi keluaran akhir, melainkan dijumlahkan kembali dengan input awal melalui *shortcut connection* yang membawa nilai identitas dari x .

Penjumlahan antara hasil transformasi dan input tersebut membentuk fungsi keluaran akhir $H(x)$, sebagaimana dirumuskan pada Persamaan (1) berikut:

$$H(x) = \mathcal{F}(x) + x \quad (1)$$

Pada Persamaan 2, $H(x)$ merupakan output dari *residual block*, $\mathcal{F}(x)$ adalah hasil transformasi dari dua weight layer, dan x merupakan input asli yang disambungkan langsung melalui jalur identitas. Pendekatan ini secara signifikan membantu jaringan dalam mengatasi permasalahan degradasi akurasi yang sering terjadi pada jaringan saraf yang sangat mendalam, dengan memungkinkan sistem lebih mudah mempelajari fungsi identitas atau koreksi kecil terhadap input.

Gambar 2 menggambarkan struktur dasar dari *residual block* yang umum digunakan dalam ResNet. Jalur utama terdiri atas dua weight layer berurutan yang memproses input x untuk menghasilkan $\mathcal{F}(x)$, sedangkan jalur *shortcut* secara langsung membawa x tanpa perubahan (*identity mapping*). Setelah kedua jalur digabung melalui operasi penjumlahan, hasilnya diproses kembali oleh fungsi aktivasi ReLU.

VGG16 adalah arsitektur CNN yang terkenal karena kesederhanaan dan konsistensinya dalam desain, serta performanya yang baik pada berbagai tugas klasifikasi citra. Arsitektur ini terdiri dari 16 lapisan yang dapat dilatih (13 konvolusi + 3 fully connected). VGG16 menekankan penggunaan filter 3x3 yang berurutan dan lapisan pooling untuk menangkap fitur spasial secara bertahap. Urutan lapisan VGG16 dapat diringkas sebagai berikut:



Gambar 3. Arsitektur VGG-16

Arsitektur dasar dari model VGG-16 seperti yang diilustrasikan pada Gambar 3, ini terdiri dari 13 layer konvolusional, 5 layer pooling, dan 3 layer fully connected. Model ini menerima citra input berukuran 64x64 piksel

dan mengolahnya melalui serangkaian blok konvolusi dengan filter berukuran 3×3 , yang semuanya menggunakan fungsi aktivasi ReLU. Setiap dua atau tiga layer konvolusional diikuti oleh operasi max pooling dengan ukuran kernel 2×2 untuk mengurangi dimensi spasial dan mencegah overfitting. Setelah lima tahap konvolusi dan pooling, fitur dikompresi menjadi bentuk vektor dan diproses oleh tiga fully connected layer di bagian akhir, yang berfungsi sebagai klasifikator untuk menentukan kelas dari citra masukan. Arsitektur ini dikenal karena kesederhanaannya, kedalaman tetap, dan performanya yang sangat baik dalam berbagai tugas klasifikasi gambar.

Setiap operasi konvolusi pada VGG16 dirumuskan secara umum sebagai:

$$Y_{i,j,k} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{c=0}^{C_{in}-1} X_{i+m,j+n,c} \cdot W_{m,n,c,k} \quad (2)$$

Pada Persamaan (2), $Y_{i,j,k}$ merepresentasikan hasil dari operasi konvolusi di posisi (i, j) pada channel ke- k dari output, sementara $X_{i+m,j+n,c}$ adalah nilai piksel dari input citra pada channel ke- c , dan $W_{m,n,c,k}$ merupakan bobot kernel pada posisi (m, n) untuk channel input ke- c dan output ke- k . Operasi ini dilakukan dengan menjumlahkan hasil perkalian antara nilai piksel dan bobot kernel pada area lokal $M \times N$, kemudian menjelajah seluruh channel input. Fungsi konvolusi ini memungkinkan model untuk mengekstraksi fitur spasial seperti tepi, pola tekstur, dan struktur penting dari citra input, yang kemudian diproses pada layer selanjutnya.

Evaluasi model adalah tahap krusial dalam proses pelatihan machine learning. Tujuannya adalah menentukan efektivitas model yang telah dikembangkan dalam menggeneralisasi informasi ke data baru yang belum pernah dianalisis sebelumnya. Dalam studi ini, kami mengevaluasi model yang dikembangkan dengan dataset citra HAM10000. Untuk penilaian, kami menggunakan beberapa metrik standar dalam tugas klasifikasi multi-kelas, termasuk *accuracy*, *precision*, *recall*, *f1-score*, dan *confusion matrix* [22].

Proses evaluasi diawali dengan melakukan prediksi terhadap data uji (*test data*) atau data validasi. Model menghasilkan keluaran berupa probabilitas dari setiap kelas untuk setiap sampel citra. Probabilitas ini kemudian dikonversi menjadi label prediksi dengan cara memilih kelas yang memiliki probabilitas tertinggi (*argmax*). Setelah didapatkan label prediksi dan label sebenarnya dari data uji, dilakukan evaluasi performa dengan beberapa metrik:

1. Accuracy (Akurasi)

Akurasi merupakan metrik evaluasi yang paling umum digunakan dalam pengukuran performa model klasifikasi. Metrik ini merepresentasikan proporsi jumlah hasil prediksi yang tepat dari total data yang diuji. Dengan kata lain, akurasi mengukur seberapa sering model memberikan prediksi yang tepat, baik untuk kelas positif maupun negatif. Akurasi menjadi indikator yang sangat intuitif karena menunjukkan persentase keberhasilan model secara keseluruhan. Perhitungan akurasi dinyatakan dalam Persamaan (3) berikut:

$$Accuracy = \frac{\sum_{i=1}^N 1(y_i = \hat{y}_i)}{N} \quad (3)$$

Dalam Persamaan (3), akurasi didefinisikan sebagai rasio jumlah prediksi yang benar terhadap total jumlah sampel. Notasi y_i menyatakan label sebenarnya dari data ke- i , sedangkan \hat{y}_i adalah hasil prediksi dari model untuk data tersebut. Fungsi indikator $1(y_i = \hat{y}_i)$ bernilai 1 jika prediksi sama dengan label sebenarnya (artinya model benar dalam mengklasifikasikan sampel ke- i), dan bernilai 0 jika tidak sama. Penjumlahan dilakukan dari $i = 1$ hingga N , di mana N merupakan jumlah total sampel.

Akurasi menjadi metrik evaluasi yang paling umum digunakan dalam tugas klasifikasi, karena memberikan gambaran seberapa sering model menghasilkan prediksi yang tepat secara keseluruhan. Pada data yang memiliki ketidakseimbangan (contohnya, jumlah class yang tidak merata), akurasi saja bisa jadi kurang representatif untuk mengukur performa model dengan tepat. Maka dari itu, akurasi umumnya disertai dengan *precision*, *recall*, serta *F1-score*. Akurasi telah lama

digunakan sebagai metrik standar untuk evaluasi klasifikasi [23].

2. Precision

Precision merupakan salah satu metrik evaluasi penting dalam sistem klasifikasi yang digunakan untuk mengukur tingkat akurasi prediction terhadap model untuk beberapa kelas. Presisi ini dapat menampilkan proporsi prediksi positif yang benar-benar relevan atau benar, sehingga metrik ini sangat berguna dalam konteks di mana kesalahan dalam memberikan prediksi positif (*false positive*) memiliki konsekuensi yang signifikan, seperti dalam klasifikasi email spam, diagnosis awal penyakit, atau sistem keamanan. *Precision* membandingkan jumlah prediksi positif tepat dengan seluruh prediksi positif dibuat oleh model, sebagaimana dirumuskan pada Persamaan (4) berikut:

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (4)$$

Pada Persamaan (4), TP_i (*True Positive*) adalah total data yang memang positif dan berhasil diklasifikasikan secara benar oleh model yang diklasifikasikan sebagai kelas i . Sedangkan FP_i (*False Positive*) adalah angka data yang faktanya tidak tergolong dalam kelas i , namun salah diklasifikasikan oleh model sebagai kelas i . Presisi mengukur seberapa akurat model dalam memprediksi suatu kelas tertentu, yaitu dari seluruh prediksi yang menyatakan “kelas i ”, berapa banyak yang benar. Nilai *precision* tinggi memperlihatkan bahwasanya model tidak sering memberikan prediksi positif palsu (*false positive*), sehingga pentingnya situasi ketika kesalahan prediksi positif (misalnya salah diagnosis kanker) harus diminimalkan.

Tingginya presisi menandakan bahwa model tidak sering memberikan prediksi positif palsu (*false positive*), sehingga pentingnya situasi ketika kesalahan prediksi positif (misalnya salah diagnosis kanker) harus diminimalkan. Presisi sering digunakan bersamaan dengan metrik *recall* untuk mengevaluasi keseimbangan antara ketepatan dan kelengkapan prediksi model, dan merupakan komponen penting dalam perhitungan F1-score [24].

3. Recall (Sensitivity)

Recall, yang juga dikenal sebagai sensitivitas atau *producer's accuracy*, satu dari beberapa metrik evaluasi yang dipakai di dalam sistem klasifikasi untuk mengukur kelengkapan identifikasi instance oleh model dari suatu kelas tertentu [25]. *Recall* sangat penting dalam konteks di mana kegagalan mendeteksi data positif (*false negative*) memiliki dampak yang serius, seperti dalam diagnosis penyakit atau deteksi penipuan. Nilai *recall* dihitung dengan membandingkan jumlah prediksi benar terhadap jumlah keseluruhan data yang seharusnya terdeteksi sebagai positif. Perhitungannya dirumuskan pada Persamaan (5) berikut:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (5)$$

Dalam Persamaan (5), TP_i (*True Positive*) menunjukkan jumlah data yang benar-benar termasuk dalam kelas i dan berhasil diklasifikasikan dengan tepat oleh model sebagai kelas tersebut. Sementara itu, FN_i (*False Negative*) adalah total dari data yang seharusnya masuk ke dalam kelas i , akan tetapi gagal dikenali oleh model (diklasifikasikan ke kelas lain).

Recall digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh instance positif yang sebenarnya ada dalam data. Dengan kata lain, metrik ini menunjukkan seberapa banyak data dari kelas i yang berhasil ditemukan oleh model dibandingkan dengan seluruh data yang benar-benar termasuk kelas i . Nilai *recall* yang tinggi berarti model memiliki kecenderungan tidak melewatkan (tidak mengabaikan) prediksi positif.

4. F1-Score

F1-score adalah salah satu evaluation matrix yang dipakai ketika menilai kinerja model klasifikasi, khususnya ketika terjadi ketidakseimbangan antara jumlah data pada masing-masing kelas. Metrik ini digunakan untuk memberikan gambaran yang lebih adil tentang performa model dengan mempertimbangkan kedua aspek utama dalam evaluasi klasifikasi, yaitu *precision* dan *recall*, sebagaimana dirumuskan pada Persamaan (6) berikut:

$$F1_i = 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i} \quad (6)$$

Dalam Persamaan (6), $Precision_i$ dan $Recall_i$ masing-masing merepresentasikan nilai presisi dan recall untuk kelas ke- i . Hubungan antara precision dan recall diukur melalui F1-score, yang merupakan harmonic mean dari keduanya bukan rata-rata aritmetika. Hal ini membuat F1-score menjadi metrik yang sensitif terhadap nilai rendah di salah satu komponen: jika precision atau recall sangat rendah, maka F1-score juga akan rendah, meskipun nilai lainnya tinggi.

F1-score sangat berguna dalam kondisi kelas tidak seimbang, di mana akurasi tidak cukup merepresentasikan performa model. Nilai F1-score berkisar antara 0 hingga 1, di mana nilai 1 menunjukkan performa sempurna (precision dan recall keduanya bernilai 1). Meskipun F1-score sangat populer, penggunaannya sebaiknya dikombinasikan dengan metrik lain karena tidak mempertimbangkan true negative [26].

4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil eksperimen yang telah dilakukan, dimulai dengan setup eksperimen yang terperinci. Selanjutnya, proses evaluasi model akan diuraikan secara komprehensif, diikuti dengan perbandingan kinerja berbagai algoritma berdasarkan metrik akurasi pelatihan, akurasi pengujian, dan nilai loss [27]. Analisis mendalam terhadap temuan-temuan ini akan disajikan untuk memberikan pemahaman menyeluruh tentang performa model yang diuji.

4.1. Setup Eksperimen

Eksperimen dalam penelitian ini diawali dengan proses *preprocessing* citra sebagai tahapan penting untuk menyiapkan input yang optimal sebelum memasuki arsitektur *Convolutional Neural Network (CNN)*. *Dataset* yang digunakan, yaitu HAM10000, berisi ribuan citra lesi kulit dalam format RGB. Setiap citra diubah ukurannya menjadi 64×64 piksel, sesuai dengan format input yang dibutuhkan oleh model ResNet50 dan VGG16.

Untuk meningkatkan kemampuan generalisasi model dan mencegah *overfitting*, diterapkan teknik augmentasi data khusus pada data pelatihan. Proses augmentasi ini

menggunakan kelas *ImageDataGenerator* dari Keras dengan sejumlah konfigurasi transformasi citra seperti rotasi, translasi, shear, dan flipping. "Detail lengkap konfigurasi augmentasi yang diimplementasikan dalam penelitian ini disajikan dalam Table 1.

Table 1. ImageDataGenerator telah diatur untuk semua model.

No	Nama Konfigurasi	Konfigurasi	Penjelasan
1	<code>preprocessing_function</code>	<code>preprocess_input</code>	Fungsi preprocessing khusus untuk model ResNet50 diterapkan
2	<code>rotation_range</code>	20	Rotasi acak hingga 20 derajat
3	<code>width_shift_range</code>	0.125	Pergeseran horizontal acak hingga 12.5% dari total lebar
4	<code>height_shift_range</code>	0.125	Pergeseran vertikal acak hingga 12.5% dari total tinggi
5	<code>vertical_flip</code>	True	Flip vertikal secara acak.
6	<code>horizontal_flip</code>	True	Flip horizontal secara acak

Augmentasi pada Table 1 ini hanya diterapkan pada data training, sedangkan data validasi dan data pengujian hanya melalui tahap *preprocessing* menggunakan fungsi `preprocess_input`, tanpa modifikasi tambahan. Hal ini bertujuan untuk memastikan bahwa evaluasi performa model dilakukan secara objektif pada data yang representatif terhadap kondisi sebenarnya, tanpa pengaruh variasi buatan. Konsistensi dalam proses *preprocessing* antara tahap pelatihan dan pengujian menjadi kunci untuk menjaga integritas eksperimen. Penggunaan augmentasi secara selektif pada data pelatihan memungkinkan model belajar dari variasi bentuk, ukuran, dan orientasi citra tanpa merusak keakuratan evaluasi.

4.2. Evaluasi Model

a. Classification Report

Evaluasi performa klasifikasi citra lesi kulit dilakukan menggunakan metrik *precision*, *recall*, dan *F1-score* pada masing-masing kelas target, sebagaimana ditampilkan pada *classification report* untuk model

ResNet-50 dan VGG-16. Hasil evaluasi menunjukkan adanya perbedaan signifikan dalam kemampuan masing-masing model dalam mengidentifikasi kelas-kelas lesi kulit.

Table 2 Classification Report ResNet50

Class	Precision	Recall	F1-score	Support
Akiec	0.69	0.55	0.61	49
bcc	0.73	0.63	0.67	77
bkl	0.58	0.72	0.64	165
Df	0.56	0.45	0.50	17
Mel	0.61	0.54	0.58	167
Nv	0.91	0.92	0.92	1006
Vasc	1.00	0.86	0.92	22
Accuracy				
Macro avg			0.82	1503
Weighted avg	0.73	0.67	0.69	1503
Weighted avg	0.82	0.82	0.82	1503

Hasil classification report yang ditunjukkan pada Table 2 menunjukkan kinerja model dalam mengenali tujuh kelas lesi kulit pada dataset HAM10000. Secara umum, model memiliki akurasi sebesar 82% dari total 1503 data uji. Kelas nevus (Nv) merupakan kelas dengan performa tertinggi, dengan precision dan recall sebesar 91% dan 92%, serta f1-score 92%, yang mencerminkan dominasi jumlah data pada kelas ini dan kemampuan model mengenalinya dengan sangat baik. Sebaliknya, performa model terhadap kelas minoritas seperti dermatofibroma (Df) dan akiec masih kurang optimal, yang terlihat dari *f1-score* yang relatif rendah, masing-masing 0.50 dan 0.61. Rata-rata makro (*macro avg*) *f1-score* hanya mencapai 0.69, mengindikasikan bahwa model belum seimbang dalam mengenali semua kelas secara merata. Namun, nilai rata-rata berbobot (*weighted avg*) tetap tinggi di angka 0.82, yang menunjukkan bahwa model lebih akurat dalam mengenali kelas-kelas mayoritas yang memiliki jumlah data lebih besar. Hal ini menegaskan pentingnya penanganan ketidakseimbangan kelas untuk meningkatkan kemampuan model pada deteksi lesi yang lebih jarang ditemukan namun kritis, seperti *melanoma* dan *dermatofibroma*.

Table 3 Classification Report VGG16

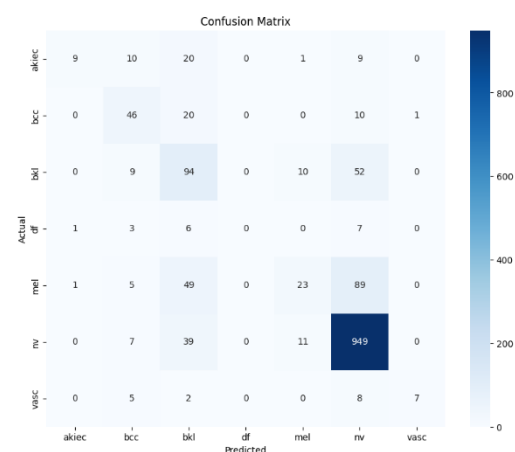
Class	Precision	Recall	F1-score	Support
Akiec	0.00	0.00	0.00	49
Bcc	0.41	0.68	0.51	77
Bkl	0.32	0.45	0.38	165
Df	0.00	0.00	0.00	17
Mel	0.00	0.00	0.00	167
Nv	0.83	0.94	0.88	1006
Vasc	0.00	0.00	0.00	22
Accuracy				
Macro avg	0.22	0.30	0.25	1503
Weighted avg	0.61	0.71	0.66	1503

Hasil classification report yang terdapat pada Table 3 menunjukkan bahwa model memiliki akurasi sebesar 71% dari total 1503 data uji pada dataset HAM10000. Kelas *nevus* (Nv) merupakan satu-satunya kelas dengan performa tinggi, dengan precision 83%, recall 94%, dan f1-score 88%, mencerminkan dominasi jumlah data pada kelas ini. Sebaliknya, performa model pada kelas-kelas minoritas seperti *akiec*, *melanoma* (Mel), dan *dermatofibroma* (Df) sangat rendah, dengan f1-score 0.00, menandakan ketidakmampuan model dalam mengenali lesi tersebut. Nilai macro average f1-score hanya 0.25, sedangkan weighted average mencapai 0.66, menunjukkan ketidakseimbangan kinerja model terhadap seluruh kelas. Temuan ini menekankan perlunya penanganan ketidakseimbangan data agar model dapat mengenali semua jenis lesi kulit secara lebih adil dan akurat..

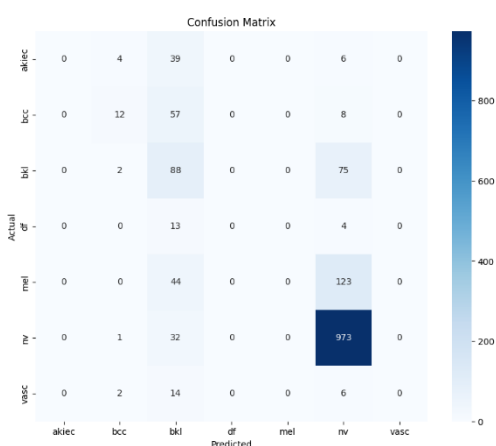
b. Confusion Matrix

Evaluasi performa model klasifikasi pada penelitian ini ditunjukkan melalui *confusion matrix* yang ditampilkan pada Gambar a dan Gambar b. Gambar 4 ini merepresentasikan perbandingan performa dua model CNN dalam mengklasifikasikan citra lesi kulit dari tujuh kelas yang berbeda, yaitu akiec, bcc, bkl, df, mel, nv, dan vasc.

Melalui analisis *confusion matrix*, kemampuan model dalam mengidentifikasi setiap kelas dapat diamati secara lebih rinci. Matriks ini tidak hanya menunjukkan jumlah prediksi yang benar, tetapi juga kesalahan klasifikasi pada tiap kategori. Visualisasi ini penting untuk memahami performa model secara menyeluruh, termasuk mengidentifikasi kelas yang sering salah diklasifikasikan. Misalnya, dapat diketahui apakah melanoma sering keliru dikenali sebagai lesi jinak. Dengan demikian, *confusion matrix* memberikan gambaran yang lebih mendalam mengenai kekuatan dan kelemahan model dalam klasifikasi multikelas. Detail lebih lanjut dapat dilihat pada Gambar a dan Gambar b.



a. Training Model Resnet50



b. Training Model VGG16

Gambar 4. Confusion Matrix

Berdasarkan Gambar 4, Model ResNet50 menunjukkan performa yang *superior*

dibandingkan VGG16 pada metrik agregat. ResNet50 mencapai *precision (weighted)* sebesar 0.81, *recall (weighted)* 0.80, dan *F1-Score (weighted)* 0.80. Sementara itu, VGG16 mencatatkan *precision (weighted)* sebesar 0.72, *recall (weighted)* 0.75, dan *F1-Score (weighted)* 0.73. Perbedaan signifikan pada ketiga metrik ini mengindikasikan bahwa ResNet50 memiliki kemampuan yang lebih baik dalam membuat prediksi positif yang relevan, mendeteksi sebagian besar *instance* positif, dan mencapai keseimbangan yang lebih baik antara *precision* dan *recall* secara keseluruhan.

Pemeriksaan lebih lanjut pada *precision*, *recall*, dan *F1-score* per kelas memberikan gambaran yang lebih rinci. Untuk ResNet50, performa pada kelas mayoritas *nv* sangat kuat (*precision: 0.91, recall: 0.92, F1-score: 0.91*). Meskipun demikian, ResNet50 juga menunjukkan peningkatan kinerja pada beberapa kelas minoritas penting dibandingkan VGG16. Sebagai contoh, *recall* untuk *melanoma (mel)* pada ResNet50 adalah 0.60 (meningkat dari 0.30 pada VGG16), dan *recall* untuk *dermatofibroma (df)* adalah 0.42 (meningkat dari 0.08 pada VGG16).

Sebaliknya, VGG16 menunjukkan *recall* yang jauh lebih rendah untuk kelas *melanoma (mel)* hanya 0.30, dan sangat rendah untuk *dermatofibroma (df)* yaitu 0.08, meskipun memiliki *recall* yang tinggi pada kelas *nv* (0.93). Perbandingan per kelas ini menegaskan bahwa ResNet50, meskipun masih menghadapi tantangan pada beberapa kelas minoritas, menunjukkan kemampuan generalisasi dan deteksi yang lebih baik secara keseluruhan, khususnya pada kelas kritis seperti melanoma. Ini menyoroti pentingnya analisis mendalam di luar metrik agregat untuk evaluasi model yang adil dan relevan secara klinis.

4.3. Perbandingan Model

Evaluasi lebih lanjut terhadap performa model dilakukan dengan membandingkan hasil dari dua arsitektur CNN populer, yaitu VGG-16 dan ResNet-50, baik versi original berdasarkan referensi sebelumnya maupun versi modifikasi yang diusulkan dalam penelitian ini [11]. Hasil perbandingan metrik akurasi validasi, akurasi pengujian, dan loss ditunjukkan pada Table 4.

Table 4 Perbandingan Algoritma Vgg-16 dan ResNet50

Approaches	Pre (%)	F1-s (%)	Acc (%)
SoftMax	86.33	88.44	85.74
Naïve Bayes	85.62	84.34	84.56
M-SVM	85.24	85.85	82.66
W-KNN	83.23	80.31	79.90
Proposed Classifier	86.33	88.44	89.30
ResNet50	82.74	81.73	82.14
VGG-16	61.07	65.66	71.32

Berdasarkan hasil perbandingan antara model yang terdapat pada Table 4 dalam penelitian ini, yaitu ResNet50 dan VGG-16, dengan pendekatan-pendekatan klasifikasi dari jurnal sebelumnya, terlihat bahwa performa kedua model menunjukkan hasil yang berbeda secara signifikan. Model ResNet50 mencatatkan precision sebesar 82,74%, f1-score sebesar 81,73%, dan akurasi sebesar 82,14%, yang menempatkannya pada posisi yang kompetitif di antara pendekatan-pendekatan dari jurnal terdahulu. Jika dibandingkan secara khusus dengan metode *Weighted K-Nearest Neighbor* (W-KNN), ResNet50 berhasil melampaui metode tersebut pada dua metrik penting, yaitu f1-score (81,73% vs 80,31%) dan akurasi (82,14% vs 79,90%). Hal ini menunjukkan bahwa pendekatan *deep learning modern* seperti ResNet50 memiliki kemampuan generalisasi yang lebih baik dibandingkan metode klasikal tertentu, meskipun tanpa optimasi ekstrem.

Walaupun performa ResNet50 masih berada sedikit di bawah metode seperti *SoftMax*, *Naïve Bayes*, dan *Proposed Classifier* dari jurnal yang dilaporkan memiliki *f1-score* dan akurasi di atas 85%, hasil yang dicapai ResNet50 tetap menunjukkan potensi yang kuat untuk digunakan dalam tugas klasifikasi multikelas, terutama jika dipadukan dengan strategi *fine-tuning* dan teknik regularisasi yang lebih optimal.

Sebaliknya, model VGG-16 menunjukkan performa yang jauh lebih rendah dibandingkan semua pendekatan yang dibandingkan, dengan *precision* hanya 61,07%, *f1-score* sebesar 65,66%, dan akurasi sebesar 71,32%. Hasil ini mengindikasikan bahwa VGG-16 belum mampu menangani kompleksitas data secara

efektif dalam konfigurasi saat ini. Rendahnya performa ini kemungkinan disebabkan oleh arsitektur VGG yang lebih sederhana dibandingkan ResNet, serta belum optimalnya proses pelatihan seperti penggunaan learning rate, dropout, atau pengaturan augmentasi data.

Secara keseluruhan, dapat disimpulkan bahwa ResNet50 merupakan model yang lebih unggul dibandingkan VGG-16, dan juga mampu mengungguli salah satu metode klasikal dalam jurnal sebelumnya, yaitu W-KNN, dalam aspek f1-score dan akurasi. Meskipun masih belum melampaui metode *Proposed Classifier* dari jurnal tersebut, ResNet50 menunjukkan performa yang menjanjikan dan relevan untuk dikembangkan lebih lanjut dalam klasifikasi citra multikelas, khususnya pada domain medis atau tugas-tugas sejenis.

5. KESIMPULAN

Dalam penelitian ini digunakan dua arsitektur deep learning, yaitu ResNet50 dan VGG-16, untuk tugas klasifikasi citra dermoskopi pada dataset HAM10000. Hasil pengujian menunjukkan bahwa ResNet50 memiliki performa klasifikasi yang lebih unggul dibandingkan VGG-16, baik dari segi akurasi, precision, maupun f1-score. ResNet50 mencatatkan akurasi sebesar 82,14%, f1-score 81,73%, dan precision 82,74%, sedangkan VGG-16 hanya mencapai akurasi 71,32%, f1-score 65,66%, dan precision 61,07%.

Ketika dibandingkan dengan metode-metode dalam jurnal sebelumnya, seperti *SoftMax*, *Naïve Bayes*, *Multi-class SVM*, dan *Weighted K-NN*, model ResNet50 menunjukkan kinerja yang kompetitif. Khususnya, ResNet50 berhasil melampaui metode W-KNN dalam hal *f1-score* dan akurasi (masing-masing 81,73% vs 80,31%, dan 82,14% vs 79,90%), menunjukkan keunggulan arsitektur *deep learning modern* dibandingkan metode klasikal tertentu. Meski demikian, performa ResNet50 masih sedikit di bawah teknik terbaik dalam jurnal, yaitu *Proposed Classifier*, yang mencapai akurasi tertinggi sebesar 89,30%.

Dari segi pelatihan, model ResNet50 membutuhkan waktu pelatihan yang lebih lama, yaitu sekitar 49 menit 33 detik (dengan rata-rata 74 detik per epoch), dibandingkan VGG-16 yang hanya membutuhkan 43 menit 2 detik (64 detik per epoch). Hal ini sejalan dengan kompleksitas arsitektur ResNet50 yang lebih dalam dan memiliki shortcut connection, yang memungkinkan pembelajaran fitur yang lebih kuat namun dengan konsekuensi waktu komputasi yang lebih tinggi.

Secara keseluruhan, penelitian ini menunjukkan bahwa ResNet50 merupakan arsitektur yang lebih andal daripada VGG-16 dalam klasifikasi kanker kulit berbasis citra dermoskopi, baik dari sisi performa metrik maupun kedalaman pemodelan. Meskipun memerlukan waktu pelatihan lebih lama, ResNet50 memberikan hasil yang lebih konsisten dan akurat, menjadikannya kandidat kuat untuk pengembangan sistem pendukung diagnosis medis berbasis kecerdasan buatan.

DAFTAR PUSTAKA

- [1] M. Berwick and A. Garcia, "Solar UV Exposure and Mortality from Skin Tumors: An Update BT - Sunlight, Vitamin D and Skin Cancer," J. Reichrath, Ed., Cham: Springer International Publishing, 2020, pp. 143–154. doi: 10.1007/978-3-030-46227-7_7.
- [2] F. J., "The Global Cancer Observatory." [Online]. Available: <https://gco.iarc.who.int/media/globocan/factsheets/populations/360-indonesia-fact-sheet.pdf>
- [3] "Analisis Perbandingan Kinerja Algoritma Multilayer Perceptron dan K-Nearest Neighbor pada Klasifikasi Tipe Migrain," *J. Teknol. dan Inf.*, vol. 13, no. 1 SE-Article, pp. 44–55, Feb. 2023.
- [4] P. Rakshit, A. Ghosh, C. Chakraborty, J. Paul, and D. Das, "Skin Cancer Detection Using Deep Learning," *Lect. Notes Electr. Eng.*, vol. 1243 LNEE, pp. 359–372, 2025, doi: 10.1007/978-981-97-6465-5_29.
- [5] and T. R. Y. Chen, H. Chen, M. Han, B. Liu, Q. Chen, "A Novel Computing Power Allocation Algorithm for Blockchain System in Multiple Mining Pools Under Withholding Attack," *IEEE Access*, vol. 8, pp. 155630–155644, 2020, doi: doi: 10.1109/ACCESS.2020.3017716.
- [6] M. T. Z. Quen, J. Mountstephens, Y. G. Teh, and J. Teo, "Medical image interpretation training with a low-cost eye tracking and feedback system: A preliminary study," *Healthc. Technol. Lett.*, vol. 8, no. 4, pp. 97–103, Aug. 2021, doi: <https://doi.org/10.1049/htl2.12014>.
- [7] A. Mahbod, G. Schaefer, C. Wang, G. Dorffner, R. Ecker, and I. Ellinger, "Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification," *Comput. Methods Programs Biomed.*, vol. 193, p. 105475, 2020, doi: <https://doi.org/10.1016/j.cmpb.2020.105475>.
- [8] X. Li, X. Zhao, H. Ma, and B. Xie, "Image Analysis and Diagnosis of Skin Diseases - A Review," *Curr. Med. Imaging*, vol. 19, no. 3, pp. 199–242, 2023, doi: <https://doi.org/10.2174/1573405618666220516114605>.
- [9] D. Meedeniya, S. De Silva, L. Gamage, and U. Isuranga, "Skin cancer identification utilizing deep learning: A survey," *IET Image Process.*, vol. 18, no. 13, pp. 3731–3749, 2024, doi: 10.1049/ipr2.13219.
- [10] T. Chagovets *et al.*, "Automation of Target Delivery and Diagnostic Systems for High Repetition Rate Laser-Plasma Acceleration," 2021. doi: 10.3390/app11041680.
- [11] R. Alabdujabbar and H. Alshamlan, "Intelligent multiclass skin cancer detection using convolution neural networks," *Comput. Mater. Contin.*, vol. 69, no. 1, pp. 831–847, 2021, doi: 10.32604/cmc.2021.018402.
- [12] and Y. L. X. Zhou, J. Feng, "Non-intrusive load decomposition based on CNN-LSTM hybrid deep learning model," *Energy Reports*, vol. 7, pp. 5762–5771, 2021, doi: doi: 10.1016/j.egyr.2021.09.001.
- [13] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, 2021, doi: doi: 10.1016/j.aci.2018.08.003.
- [14] T. Chagovets *et al.*, "Automation of Target Delivery and Diagnostic Systems for High Repetition Rate Laser-Plasma Acceleration," 2021, doi: doi: 10.3390/app11041680.
- [15] and B. X. X. Li, X. Zhao, H. Ma, "Image Analysis and Diagnosis of Skin Diseases - A Review," *Curr. Med. Imaging*, vol. 19, no. 3, pp. 199–242, 2023, doi: <https://doi.org/10.2174/1573405618666220516114605>.
- [16] A. D. A. N. Sgd, "KLASIFIKASI CITRA PENYAKIT KANKER MULUT MENGGUNAKAN ARSITEKTUR RESNET50 OPTIMASI," vol. 12, no. 3, 2024.
- [17] "Studi Perbandingan Deteksi Intrusi Jaringan Menggunakan Machine Learning: (Metode SVM dan ANN)," *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 152–164, 2023, doi: doi:

- 10.34010/jati.v13i2.10484.
- [18] U. Ali, I. H. Lee, and M. T. Mahmood, "Guided image filtering in shape-from-focus: A comparative analysis," *Pattern Recognit.*, vol. 111, p. 107670, 2021, doi: <https://doi.org/10.1016/j.patcog.2020.107670>.
- [19] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nat. Med.*, vol. 25, no. 1, pp. 24–29, 2019, doi: 10.1038/s41591-018-0316-z.
- [20] S. R. Yang, H. C. Yang, F. R. Shen, and J. Zhao, "Image Data Augmentation for Deep Learning: A Survey," *Ruan Jian Xue Bao/Journal Softw.*, vol. 36, no. 3, pp. 1390–1412, 2025, doi: 10.13328/j.cnki.jos.007263.
- [21] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, no. xxxx, p. 105524, 2020, doi: 10.1016/j.asoc.2019.105524.
- [22] S. Sharma and K. Guleria, "A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks," *Procedia Comput. Sci.*, vol. 218, pp. 357–366, 2022, doi: 10.1016/j.procs.2023.01.018.
- [23] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.
- [24] X. Zhou, J. Feng, and Y. Li, "Non-intrusive load decomposition based on CNN–LSTM hybrid deep learning model," *Energy Reports*, vol. 7, pp. 5762–5771, 2021, doi: 10.1016/j.egyr.2021.09.001.
- [25] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, 2021, doi: <https://doi.org/10.1016/j.isprsjprs.2020.12.010>.
- [26] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, 2020, doi: 10.1186/s12864-019-6413-7.
- [27] "Studi Perbandingan Deteksi Intrusi Jaringan Menggunakan Machine Learning: (Metode SVM dan ANN)," *J. Teknol. dan Inf.*, vol. 13, no. 2 SE-Article, pp. 152–164, Aug. 2023, doi: 10.34010/jati.v13i2.10484.