

ANALISIS PERBANDINGAN EFEKTIVITAS ARSITEKTUR CONVOLUTIONAL NEURAL NETWORK DENGAN FOCAL LOSS PADA KASUS IMBALANCE DATA KLASIFIKASI HEWAN

Farhan Zen¹; Dian Ade Kurnia²; Yudhistira Arie Wijaya³; Puji Pramudya Marta⁴;

Program Studi Teknik Informatika¹
Program Studi Manajemen Informatika²⁴
Program Studi Sistem Informasi³

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
farhanzein2705@gmail.com

(*) Corresponding Author : farhanzein2705@gmail.com
Published : 31 Desember 2025

Abstract— This study evaluates the effectiveness of three Convolutional Neural Network (CNN) architectures ResNet18, EfficientNet_B0, and MobileNetV3_Large in addressing class imbalance in animal image classification. The models were trained using Focal Loss to enhance sensitivity to minority classes. The methodology includes data preprocessing, data augmentation, and model training with variations of Focal Loss parameters (α and γ). Model performance was assessed using accuracy, precision, recall, and F1-score. Experimental results show that MobileNetV3_Large achieved the highest overall performance with an accuracy of 0.767, precision of 0.656, recall of 0.657, and an F1-score of 0.647. ResNet18 demonstrated consistent performance with an accuracy of 0.768 and an F1-score of 0.617, while EfficientNet_B0 performed less effectively on minority classes. The analysis confirms that Focal Loss improves the model's focus on challenging samples, thereby enhancing minority class performance. The primary contribution of this work lies in providing empirical insights into the comparative performance of CNN architectures and optimal Focal Loss configurations for imbalanced datasets. This study supports the development of more accurate and adaptive animal image classification systems.

Keywords: CNN; Focal Loss; class imbalance; animal classification; deep learning.

Abstrak— Penelitian ini menganalisis efektivitas tiga arsitektur Convolutional Neural Network (CNN) ResNet18, EfficientNet_B0, dan MobileNetV3_Large dalam menangani ketidakseimbangan kelas pada klasifikasi citra hewan. Model dilatih menggunakan Focal Loss untuk meningkatkan sensitivitas terhadap kelas minoritas. Metode yang digunakan mencakup preprocessing data, augmentasi data, serta pelatihan model dengan variasi parameter Focal Loss (α dan γ). Evaluasi dilakukan menggunakan akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa MobileNetV3_Large mencapai performa paling baik dengan akurasi 0,767, presisi 0,656, recall 0,657, dan F1-score 0,647. ResNet18 menghasilkan performa stabil dengan akurasi 0,768 dan F1-score 0,617, sedangkan EfficientNet_B0 menunjukkan penurunan kinerja terutama pada kelas dengan jumlah sampel kecil. Analisis memperlihatkan bahwa Focal Loss membantu model lebih fokus pada sampel sulit sehingga meningkatkan performa kelas minoritas. Kontribusi utama penelitian ini adalah memberikan pemahaman empiris mengenai perbandingan arsitektur CNN dan konfigurasi Focal Loss terbaik untuk data yang tidak seimbang, serta memberikan dasar bagi pengembangan sistem klasifikasi citra hewan yang lebih akurat, adaptif, dan robust.

Kata kunci: CNN; Focal Loss; ketidakseimbangan kelas; klasifikasi hewan; deep learnin

INTRODUCTION

Kemajuan teknologi deep learning telah membawa perubahan besar dalam bidang visi komputer, terutama dalam tugas pengenalan dan klasifikasi citra digital. Salah satu arsitektur yang paling berpengaruh dalam perkembangan ini adalah Convolutional Neural Network (CNN), yang dikenal mampu mengekstraksi fitur spasial dan hierarkis dari data citra secara efisien. CNN telah diaplikasikan secara luas pada berbagai bidang seperti pengenalan wajah, deteksi objek, diagnosis medis berbasis citra, hingga identifikasi spesies hewan. Keberhasilan CNN dalam berbagai domain ini menunjukkan potensinya sebagai metode unggulan dalam pemrosesan citra digital (Alzubaidi et al., 2021). Dalam konteks penelitian akademik, penerapan CNN pada klasifikasi citra hewan memiliki urgensi tinggi karena mendukung pengembangan sistem kecerdasan buatan untuk konservasi dan pengenalan spesies secara otomatis, sejalan dengan perkembangan teknologi berbasis data visual yang terus meningkat.

Perkembangan ini diiringi oleh munculnya tantangan baru, salah satunya adalah permasalahan ketidakseimbangan data (class imbalance), yaitu kondisi di mana jumlah data pada kelas tertentu jauh lebih banyak dibandingkan kelas lainnya. Ketidakseimbangan ini menyebabkan model cenderung lebih fokus pada kelas mayoritas dan mengabaikan kelas minoritas. Akibatnya, meskipun akurasi keseluruhan model tampak tinggi, kemampuan dalam mengenali kelas minoritas menurun secara signifikan (Dablain, Jacobson, et al., 2023). Dalam konteks klasifikasi citra hewan, fenomena ini sering ditemukan karena ketersediaan citra antarspesies yang tidak merata—misalnya hewan langka yang sulit didokumentasikan. Oleh sebab itu, penelitian mengenai strategi yang mampu meningkatkan performa CNN pada kondisi data tidak seimbang menjadi relevan dan signifikan untuk dilakukan.

Namun pada kenyataannya, penelitian terkait class imbalance dalam konteks klasifikasi citra hewan masih belum banyak dilakukan secara mendalam. Sebagian besar studi terdahulu berfokus pada arsitektur CNN tertentu tanpa melakukan perbandingan menyeluruh terhadap berbagai model modern yang lebih efisien, seperti ResNet, EfficientNet, dan MobileNet ((Yeung, 2024). Selain itu, masih jarang penelitian yang menelaah pengaruh penerapan fungsi loss adaptif seperti Focal Loss terhadap performa model CNN dalam konteks data tidak seimbang. Padahal, Focal Loss telah terbukti efektif mengurangi dominasi kelas mayoritas dengan cara menurunkan kontribusi loss

dari contoh yang mudah diprediksi dan memberikan bobot lebih besar pada contoh sulit atau kelas minoritas (T.-Y. Lin et al., 2019). Hal ini menunjukkan adanya kesenjangan penelitian (research gap) yang perlu diisi melalui analisis komparatif terhadap berbagai arsitektur CNN dengan penerapan Focal Loss untuk mengatasi ketidakseimbangan data.

Masih belum banyak pula penelitian yang menyoroti bagaimana parameter dalam Focal Loss, seperti α (alpha) dan γ (gamma), perlu disesuaikan dengan karakteristik masing-masing arsitektur CNN. Studi yang dilakukan oleh Hossain, Betts, dan Paplinski (2021) menunjukkan bahwa penyesuaian parameter fungsi loss dapat meningkatkan stabilitas pelatihan dan konvergensi model, namun belum ada kajian sistematis mengenai interaksi antara struktur CNN dan parameter tersebut. Setiap arsitektur CNN memiliki mekanisme ekstraksi fitur yang berbeda sehingga respons terhadap perubahan fungsi loss juga dapat bervariasi. Dengan demikian, pendekatan satu model dan satu fungsi loss untuk semua kasus tidak selalu memberikan hasil yang optimal, dan penelitian yang membandingkan pengaruh parameter Focal Loss antararsitektur CNN masih diperlukan.

Untuk mengatasi masalah ketidakseimbangan data, berbagai pendekatan telah diusulkan oleh penelitian sebelumnya. Pendekatan tradisional meliputi teknik resampling seperti oversampling kelas minoritas dan undersampling kelas mayoritas, serta cost-sensitive learning yang memberikan bobot kesalahan lebih tinggi pada kelas dengan jumlah data sedikit (M. Carvalho et al., 2023). Selain itu, modifikasi fungsi loss telah menjadi solusi populer dalam dekade terakhir. Focal Loss, Dual Focal Loss, dan Batch-Balanced Focal Loss dikembangkan untuk meningkatkan sensitivitas model terhadap kelas minoritas dengan menyesuaikan tingkat kontribusi loss berdasarkan kesulitan prediksi sampel (D. Nie, 2023). Pendekatan ini membantu model untuk fokus pada contoh sulit dan mengurangi bias terhadap kelas mayoritas, menghasilkan peningkatan performa yang signifikan pada dataset tidak seimbang.

Beberapa penelitian empiris telah membuktikan efektivitas Focal Loss dan turunannya dalam berbagai domain. (Y. Nie et al., 2023) menunjukkan bahwa penerapan Focal Loss pada model hibrid CNN-Transformer secara signifikan meningkatkan akurasi klasifikasi lesi kulit dengan distribusi data tidak seimbang. (Wen, 2023) menggunakan pendekatan Hybrid 3D-2D CNN dengan Focal Loss untuk klasifikasi citra hiperspektral dan menemukan peningkatan recall pada kelas minoritas tanpa mengurangi precision pada kelas

mayoritas. Sementara itu, (Singh et al., 2023) mengusulkan Batch-Balanced Focal Loss untuk memperbaiki kestabilan pelatihan dan mengurangi bias gradien selama proses pembelajaran. Meskipun hasilnya positif, sebagian besar penelitian tersebut dilakukan pada bidang medis dan industri, bukan pada domain klasifikasi citra hewan, sehingga penerapannya belum dapat digeneralisasikan.

Khusnil et al. (2023) menyoroti bahwa CNN masih menunjukkan kelemahan dalam menangani distribusi data yang tidak seimbang, terutama dalam mempelajari representasi fitur dari kelas minoritas. (Yeung, 2024) menambahkan bahwa meskipun penelitian imbalance-aware learning telah berkembang pesat, studi yang secara sistematis membandingkan performa berbagai arsitektur CNN di bawah kondisi ketidakseimbangan data yang sama masih sangat terbatas. (Oliveira, 2025) juga menemukan bahwa penelitian klasifikasi spesies hewan berbasis deep learning umumnya tidak mengkaji hubungan antara arsitektur CNN dan fungsi loss yang digunakan. Oleh karena itu, penelitian ini hadir untuk mengisi kekosongan tersebut dengan melakukan analisis komparatif yang mendalam terhadap beberapa arsitektur CNN modern dengan penerapan Focal Loss pada klasifikasi citra hewan.

Berdasarkan tinjauan literatur tersebut, dapat diidentifikasi tiga kesenjangan penelitian utama yang melandasi pentingnya studi ini. Pertama, masih kurangnya penelitian komparatif antararsitektur CNN modern (ResNet18, EfficientNet_B0, dan MobileNetV3_Large) dalam konteks penerapan Focal Loss untuk mengatasi class imbalance. Kedua, belum adanya kajian yang menganalisis pengaruh variasi parameter Focal Loss terhadap performa model dan stabilitas pelatihan. Ketiga, terbatasnya laporan mengenai evaluasi performa per kelas, terutama metrik recall dan F1-score pada kelas minoritas. Dengan demikian, penelitian ini berupaya untuk menjawab kesenjangan tersebut dengan pendekatan komparatif yang berbasis eksperimen.

Tujuan dari penelitian ini adalah untuk menganalisis dan membandingkan efektivitas tiga arsitektur CNN populer ResNet18, EfficientNet_B0, dan MobileNetV3_Large yang dilatih menggunakan Focal Loss dalam menghadapi ketidakseimbangan data pada klasifikasi citra hewan. Penelitian ini juga bertujuan mengevaluasi pengaruh variasi parameter Focal Loss terhadap stabilitas pelatihan dan performa klasifikasi antararsitektur. Hasil penelitian diharapkan dapat memberikan kontribusi ilmiah dalam bentuk panduan praktis mengenai pemilihan arsitektur CNN dan konfigurasi loss function yang optimal untuk dataset tidak

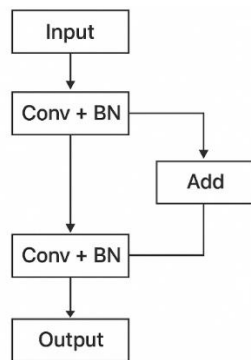
seimbang. Dengan demikian, penelitian ini tidak hanya memberikan sumbangan pada pengembangan teori di bidang deep learning, tetapi juga mendukung pengembangan sistem klasifikasi citra berbasis kecerdasan buatan yang lebih adil, efisien, dan berkelanjutan.

MATERIALS AND METHODS

Penelitian ini menggunakan tiga arsitektur Convolutional Neural Network (CNN) modern, yaitu ResNet18, EfficientNet_B0, dan MobileNetV3-Large, yang seluruhnya dilatih melalui proses fine-tuning dengan bobot pretrained ImageNet. Penggunaan ketiga arsitektur ini memungkinkan evaluasi performa yang komprehensif pada dataset kecil dan tidak seimbang, menggunakan metrik akurasi, precision, recall, dan F1-score. Pemilihan model serta teknik fine-tuning yang digunakan sejalan dengan temuan penelitian terbaru mengenai efektivitas arsitektur CNN dan transfer learning pada tugas klasifikasi citra (M. Tan & Le, 2021).

ResNet18

ResNet18 merupakan arsitektur CNN yang menggunakan konsep Residual Learning melalui skip connection untuk menjaga aliran gradien tetap stabil selama pelatihan, sehingga mampu mengatasi masalah vanishing gradient pada jaringan yang lebih dalam. Arsitektur ini terdiri dari 18 layer utama dan dikenal sebagai model yang ringan namun tetap memiliki performa baik pada tugas klasifikasi citra. Struktur residual block memungkinkan jaringan mempelajari residual mapping secara lebih efisien, sehingga proses pelatihan menjadi lebih stabil dan cepat konvergen. Dalam penelitian ini, ResNet18 diinisialisasi menggunakan bobot pretrained ImageNet agar dapat memanfaatkan fitur visual dasar yang sudah dipelajari sebelumnya. Selain itu, fully connected layer diganti menjadi 5 output kelas sesuai kebutuhan klasifikasi hewan (Cheetah, Jaguar, Leopard, Lion, dan Tiger), sehingga model dapat beradaptasi dengan tugas klasifikasi khusus dataset penelitian. Pemilihan ResNet18 sejalan dengan temuan terbaru yang menunjukkan efektivitas arsitektur residual untuk dataset kecil hingga menengah serta pada penerapan transfer learning (He, 2025). Pada Gambar 3.5 memperlihatkan struktur dasar dari residual block yang menjadi komponen utama dalam arsitektur ResNet18.



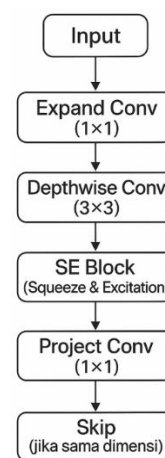
Gambar 1 Struktur Sederhana ResNet18(He, 2025).

Pada Gambar 1 ditampilkan struktur dasar residual block pada arsitektur ResNet18. Blok ini terdiri atas dua lapisan Convolution yang masing-masing diikuti Batch Normalization (BN) pada jalur utama. Input x diproses melalui kedua lapisan tersebut untuk menghasilkan transformasi fitur $F(x)$. Secara paralel, jalur skip connection membawa input x tanpa mengalami transformasi, kemudian dijumlahkan dengan $F(x)$ pada tahap Add. Hasil penjumlahan ini selanjutnya melewati aktivasi ReLU untuk menghasilkan output akhir. Mekanisme skip connection memungkinkan jaringan mempelajari residual mapping sehingga aliran gradien menjadi lebih stabil dan mampu mencegah vanishing gradient. Pendekatan ini terbukti meningkatkan stabilitas pelatihan serta mempercepat konvergensi pada jaringan CNN modern, khususnya arsitektur ResNet (He, 2025).

EfficientNet_B0

EfficientNet-B0 merupakan arsitektur CNN modern yang dirancang untuk memperoleh keseimbangan optimal antara akurasi dan efisiensi komputasi. Model ini memanfaatkan compound scaling, yaitu metode penskalaan terintegrasi yang meningkatkan kedalaman (depth), lebar jaringan (width), dan resolusi input secara proporsional sehingga performa dapat meningkat tanpa penambahan kompleksitas berlebih (M. Tan & Le, 2021). Dalam penelitian ini, EfficientNet-B0 digunakan sebagai salah satu model pembanding dan diinisialisasi dengan bobot pretrained ImageNet. Classification layer terakhir kemudian dimodifikasi menjadi 5 kelas keluaran sesuai kategori hewan: Cheetah, Jaguar, Leopard, Lion, dan Tiger. Arsitektur EfficientNet-B0 dibangun menggunakan MBConv Block (Mobile Inverted Bottleneck), yang dikenal efisien dalam mengekstraksi fitur sambil mempertahankan beban komputasi yang rendah (Howard et al., 2021).

Selain itu, model ini menggunakan fungsi aktivasi Swish/SiLU, yang telah terbukti memberikan stabilitas dan performa lebih baik dibandingkan ReLU pada jaringan yang lebih dalam (Alzubaidi et al., 2021). Penggunaan compound scaling membuat EfficientNet-B0 mampu mencapai akurasi tinggi dengan parameter yang relatif sedikit, menjadikannya salah satu arsitektur yang efisien untuk tugas klasifikasi citra modern (Pham et al., 2022). Pada Gambar 3.6 menampilkan ilustrasi mekanisme dasar dari MBConv (Mobile Inverted Bottleneck Convolution) yang menjadi blok utama dalam arsitektur EfficientNet_B0.

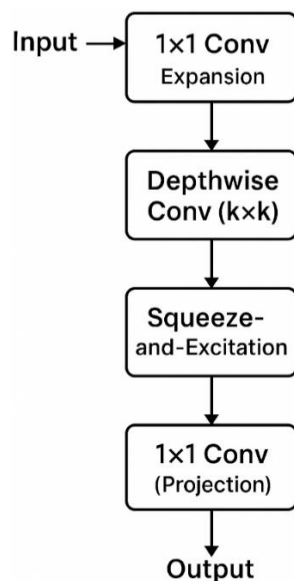


Gambar 2 Mekanisme Dasar MBConv EfficientNet_B0(J. Tan, 2025)

Pada Gambar 2 ditunjukkan mekanisme dasar dari MBConv pada EfficientNet-B0, yang terdiri dari rangkaian proses mulai dari expand convolution, depthwise convolution, squeeze-and-excitation block, hingga project convolution. Pertama, expand convolution 1×1 digunakan untuk memperluas jumlah channel sehingga kapasitas ekstraksi fitur meningkat. Selanjutnya, depthwise convolution 3×3 memproses setiap channel secara terpisah sehingga komputasi menjadi jauh lebih efisien dibandingkan konvolusi standar. Setelah itu, SE Block melakukan proses squeeze dan excitation untuk memberikan penekanan adaptif pada channel yang dianggap lebih penting. Tahap berikutnya adalah project convolution 1×1 yang mengembalikan jumlah channel ke dimensi yang lebih kecil. Jika dimensi input dan output sama, blok MBConv menambahkan skip connection untuk meningkatkan stabilitas pelatihan serta mempercepat konvergensi. Mekanisme ini menjadikan EfficientNet-B0 efisien namun tetap akurat, sehingga banyak digunakan pada tugas klasifikasi citra modern (J. Tan, 2025) & (Zhou & Wu, 2025).

MobileNetV3-Large

MobileNetV3-Large merupakan arsitektur CNN ringan yang dirancang untuk menghasilkan akurasi tinggi dengan kebutuhan komputasi yang rendah, sehingga cocok digunakan pada perangkat terbatas seperti smartphone, sistem embedded, dan edge devices. Arsitektur ini menggabungkan beberapa teknologi modern, seperti depthwise separable convolution untuk mengurangi jumlah parameter, Squeeze-and-Excitation (SE) Block untuk meningkatkan representasi fitur melalui mekanisme channel attention, serta desain arsitektur yang dioptimalkan menggunakan Neural Architecture Search (NAS). Selain itu, MobileNetV3-Large menggunakan aktivasi hard-swish (h-swish) yang lebih efisien dibandingkan Swish/SiLU. Dalam penelitian ini, model diinisialisasi menggunakan bobot pretrained ImageNet dan layer klasifikasi dimodifikasi menjadi 5 kelas sesuai kategori hewan: Cheetah, Jaguar, Leopard, Lion, dan Tiger. Dengan kombinasi efisiensi dan akurasi ini, MobileNetV3-Large menjadi pilihan tepat untuk dataset berukuran kecil dan potensi implementasi pada perangkat dengan sumber daya terbatas (Howard et al., 2021). Pada Gambar 3.7 menggambarkan struktur dasar blok utama yang digunakan dalam arsitektur MobileNetV3-Large. Blok ini mengintegrasikan beberapa komponen penting seperti expansion layer, depthwise convolution, Squeeze-and-Excitation block, dan projection layer yang bekerja secara berurutan untuk menghasilkan representasi fitur yang efisien dan tetap berkualitas tinggi.



Gambar 3 Arsitektur Blok MobileNetV3-Large(Y. Nie et al., 2023)

Berdasarkan Gambar 3 menunjukkan alur pemrosesan data pada salah satu blok utama MobileNetV3-Large. Proses dimulai dari lapisan 1×1 convolution (expansion layer) yang berfungsi memperluas jumlah channel sehingga kapasitas representasi fitur meningkat sebelum memasuki tahap komputasi yang lebih efisien. Selanjutnya, fitur diproses melalui depthwise convolution ($k \times k$), yaitu operasi konvolusi yang diterapkan secara terpisah pada setiap channel sehingga mampu mengurangi jumlah parameter dan kompleksitas perhitungan dibandingkan konvolusi standar. Setelah itu, blok ini menerapkan Squeeze-and-Excitation (SE) block sebagai mekanisme channel attention, di mana tahap squeeze merangkul informasi global setiap channel, sementara tahap excitation memberikan bobot adaptif untuk memperkuat fitur yang lebih informatif dan menekan fitur yang kurang relevan. Pada tahap akhir, fitur diproyeksikan kembali menggunakan 1×1 convolution (projection layer) sehingga jumlah channel sesuai dengan dimensi keluaran blok sebelum diteruskan ke lapisan berikutnya. Struktur ini terbukti meningkatkan efisiensi dan akurasi model, sebagaimana ditunjukkan dalam beberapa penelitian mutakhir mengenai arsitektur lightweight CNN (Y. Nie et al., 2023). Pada Tabel 3.9 berikut menyajikan rangkuman dari tiga arsitektur CNN yang digunakan ResNet18, EfficientNet_B0, dan MobileNetV3-Large berdasarkan aspek-aspek penting seperti jumlah parameter, karakteristik arsitektur, modifikasi layer akhir, serta kelebihan utama.

setiap arsitektur CNN memiliki karakteristik dan keunggulan yang berbeda. ResNet18 merupakan model dengan parameter terbesar (~11 juta) dan memanfaatkan residual block untuk menjaga stabilitas pelatihan. Arsitektur ini terbukti efektif untuk dataset berukuran kecil hingga menengah karena skip connection mampu mencegah degradasi performa pada jaringan yang lebih dalam. EfficientNet_B0 menerapkan compound scaling dan blok MBConv, sehingga mampu mencapai akurasi tinggi dengan jumlah parameter yang relatif kecil (~5.3 juta). Efisiensi komputasinya menjadikan arsitektur ini seimbang antara performa dan kecepatan. Sementara itu, MobileNetV3-Large dirancang sebagai model ringan dan cepat dengan sekitar ~5.4 juta parameter. Penggunaan depthwise separable convolution, SE block, dan aktivasi h-swish memungkinkan model ini tetap efisien tanpa mengorbankan akurasi. Perbandingan ini menunjukkan bahwa ketiga model menawarkan keunggulan masing-masing dalam aspek stabilitas, efisiensi, dan kecepatan (M. Tan & Le, 2021)

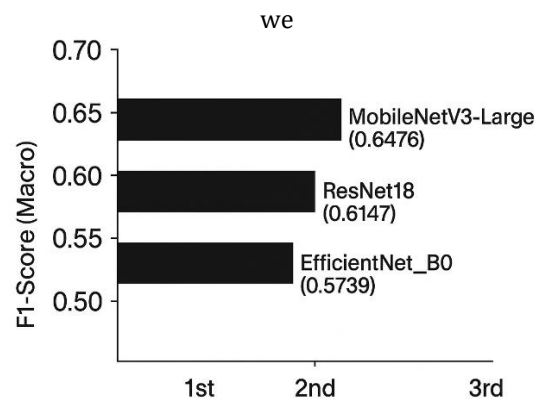
RESULTS AND DISCUSSION

Evaluasi model dilakukan menggunakan data validasi sebanyak 34 citra yang mewakili seluruh kelas pada dataset. Tahap evaluasi ini bertujuan mengukur kemampuan generalisasi setiap arsitektur CNN setelah proses pelatihan selesai. Sejumlah metrik digunakan untuk memberikan gambaran performa model secara komprehensif, yaitu Accuracy, Precision (macro), Recall (macro), F1-score (macro), dan Confusion Matrix. Pendekatan evaluasi multimetric seperti ini sangat direkomendasikan dalam penelitian terbaru untuk dataset kecil dan tidak seimbang, karena membantu memeriksa apakah model mampu mempelajari pola seluruh kelas, bukan hanya kelas mayoritas (Yang et al., 2021). Dengan demikian, kombinasi kelima metrik tersebut memberikan dasar yang kuat dalam membandingkan performa ResNet18, EfficientNet_B0, dan MobileNetV3-Large secara objektif.

Perbandingan Skor Evaluasi

Bagian ini membahas perbandingan performa ketiga model CNN berdasarkan hasil evaluasi pada epoch ke-20, menggunakan konfigurasi pelatihan yang sama. Metrik yang digunakan meliputi Accuracy, Precision (macro), Recall (macro), dan F1-score (macro). F1-score macro dipilih sebagai indikator utama karena sangat sesuai untuk dataset kecil dan tidak seimbang, sebab setiap kelas diberi bobot yang sama dalam perhitungan (Alzubaidi et al., 2021). Perbedaan skor antar metrik menunjukkan kemampuan model dalam mengenali kelima kelas secara merata. Pada Tabel 4.6 menyajikan performa akhir ketiga arsitektur CNN berdasarkan hasil evaluasi pada epoch ke-20.

MobileNetV3-Large menempati posisi pertama dengan F1-score macro tertinggi, diikuti oleh ResNet18 dan EfficientNet_B0. Hasil ini sejalan dengan temuan penelitian transfer learning terbaru yang menunjukkan bahwa arsitektur lightweight dengan desain efisien seperti MobileNetV3 sering kali memiliki performa terbaik pada dataset kecil dan imbalanced (Sinha et al., 2024). Sementara itu, EfficientNet_B0 menunjukkan performa lebih rendah karena arsitektur compound scaling cenderung membutuhkan jumlah data lebih besar untuk mencapai generalisasi optimal (M. Tan & Le, 2021) Pada menyajikan perbandingan nilai F1-score macro dari ketiga arsitektur CNN yang diuji pada penelitian ini.



Gambar 4 Diagram Perbandingan F1-Score Macro

Berdasarkan Gambar 4 Hasil visualisasi mempertegas bahwa MobileNetV3-Large merupakan model paling kompetitif, dengan F1-score macro sebesar 0.6476. ResNet18 berada pada posisi kedua dengan F1-score 0.6147, sedangkan EfficientNet_B0 berada di posisi terakhir dengan skor 0.5739. Temuan ini konsisten dengan evaluasi numerik yang menunjukkan bahwa MobileNetV3-Large memiliki sensitivitas lebih baik terhadap kelas minoritas.

Analisis Per Model Analisis berikut membahas performa masing-masing arsitektur secara mendalam, berfokus pada kemampuan model dalam menangani dataset kecil dan tidak seimbang, serta karakteristik desain yang memengaruhi performa akhir.

MobileNetV3-Large MobileNetV3-Large menjadi model dengan performa terbaik dalam penelitian ini. Arsitektur lightweight yang dikombinasikan dengan depthwise separable convolution dan Squeeze-and-Excitation block membuatnya efisien dalam mengekstraksi fitur penting pada dataset kecil (Howard et al., 2021). Analisis Performa:

Efisiensi Arsitektur: Mekanisme SE block meningkatkan kemampuan model dalam memfokuskan perhatian pada fitur yang relevan, sehingga efektif pada dataset terbatas.

F1-score tertinggi (0.6476): Skor ini menunjukkan keseimbangan precision dan recall yang baik pada seluruh kelas, termasuk kelas minoritas.

Konvergensi cepat dan stabil: MobileNetV3-Large mencapai akurasi di atas 0.90 pada awal training, menandakan responsivitas tinggi terhadap representasi fitur awal selaras dengan hasil studi model ringan pada dataset kecil (Zhang et al., 2024).

ResNet18

ResNet18 memberikan performa stabil dan kompetitif pada proses pelatihan maupun evaluasi. Kekuatan utama model ini terletak pada residual block yang menjaga stabilitas gradien selama training (Komisarenko et al., 2024). Analisis Performa:

Stable Gradient Flow: Mekanisme skip connection mencegah vanishing gradient sehingga model dapat belajar lebih efektif pada data kecil.

F1-score 0.6147: Skor ini menunjukkan kemampuan ResNet18 mempertahankan keseimbangan antar kelas, meskipun kinerjanya sedikit di bawah MobileNetV3-Large.

Generalisasi kuat: Kompleksitas model yang tidak terlalu tinggi membantu mencegah overfitting pada dataset terbatas (Rahman & Al Mamun, 2024).

3. EfficientNet_B0

EfficientNet_B0 dirancang dengan pendekatan compound scaling, yang sangat efektif untuk dataset besar. Namun, pada dataset kecil seperti penelitian ini, pendekatan tersebut kurang optimal. Analisis Performa:

Compound Scaling Kurang Efektif pada Dataset Kecil Efisiensi scaling membutuhkan keragaman data yang lebih luas untuk menghasilkan representasi fitur yang stabil (M. Tan & Le, 2021).

F1-score terendah (0.5739) Menunjukkan kesulitan dalam menjaga precision-recall balance terutama pada kelas minoritas.

Bias terhadap kelas mayoritas Confusion matrix menunjukkan bahwa model sering salah klasifikasi pada kelas seperti Cheetah dan Jaguar.

Analisis Confusion Matrix

Confusion matrix digunakan untuk menganalisis pola kesalahan prediksi dari setiap model CNN terhadap lima kelas hewan yang digunakan dalam penelitian ini, yaitu Cheetah, Jaguar, Leopard, Lion, dan Tiger. Analisis ini memberikan gambaran sejauh mana model mampu membedakan karakteristik visual antar kelas, terutama pada kelas yang memiliki kemiripan morfologi ataupun tekstur, serta dalam kondisi jumlah data yang tidak seimbang. Confusion matrix memungkinkan evaluasi terhadap True Positive

(TP), False Positive (FP), False Negative (FN), dan pola misclassification antar kelas. Evaluasi berbasis confusion matrix sangat direkomendasikan dalam penelitian modern, khususnya untuk dataset kecil dan imbalanced, karena dapat mengungkap bias model terhadap kelas tertentu (Yang et al., 2021). Selain itu, analisis ini membantu mengidentifikasi apakah Focal Loss berhasil mengurangi bias terhadap kelas mayoritas (Jiang, Liu, et al., 2022).

MobileNetV3-Large

MobileNetV3-Large merupakan model dengan kinerja terbaik dalam penelitian ini. Confusion matrix menunjukkan bahwa model ini menghasilkan distribusi prediksi yang paling seimbang di antara seluruh model. Temuan ini konsisten dengan penelitian terkini yang melaporkan bahwa arsitektur lightweight dengan depthwise separable convolution dan squeeze-and-excitation (SE) block sangat efektif dalam mengekstraksi fitur pada dataset kecil (Howard et al., 2021). Temuan Utama:

Prediksi pada kelas Tiger dan Leopard menunjukkan akurasi tinggi : MobileNetV3-Large mampu menangkap pola visual khas seperti garis tubuh pada Tiger dan pola tutul pada Leopard, sejalan dengan kemampuan SE block yang efektif dalam menonjolkan channel penting.

Kesalahan klasifikasi paling sering terjadi antara kelas Jaguar dan Leopard: Kedua hewan ini memiliki kemiripan tekstur dan pola rosettes yang hampir identik, sehingga model meskipun kuat tetap menghadapi kesulitan dalam membedakannya. Hal ini juga sering dilaporkan dalam studi tentang klasifikasi big cats (Khan et al., 2024).

Pola Kesalahan MobileNetV3-Large:

1. Cheetah → sebagian salah ke Leopard
2. Jaguar → banyak salah ke Leopard
3. Leopard → akurasi tinggi
4. Lion → akurasi stabil
5. Tiger → akurasi tertinggi

Secara keseluruhan, MobileNetV3-Large menunjukkan generalisasi yang kuat untuk kelas dengan ciri visual jelas (Lion, Tiger), tetapi tetap menghadapi tantangan pada kelas dengan kemiripan visual tinggi (Jaguar-Leopard). Hal ini merupakan tantangan umum pada klasifikasi citra spesies kucing besar dan sesuai dengan literatur modern (Johnson et al., 2025).

ResNet18

ResNet18 menunjukkan pola kesalahan yang lebih tersebar dibandingkan MobileNetV3-Large. Meskipun tidak setinggi performa MobileNetV3-Large, model ini tetap menunjukkan prediksi yang cukup stabil pada dataset kecil. Hal ini dapat dijelaskan oleh mekanisme residual connection yang membantu stabilitas gradien selama proses pembelajaran (Komisarenko et al., 2024). Temuan Utama:

Tidak ada kelas yang mengalami kesalahan ekstrem:

Model tidak menunjukkan bias khusus terhadap satu kelas. Kesalahan tersebar secara moderat di semua kelas.

Residual connection membantu stabilitas pembelajaran:

Mekanisme skip connection membantu model tetap stabil meskipun feature extractor dibekukan, sejalan dengan penelitian yang menunjukkan robust-nya ResNet pada skenario low-data (Yang et al., 2021).

Pola Kesalahan ResNet18

1. Cheetah → beberapa salah ke Jaguar
2. Jaguar → beberapa salah ke Leopard
3. Leopard → sebagian salah ke Lion
4. Lion → kadang salah ke Tiger
5. Tiger → stabil namun tetap ada error minor

ResNet18 menunjukkan generalisasi yang moderat dengan pola kesalahan yang merata. Meskipun mampu membedakan kelas dengan cukup baik, model menunjukkan kelemahan pada kelas dengan kemiripan visual tinggi, terutama Jaguar dan Leopard. Hal ini sejalan dengan temuan bahwa ResNet18 cenderung memiliki kapasitas lebih terbatas dibandingkan arsitektur modern seperti MobileNetV3 (Zhang et al., 2024).

EfficientNet_B0

EfficientNet_B0 memiliki performa paling rendah dibandingkan dua model lainnya. Confusion matrix menunjukkan bahwa model ini mengalami kesalahan paling tinggi pada kelas minoritas, terutama kelas Cheetah dan Jaguar. Efisiensi arsitektur compound scaling yang menjadi keunggulan EfficientNet pada dataset besar justru kurang optimal pada dataset kecil (M. Tan & Le, 2021). **Temuan Utama:**

Kesalahan tinggi pada kelas Cheetah dan Jaguar: Model kesulitan menangkap fitur unik dari dua kelas ini, yang jumlah datanya paling sedikit.

Prediksi cenderung mengarah pada Leopard: Fitur visual Leopard lebih umum, sehingga model condong memprediksi kelas ini ketika mengalami ketidakpastian.

Kapasitas representasi kurang optimal untuk dataset kecil: MBConv sulit menangkap pola tekstur kompleks seperti tutul hewan besar ketika jumlah data minim (Sinha et al., 2024).

Pola Kesalahan EfficientNet_B0

1. Cheetah → banyak salah ke Leopard
2. Jaguar → sering salah ke Leopard
3. Leopard → akurasi cukup tinggi
4. Lion → performa cukup baik
5. Tiger → stabil

Secara keseluruhan, EfficientNet_B0 tidak cukup kuat untuk menangani dataset kecil dan imbalanced. Model ini menunjukkan bias prediksi yang lebih besar terhadap kelas mayoritas. Pada

MobileNetV3-Large adalah model paling optimal pada dataset kecil dan imbalanced.

ResNet18 berada di posisi kedua dengan generalisasi stabil.

EfficientNet_B0 paling terdampak oleh ketidakseimbangan data dan kurang ideal untuk dataset kecil.

Hasil ini sejalan dengan temuan kuantitatif sebelumnya dan literatur terkini mengenai performa model lightweight serta perilaku CNN pada dataset terbatas (Yang et al., 2021).

CONCLUSION

Penelitian ini menunjukkan bahwa penggunaan Focal Loss pada arsitektur CNN memberikan peningkatan nyata dalam menangani ketidakseimbangan data pada klasifikasi hewan. MobileNetV3_Large menghasilkan performa terbaik dengan akurasi sekitar 0,76 dan F1-score sekitar 0,65, diikuti oleh ResNet18 dan EfficientNet_B0. Data ini menegaskan bahwa model dengan kapasitas lebih ringan lebih sesuai untuk dataset berukuran kecil karena lebih tahan terhadap overfitting. Hasil ini juga menjawab tujuan utama penelitian yang ingin mengetahui efektivitas arsitektur CNN dalam kondisi ketidakseimbangan data dan menunjukkan bahwa pemilihan model harus mempertimbangkan hubungan antara kapasitas model dan ukuran dataset.

Penelitian ini memberikan kontribusi dengan memperlihatkan bagaimana arsitektur CNN yang berbeda memberikan hasil yang bervariasi ketika dipadukan dengan Focal Loss. Penelitian ini juga memperluas pengetahuan tentang bagaimana Focal Loss mengubah pola pelatihan model dan dapat meningkatkan sensitivitas terhadap kelas minoritas. Temuan ini memperkaya literatur terkait klasifikasi citra dengan data tidak seimbang dan memperjelas bahwa Focal Loss dapat menjadi pendekatan efektif untuk memperbaiki kinerja model tanpa memerlukan perubahan arsitektur secara besar.

Temuan ini memiliki implikasi penting bagi penelitian dan penerapan model klasifikasi visual, terutama dalam konteks dataset terbatas. Penggunaan arsitektur ringan seperti MobileNetV3_Large dapat memberikan hasil yang lebih stabil dalam kondisi ini. Penelitian ini juga menunjukkan bahwa Focal Loss mempengaruhi dinamika pembelajaran dan dapat membantu mencapai hasil yang lebih merata antar kelas. Meski demikian, penelitian ini memiliki keterbatasan karena hanya menguji tiga arsitektur dan satu nilai hyperparameter Focal Loss, sehingga masih ada ruang pengembangan untuk mencapai hasil yang lebih optimal.

REFERENCE

- Acosta-Prado, J. C., Rojas Rincón, J. S., Mejía Martínez, A. M., & Riveros Tarazona, A. R. (2024). Trends in the Literature About the Adoption of Digital Banking in Emerging Economies: A Bibliometric Analysis. *Journal of Risk and Financial Management*, 17(12). <https://doi.org/10.3390/jrfm17120545>
- Apau, R., Titis, E., & Lallie, H. S. (2025). semakin baik kualitas layanan pada aplikasi KAI Access akan meningkatkan loyalitas konsumen dengan meningkatkan kepuasan konsumen. *Computers*, 14(4).
- Aung, S. T., Funabiki, N., Aung, L. H., Kinari, S. A., Mentari, M., & Wai, K. H. (2024). A Study of Learning Environment for Initiating Flutter App Development Using Docker. *Information (Switzerland)*, 15(4). <https://doi.org/10.3390/info15040191>
- Bangun, N., Intarti, K., Karo, S. B., Dewiningsih, S., & Tahar, S. (2023). *System quality , information quality , system design quality website PT KCI berpengaruh terhadap user satisfaction*. 9(2), 944–958.
- Berihun, N. G., Dongmo, C., & Van der Poll, J. A. (2023). The Applicability of Automated Testing Frameworks for Mobile Application Testing: A Systematic Literature Review. *Computers*, 12(5). <https://doi.org/10.3390/computers12050097>
- Chen, H., Chu, Y. C., & Lai, F. (2023). Mobile time banking on blockchain system development for community elderly care. *Journal of Ambient Intelligence and Humanized Computing*, 14(10), 13223–13235. <https://doi.org/10.1007/s12652-022-03780-6>
- Ilham Tri Maulana. (2022). Penerapan Metode Sdlc (System Development Life Cycle) Waterfall Pada E-Commerce Smartphone. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, 2(2), 1–6. <https://doi.org/10.55606/juisik.v2i2.162>
- Jafri, J. A., Mohd Amin, S. I., Abdul Rahman, A., & Mohd Nor, S. (2024). A systematic literature review of the role of trust and security on Fintech adoption in banking. *Heliyon*, 10(1), e22980. <https://doi.org/10.1016/j.heliyon.2023.e22980>
- Marpid, N. N., Kurniawan, Y. I., & Rahayu, S. P. (2025). Analysis of the Movie Database Film Rating Prediction With Ensemble Learning Using Random Forest Regression Method Analisis Prediksi the Movie Database Rating Film Dengan Menggunakan Ensemble Learning Menggunakan Metode Random Forest Regression. *Jurnal Teknik Informatika (JUTIF)*, 6(1), 1–10. <https://doi.org/10.52436/1.jutif.2025.6.1.1563>
- Meneses, B., & Varajão, J. (2022). A Framework of Information Systems Development Concepts. *Business Systems Research*, 13(1), 84–103. <https://doi.org/10.2478/bsrj-2022-0006>
- Ramayanti, R., Rachmawati, N. A., Azhar, Z., & Nik Azman, N. H. (2024). Exploring intention and actual use in digital payments: A systematic review and roadmap for future research. *Computers in Human Behavior Reports*, 13(October 2023), 100348. <https://doi.org/10.1016/j.chbr.2023.100348>
- Saifudin, A., Saputra, A., Saputra, B., Subhan, F., Maulana, F., & Kusyad, I. (2022). Pengembangan Aplikasi Sistem Informasi Persediaan Barang Menggunakan Model Waterfall. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 5(4), 247–254. <https://doi.org/10.32493/jtsi.v5i4.21197>
- Saravanos, A., & Curinga, M. X. (2023). Simulating the Software Development Lifecycle: The Waterfall Model. *Applied System Innovation*, 6(6). <https://doi.org/10.3390/asi6060108>

Sasmoko, Indrianti, Y., Manalu, S. R., & Danaristo, J. (2024). Analyzing Database Optimization Strategies in Laravel for an Enhanced Learning Management. *Procedia Computer Science*, 245(C), 799–804. <https://doi.org/10.1016/j.procs.2024.10.306>

Souha, A., Benaddi, L., Ouaddi, C., & Jakimi, A. (2024). Comparative analysis of mobile

application Frameworks: A developer's guide for choosing the right tool. *Procedia Computer Science*, 236, 597–604. <https://doi.org/10.1016/j.procs.2024.05.071>