

Perbandingan Kinerja Model ResNet-50 dan MobileNetV2 untuk Klasifikasi Citra Menggunakan Metode CNN

Muhammad Nalendra Zaradiva¹, Irma Amelia Dewi²

^{1,2}Program Studi Informatika, Fakultas Teknologi Industri, Institut Teknologi Nasional Bandung, Indonesia
Email: * muhammad.nalendra@mhs.itenas.ac.id

Abstract—Convolutional Neural Network (CNN) architectures such as ResNet-50 and MobileNetV2 are widely applied for automated medical image classification, yet their comparative performance on breast ultrasound imaging, particularly regarding the trade-off between classification accuracy and computational efficiency, remains underexplored. This study conducts a comparative analysis of ResNet-50 and MobileNetV2 for three-class classification (normal, benign, malignant) of the Breast Ultrasound Images Dataset (BUSI), consisting of 780 images. Both models were trained using an identical transfer learning scheme (feature extraction followed by fine-tuning), the same hyperparameter configuration, and the same computing environment to ensure a fair comparison. Model performance was evaluated using classification metrics (accuracy, precision, recall, F1-score) and computational efficiency metrics (number of parameters, model size, inference time, and throughput). The results show that ResNet-50 consistently outperforms MobileNetV2 across all classification metrics, with an accuracy margin of 10.26 percentage points, owing to its deeper residual learning capacity. Conversely, MobileNetV2 demonstrates substantial superiority in computational efficiency, requiring only about 10% of ResNet-50's parameters, achieving roughly nine times smaller model size, and delivering about ~31% faster inference and higher throughput. These findings confirm a clear accuracy, efficiency trade-off and suggest that ResNet-50 is more suitable for server-based clinical decision support systems prioritizing diagnostic accuracy, while MobileNetV2 is better suited for mobile or edge-device deployment in resource-constrained healthcare settings, provided that its lower recall on the malignant class is further mitigated.

Keywords—Breast Ultrasound Classification, Convolutional Neural Network, MobileNetV2, ResNet-50, Transfer Learning.

Abstrak—Arsitektur Convolutional Neural Network (CNN) seperti ResNet-50 dan MobileNetV2 banyak digunakan untuk klasifikasi citra medis secara otomatis. Namun, perbandingan kinerjanya pada klasifikasi citra ultrasonografi (USG) payudara, khususnya terkait kompromi antara akurasi klasifikasi dan efisiensi komputasi, masih belum banyak dikaji. Penelitian ini melakukan analisis komparatif antara ResNet-50 dan MobileNetV2 untuk klasifikasi tiga kelas (normal, jinak, dan ganas) menggunakan Breast Ultrasound Images Dataset (BUSI) yang terdiri atas 780 citra. Kedua model dilatih menggunakan skema transfer learning yang sama, yaitu tahap feature extraction yang dilanjutkan dengan fine-tuning, serta menerapkan konfigurasi hiperparameter dan lingkungan komputasi yang identik untuk memastikan perbandingan yang adil. Evaluasi dilakukan berdasarkan metrik klasifikasi yang meliputi akurasi, presisi, recall, dan F1-score, serta metrik efisiensi komputasi yang mencakup jumlah parameter, ukuran model, waktu inferensi, dan throughput. Hasil penelitian menunjukkan bahwa ResNet-50 secara konsisten memberikan performa klasifikasi yang lebih baik dibandingkan MobileNetV2 pada seluruh metrik evaluasi, dengan selisih akurasi sebesar 10,26 poin persentase, yang didukung oleh kemampuan pembelajaran residual yang lebih mendalam. Sebaliknya, MobileNetV2 menunjukkan efisiensi komputasi yang jauh lebih tinggi dengan hanya memerlukan sekitar 10% jumlah parameter ResNet-50, ukuran model sekitar sembilan kali lebih kecil, serta waktu inferensi sekitar ~31% lebih cepat dan throughput yang lebih tinggi. Temuan ini mengonfirmasi adanya kompromi yang jelas antara akurasi dan efisiensi komputasi. Oleh karena itu, ResNet-50 lebih sesuai untuk sistem pendukung keputusan klinis berbasis server yang mengutamakan akurasi diagnosis, sedangkan MobileNetV2 lebih tepat diterapkan pada perangkat bergerak atau edge device di lingkungan layanan kesehatan dengan keterbatasan sumber daya, dengan catatan bahwa performa recall pada kelas ganas masih perlu ditingkatkan. recall on the malignant class is further mitigated.

Kata Kunci—Klasifikasi USG Payudara, Convolutional Neural Network, MobileNetV2, ResNet-50, Transfer Learning.

I. PENDAHULUAN

Klasifikasi citra medis berbasis *deep learning*, khususnya menggunakan arsitektur *Convolutional Neural Network* (CNN), telah menjadi pendekatan utama dalam mendukung diagnosis otomatis di bidang kesehatan, termasuk deteksi dini

kanker payudara melalui citra ultrasonografi (USG). Kemampuan CNN dalam mengekstraksi fitur visual secara hierarkis tanpa memerlukan *feature engineering* manual menjadikannya sangat relevan untuk mengenali pola-pola diagnostik pada citra USG, seperti bentuk dan tekstur lesi, yang seringkali sulit dibedakan secara visual akibat karakteristik

citra USG yang memiliki *speckle noise* dan kontras rendah. Namun, penerapan CNN pada konteks klinis tidak hanya menuntut akurasi klasifikasi yang tinggi, tetapi juga efisiensi komputasi yang memadai, mengingat sistem diagnosis berbasis citra dapat diimplementasikan pada infrastruktur yang beragam, mulai dari *server* rumah sakit dengan sumber daya komputasi besar hingga perangkat *mobile* atau *edge device* di fasilitas kesehatan dengan keterbatasan daya dan memori. Kebutuhan akan keseimbangan antara akurasi dan efisiensi inilah yang mendorong pentingnya kajian komparatif terhadap arsitektur CNN yang mewakili kedua orientasi tersebut, sehingga pemilihan model dapat disesuaikan dengan konteks penerapannya secara tepat.

Di antara berbagai arsitektur CNN yang telah diusulkan, ResNet-50 dan MobileNetV2 merupakan dua arsitektur yang banyak digunakan dan menjadi rujukan dalam berbagai penelitian maupun implementasi industri. ResNet-50 menghadirkan konsep *residual learning* melalui *skip connection* yang memungkinkan pelatihan jaringan yang sangat dalam tanpa mengalami masalah *vanishing gradient*, sehingga menghasilkan akurasi klasifikasi yang tinggi[1]. Namun, arsitektur ini memiliki jumlah parameter dan kompleksitas komputasi yang relatif besar, sehingga kurang ideal untuk diterapkan pada perangkat dengan sumber daya terbatas. Di sisi lain, MobileNetV2 dirancang khusus untuk efisiensi komputasi melalui penggunaan *depthwise separable convolution* dan struktur *inverted residual with linear bottleneck*, menjadikannya arsitektur yang ringan dan cocok untuk diimplementasikan pada perangkat *mobile* maupun *edge device* dengan keterbatasan daya komputasi dan memori.

Perbedaan karakteristik mendasar antara kedua arsitektur ini yaitu ResNet-50 yang mengutamakan kedalaman jaringan dan akurasi, serta MobileNetV2 yang mengutamakan efisiensi dan kecepatan, menimbulkan pertanyaan penting mengenai *trade-off* antara akurasi klasifikasi dan efisiensi komputasi pada kasus penggunaan tertentu. Pemilihan arsitektur yang tepat menjadi krusial, terutama ketika sistem klasifikasi citra akan diterapkan pada lingkungan produksi dengan batasan sumber daya perangkat keras maupun kebutuhan *real-time processing*.

Meskipun telah banyak penelitian yang mengevaluasi performa ResNet-50 dan MobileNetV2 secara terpisah pada berbagai *dataset*, penelitian yang secara spesifik membandingkan kedua arsitektur ini secara komprehensif, tidak hanya dari segi akurasi, tetapi juga dari sisi efisiensi komputasi seperti waktu inferensi, jumlah parameter, dan penggunaan sumber daya pada Breast Ultrasound Images Dataset (BUSI) masih relatif terbatas. Penelitian ini menghadirkan analisis komparatif yang tidak hanya berfokus pada metrik akurasi klasifikasi, tetapi juga mempertimbangkan aspek efisiensi model, sehingga memberikan gambaran yang lebih menyeluruh mengenai *trade-off* antara performa dan kompleksitas komputasi dari kedua arsitektur tersebut. Hasil penelitian ini diharapkan dapat menjadi acuan praktis dalam pemilihan arsitektur CNN yang sesuai dengan kebutuhan spesifik suatu sistem, baik yang mengutamakan akurasi tinggi maupun efisiensi komputasi.

Dokumen ini adalah *template*. Sebuah salinan elektronik yang dapat diunduh dari situs web Digital Nexus Systematic Journal Program Studi Teknologi Informasi Fakultas Sains & Humaniora UNIMUGO (DNS). Untuk pertanyaan atas kertas panduan, silakan hubungi panitia publikasi jurnal seperti yang ditunjukkan pada situs web. Informasi tentang makalah akhir penyerahan tersedia dalam situs web Digital Nexus Systematic.

II. TINJAUAN PUSTAKA

A. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) merupakan salah satu arsitektur *deep learning* yang dirancang khusus untuk mengolah data berbentuk grid seperti citra digital. CNN bekerja dengan memanfaatkan operasi konvolusi untuk mengekstraksi fitur-fitur visual secara hierarkis, mulai dari fitur tingkat rendah (tepi, tekstur) hingga fitur tingkat tinggi (bentuk objek, pola kompleks), tanpa memerlukan proses ekstraksi fitur manual. Kemampuan ini menjadikan CNN sebagai fondasi utama dalam berbagai tugas *computer vision*, termasuk klasifikasi citra, deteksi objek, dan segmentasi semantik.

B. Arsitektur MobileNetV2

MobileNetV2 dikembangkan sebagai penyempurnaan dari MobileNetV1, dengan memperkenalkan struktur *inverted residual* dengan *linear bottleneck* sebagai modul baru untuk mengurangi komputasi sekaligus meningkatkan akurasi[2]. Modul ini bekerja dengan memperluas representasi input berdimensi rendah ke dimensi tinggi, melakukan penyaringan melalui *depthwise convolution*, kemudian memproyeksikannya kembali ke dimensi rendah, berkebalikan dengan blok residual konvensional yang melakukan kompresi terlebih dahulu. Arsitektur ini juga menerapkan fungsi aktivasi ReLU6 sebagai pengganti ReLU standar guna membatasi beban komputasi, sehingga menjadikan MobileNetV2 sangat sesuai untuk diimplementasikan pada perangkat *mobile* dan *edge device* dengan sumber daya terbatas.

C. Arsitektur ResNet-50

ResNet-50 merupakan varian dari keluarga *Residual Network* (ResNet)[1] yang diperkenalkan oleh He et al. pada tahun 2016. Arsitektur ini mengatasi permasalahan *vanishing gradient* pada jaringan yang sangat dalam melalui penerapan *residual learning*, yaitu penggunaan *skip connection* yang memungkinkan sinyal gradien mengalir langsung antar-lapisan tanpa mengalami degradasi signifikan. ResNet-50 tersusun atas 50 lapisan dengan struktur *bottleneck block*, sehingga mampu mempelajari representasi fitur yang sangat dalam sekaligus mempertahankan efisiensi komputasi dibandingkan jaringan sedalam itu tanpa mekanisme residual.

D. Penelitian Terdahulu

Sejumlah penelitian telah membandingkan performa ResNet-50 dan MobileNetV2 pada berbagai domain klasifikasi citra dengan hasil yang bervariasi tergantung karakteristik *dataset* yang digunakan.

Pada studi klasifikasi kepadatan lalu lintas, penelitian yang membandingkan EfficientNetB0, MobileNetV2, dan ResNet50 menggunakan pendekatan transfer learning pada dataset Traffic Density Singapore menunjukkan bahwa ResNet50 memberikan performa validasi terbaik sekitar 85%, sementara EfficientNetB0 dan MobileNetV2 masing-masing mencapai 81% dan 79%, sehingga menyimpulkan ResNet50 sebagai arsitektur yang paling optimal untuk kasus tersebut[3].

Sebaliknya, pada studi klasifikasi citra hasil AI-generation, MobileNetV2 justru mencapai akurasi tertinggi sebesar 89% pada epoch ke-10, namun mengalami penurunan performa pada epoch ke-30 dan ke-50 yang mengindikasikan overfitting, sedangkan ResNet50 menunjukkan performa yang lebih stabil di seluruh variasi epoch yang diuji[4]. Temuan ini mengindikasikan bahwa MobileNetV2 cenderung lebih cepat konvergen namun rentan terhadap *overfitting* pada pelatihan jangka panjang, sementara ResNet50 lebih stabil meski membutuhkan sumber daya lebih besar.

Pada domain citra medis, perbandingan MobileNetV2, VGG16, dan ResNet50 untuk klasifikasi tumor otak menghasilkan akurasi berturut-turut 92,05%, 91,13%, dan 85,32%, di mana MobileNetV2 justru menunjukkan performa terbaik dan ResNet-50 performa terendah[5], bertentangan dengan temuan pada domain lalu lintas, yang menegaskan bahwa keunggulan suatu arsitektur sangat bergantung pada karakteristik *dataset* yang digunakan.

Penelitian lain pada konteks pengenalan objek berbasis video menemukan bahwa arsitektur dengan akurasi tertinggi adalah ResNet-50 sebesar 97,81%, diikuti oleh MobileNetV2 sebesar 96,36%, sementara VGG-16 hanya mencapai 52,25%, menunjukkan bahwa kedua arsitektur (ResNet-50 dan MobileNetV2) secara konsisten mengungguli arsitektur CNN klasik seperti VGG-16 meskipun terdapat selisih akurasi di antara keduanya[6].

Studi perbandingan lain pada klasifikasi Bahasa Isyarat Indonesia juga turut mengonfirmasi bahwa performa ResNet50 dan MobileNetV2 perlu dievaluasi secara spesifik terhadap domain aplikasi, karena karakteristik data seperti variasi pencahayaan, sudut pengambilan gambar, dan jumlah kelas turut memengaruhi hasil akhir kedua model tersebut[7].

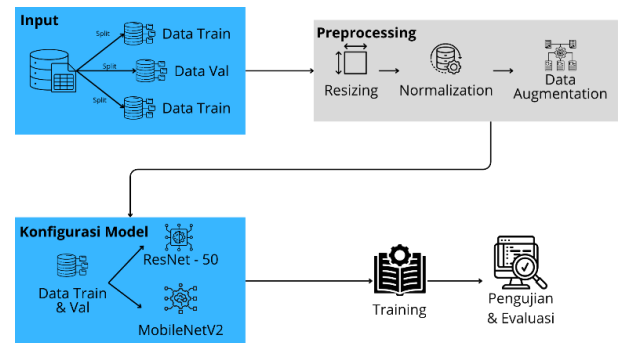
Secara umum, kajian-kajian terdahulu memperlihatkan adanya inkonsistensi hasil antara akurasi ResNet-50 dan MobileNetV2 pada domain yang berbeda-beda, serta minimnya penelitian yang mengevaluasi *trade-off* antara akurasi dan efisiensi komputasi (waktu inferensi, jumlah parameter) secara bersamaan pada satu domain yang sama. Hal inilah yang menjadi celah (*gap*) yang coba diisi oleh penelitian ini.

III. METODE

A. Tahapan Penelitian

Penelitian ini menggunakan pendekatan eksperimental kuantitatif, yaitu suatu desain penelitian yang bertujuan untuk mengukur dan membandingkan performa dua model *Convolutional Neural Network* (ResNet-50 dan MobileNetV2) secara objektif berdasarkan metrik evaluasi yang terukur. Pendekatan eksperimental dipilih karena penelitian ini

melibatkan manipulasi variabel independen (arsitektur model) untuk mengamati pengaruhnya terhadap variabel dependen (akurasi, *precision*, *recall*, *F1-score*, waktu inferensi, dan jumlah parameter), dengan seluruh kondisi pelatihan lainnya (dataset, *hyperparameter*, lingkungan komputasi) dikendalikan agar tetap konsisten pada kedua model, sehingga perbedaan hasil yang diperoleh dapat diatribusikan secara valid pada perbedaan arsitektur.



Gambar 1. Diagram Tahapan Penelitian

Secara umum, penelitian ini dilaksanakan melalui beberapa tahapan sistematis seperti pada gambar 1, sebagai berikut:

- 1) *Pengumpulan Dataset*: Dataset citra yang digunakan dalam penelitian ini adalah *Breast Ultrasound Images Dataset* (BUSI), yang diunduh dari platform Kaggle melalui tautan <https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset>. Dataset yang telah diunduh kemudian diperiksa kelengkapan dan kualitasnya, sebelum dibagi menjadi data latih (*training*), data validasi (*validation*), dan data uji (*testing*).
- 2) *Praproses Data*: Data citra yang telah dikumpulkan melalui proses *resizing* ke ukuran input standar kedua arsitektur, normalisasi nilai piksel, penanganan ketidakseimbangan kelas, serta augmentasi data pada subset data latih. Rincian lengkap tahapan praproses dijelaskan pada bagian C. Praproses Data.
- 3) *Perancangan dan Konfigurasi Model*: Menyiapkan arsitektur ResNet-50 dan MobileNetV2 dengan pendekatan *transfer learning*, termasuk penggantian *classifier head* sesuai jumlah kelas pada penelitian ini, serta menentukan konfigurasi *hyperparameter* yang identik untuk kedua model agar perbandingan bersifat adil (*fair comparison*). Rincian skenario eksperimen dan konfigurasi *hyperparameter* dijelaskan pada bagian D. Skenario Eksperimen.
- 4) *Pelatihan Model (Training)*: Melatih kedua model melalui dua tahap, yaitu *feature extraction* yang dilanjutkan dengan *fine-tuning*, pada dataset dan kondisi lingkungan komputasi (perangkat keras/GPU) yang sama, agar perbedaan hasil yang diperoleh dapat diatribusikan pada perbedaan arsitektur.
- 5) *Pengujian dan Evaluasi Model*: Menguji kedua model pada data uji yang belum pernah dilihat selama pelatihan maupun validasi, kemudian mengukur performanya

menggunakan metrik evaluasi klasifikasi (akurasi, *precision*, *recall*, *F1-score*) serta metrik efisiensi komputasi (jumlah parameter, ukuran model, waktu inferensi, *throughput*).

- 6) *Analisis dan Perbandingan Hasil*: Membandingkan hasil evaluasi kedua model secara kuantitatif, baik pada metrik performa klasifikasi maupun efisiensi komputasi, untuk mengidentifikasi *trade-off* di antara keduanya serta menjawab rumusan masalah penelitian.
- 7) *Penarikan Kesimpulan*: Menyimpulkan arsitektur mana yang lebih unggul berdasarkan konteks *trade-off* akurasi dan efisiensi, serta merumuskan rekomendasi penerapan masing-masing arsitektur sesuai kebutuhan.

B. Dataset Penelitian

Penelitian ini menggunakan Breast Ultrasound Images Dataset (BUSI) yang tersedia secara publik melalui platform Kaggle pada tautan yang telah dibagikan. Dataset ini pertama kali diperkenalkan oleh Al-Dhabyani et al. (2020) dan dikumpulkan dari Baheya Hospital for Early Detection and Treatment of Women's Cancer di Kairo, Mesir, pada tahun 2018, dengan data yang diperoleh dari 600 pasien wanita berusia 25 hingga 75 tahun. Dataset ini banyak dijadikan rujukan (*benchmark*) dalam berbagai penelitian klasifikasi maupun segmentasi citra medis di bidang deteksi kanker payudara[8].

Secara keseluruhan, dataset BUSI terdiri dari 780 citra ultrasound payudara, mencakup 437 citra dengan lesi jinak (*benign*), 210 citra dengan lesi ganas (*malignant*), dan 133 citra normal tanpa lesi. Setiap citra pada dataset ini berformat PNG dengan ukuran rata-rata 500×500 piksel, dan disertai dengan anotasi *ground truth* pada tingkat piksel untuk area lesi.

TABEL 1. STATISTIK DATASET

Kelas	Keterangan	Jumlah Citra
Normal	Tidak terdapat lesi/tumor	133
Benign (Jinak)	Lesi bersifat non-kanker	437
Malignant (ganas)	Lesi bersifat kanker	210
Total		780

Dataset ini bersifat tidak seimbang (*imbalanced*), dengan kelas *benign* mendominasi jumlah data dibandingkan kelas *normal* dan *malignant*. Kondisi ini menjadi salah satu pertimbangan penting dalam tahap *praproses data*, khususnya dalam penerapan teknik *augmentasi* maupun strategi penyeimbangan kelas (*class balancing*) agar tidak terjadi bias model terhadap kelas mayoritas.

Meskipun dataset BUSI pada dasarnya dirancang untuk mendukung tugas segmentasi maupun klasifikasi, penelitian ini akan memanfaatkan label kelas (*normal*, *benign*, *malignant*) yang tersedia untuk membangun tugas klasifikasi citra tiga kelas, sesuai dengan tujuan penelitian dalam membandingkan performa ResNet-50 dan MobileNetV2. Anotasi *ground truth mask* yang disediakan pada dataset ini tidak digunakan dalam

penelitian ini karena fokus penelitian adalah pada klasifikasi, bukan segmentasi citra.

Penggunaan dataset citra medis dalam ranah ultrasound ini juga relevan untuk menguji kemampuan generalisasi kedua arsitektur pada domain yang memiliki karakteristik citra berbeda dari dataset ImageNet[9] (yang umumnya digunakan sebagai basis *pretrained weight*), yaitu citra grayscale dengan tekstur derau (*speckle noise*) khas ultrasound, kontras rendah, dan batas objek yang tidak selalu tegas.

Mengacu pada praktik umum yang digunakan pada penelitian-penelitian sejenis menggunakan dataset BUSI, dataset akan dibagi menjadi tiga subset dengan rasio 70% data latih : 15% data validasi : 15% data uji, dengan pembagian dilakukan secara *stratified* agar proporsi ketiga kelas tetap terjaga pada setiap subset. Pembagian ini bertujuan untuk memastikan model dievaluasi pada data yang belum pernah dilihat selama proses pelatihan, sekaligus menyediakan data validasi untuk pemantauan performa dan pencegahan *overfitting* selama pelatihan berlangsung.

C. Praproses Data

Sebelum digunakan dalam proses pelatihan model, seluruh citra pada dataset BUSI perlu melalui serangkaian tahapan *praproses (preprocessing)* agar sesuai dengan kebutuhan arsitektur ResNet-50 dan MobileNetV2, sekaligus untuk meningkatkan kualitas dan variasi data guna mendukung generalisasi model. Tahapan *praproses* yang dilakukan meliputi *resizing*, normalisasi, penanganan ketidakseimbangan kelas, *augmentasi data*, dan pembagian dataset.

- 1) *Penyesuaian Ukuran Citra (Resizing)*: Citra pada dataset BUSI memiliki ukuran yang bervariasi dengan rata-rata 500×500 piksel, sedangkan kedua arsitektur yang digunakan dalam penelitian ini pada dasarnya dirancang dan dilatih awal (*pretrained*) pada dataset ImageNet dengan ukuran input standar 224×224 piksel. Oleh karena itu, seluruh citra diubah ukurannya (*resize*) menjadi 224×224 piksel menggunakan interpolasi bilinear agar sesuai dengan dimensi input yang disyaratkan oleh kedua model, sekaligus menjaga konsistensi ukuran input antar kedua arsitektur agar perbandingan tetap adil.

- 2) *Normalisasi Citra*: Setelah proses *resizing*, nilai piksel citra yang semula berada pada rentang 0–255 dinormalisasi ke rentang skala yang sesuai dengan kebutuhan masing-masing arsitektur menggunakan nilai mean dan standar deviasi dari dataset ImageNet, yaitu:

a) Mean: [0.485, 0.456, 0.406]

b) Standar deviasi: [0.229, 0.224, 0.225]

Normalisasi ini penting dilakukan karena kedua model menggunakan bobot hasil *pretraining* pada ImageNet, sehingga distribusi nilai piksel input perlu disesuaikan agar sejalan dengan distribusi data yang digunakan pada saat *pretraining* berlangsung. Mengingat citra ultrasound pada dataset BUSI aslinya berformat grayscale, citra tersebut terlebih dahulu dikonversi menjadi format 3-kanal (RGB) melalui duplikasi kanal, agar kompatibel dengan struktur input kedua arsitektur yang dirancang untuk citra berwarna.

- 3) *Penanganan Ketidakseimbangan Kelas (Class Imbalance)*: Sebagaimana telah dijelaskan pada bagian Dataset Penelitian, dataset BUSI memiliki distribusi kelas yang tidak seimbang, dengan kelas *benign* (437 citra) jauh lebih banyak dibandingkan kelas *normal* (133 citra) dan *malignant* (210 citra). Ketidakseimbangan ini berpotensi menyebabkan model bias dalam memprediksi kelas mayoritas, sebagaimana banyak diulas dalam kajian tentang penanganan ketidakseimbangan kelas pada *deep learning*[10]. Untuk mengatasi permasalahan ini, penelitian ini menerapkan kombinasi strategi berikut:
 - a) Augmentasi data terarah pada kelas minoritas (normal dan malignant) dengan proporsi augmentasi yang lebih besar dibandingkan kelas mayoritas, guna menyeimbangkan jumlah sampel efektif antar kelas pada data latih.
 - b) *Class weighting* pada fungsi *loss*, yaitu memberikan bobot lebih besar pada kelas dengan jumlah sampel lebih sedikit selama proses perhitungan *loss*, sehingga kesalahan prediksi pada kelas minoritas memberikan kontribusi *loss* yang lebih signifikan terhadap proses pembaruan bobot model.
- 4) *Augmentasi Data*: Augmentasi data diterapkan hanya pada data latih untuk meningkatkan variasi data dan mencegah *overfitting*, sebagaimana ditegaskan dalam berbagai kajian tentang augmentasi citra pada *deep learning*[11], tanpa mengubah distribusi data validasi maupun data uji. Teknik augmentasi yang digunakan meliputi:

TABEL 2. TEKNIK AUGMENTASI DATA

Teknik Augmentasi	Parameter	Tujuan
Random horizontal flip	Probabilitas 0.5	Variasi orientasi citra
Random rotation	$\pm 15^\circ$	Simulasi variasi sudut pengambilan
Random brightness/contrast	$\pm 10\text{--}20\%$	Simulasi variasi kualitas pencitraan
Random zoom/crop	0.9–1.1x	Variasi skala objek

Pemilihan teknik augmentasi ini mengacu pada pendekatan augmentasi data yang telah diterapkan pada klasifikasi citra ultrasound payudara[12], dengan mempertimbangkan karakteristik khas citra ultrasound, seperti tekstur derau (*speckle noise*) dan kontras rendah, sehingga teknik augmentasi yang digunakan tidak merusak informasi diagnostik penting pada citra, seperti bentuk dan batas lesi.

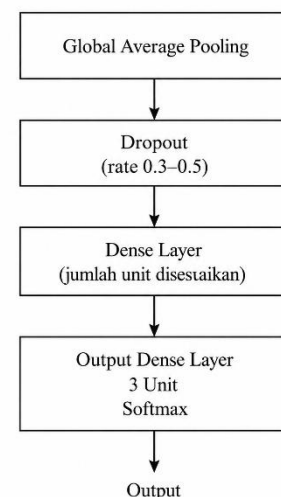
- 5) *Pembagian Dataset*: Sesuai dengan yang telah dijelaskan pada bagian sebelumnya, dataset dibagi menjadi tiga subset dengan rasio 70% data latih, 15% data validasi, dan 15% data uji, menggunakan metode *stratified split* agar proporsi masing-masing kelas (*normal*, *benign*, *malignant*) tetap konsisten pada setiap subset. Pembagian dilakukan sebelum tahap augmentasi diterapkan, untuk memastikan

tidak ada kebocoran data (*data leakage*) antara data latih dan data uji akibat augmentasi.

D. Skenario Eksperimen

Bagian ini menjelaskan rancangan eksperimen yang dilakukan untuk melatih dan menguji kedua model, ResNet-50 dan MobileNetV2, pada dataset BUSI yang telah melalui tahap praproses. Skenario eksperimen dirancang agar kedua model diperlakukan secara setara (*fair comparison*), sehingga perbedaan hasil yang diperoleh dapat diatribusikan pada perbedaan arsitektur, bukan pada perbedaan kondisi pelatihan.

- 1) *Strategi Transfer Learning*: Kedua model menggunakan pendekatan *transfer learning*[13] dengan memanfaatkan bobot hasil *pretraining* pada dataset ImageNet. Pendekatan ini dipilih karena ukuran dataset BUSI yang relatif kecil (780 citra) tidak memadai untuk melatih CNN berukuran besar dari awal (*training from scratch*) tanpa risiko *overfitting* yang tinggi. Skema *transfer learning* yang diterapkan meliputi dua tahap:
 - a) Feature extraction. Membekukan (*freeze*) seluruh lapisan konvolusi dasar (*backbone*) dan hanya melatih lapisan klasifikasi (*fully connected layer*) yang baru ditambahkan di bagian akhir model.
 - b) Fine-tuning. Setelah tahap feature extraction mencapai konvergensi awal, sebagian lapisan akhir *backbone* dibuka (*unfreeze*) dan dilatih ulang dengan *learning rate* yang lebih kecil, untuk menyesuaikan fitur yang telah dipelajari dari ImageNet terhadap karakteristik citra ultrasound.
- 2) *Modifikasi Arsitektur (Classifier Head)*: Lapisan klasifikasi asli pada kedua model (yang dirancang untuk 1000 kelas ImageNet) digantikan dengan *classifier head* baru yang disesuaikan dengan jumlah kelas pada penelitian ini (3 kelas: *normal*, *benign*, *malignant*). Struktur *classifier head* yang digunakan pada kedua model dibuat identik, terdiri atas:



Gambar 2. Struktur Classifier Head

Penyamaan struktur classifier head ini penting agar perbedaan performa yang muncul murni disebabkan oleh perbedaan backbone (ResNet-50 vs MobileNetV2), bukan oleh perbedaan desain lapisan klasifikasi.

- 3) *Konfigurasi Hyperparameter*: Untuk menjamin perbandingan yang adil, kedua model dilatih menggunakan konfigurasi *hyperparameter* yang identik sebagai berikut:

TABEL 3. KONFIGURASI HYPERPARAMETER

Hyperparameter	Nilai
Ukuran input	$224 \times 224 \times 3$
Batch size	32
Optimizer	Adam
Learning rate (feature extraction)	0.001
Learning rate (fine-tuning)	0.00001
Jumlah epoch	50 (dengan early stopping)
Loss function	Categorical Cross-Entropy (dengan class weighting)
Early stopping	Patience 10 epoch, dipantau dari validation loss
Learning rate scheduler	ReduceLROnPlateau, factor 0.5, patience 5

Nilai-nilai *hyperparameter* di atas bersifat awal (*baseline*) dan dapat disesuaikan lebih lanjut berdasarkan hasil *validation loss* pada tahap eksperimen pendahuluan (*preliminary experiment*), namun tetap diterapkan secara identik pada kedua model demi menjaga validitas perbandingan.

- 4) *Reproduksibilitas Eksperimen*: Untuk menjamin reproduksibilitas hasil, seluruh proses eksperimen menggunakan *random seed* tetap (*seed* = 42) yang diterapkan pada inialisasi bobot model, proses *shuffling* data latih, serta pembagian dataset. Pembagian dataset menjadi data latih, validasi, dan uji (rasio 70:15:15) dilakukan satu kali menggunakan metode *stratified split* dengan *random state* tetap, tanpa pengulangan pembagian data (*data splitting*) maupun validasi silang (*cross-validation*). Hasil performa yang dilaporkan pada penelitian ini merupakan hasil dari satu kali proses pelatihan dan pengujian (*single run*) untuk masing-masing model, bukan merupakan nilai rata-rata dari beberapa kali percobaan dengan *seed* atau inialisasi yang berbeda.
- 5) *Lingkungan Eksperimen (Environment)*: Seluruh proses pelatihan dan pengujian kedua model dilakukan pada lingkungan komputasi yang sama untuk menjamin konsistensi pengukuran, khususnya pada metrik efisiensi seperti waktu pelatihan dan waktu inferensi. Spesifikasi lingkungan eksperimen meliputi:

TABEL 4. ENVIRONMENT PELATIHAN DAN PENGUJIAN

Komponen	Spesifikasi
----------	-------------

Platform	Google Colab
GPU	NVIDIA T4 GPU
Framework	TensorFlow, Keras
Bahasa pemrograman	Python 3.11
Library pendukung	NumPy, Pandas, Scikit-learn, Matplotlib

- 6) *Skenario Pengujian*: Setelah proses pelatihan selesai, kedua model diuji menggunakan data uji (*test set*) yang sama sekali belum pernah dilihat model selama pelatihan maupun validasi. Pengujian dilakukan dalam dua aspek utama:

- a) Pengujian performa klasifikasi. Mengukur kemampuan model dalam mengklasifikasikan citra ke dalam tiga kelas menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan *confusion matrix*.
- b) Pengujian efisiensi komputasi. Mengukur waktu inferensi per citra (*inference time*), jumlah parameter model (*total trainable parameters*), ukuran model (*model size* dalam MB), dan penggunaan memori selama proses inferensi.

Kedua aspek pengujian ini dilakukan pada kondisi *hardware* yang sama untuk kedua model, guna memastikan hasil pengukuran efisiensi bersifat dapat dibandingkan secara langsung.

- 7) *Metrik Evaluasi*: Evaluasi model dilakukan menggunakan dua kelompok metrik, yaitu metrik performa klasifikasi dan metrik efisiensi komputasi, agar analisis *trade-off* antara akurasi dan efisiensi dapat dilakukan secara komprehensif.

Metrik performa klasifikasi dihitung berdasarkan *confusion matrix*, yang menghasilkan komponen *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) untuk tiap kelas. Berdasarkan komponen tersebut, dihitung:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Mengingat penelitian ini merupakan klasifikasi multi-kelas dengan distribusi data tidak seimbang, *precision*, *recall*, dan *F1-score* dilaporkan dalam bentuk *macro average* (rata-rata antar kelas tanpa bobot, mencerminkan performa pada kelas minoritas) dan *weighted average* (rata-rata berbobot jumlah sampel, mencerminkan performa keseluruhan model).

Metrik efisiensi komputasi meliputi jumlah parameter (*total trainable parameters*), ukuran model (*model size* dalam MB), waktu inferensi rata-rata per citra (*inference time* dalam ms), dan *throughput* (jumlah citra yang diproses per detik).

IV. HASIL DAN DISKUSI

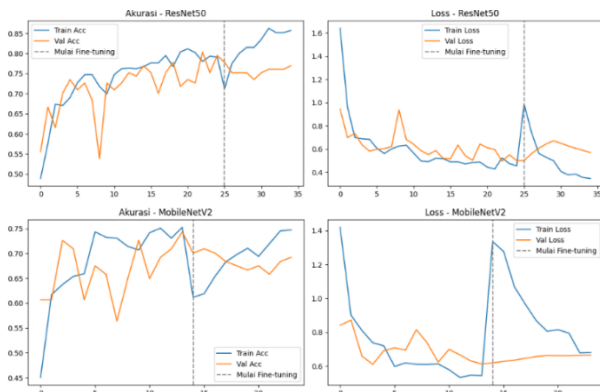
A. Hasil Pelatihan Model

Proses pelatihan kedua model dilakukan melalui dua tahap sesuai skenario yang telah dirancang pada 2.4 Skenario Eksperimen, yaitu tahap *feature extraction* (maksimal 25 epoch) yang dilanjutkan dengan tahap *fine-tuning* (maksimal 25 epoch), dengan mekanisme *early stopping* (patience 10, dipantau dari *val_loss*) dan *ReduceLROnPlateau* aktif pada kedua tahap.

ResNet-50 menjalani 25 epoch penuh pada tahap *feature extraction*, dengan akurasi validasi meningkat dari 55,56% pada epoch pertama hingga mencapai puncak 80,34% pada epoch ke-23, kemudian dilanjutkan *fine-tuning* yang dihentikan lebih awal oleh *early stopping* pada epoch ke-10 (dari maksimal 25 epoch). Kurva pada Gambar 2 (kiri atas) menunjukkan bahwa *training accuracy* ResNet-50 terus meningkat secara konsisten hingga mendekati 86%, sementara *val accuracy* relatif tertahan di kisaran 73–79% dengan sedikit fluktuasi menjelang awal *fine-tuning* (garis putus-putus abu-abu), mengindikasikan mulai munculnya sedikit *gap* antara performa data latih dan data validasi meskipun belum mengarah pada *overfitting* yang parah.

MobileNetV2 menunjukkan pola yang berbeda: tahap *feature extraction* dihentikan lebih awal oleh *early stopping* pada epoch ke-14 (dari maksimal 25 epoch), dengan akurasi validasi sempat mencapai puncak 75,21% pada epoch ke-6. Kurva pada Gambar 2 memperlihatkan fluktuasi *val accuracy* yang lebih tajam dibandingkan ResNet-50, serta penurunan cukup signifikan pada *val loss* dan akurasi tepat setelah *fine-tuning* dimulai (epoch ke-14), sebelum berangsur stabil kembali di kisaran akurasi 65–70%. Pola ini mengindikasikan bahwa MobileNetV2 lebih sensitif terhadap perubahan *learning rate* saat transisi ke tahap *fine-tuning* dibandingkan ResNet-50.

Secara umum, kurva *loss* ResNet-50 menunjukkan tren penurunan yang lebih halus dan stabil dibandingkan MobileNetV2, yang mengalami lonjakan *train loss* cukup tajam tepat pada awal *fine-tuning* sebelum kembali menurun.



Gambar 2. History Plot Hasil Pelatihan

B. Hasil Evaluasi Performa Klasifikasi

Evaluasi dilakukan pada 117 citra data uji (20 normal, 66 benign, 31 malignant) yang tidak digunakan dalam proses pelatihan maupun validasi.

1) Metrik Keseluruhan:

TABEL 5. MATRIX KESELURUHAN

Metrik	ResNet-50	MobileNetV2	Selisih
Akurasi	85,47%	75,21%	+10,26%
Precision (macro)	84,11%	70,93%	+13,18%
Recall (macro)	85,69%	73,25%	+12,44%
F1-Score (macro)	84,34%	71,92%	+12,42%
Precision (weighted)	86,00%	76,16%	+9,84%
Recall (weighted)	85,47%	75,21%	+10,26%
F1-Score (weighted)	85,39%	75,54%	+9,85%

ResNet-50 unggul secara konsisten pada seluruh metrik performa klasifikasi, dengan selisih akurasi sebesar 10,26 poin persentase dibandingkan MobileNetV2.

2) Metrik Per Kelas:

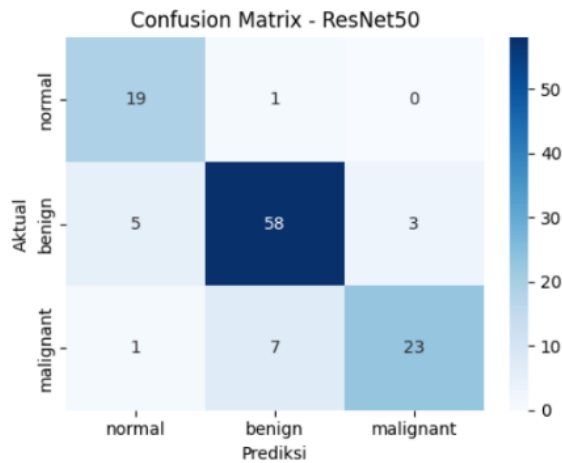
TABEL 6. MATRIX PER KELAS RESNET50

Kelas	Precision	Recall	F1-Score	Support
Normal	0,76	0,95	0,84	20
Benign	0,88	0,88	0,88	66
Malignant	0,88	0,74	0,81	31

TABEL 7. MATRIX PER KELAS MOBILENETV2

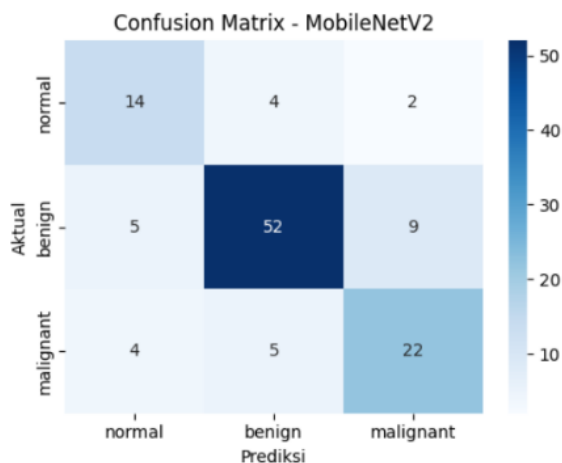
Kelas	Precision	Recall	F1-Score	Support
Normal	0,61	0,70	0,65	20
Benign	0,85	0,79	0,82	66
Malignant	0,67	0,71	0,69	31

C. Analisis Confusion Matrix



Gambar 3. Confusion Matrix – ResNet50

Pada model ResNet-50 dari 31 citra *malignant*, model berhasil mengenali 23 citra dengan benar, namun 7 citra *malignant* justru salah diklasifikasikan sebagai *benign* dan 1 citra sebagai *normal*. Sementara pada kelas *normal*, ResNet-50 hanya melakukan 1 kesalahan (diklasifikasikan sebagai *benign*), menghasilkan *recall* normal tertinggi di antara kedua model (95%).



Gambar 4. Confusion Matrix – MobileNetV2

MobileNetV2: dari 31 citra *malignant*, model berhasil mengenali 22 citra dengan benar, namun kesalahan tersebar lebih merata — 4 citra ke kelas *normal* dan 5 citra ke kelas *benign*. Pada kelas *normal*, MobileNetV2 melakukan 6 kesalahan (4 ke *benign*, 2 ke *malignant*), jauh lebih banyak dibandingkan ResNet-50.

Yang perlu digarisbawahi, kesalahan klasifikasi *malignant* menjadi *benign* atau *normal* merupakan jenis kesalahan paling kritis dalam konteks diagnosis kanker, karena berpotensi menyebabkan kasus ganas tidak terdeteksi. Pada aspek ini, ResNet-50 (*recall* *malignant* 74%) maupun MobileNetV2 (*recall* *malignant* 71%) sama-sama masih memiliki proporsi *false negative* yang cukup besar pada kelas paling kritis ini, meskipun ResNet-50 sedikit lebih baik.

D. Hasil Evaluasi Efisiensi Komputasi

TABEL 8. EVALUASI EFISIENSI KOMPUTASI

Metrik	ResNet-50	MobileNet V2	Rasio (ResNet50/MobileNetV2)
Total Parameter	23.850.371	2.422.339	≈9,8× lebih besar
Parameter Terlatih	17.476.099	1.963.971	≈8,9× lebih besar
Ukuran Model	224,75 MB	24,62 MB	≈9,1× lebih besar
Waktu Inferensi	5,754 ms/citra	3,987 ms/citra	≈1,44× lebih lambat
Throughput	173,78 citra/detik	250,82 citra/detik	≈0,69× lebih rendah

MobileNetV2 unggul telak dalam seluruh aspek efisiensi komputasi: jumlah parameternya hanya sekitar 10% dari ResNet-50, ukuran model sekitar 9 kali lebih kecil (24,62 MB berbanding 224,75 MB), serta mampu memproses citra sekitar 31% lebih cepat per citra, serta menghasilkan throughput sekitar 44% lebih tinggi dibandingkan ResNet-50.

E. Pembahasan

- 1) *Trade-off Akurasi vs Efisiensi*: Hasil penelitian ini secara jelas menunjukkan adanya trade-off antara akurasi dan efisiensi komputasi sesuai dengan yang diprediksi pada tinjauan pustaka. ResNet-50 unggul dalam seluruh metrik performa klasifikasi berkat kedalaman arsitektur dan mekanisme *residual learning* yang memungkinkannya mempelajari representasi fitur yang lebih kaya dari citra ultrasound, meskipun harus dibayar dengan kebutuhan parameter, ukuran model, dan waktu inferensi yang jauh lebih besar. Sebaliknya, MobileNetV2, dengan struktur *depthwise separable convolution* dan *inverted residual* yang dirancang untuk efisiensi, mencapai kecepatan dan ukuran model yang jauh lebih ringan, namun dengan konsekuensi penurunan akurasi dan F1-score yang cukup signifikan.
- 2) *Perbandingan dengan Penelitian Terdahulu*: Temuan bahwa ResNet-50 mengungguli MobileNetV2 pada dataset BUSI ini konsisten dengan beberapa penelitian terdahulu pada domain lalu lintas yang dibahas[3], namun bertentangan dengan studi lain pada domain tumor otak yang justru menunjukkan MobileNetV2 lebih unggul[5]. Perbedaan ini mempertegas simpulan pada tinjauan pustaka bahwa keunggulan relatif kedua arsitektur sangat bergantung pada karakteristik dataset, dalam kasus BUSI, jumlah data yang relatif kecil (780 citra) dan tekstur citra ultrasound yang kompleks (*speckle noise*, kontras rendah) tampaknya lebih diuntungkan oleh kapasitas representasi ResNet-50 yang lebih besar, dibandingkan efisiensi MobileNetV2 yang berorientasi pada kecepatan.

- 3) *Indikasi Overfitting dan Keterbatasan Dataset*: Pola kurva pelatihan pada Gambar 2, khususnya fluktuasi *val_accuracy* MobileNetV2 dan *gap* yang mulai muncul antara *train* dan *val accuracy* pada ResNet-50 menjelang akhir pelatihan, mengindikasikan bahwa ukuran dataset BUSI yang relatif kecil (546 citra latih) merupakan salah satu keterbatasan utama dalam penelitian ini. Kondisi ini konsisten dengan catatan pada tinjauan pustaka bahwa dataset BUSI, meskipun banyak dijadikan *benchmark*, memiliki keterbatasan dalam variasi dan jumlah data untuk melatih model yang sangat *generalizable*. Ketidakseimbangan kelas (kelas *benign* mendominasi dengan 437 dari 780 citra) turut berkontribusi terhadap performa yang lebih rendah pada kelas minoritas, terutama terlihat dari rendahnya *precision* MobileNetV2 pada kelas *normal* (0,61). Perlu dicatat bahwa hasil evaluasi pada penelitian ini diperoleh dari satu kali proses pelatihan dan pengujian (*single run*) untuk masing-masing model, dengan *random seed* tetap (*seed* = 42) serta satu kali *stratified split* data (70:15:15). Pendekatan ini memungkinkan hasil dapat direproduksi secara konsisten, namun belum mencerminkan variabilitas performa model yang mungkin timbul akibat perbedaan inisialisasi bobot, urutan *batch* pelatihan, maupun komposisi data latih/validasi/uji pada *seed* yang berbeda. Oleh karena itu, selisih performa antara ResNet-50 dan MobileNetV2 yang dilaporkan pada penelitian ini belum disertai ukuran variabilitas (seperti standar deviasi) yang biasanya diperoleh dari beberapa kali pengulangan eksperimen (*multiple runs*) atau validasi silang (*k-fold cross-validation*). Penelitian selanjutnya disarankan untuk melakukan beberapa kali pengulangan eksperimen dengan *seed* yang bervariasi guna memperkuat validitas statistik dari perbandingan kedua arsitektur.
- 4) *Implikasi Praktis*: Berdasarkan hasil ini, pemilihan arsitektur sebaiknya disesuaikan dengan konteks penerapan:
 - a) ResNet-50 lebih sesuai digunakan pada skenario yang mengutamakan akurasi diagnosis maksimal, seperti sistem pendukung keputusan klinis (*clinical decision support*) yang dijalankan pada infrastruktur server dengan sumber daya komputasi memadai, di mana akurasi yang lebih tinggi memberikan nilai lebih besar dibandingkan kecepatan inferensi.
 - b) MobileNetV2 lebih sesuai untuk skenario yang mengutamakan efisiensi dan portabilitas, misalnya implementasi pada perangkat mobile atau edge device di fasilitas kesehatan dengan sumber daya terbatas, dengan catatan bahwa penurunan akurasi, khususnya *recall* pada kelas *malignant*, perlu dimitigasi lebih lanjut, misalnya melalui augmentasi data tambahan, penyesuaian *threshold* klasifikasi, atau kombinasi dengan model ensemble

sebagaimana telah diterapkan pada studi klasifikasi kanker payudara lainnya[14].

V. KESIMPULAN

Penelitian ini berhasil membandingkan performa ResNet-50 dan MobileNetV2 dalam tugas klasifikasi tiga kelas (*normal*, *benign*, *malignant*) pada dataset Breast Ultrasound Images (BUSI) menggunakan pendekatan transfer learning dengan kondisi eksperimen yang identik untuk kedua model. Hasil evaluasi menunjukkan bahwa ResNet-50 unggul secara konsisten pada seluruh metrik performa klasifikasi, dengan selisih akurasi sebesar 10,26 poin persentase dibandingkan MobileNetV2, berkat kapasitas representasi fitur yang lebih dalam melalui mekanisme residual learning. Sebaliknya, MobileNetV2 unggul telak dalam seluruh aspek efisiensi komputasi, dengan jumlah parameter hanya sekitar 10% dari ResNet-50, ukuran model sekitar sembilan kali lebih kecil, serta waktu inferensi dan throughput yang sekitar ~31% lebih baik. Temuan ini mengonfirmasi adanya trade-off yang jelas antara akurasi dan efisiensi komputasi, sekaligus menegaskan bahwa keunggulan relatif suatu arsitektur CNN sangat bergantung pada karakteristik dataset yang digunakan. Berdasarkan hasil ini, ResNet-50 lebih direkomendasikan untuk skenario yang mengutamakan akurasi diagnosis maksimal pada infrastruktur server dengan sumber daya memadai, sementara MobileNetV2 lebih sesuai untuk implementasi pada perangkat mobile atau edge device dengan keterbatasan sumber daya, dengan catatan bahwa *recall* pada kelas *malignant* masih perlu ditingkatkan lebih lanjut. Penelitian selanjutnya disarankan untuk memperluas ukuran dataset, menguji arsitektur CNN atau Vision Transformer lain yang telah diterapkan pada citra ultrasound payudara[15], serta menerapkan strategi mitigasi ketidakseimbangan kelas yang lebih optimal guna meningkatkan sensitivitas deteksi kasus ganas.

DAFTAR PUSTAKA

- [1] K. He, X. Zhang, S. Ren, dan J. Sun, "Deep Residual Learning for Image Recognition," dalam 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Jun 2016, hlm. 770–778. doi: 10.1109/CVPR.2016.90.
- [2] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, dan L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," dalam 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Jun 2018, hlm. 4510–4520. doi: 10.1109/CVPR.2018.00474.
- [3] D. R. Fauzi dan G. A. Haqdu D, "Comparison of CNN Models Using EfficientNetB0, MobileNetV2, and ResNet50 for Traffic Density with Transfer Learning," Journal of Intelligent Systems Technology and Informatics, vol. 1, no. 1, hlm. 22–30, Jun 2025, doi: 10.64878/jistics.v1i1.6.
- [4] I. G. N. A. E. P. Pramudita, I. G. I. Sudipa, Y. P. Fitryani, I. B. A. I. Iswara, dan I. G. A. A. M. Aristamy, "A Comparative Study of MobileNetV2 and ResNet50 for Multi-Class AI- Generated and Real Image Classification," sinkron, vol. 10, no. 1, hlm. 536–549, Jan 2026, doi: 10.33395/sinkron.v10i1.15682.
- [5] S. Shi, "A comparison of MobileNetV2, VGG16, and ResNet50 for classifying brain tumors," dalam 2024 IEEE 2nd International Conference on Electrical, Automation and Computer Engineering (ICEACE), IEEE, Des 2024, hlm. 176–180. doi:

- 10.1109/ICEACE63551.2024.10898757.
- [6] H. Hartatik dan M. K. Anam, "Comparison of Convolutional Neural Network Architecture on Detection of Helmet Use by Humans," *Elinvo (Electronics, Informatics, and Vocational Education)*, vol. 8, no. 1, hlm. 44–54, Jun 2023, doi: 10.21831/elinvo.v8i1.52104.
 - [7] F. P. Daviana, A. Aryanti, dan N. Anugraha, "Performance Comparison Between ResNet50 and MobileNetV2 for Indonesian Sign Language Classification," *JURIKOM (Jurnal Riset Komputer)*, vol. 12, no. 3, hlm. 319–328, Jun 2025, doi: 10.30865/jurikom.v12i3.8667.
 - [8] W. Al-Dhabyani, M. Gomaa, H. Khaled, dan A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, hlm. 104863, Feb 2020, doi: 10.1016/j.dib.2019.104863.
 - [9] O. Russakovsky dkk., "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, hlm. 211–252, Des 2015, doi: 10.1007/s11263-015-0816-y.
 - [10] J. M. Johnson dan T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, Des 2019, doi: 10.1186/s40537-019-0192-5.
 - [11] C. Shorten dan T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, Des 2019, doi: 10.1186/s40537-019-0197-0.
 - [12] W. Al-Dhabyani, M. Gomaa, H. Khaled, dan A. Fahmy, "Deep Learning Approaches for Data Augmentation and Classification of Breast Masses using Ultrasound Images," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, 2019, doi: 10.14569/IJACSA.2019.0100579.
 - [13] K. Weiss, T. M. Khoshgoftaar, dan D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, hlm. 9, Des 2016, doi: 10.1186/s40537-016-0043-6.
 - [14] M. D. Ali dkk., "Breast Cancer Classification through Meta-Learning Ensemble Technique Using Convolution Neural Networks," *Diagnostics*, vol. 13, no. 13, hlm. 2242, Jun 2023, doi: 10.3390/diagnostics13132242.
 - [15] G. Ayana dan S. Choe, "BUViTNet: Breast Ultrasound Detection via Vision Transformers," *Diagnostics*, vol. 12, no. 11, hlm. 2654, Nov 2022, doi: 10.3390/diagnostics12112654.