

KOMPARASI ARSITEKTUR CNN LIGHTWEIGHT MOBILENETV2, SHUFFLENETV2, DAN EFFICIENTNETB0 UNTUK KLASIFIKASI KONDISI BAN

Gloria Missionary Tumundo, Nasya Gloria Najooan, Nofia Komala Dewi,
Pujo Hari Saputro, Sherwin Reinaldo Unsratdianto Sompie

Teknik Informatika, Universitas Sam Ratulangi
Jalan Kampus Bahu, Kecamatan Malalayang, Kota Manado, Indonesia
gloriatumundo026@student.unsrat.ac.id

ABSTRAK

Kecelakaan lalu lintas akibat kondisi ban yang tidak layak masih menjadi ancaman keselamatan yang serius, sementara inspeksi manual dinilai kurang akurat karena bersifat subjektif. Penelitian sebelumnya yang menggunakan Xception tercatat memiliki waktu inferensi 9,19 milidetik per citra saat diuji pada GPU kelas *workstation* (RTX 3060 Ti), performa yang belum terbukti dapat dicapai pada perangkat *mobile* dengan sumber daya yang jauh lebih terbatas. Penelitian ini membandingkan tiga arsitektur CNN *lightweight* yaitu MobileNetV2, ShuffleNetV2, dan EfficientNetB0 untuk klasifikasi kondisi ban menjadi kelas *defective* dan *good*, guna mencari arsitektur yang paling sesuai untuk aplikasi inspeksi ban berbasis ponsel. Dataset yang digunakan merupakan gabungan dataset publik Kaggle dan foto lapangan dengan total 3.664 citra yang mencakup kondisi pencahayaan siang dan malam. Semua model dilatih menggunakan strategi *Two-Stage Training* dengan optimizer AdamW, *Cosine Annealing*, praproses CLAHE, dan augmentasi data. Hasil pengujian pada 368 citra menunjukkan MobileNetV2 sebagai model terbaik dengan akurasi 97,83%, F1-Score 0,9782, dan hanya melewati 2 ban rusak (FN=2) dengan ukuran model 8,73 MB. EfficientNetB0 mencatat peringatan tidak perlu paling sedikit (FP=4), sedangkan ShuffleNetV2 paling cepat dalam inferensi (0,727 ms/gambar) namun dengan risiko keselamatan lebih tinggi. Validasi Grad-CAM mengonfirmasi MobileNetV2 dan EfficientNetB0 mengambil keputusan berdasarkan fitur fisik ban yang sesungguhnya.

Kata kunci : CNN lightweight, klasifikasi kondisi ban, MobileNetV2, ShuffleNetV2, EfficientNetB0

1. PENDAHULUAN

Ban adalah satu-satunya bagian kendaraan yang langsung menyentuh permukaan jalan, sehingga kondisinya sangat menentukan keselamatan saat berkendara. Secara global, kecelakaan lalu lintas telah mengakibatkan sekitar 1,19 juta kematian per tahun, di mana 92% di antaranya terpusat di negara-negara berpenghasilan rendah dan menengah, termasuk Indonesia [1]. Di antara berbagai jenis kegagalan teknis kendaraan, malfungsi pada sistem rem dan ban merupakan jenis kerusakan kendaraan yang paling dominan, menyumbang sekitar 65% dari seluruh kasus kerusakan kendaraan yang tercatat pada kecelakaan [2]. Kondisi ban yang mengalami malfungsi saat melaju terbukti secara signifikan memperburuk tingkat keparahan cedera yang dialami pengendara [3]. Masalah ini juga terjadi di dalam negeri: Korlantas Polri mencatat 110.528 kejadian kecelakaan lalu lintas sepanjang 2023 dengan 18.357 korban meninggal dunia, di mana sebagian besar korban berada pada usia produktif, dan kondisi kendaraan yang tidak layak menjadi salah satu penyebabnya [4]. Khusus untuk ban, risiko ini semakin tinggi karena cara inspeksi yang umum dilakukan saat ini masih mengandalkan pengamatan visual secara manual, yang tidak selalu akurat, mudah dipengaruhi penilaian subjektif, dan sering menghasilkan kesimpulan yang keliru tentang kondisi ban yang sebenarnya [5], [6].

Keterbatasan tersebut mendorong penggunaan teknologi deep learning berbasis Convolutional Neural Network (CNN) untuk mengotomatisasi inspeksi

kondisi ban melalui analisis citra visual [7]. Namun studi-studi terdahulu masih mengandalkan arsitektur berat seperti Xception [8], [9] dan VGG19 [9] yang tidak kompatibel untuk *deployment* pada perangkat *mobile*, sementara arsitektur *transfer learning* berukuran besar secara umum memerlukan konsumsi memori dan komputasi yang tinggi [10]. Hal ini mendorong penelitian untuk mengeksplorasi tiga arsitektur CNN ringan (*lightweight*): MobileNetV2 [11], ShuffleNetV2 [12], dan EfficientNetB0 [13], yang dipilih karena masing-masing dibangun melalui strategi efisiensi yang berbeda, sehingga perbandingan ketiganya dapat memberikan gambaran yang lebih menyeluruh dibandingkan jika hanya membandingkan varian dari satu pendekatan desain yang sama. Ketiganya juga terbukti bersaing ketat dalam klasifikasi citra pada lingkungan dengan keterbatasan sumber daya [14], namun belum ada studi yang membandingkan ketiganya secara langsung dalam konteks klasifikasi kondisi ban. Studi terdahulu mencakup berbagai pendekatan (dari arsitektur tunggal [7], [15] hingga komparasi arsitektur berat berbasis X-ray [8], [9] dan evaluasi lokal [5], [6]), namun seluruhnya belum membandingkan ketiga arsitektur *lightweight* sekaligus dengan metrik keselamatan kritis dan validasi Grad-CAM pada citra visual permukaan ban.

Studi komparasi *lightweight* CNN pada bidang lain [14], [16], [17] menunjukkan bahwa keunggulan suatu arsitektur sangat bergantung pada *dataset* yang digunakan, sehingga evaluasi spesifik pada domain

ban diperlukan. Dari keseluruhan kajian, terdapat kesenjangan yang signifikan: belum ada studi yang sekaligus membandingkan tiga arsitektur *lightweight* CNN dalam domain klasifikasi kondisi ban dengan metrik *Missed Defective* sebagai indikator keselamatan kritis, validasi Grad-CAM pada citra visual permukaan ban, dan *dataset* hibrida yang mencerminkan kondisi nyata Indonesia.

Berdasarkan kesenjangan tersebut, penelitian ini memiliki empat kontribusi utama: (1) perbandingan tiga arsitektur *lightweight* CNN (MobileNetV2, ShuffleNetV2, dan EfficientNetB0) dalam kondisi pelatihan yang sama menggunakan strategi *Two-Stage Training* dengan AdamW, *Cosine Annealing*, *Square Padding* dan CLAHE, agar evaluasi antararsitektur berlangsung secara adil; (2) penerapan metrik *Missed Defective* sebagai indikator keselamatan vital untuk mengukur proporsi ban rusak yang lolos tidak terdeteksi, melengkapi metrik akurasi global dan F1-Score; (3) implementasi Grad-CAM pada ketiga arsitektur *lightweight* untuk menghasilkan peta aktivasi visual yang memvalidasi secara ilmiah bahwa keputusan klasifikasi didasarkan pada fitur kerusakan permukaan ban yang sesungguhnya; serta (4) pembangunan *dataset* hibrida yang menggabungkan *dataset* publik Kaggle dengan citra pengambilan mandiri di lapangan untuk menangkap variasi kondisi pencahayaan ekstrem di Indonesia sebagai landasan evaluasi model yang representatif.

2. TINJAUAN PUSTAKA

2.1. Inspeksi Kondisi Ban dan Penelitian Terdahulu

Cara inspeksi ban yang umum dilakukan saat ini masih mengandalkan pengamatan visual oleh teknisi atau pemilik kendaraan. Pendekatan ini pada dasarnya

bersifat subjektif, hasilnya tidak selalu sama antar orang, dan sulit diimplementasikan secara luas [5], [6]. Padahal, degradasi fisik pada tapak ban yang tidak terdeteksi secara langsung berkontribusi pada kerusakan kendaraan yang memperburuk tingkat keparahan cedera dan meningkatkan risiko fatalitas pengendara [2], [3].

Dua pendekatan teknologi telah dikembangkan sebagai respons atas keterbatasan tersebut. Pendekatan pertama memanfaatkan inspeksi radiografi berbasis citra X-ray yang mampu mengungkap cacat pada struktur internal ban dengan akurasi tinggi [8], [9]; namun pendekatan ini bergantung pada peralatan industri berbiaya sangat tinggi yang tidak dapat diakses oleh pengguna kendaraan pribadi maupun bengkel kecil. Pendekatan kedua menggunakan analisis citra visual permukaan luar ban yang dapat diperoleh menggunakan kamera *smartphone*, menjadikannya jauh lebih praktis untuk implementasi pada perangkat *mobile* pengguna di lapangan. Pada pendekatan kedua ini, studi-studi terdahulu masih menunjukkan kesenjangan yang signifikan: sebagian mengadopsi arsitektur CNN masif yang tidak kompatibel untuk *deployment* pada perangkat *edge* seperti ResNet-34 [5] dan Xception [9], sementara sebagian lainnya hanya menguji satu arsitektur ringan tunggal tanpa komparasi simultan dan tanpa mempertimbangkan metrik keselamatan [6], [7], sedangkan studi yang mengevaluasi beberapa arsitektur sekalipun belum membandingkan ketiga arsitektur CNN *lightweight* secara simultan dalam domain klasifikasi kondisi ban dengan metrik keselamatan kritis [18]. Pemetaan sistematis terhadap penelitian terdahulu beserta celah riset yang teridentifikasi disajikan pada Tabel 1.

Tabel 1. Pemetaan Penelitian Terdahulu dan Celah Riset (*Research Gap*)

Peneliti & Tahun	Arsitektur / Pendekatan	Performa
Pathmanaban et al. [15] (2023)	CNN tunggal, citra visual	Presisi 82,39%, latensi 1,158 det/citra
Lin [7] (2023)	ShuffleNet modifikasi	<i>Detection rate</i> 94,7%
Saleh et al. [8] (2024)	Xception + Grad-CAM, citra X-ray	Akurasi 99,19%, <i>recall</i> 98,75%
Saleh & Konyar [9] (2024)	9 arsitektur masif, citra X-ray	Akurasi 94,75%, <i>recall</i> 73,7%
Mayana & Leni [5] (2023)	ResNet-34, citra visual lokal	Akurasi 91,6%
Rizky [6] (2025)	MobileNetV2, citra visual ban motor lokal	Akurasi 96,67%
Conv-Tire [18] (2022)	ResNet-50, MobileNetV2, InceptionV3, DenseNet-121	Akurasi terbaik 92,62% (DenseNet-121)
Penelitian Ini (2026)	MobileNetV2, ShuffleNetV2, EfficientNetB0	Akurasi, F1, <i>Missed Defective</i> , Latensi, Ukuran Model

Berdasarkan Tabel 1, belum ada studi yang sekaligus membandingkan ketiga arsitektur *lightweight* CNN dalam kondisi terkontrol dengan metrik *Missed Defective* dan validasi Grad-CAM pada citra visual ban, khususnya dengan *dataset* hibrida pencahayaan ekstrem khas Indonesia. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut.

2.2. Arsitektur CNN *Lightweight*

Arsitektur CNN konvensional seperti VGG-16 (138 juta parameter) dan ResNet-50 (25 juta parameter) mencapai akurasi tinggi namun memerlukan sumber daya komputasi dan memori yang substansial, menjadikannya tidak kompatibel untuk *deployment* pada perangkat *mobile* atau *edge* dengan keterbatasan RAM, kapasitas baterai, dan kecepatan pemrosesan [10]. Studi komparatif pada *benchmark*

CIFAR-100 menunjukkan bahwa MobileNetV2, ShuffleNetV2, dan EfficientNetB0 mampu mencapai akurasi kompetitif dengan jumlah parameter yang jauh lebih efisien dibandingkan arsitektur konvensional, menunjukkan relevansinya sebagai kandidat untuk sistem inspeksi ban berbasis perangkat *mobile* [14]. MobileNetV2 [11] menggunakan struktur *inverted residual* dengan lapisan *bottleneck linear* dan *depthwise separable convolution* yang memfaktorkan konvolusi standar menjadi dua operasi terpisah (konvolusi per-kanal (*depthwise*) diikuti konvolusi 1×1 (*pointwise*)) sehingga meminimalkan jumlah parameter sekaligus mempertahankan kapasitas representasi fitur. ShuffleNetV2 [12] dirancang dengan mempertimbangkan *Memory Access Cost* (MAC) secara langsung, bukan hanya jumlah operasi (FLOPs) seperti yang umum digunakan; operasi *Channel Split* dan *Channel Shuffle* memungkinkan pertukaran informasi antar-kanal secara efisien sehingga menghasilkan waktu inferensi yang sangat cepat dengan ukuran model yang kecil. EfficientNetB0 [13] menerapkan metode *Compound Scaling* yang menyeimbangkan kedalaman, lebar, dan resolusi jaringan secara simultan menggunakan koefisien tunggal hasil *Neural Architecture Search*. Pada studi Fauzan et al. [14], EfficientNetB0 menunjukkan akurasi tertinggi di antara arsitektur *lightweight* (MobileNetV2, ShuffleNetV2, EfficientNetB0) pada dataset CIFAR-100 yang digunakan sebagai *benchmark*, meskipun dengan jumlah parameter lebih besar.

Ketiga arsitektur ini dipilih karena masing-masing mewakili pendekatan desain *lightweight* CNN yang berbeda: efisiensi berbasis faktoring konvolusi (MobileNetV2), efisiensi berbasis optimasi MAC (ShuffleNetV2), dan efisiensi berbasis penskalaan sistematis (EfficientNetB0). Dengan membandingkan ketiganya dalam kondisi pelatihan yang sama, hasil analisis yang diperoleh menjadi lebih adil dan tidak condong ke salah satu pendekatan desain tertentu.

2.3. Interpretabilitas Model dan Metrik Keselamatan

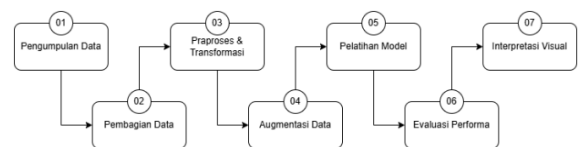
Gradient-weighted Class Activation Mapping (Grad-CAM) [19] menghasilkan peta panas (*heatmap*) yang memvisualisasikan area pada citra masukan yang paling berpengaruh terhadap keputusan klasifikasi model CNN, tanpa memerlukan modifikasi arsitektur maupun pelatihan ulang. Dalam konteks inspeksi ban, Grad-CAM digunakan untuk memvalidasi apakah keputusan model benar-benar didasarkan pada fitur kerusakan fisik ban yang sebenarnya, seperti retakan atau area keausan pada tapak ban dan bukan pada artefak latar belakang seperti tanah, velg, atau bayangan [8]. Meskipun Grad-CAM telah diterapkan pada deteksi cacat ban berbasis citra X-ray [8], penerapannya secara komparatif pada model *lightweight* untuk klasifikasi citra visual permukaan ban belum dieksplorasi secara sistematis.

Dari perspektif metrik keselamatan, penelitian ini menetapkan *Missed Defective*, didefinisikan sebagai jumlah *False Negative* (FN), yaitu ban rusak yang diprediksi sistem sebagai ban dalam kondisi baik, sebagai kriteria keputusan utama pemilihan model. Pada sistem *safety-critical* seperti inspeksi ban, kesalahan FN memiliki dampak yang jauh lebih berat dibandingkan *False Positive*: model yang meloloskan ban rusak sebagai "aman" secara langsung meningkatkan risiko kecelakaan [2], [3], sementara model yang terlalu konservatif hanya mengakibatkan penggantian ban yang tidak perlu. Pendekatan ini secara eksplisit berbeda dari studi terdahulu [5], [6], [7], [15] yang menempatkan nilai FN sebagai angka pelengkap tabel, bukan sebagai penentu rekomendasi model.

3. METODE PENELITIAN

3.1. Alur Penelitian

Penelitian ini dilakukan menggunakan platform Google Colaboratory dengan *framework* PyTorch dan akselerasi GPU Tesla T4. Keseluruhan proses terdiri dari tujuh tahap yang dilakukan secara berurutan.



Gambar 1. Aluvr penelitian

3.2. Dataset

Dataset yang digunakan merupakan kombinasi dua sumber. Sumber pertama berasal dari Kaggle (1.028 citra *defective*, 828 citra *good*), sumber kedua merupakan dataset mandiri menggunakan kamera ponsel sebanyak 1.808 citra *outdoor* dengan variasi pencahayaan siang dan malam. Total 3.664 citra (1.943 *defective*, 1.721 *good*).

Pembagian data dilakukan dengan rasio 80:10:10 menggunakan *random seed* tetap (SEED=42) agar hasil penelitian dapat direproduksi. Distribusi lengkap setelah pembagian ditampilkan pada Tabel 2.

Tabel 2. Distribusi *Dataset* Setelah Pembagian Data

<i>Split</i>	<i>Good</i>	<i>Defective</i>	<i>Total</i>
<i>Train</i>	1.376	1.554	2.930
<i>Validasi</i>	172	194	366
<i>Test</i>	173	195	368
<i>Total</i>	1.721	1.943	3.664

Set validasi dan *test* menggunakan citra asli tanpa augmentasi sehingga ukurannya proporsional terhadap data mentah, bukan data hasil augmentasi, dan evaluasi yang dihasilkan lebih mencerminkan kondisi nyata.

3.3. Praproses Citra

Praproses diterapkan pada semua *split* (*train*, *validasi*, dan *test*) untuk mencegah *data leakage*. Praproses dilakukan dalam tiga tahap. Pertama,

peningkatan kontras menggunakan Contrast Limited Adaptive Histogram Equalization (CLAHE) dengan parameter clipLimit=2,0 dan tileGridSize=8×8 pada kanal L ruang warna LAB; teknik ini meningkatkan kontras lokal tekstur ban sekaligus mereduksi *glare* akibat pantulan cahaya pada kondisi pengambilan gambar yang bervariasi. Kedua, *square padding* dengan penambahan piksel bernilai nol pada sisi terpendek gambar guna mencegah distorsi geometris saat proses *resize*. Ketiga, *resize* ke dimensi 224×224 piksel sesuai standar masukan CNN berbasis *pretrained* ImageNet.

3.4. Augmentasi Data

Augmentasi data hanya diterapkan pada *train set* untuk mencegah *data leakage* pada set validasi dan *test* sekaligus mengatasi ketidakseimbangan jumlah citra antar kelas dan mengurangi risiko *overfitting* [20]. Setiap kelas ditargetkan memiliki 5.000 citra hasil augmentasi menggunakan beberapa teknik yaitu rotasi ±20°, *horizontal flip*, *brightness range* [0,85–1,15], pergeseran lebar dan tinggi masing-masing 8%, serta *fill_mode=nearest*. Hasil akhir *train set* berjumlah 10.000 citra (5.000 kelas *good* dan 5.000 kelas *defective*). Set validasi (366 citra) dan *test* (368 citra) tidak dikenai augmentasi.

3.5. Arsitektur Model dan Transfer Learning

Tiga arsitektur CNN *lightweight* dipilih sebagai objek komparasi: MobileNetV2 [11], ShuffleNetV2 [12], dan EfficientNetB0 [13]. Ketiga model dipilih karena dirancang khusus untuk komputasi terbatas pada *edge device* dan perangkat bergerak, sehingga relevan untuk skenario inspeksi ban berbasis ponsel [10], [14]. Seluruh model diinisialisasi menggunakan bobot *pretrained* ImageNet1K_V1. *Classifier head* bawaan masing-masing model digantikan dengan lapisan Dropout(p=0,5) [21] yang diikuti *fully connected layer* dengan dua keluaran sesuai jumlah kelas. Masukan model dinormalisasi menggunakan mean [0,485; 0,456; 0,406] dan std [0,229; 0,224; 0,225] sesuai statistik ImageNet.

Pelatihan menggunakan strategi dua tahap (*Two-Stage Training*) [10]. Pada Stage 1 (*feature extraction*), seluruh *backbone* dibekukan (*frozen*) dan hanya *classifier* yang dilatih menggunakan optimizer AdamW [22] dengan *learning rate* 1×10^{-3} dan *weight decay* 1×10^{-4} , serta penjadwal StepLR (step=5, $\gamma=0,5$), maksimum 30 *epoch*. Pada Stage 2 (*fine-tuning*), seluruh lapisan dibuka (*unfrozen*) dan dilatih kembali dengan *learning rate* 1×10^{-4} menggunakan penjadwal *Cosine Annealing* [23], maksimum 70 *epoch*. Kedua tahap menerapkan *early stopping* dengan *patience*=20 [24] untuk menghentikan pelatihan saat tidak ada perbaikan pada *validation loss*. Konfigurasi hiperparameter yang identik untuk ketiga model ditampilkan pada Tabel 3.

Tabel 3. Konfigurasi Hiperparameter

Hiperparameter	Stage 1	Stage 2
Maks. Epoch	30	70
Learning Rate	1×10^{-3}	1×10^{-4}
Weight Decay	1×10^{-4}	1×10^{-4}
Batch Size	32	32
Dropout	0,5	0,5
Scheduler	StepLR (step=5, $\gamma=0,5$)	CosineAnnealingLR
Early Stopping	patience=20	patience=20
Optimizer	AdamW	AdamW
Status Backbone	Frozen	Unfrozen

3.6. Evaluasi Model

Evaluasi dilakukan pada *test set* sebanyak 368 citra yang sepenuhnya terpisah dari proses pelatihan maupun validasi. Metrik yang digunakan meliputi *accuracy*, *precision*, *recall*, dan F1-Score dengan pendekatan *macro average*, mengingat distribusi *test set* mendekati seimbang (195 *defective* : 173 *good*). Persamaan *recall* dan F1-Score adalah sebagai berikut:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1)$$

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

Pada sistem inspeksi keselamatan, False Negative (FN) atau *missed defective* ditetapkan sebagai kriteria keputusan utama, karena ban rusak yang lolos deteksi berpotensi lebih berbahaya daripada kesalahan deteksi pada ban baik [3]. Selain metrik klasifikasi, efisiensi model diukur melalui *inference time* (ms/gambar), ukuran model (MB), dan jumlah parameter. Pengukuran *inference time* dilakukan dengan protokol *warmup* 5 batch dan rata-rata dari 20 batch berikutnya pada GPU Tesla T4 untuk memperoleh estimasi yang stabil.

3.7. Analisis Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) [19] diterapkan untuk memvalidasi bahwa keputusan ketiga model didasarkan pada fitur kerusakan fisik ban, bukan artefak latar belakang. Target layer yang digunakan adalah model.features[-1] untuk MobileNetV2, model.conv5 untuk ShuffleNetV2, dan model.features[-1] untuk EfficientNetB0. Visualisasi diterapkan pada satu citra *defective* dan satu citra *good* untuk masing-masing model.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengujian ketiga arsitektur CNN *lightweight* pada *test set* yang sama sebanyak 368 citra (195 *defective*, 173 *good*) independen dari pelatihan dan validasi, menggunakan konfigurasi pelatihan dan metrik evaluasi identik agar komparasi adil, dengan mempertimbangkan metrik klasifikasi dan keselamatan model. Pembahasan mencakup proses pelatihan dan konvergensi model, performa klasifikasi, validasi interpretabilitas melalui Grad-CAM, serta perbandingan dengan penelitian terdahulu.

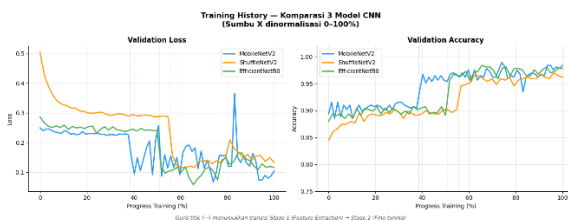
4.1. Pelatihan dan Konvergensi Model

Ketiga model dilatih menggunakan strategi *Two-Stage Training* pada platform Google Colaboratory dengan GPU Tesla T4. Hasil pelatihan ketiga model ditampilkan pada Tabel 4.

Tabel 4. Hasil Pelatihan Ketiga Model

Model	Ep. Stage 1	Ep. Stage 2	Best Val Loss	Best Val Acc (%)	Waktu (mnt)
MobileNetV2	30/30	48/70	0,0669	98,91	69,7
ShuffleNetV2	30/30	24/70	0,1150	96,45	98,4
EfficientNetB0	30/30	29/70	0,0580	98,36	105,1

Dari Tabel 4 dapat dilihat bahwa ketiga model menyelesaikan Stage 1 secara penuh selama 30 *epoch*, sementara Stage 2 dihentikan lebih awal melalui mekanisme *early stopping* dengan *patience*=20 [24]. Pola konvergensi ketiga model ditunjukkan pada Gambar 2.



Gambar 2. Training history

Dari Gambar 2 terlihat bahwa pada Stage 1 (*feature extraction*, sebelum garis titik), ketiga model menunjukkan penurunan *validation loss* yang bertahap dan stabil dengan *validation accuracy* berkisar di 88%–92%. ShuffleNetV2 memulai Stage 1 dengan *validation loss* awal yang jauh lebih tinggi (~0,50) dibandingkan MobileNetV2 (~0,25) dan EfficientNetB0 (~0,29), menunjukkan bahwa arsitektur berbasis *channel split* memerlukan lebih banyak iterasi awal untuk mengadaptasi fitur *pretrained* ImageNet pada domain tekstur ban. Pola ini secara keseluruhan sesuai dengan cara kerja *transfer learning* yang membekukan *backbone* pada tahap awal dan hanya melatih *classifier head* [10]. Transisi ke Stage 2 (*fine-tuning*, setelah garis titik) menghasilkan penurunan *validation loss* yang tajam dan peningkatan *validation accuracy* secara signifikan pada ketiga model, mengonfirmasi efektivitas strategi pembelajaran dua tahap dalam memanfaatkan representasi fitur ImageNet secara bertahap.

Total waktu *training* pada Tabel 4 mencerminkan durasi kumulatif per sesi Colaboratory termasuk inisialisasi, sehingga *inference time* per gambar (0,727 ms) lebih tepat merepresentasikan efisiensi aktual ShuffleNetV2.

Terdapat perbedaan pola konvergensi yang menarik di antara ketiga model pada Stage 2. MobileNetV2 menunjukkan fluktuasi *validation loss* pada pertengahan Stage 2 (karakteristik Cosine Annealing [23]) namun berkonvergensi ke best

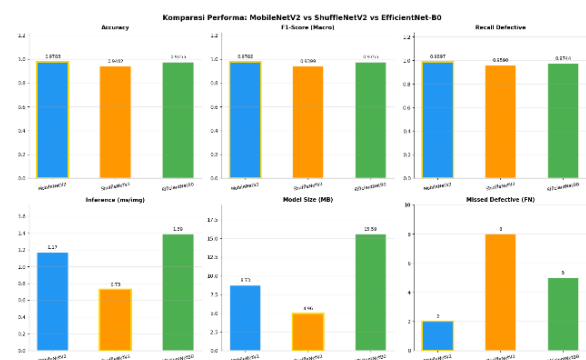
validation loss 0,0669 sebelum *early stopping epoch* ke-48. ShuffleNetV2 mencapai *early stopping* paling awal (*epoch* ke-24), yang menunjukkan bahwa kapasitas representasinya lebih terbatas dibanding dua model lainnya. EfficientNetB0 mencatat best *validation loss* terendah (0,0580) dengan best *validation accuracy* 98,36%, konsisten dengan *Compound Scaling* [13].

4.2. Evaluasi Performa dan Metrik Keselamatan

Hasil evaluasi lengkap disajikan pada Tabel 5 dan Gambar 3.

Tabel 5. Hasil Evaluasi Performa pada Test Set

Metrik	MobileNet V2	ShuffleNet V2	EfficientNet B0
Accuracy	0,9783	0,9402	0,9755
Precision (Macro)	0,9790	0,9412	0,9753
Recall (Macro)	0,9775	0,9390	0,9756
F1-Score (Macro)	0,9782	0,9399	0,9755
Recall Defective	0,9897	0,9590	0,9744
Missed Defective (FN)	2	8	5
False Alarm (FP)	6	14	4
Inference (ms/gambar)	1,170	0,727	1,388
Ukuran Model (MB)	8,73	4,96	15,59
Parameter (juta)	2,23	1,26	4,01



Gambar 3. Bar chart

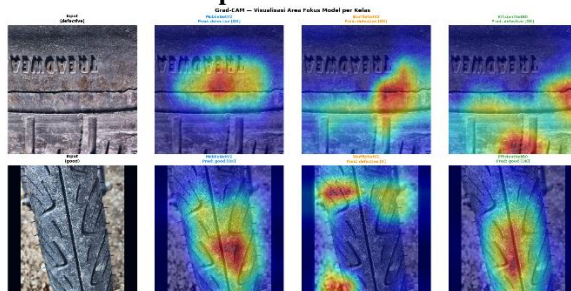
Dari Tabel 5 dan Gambar 3 dapat dilihat bahwa MobileNetV2 dan EfficientNetB0 mencapai performa klasifikasi global yang kompetitif, masing-masing dengan akurasi 97,83% dan 97,55% serta F1-Score *macro* 0,9782 dan 0,9755. ShuffleNetV2 berada di bawah kedua model tersebut dengan akurasi 94,02% dan F1-Score *macro* 0,9399. Selisih akurasi sebesar 3,8 poin persentase antara MobileNetV2 dan ShuffleNetV2 menunjukkan bahwa optimasi berbasis

Memory Access Cost (MAC) pada ShuffleNetV2 [12] menghasilkan kemampuan representasi fitur yang lebih terbatas dibandingkan pendekatan *inverted residual* pada MobileNetV2 [11] maupun *Compound Scaling* pada EfficientNetB0 [13] untuk domain klasifikasi kondisi ban.

Dari sisi keselamatan yang menjadi kriteria utama penelitian ini, indikator terpenting adalah *Missed Defective*, yaitu jumlah ban rusak yang diprediksi sistem sebagai kondisi baik (False Negative pada kelas *defective*). Dari 195 ban *defective* pada *test set*, MobileNetV2 hanya meloloskan 2 sampel (FN=2, *recall defective* 98,97%), diikuti EfficientNetB0 dengan 5 sampel (FN=5, *recall defective* 97,44%), dan ShuffleNetV2 dengan 8 sampel (FN=8, *recall defective* 95,90%). Dampak dari kesalahan ini sangat signifikan dalam konteks keselamatan berkendara: ban rusak yang lolos terdeteksi secara langsung menempatkan pengendara pada risiko kegagalan ban saat berkendara [3], sementara *false alarm* hanya menghasilkan ketidaknyamanan berupa pengecekan ulang ban yang sebenarnya masih layak pakai. MobileNetV2 merupakan model yang paling aman untuk skenario inspeksi mandiri. Di sisi lain, EfficientNetB0 mencatat *false alarm* terendah (FP=4) dibandingkan MobileNetV2 (FP=6) dan ShuffleNetV2 (FP=14), yang berarti EfficientNetB0 paling jarang memberikan peringatan yang tidak diperlukan kepada pengguna.

Dari sisi efisiensi komputasi, ShuffleNetV2 mengungguli kedua model lainnya dengan *inference time* 0,727 ms/gambar, ukuran model 4,96 MB, dan jumlah parameter 1,26 juta. Seluruh model berada jauh di bawah ambang latensi praktis yang dilaporkan Pathmanaban et al. [15] (1.158 ms/gambar), sehingga perbedaan *inference time* antar model tidak berdampak signifikan pada pengalaman pengguna. Menariknya, ukuran model tidak berbanding lurus dengan performa keselamatan maupun kecepatan. EfficientNetB0 dengan parameter dan ukuran hampir dua kali lipat MobileNetV2 justru mencatat *Missed Defective* lebih tinggi (FN=5 berbanding FN=2) dan *inference time* lebih lambat (1,388 berbanding 1,170 ms/gambar), menunjukkan bahwa penambahan kapasitas parameter pada EfficientNetB0 tidak sejalan dengan peningkatan performa keselamatan maupun kecepatan pada domain klasifikasi kondisi ban ini.

4.3. Validasi Interpretabilitas via Grad-CAM



Gambar 4. Grad-CAM

Untuk memvalidasi bahwa keputusan klasifikasi ketiga model didasarkan pada fitur kerusakan fisik ban dan bukan pada artefak latar belakang, analisis Grad-CAM [19] diterapkan pada satu sampel citra *defective* dan satu sampel citra *good* dari *test set*. Hasil visualisasi disajikan pada Gambar 4.

Dari Gambar 4 baris atas (citra *defective*), terlihat bahwa ketiga model berhasil mengklasifikasikan ban rusak dengan benar. MobileNetV2 menghasilkan peta aktivasi yang terlokalisasi secara presisi pada area retakan horizontal di bagian tengah citra, tepat di sepanjang garis keretakan yang terlihat pada permukaan ban. ShuffleNetV2 menunjukkan *hotspot* aktivasi yang berpusat di atas retakan, tepat pada area teks *embossed TREADWEAR*, dengan aktivasi sekunder yang meluas ke bawah menuju garis retakan (menunjukkan bahwa model mengandalkan kombinasi fitur tekstur *embossed* dan retakan secara bersamaan, alih-alih terlokalisasi secara presisi pada area kerusakan fisik semata). EfficientNetB0 menghasilkan dua area aktivasi yang terpisah, satu berpusat di area retakan tengah dan satu lagi di pojok kanan bawah citra, menunjukkan pola perhatian yang lebih terdistribusi namun tetap relevan pada area permukaan ban. Pola aktivasi ketiga model pada kelas *defective* secara keseluruhan terfokus pada area fisik ban dan bukan pada latar belakang, yang mengonfirmasi validitas keputusan klasifikasi sebagaimana yang ditekankan dalam pendekatan Explainable AI untuk inspeksi ban [8].

Perbedaan yang lebih signifikan dan informatif terlihat pada baris bawah (citra *good*). MobileNetV2 menghasilkan aktivasi yang terlokalisasi pada bagian tengah-bawah citra, dengan *hotspot* merah yang berpusat pada area perpotongan alur tapak ban berbentuk L (fitur yang mencerminkan kedalaman dan kelengkapan *groove* yang menjadi indikator kondisi ban masih layak pakai) dan mengklasifikasikan citra dengan benar sebagai *good*. EfficientNetB0 menghasilkan peta aktivasi yang tersebar mengikuti pola *groove* ban dari atas hingga bawah citra, dengan intensitas tinggi yang terdistribusi di sepanjang alur tapak ban secara vertikal, dan mengklasifikasikan citra dengan benar sebagai *good*, yang berarti model mengenali keseluruhan struktur permukaan tapak sebagai fitur diskriminatif kondisi ban yang masih layak. Sebaliknya, ShuffleNetV2 menunjukkan peta aktivasi dengan intensitas tertinggi justru pada area latar belakang berupa permukaan tanah berbatu di bagian bawah gambar, bukan pada permukaan ban itu sendiri, dan salah mengklasifikasikan ban baik tersebut sebagai *defective*. Pola aktivasi yang tidak terfokus ini secara visual mengonfirmasi bahwa ShuffleNetV2 memiliki kemampuan diskriminasi fitur yang lebih lemah antara tekstur permukaan ban dan artefak latar belakang, yang turut menjelaskan mengapa model ini mencatat *false alarm* (FP) tertinggi sebesar 14 sampel dan *missed defective* (FN) tertinggi sebesar 8 sampel pada *test set*. Temuan ini sesuai dengan karakteristik arsitektur ShuffleNetV2 yang

mengutamakan kecepatan dengan membatasi kapasitas kanal, sehingga representasi fitur spasial yang dihasilkan kurang mampu membedakan detail tekstur lokal yang halus [12].

4.4. Perbandingan dengan Penelitian Terdahulu dan Rekomendasi Model

Hasil penelitian ini menunjukkan peningkatan performa yang signifikan dibandingkan studi-studi terdahulu dalam domain klasifikasi kondisi ban. Dibandingkan studi terdahulu pada domain ban, ketiga model penelitian ini melampaui presisi Pathmanaban et al. [15] (82,39%), akurasi Mayana & Leni [5] (91,6%) dan Rizky [6] (96,67%), serta sebanding dengan Lin [7] (94,7%), dengan keunggulan tambahan berupa metrik *Missed Defective* dan validasi Grad-CAM yang tidak dimiliki studi-studi tersebut. Secara khusus, MobileNetV2 mencapai *recall defective* 98,97%, melampaui *recall* 73,7% yang dilaporkan Saleh & Konyar [9] sebesar lebih dari 25 poin persentase, sekaligus menggunakan citra visual permukaan tanpa peralatan X-ray industri [8], [9]. Conv-Tire [18] mencatat akurasi terbaik 92,62% melalui DenseNet-121 tanpa metrik *Missed Defective*; penelitian ini menunjukkan bahwa arsitektur *lightweight* yang diuji mampu melampaui akurasi tersebut hingga 5,2 poin persentase meski dengan ukuran model yang jauh lebih kecil.

Temuan ini mengonfirmasi fenomena *dataset-dependent* [14], [16], [17]: MobileNetV2 unggul menyeluruh pada domain ban, namun EfficientNetB0 mengungguli pada metrik *false alarm* yang menunjukkan keunggulan arsitektur bergantung pada karakteristik dataset dan metrik yang diprioritaskan. Keunggulan lain penelitian ini adalah penggunaan dataset hibrida dengan variasi pencahayaan ekstrem *outdoor* yang tidak ditemukan pada studi terdahulu, sehingga hasil evaluasi lebih representatif terhadap kondisi nyata di Indonesia. Hasil ini juga sejalan dengan Shin & Yoo [25] yang menunjukkan generalisabilitas *lightweight* CNN lintas domain inspeksi visual industri.

Berdasarkan keseluruhan analisis, penelitian ini merumuskan tiga rekomendasi skenario penerapan. Pertama, untuk skenario inspeksi mandiri berbasis keselamatan di mana minimalisasi ban rusak yang lolos menjadi prioritas utama, MobileNetV2 merupakan pilihan terbaik karena menggabungkan *missed defective* terendah (FN=2), akurasi tertinggi (97,83%), dan F1-Score terbaik (0,9782) dengan ukuran model yang moderat (8,73 MB). Kedua, untuk skenario yang mengutamakan kenyamanan pengguna dengan meminimalkan peringatan tidak perlu sambil tetap mempertahankan keamanan yang memadai, EfficientNetB0 menjadi alternatif dengan *false alarm* terendah (FP=4) meskipun dengan beban model lebih besar (15,59 MB). Ketiga, untuk skenario *deployment* pada perangkat dengan kapasitas penyimpanan sangat terbatas, ShuffleNetV2 dapat dipertimbangkan dengan pemahaman bahwa terdapat *trade-off* keselamatan

yang lebih tinggi (FN=8, *recall defective* 95,90%) [12].

5. KESIMPULAN DAN SARAN

Penelitian ini membandingkan tiga arsitektur CNN *lightweight* (MobileNetV2, ShuffleNetV2, dan EfficientNetB0) untuk klasifikasi kondisi ban (*defective/good*) menggunakan 3.664 citra dengan variasi pencahayaan ekstrem *outdoor* khas Indonesia, melalui kondisi pelatihan terkontrol berbasis strategi *Two-Stage Training* dengan AdamW dan *Cosine Annealing*. Hasil evaluasi pada *test set* 368 citra menunjukkan bahwa MobileNetV2 merupakan arsitektur terbaik untuk skenario inspeksi ban berbasis perangkat *mobile*, dengan akurasi tertinggi 97,83%, F1-Score *macro* 0,9782, *Missed Defective* terendah (FN=2, *recall defective* 98,97%), dan ukuran model 8,73 MB yang kompatibel dengan kapasitas penyimpanan *smartphone* umum. EfficientNetB0 menjadi alternatif dengan *false alarm* terendah (FP=4), sementara ShuffleNetV2 menawarkan inferensi tercepat (0,727 ms/gambar) dengan *trade-off* keselamatan lebih tinggi (FN=8). Validasi Grad-CAM mengonfirmasi fokus MobileNetV2 dan EfficientNetB0 pada fitur fisik ban yang relevan, sedangkan ShuffleNetV2 menunjukkan kelemahan diskriminasi terhadap artefak latar belakang. Penelitian ini memiliki keterbatasan pada pengujian *inference time* yang belum dilakukan langsung pada perangkat *smartphone*, serta validasi Grad-CAM yang baru diterapkan pada satu sampel per kelas. Berdasarkan keterbatasan tersebut, penelitian selanjutnya disarankan menguji model langsung pada perangkat *mobile* disertai teknik kompresi seperti *quantization* dan *pruning*, serta memperluas validasi Grad-CAM pada sampel yang lebih representatif.

DAFTAR PUSTAKA

- [1] World Health Organization, "Global status report on road safety 2023," World Health Organization, 2023. Accessed: May 19, 2026. [Online]. Available: <https://www.who.int/publications/i/item/9789240086517>
- [2] E. K. Adanu, R. Dzinyela, S. Okafor, and S. Jones, "Injury-severity analysis of crashes involving defective vehicles and accounting for the underlying socioeconomic mediators," *Heliyon*, vol. 10, no. 5, p. e26944, 2024, doi: 10.1016/j.heliyon.2024.e26944.
- [3] M. T. Haq, M. Zlatkovic, and K. Ksaibati, "Assessment of tire failure related crashes and injury severity on a mountainous freeway: Bayesian binary logit approach," *Accid. Anal. Prev.*, vol. 145, p. 105693, Sep. 2020, doi: 10.1016/j.aap.2020.105693.
- [4] Korlantas Polri, "Korban Meninggal Kecelakaan Lalu Lintas Kematian Produktif." 2024. Accessed: May 22, 2026. [Online]. Available: <https://korlantas.polri.go.id/korban-meninggal->

- kecelakaan-lalu-lintas-mayoritas-usia-produktif/
- [5] Hendri Candra Mayana and Desmarita Leni, "Deteksi Kerusakan Ban Mobil Menggunakan Convolutional Neural Network dengan Arsitektur ResNet-34," *J. Surya Tek.*, vol. 10, no. 2, pp. 842–851, 2023, doi: 10.37859/jst.v10i2.6336.
 - [6] A. Rizky, W. Widhiarso, and A. Rahman, "Klasifikasi Kelayakan Ban Sepeda Motor Menggunakan Metode Convolutional Neural Network," *Progresif J. Ilm. Komput.*, vol. 21, no. 2, p. 623, 2025, doi: 10.35889/progresif.v21i2.2917.
 - [7] S. L. Lin, "Research on tire crack detection using image deep learning method," *Sci. Rep.*, vol. 13, no. 1, pp. 1–17, 2023, doi: 10.1038/s41598-023-35227-z.
 - [8] R. A. A. Saleh, F. Al-Areqi, M. Z. Konyar, K. Kaplan, S. Öngir, and H. M. Ertunc, "Advancing Tire Safety: Explainable Artificial Intelligence-Powered Foreign Object Defect Detection with Xception Networks and Grad-CAM Interpretation," *Appl. Sci.*, vol. 14, no. 10, 2024, doi: 10.3390/app14104267.
 - [9] R. A. A. Saleh, M. Z. Konyar, K. Kaplan, and H. M. Ertunc, "End-to-end tire defect detection model based on transfer learning techniques," *Neural Comput. Appl.*, vol. 36, no. 20, pp. 12483–12503, 2024, doi: 10.1007/s00521-024-09664-4.
 - [10] M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," *Technologies*, vol. 11, no. 2, pp. 1–14, 2023, doi: 10.3390/technologies11020040.
 - [11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Jan. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.
 - [12] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "Shufflenet V2: Practical guidelines for efficient cnn architecture design," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, pp. 122–138. doi: 10.1007/978-3-030-01264-9_8.
 - [13] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *36th International Conference on Machine Learning, ICML 2019*, 2019, pp. 10691–10700. [Online]. Available: <https://arxiv.org/abs/1905.11946>
 - [14] M. R. Fauzan, N. Nadhif, R. Iskandar, and R. Z. Fauzi, "Rethinking Efficiency: A Comparative Study of Lightweight CNN Architectures for Image Classification," *JISTICS*, vol. 2, no. 1, pp. 13–21, 2026, doi: 10.64878/jistics.v2i1.167.
 - [15] P. Pathmanaban, S. Sunil, A. S. S. Jeyangel, and S. Chermadurai, "Tyre Defect Detection using Deep Learning Technique," *Int. J. Veh. Struct. Syst.*, vol. 15, no. 5, pp. 699–703, 2023, doi: 10.4273/ijvss.15.5.20.
 - [16] Y. A. Sitorus, Y. A. Sari, and S. Adinugroho, "Evaluasi Komparatif Arsitektur Lightweight CNN, MobileNetV2, dan EfficientNetB0 dalam Deteksi Penyakit Daun Jagung," *J-Ptiik*, vol. 9, no. 8, pp. 2548–964, 2025, [Online]. Available: <http://j-ptiik.ub.ac.id>
 - [17] C. Nastitie and I. Handayani, "Perbandingan Metode Mobilenetv2 Dan Efficientnetb0 Untuk Klasifikasi Jenis Pisang," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 8, no. 6, pp. 1990–1999, 2025, doi: 10.31539/0240cx09.
 - [18] L. Listyalina, I. Buyung, A. Q. Munir, I. Mustiadi, and D. A. Dharmawan, "Conv-tire: Tire feasibility assessment using convolutional neural networks," *J. Inform. dan Teknol. Inf.*, vol. 19, no. 3, pp. 323–336, 2022, doi: 10.31515/telematika.v19i3.7697.
 - [19] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020, doi: 10.1007/s11263-019-01228-7.
 - [20] K. Alomar, H. I. Aysel, and X. Cai, "Data Augmentation in Classification and Segmentation: A Survey and New Strategies," *J. Imaging*, vol. 9, no. 2, 2023, doi: 10.3390/jimaging9020046.
 - [21] I. Salehin and D. K. Kang, "A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain," *Electron.*, vol. 12, no. 14, 2023, doi: 10.3390/electronics12143106.
 - [22] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *7th Int. Conf. Learn. Represent. ICLR 2019*, 2019, [Online]. Available: <https://arxiv.org/abs/1711.05101>
 - [23] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017, pp. 1–16. [Online]. Available: <https://doi.org/10.48550/arXiv.1608.03983>
 - [24] M. K. Anam, S. Defit, Haviluddin, L. Efrizoni, and M. B. Firdaus, "Early Stopping on CNN-LSTM Development to Improve Classification Performance," *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1175–1188, 2024, doi: 10.47738/jads.v5i3.312.
 - [25] E. Shin and C. D. Yoo, "Efficient Convolutional Neural Networks for Semiconductor Wafer Bin Map Classification," *Sensors*, vol. 23, no. 4, 2023, doi: 10.3390/s23041926.