

KLASIFIKASI JENIS DAN TINGKAT KERUSAKAN JALAN MENGGUNAKAN ARSITEKTUR VGG19 (STUDI KASUS DPUPR KOTA SAMARINDA)

Muhammad Alif, Anindita Anindita Septiarini, Reza Ramadhana

Informatika, Universitas Mulawarman

Jalan Kuaro, Gunung Kelua, Kecamatan Samarinda Ulu,

Kota Samarinda, Provinsi Kalimantan Timur, 75119

m.alif7890@gmail.com

ABSTRAK

Kerusakan jalan di Kota Samarinda menjadi permasalahan yang memerlukan perhatian serius, terutama dalam proses validasi laporan yang selama ini dilakukan secara manual oleh petugas DPUPR melalui platform *Jalan Peduli*. Proses penilaian secara manual bersifat subjektif, rentan terhadap bias penilai, dan tidak efisien ketika diterapkan pada skala pelaporan besar. Arsitektur VGG19 dipilih karena susunan filter konvolusi 3×3 yang seragam dan bertumpuk menghasilkan representasi tekstur permukaan aspal yang lebih rinci dibandingkan arsitektur berfilter besar, memiliki alur ekstraksi fitur tunggal tanpa koneksi residual maupun percabangan multi-skala sehingga lebih stabil dan tidak mudah *overfitting* pada dataset lokal berskala kecil seperti pada ResNet dan Inception, serta mempertahankan detail tekstur halus yang cenderung hilang pada arsitektur ringan berbasis *depthwise separable convolution* seperti MobileNet, padahal detail tersebut menjadi penentu utama dalam membedakan tingkat keparahan kerusakan. Penelitian ini mengembangkan model klasifikasi otomatis berbasis arsitektur VGG19 dengan pendekatan *transfer learning* untuk mengidentifikasi jenis dan tingkat keparahan kerusakan jalan dari citra permukaan aspal. Dataset lokal sebanyak 1.731 citra dikumpulkan dari wilayah Kota Samarinda, mencakup sembilan kelas yang merupakan kombinasi tiga jenis kerusakan (lubang, retak, dan permukaan tidak rata) dengan tiga tingkat keparahan (ringan, sedang, dan berat). Data dibagi dengan rasio 70:15:15, dan augmentasi dinamis diterapkan pada data latih hingga target 800 citra per kelas. Model dilatih dalam dua tahap: *feature extraction* dan *fine-tuning* pada lapisan *block5_conv4*. Hasil pengujian menunjukkan bahwa skenario *fine-tuning* menghasilkan akurasi terbaik sebesar 66,54% dengan *macro-average F1-score* 0,5970, meningkat 5,57% dibandingkan skenario *feature extraction*. Waktu inferensi rata-rata 3.479,40 milidetik per citra, jauh lebih cepat dari validasi manual petugas yang membutuhkan 3-5 menit per laporan.

Kata kunci : klasifikasi kerusakan jalan, VGG19, transfer learning, convolutional neural network, citra digital, Kota Samarinda

1. PENDAHULUAN

Infrastruktur jalan memiliki peran penting dalam mendukung mobilitas masyarakat, distribusi barang, dan pertumbuhan ekonomi [1]. Kualitas jalan yang baik meningkatkan efisiensi transportasi serta mendukung keselamatan pengguna jalan. Sebaliknya, kerusakan jalan dapat menimbulkan hambatan aktivitas, kerugian ekonomi, dan meningkatkan risiko kecelakaan lalu lintas [2].

Di Kota Samarinda, kerusakan jalan masih menjadi masalah yang memerlukan perhatian serius. Faktor seperti beban kendaraan berlebih, cuaca ekstrem, dan pemeliharaan yang tidak teratur dapat mempercepat timbulnya kerusakan [3]. Bentuk kerusakan yang paling umum mencakup lubang, retak, dan permukaan jalan tidak rata. Jika tidak ditangani dengan cepat, kerusakan dapat semakin parah dan menyebabkan biaya perbaikan yang lebih tinggi [4].

Untuk menangani permasalahan ini, DPUPR Kota Samarinda mengembangkan sistem pelaporan berbasis *web* bernama *Jalan Peduli*, yang memungkinkan masyarakat melaporkan kerusakan jalan dengan mengunggah foto kondisi jalan. Namun, sistem pelaporan sebelumnya melalui formulir daring memiliki keterbatasan, yaitu tidak adanya validasi foto

otomatis dan minim transparansi, yang kerap menyebabkan keterlambatan penanganan serta ketidakteraturan data [5]. Meski sistem *Jalan Peduli* telah mengatasi permasalahan alur pelaporan, kelemahan mendasar masih terdapat pada proses validasi yang sepenuhnya bergantung pada penilaian visual petugas secara manual [6]. Proses ini bersifat subjektif, rentan terhadap bias penilai [7], dan tidak efisien pada skala pelaporan yang besar.

Oleh karena itu, diperlukan pendekatan otomatis berbasis kecerdasan buatan yang mampu menganalisis foto kerusakan secara langsung. Klasifikasi otomatis berbasis *machine learning* terbukti mampu mempercepat pengambilan keputusan dan meningkatkan efisiensi pemantauan infrastruktur [8], [9]. *Convolutional Neural Network* (CNN) adalah arsitektur *deep learning* yang dirancang khusus untuk memproses data gambar dengan cara mengekstraksi fitur visual secara bertingkat, mulai dari pola sederhana hingga representasi kompleks, sehingga mampu mengklasifikasikan objek berdasarkan pola visual secara otomatis [11]. Penerapan CNN pada sistem *Jalan Peduli* bertujuan mengidentifikasi jenis dan tingkat keparahan kerusakan jalan berdasarkan foto permukaan aspal yang dilaporkan masyarakat,

mencakup tiga jenis kerusakan utama, yaitu lubang, retak, dan permukaan tidak rata, dengan masing-masing tiga tingkat keparahan: ringan, sedang, dan berat [12].

Klasifikasi pada penelitian ini dirancang ke dalam sembilan kelas yang merupakan kombinasi tiga jenis kerusakan dengan tiga tingkat keparahan karena kebutuhan validasi laporan di DPUPR tidak berhenti pada pengenalan jenis kerusakan saja. Jenis kerusakan menentukan metode penanganan yang berbeda, yaitu penambalan pada lubang, penutupan celah atau pelapisan ulang pada retak, dan perataan ulang lapisan permukaan pada permukaan tidak rata, sedangkan tingkat keparahan menentukan urgensi penanganan, volume pekerjaan, dan estimasi biaya perbaikan [3]. Label yang hanya memuat salah satu dari kedua aspek tersebut belum cukup untuk mendukung penentuan prioritas penanganan. Kedua aspek digabungkan menjadi satu label tunggal, bukan dipisah menjadi dua model bertingkat, agar kesalahan prediksi jenis kerusakan pada tahap pertama tidak merambat ke prediksi tingkat keparahan pada tahap berikutnya, dan agar keluaran model selaras dengan format satu label per laporan yang digunakan pada sistem *Jalan Peduli*. Konsekuensi dari perancangan ini adalah meningkatnya kompleksitas tugas klasifikasi, karena model harus membedakan kelas-kelas yang berbagi jenis kerusakan yang sama dengan perbedaan visual yang bertingkat dan tidak selalu tegas.

Penelitian ini menggunakan arsitektur VGG19 karena terbukti stabil dan memiliki akurasi unggul untuk klasifikasi citra kerusakan jalan [14]. Metode *transfer learning* diterapkan untuk mengatasi keterbatasan data lokal, mempercepat konvergensi, dan mencegah *overfitting* [15]. Penggunaan dataset lokal dari Kota Samarinda juga menjadi fokus utama penelitian ini, mengingat adanya variasi karakteristik jalan seperti tekstur, pencahayaan, dan pola kerusakan yang berbeda antarwilayah [16]. Tujuan penelitian ini adalah membangun model klasifikasi otomatis yang berpotensi mendukung proses validasi laporan kerusakan jalan secara lebih efisien dan objektif pada konteks sistem pelaporan *Jalan Peduli*.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Sejumlah penelitian terdahulu telah mengembangkan berbagai pendekatan untuk mendeteksi dan mengklasifikasikan kerusakan jalan berbasis citra digital. Utaminigrum et al. [17] menggunakan kombinasi *Gray-Level Co-occurrence Matrix* (GLCM), *Genetic Algorithm*, dan *Extreme Learning Machine* untuk deteksi lubang jalan, menghasilkan akurasi 88,65% dengan 57 fitur terpilih. Rachman et al. [18] menerapkan *Faster R-CNN* dengan tiga *backbone* arsitektur (VGG-16, ResNet-50, dan MobileNetV2) untuk deteksi kerusakan jalan, di mana MobileNetV2 menghasilkan performa terbaik dengan nilai *mAP* sebesar 79,7%. Vishwakarma &

Vennelakanti [19] membandingkan *Faster R-CNN* dan YOLOv5 pada dataset global *Road Damage*, dengan *Faster R-CNN* berbasis ResNet-50 mencapai *mean F1-score* sebesar 0,542.

Tang et al. [20] menerapkan VGG-19 dengan pendekatan *transfer learning* dan klasifier *Softmax* untuk klasifikasi jenis kerusakan jalan, menghasilkan akurasi sebesar 86,2%, meningkat signifikan dibandingkan model CNN lainnya. Chen et al. [22] mengembangkan metode *Grayscale-Weighted HOG* (GHOG) yang dikombinasikan dengan CNN untuk deteksi kerusakan jalan, mencapai akurasi 95% dengan *F1-score* 93,8%. Wang et al. [23] mengintegrasikan modul *Edge-Enhanced Multi-Scale Feature Fusion* (EE-MSFF) pada ResNet18 untuk meningkatkan deteksi kerusakan, menghasilkan akurasi 88,68% pada dataset RDD2020. Saha et al. [24] mengembangkan model *Federated Learning* berbasis YOLOv5 untuk deteksi kerusakan lintas negara tanpa berbagi data mentah antarserver.

Berdasarkan kajian tersebut, perbedaan utama penelitian ini terletak pada penggunaan dataset lokal yang dikumpulkan dari kondisi jalan Kota Samarinda, tujuan integrasi pada sistem pelaporan *Jalan Peduli*, serta pendekatan klasifikasi ganda yang mengklasifikasikan sekaligus jenis kerusakan dan tingkat keparahannya ke dalam sembilan kelas secara simultan.

2.2. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah arsitektur *deep learning* yang dirancang khusus untuk memproses data dengan struktur *grid* seperti citra digital. CNN menggunakan operasi konvolusi matematis sebagai mekanisme fundamental untuk mengekstraksi fitur spasial dari *input*, dengan keunggulan *weight sharing* dan sifat *translation invariance* yang memungkinkan model mengenali objek di berbagai posisi dalam citra [32]. Operasi konvolusi direpresentasikan sebagai:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) \cdot K(m, n) \quad (1)$$

di mana $S(i, j)$ adalah nilai *output* pada posisi (i, j) dalam *feature map*, I adalah matriks citra *input*, dan K adalah *kernel* konvolusi yang dapat dipelajari.

Arsitektur CNN terdiri dari *convolutional layer* untuk mengekstraksi fitur, *pooling layer* untuk mereduksi dimensi spasial, *flatten layer* sebagai penghubung, dan *fully connected layer* untuk klasifikasi akhir. Fungsi aktivasi ReLU (Rectified Linear Unit) mendefinisikan keluaran $f(x) = \max(0, x)$ dan berperan memperkenalkan non-linearitas ke dalam model, sedangkan fungsi *Softmax* pada lapisan keluaran mengubah vektor logit menjadi distribusi probabilitas untuk klasifikasi multi-kelas [32]. Teknik *dropout* diterapkan sebagai regularisasi untuk mencegah *overfitting* dengan secara acak menonaktifkan sejumlah neuron selama pelatihan [32].

2.3. Arsitektur VGG19

VGG19 adalah arsitektur CNN sangat dalam yang dikembangkan oleh Simonyan & Zisserman [37] untuk mengkaji pengaruh kedalaman jaringan terhadap akurasi pengenalan citra skala besar. Arsitektur ini menggunakan filter konvolusi kecil berukuran 3×3 yang disusun bertingkat hingga mencapai 19 lapisan berbobot. Masukan jaringan berupa citra RGB berukuran 224×224 piksel yang kemudian diproses melalui lima blok konvolusi dengan *max-pooling* 2×2 , lalu dilanjutkan ke tiga lapisan *fully connected* (FC-4096, FC-4096, FC-1000) dengan aktivasi *Softmax* pada keluaran [37]. Desain *stacking* filter 3×3 memiliki *receptive field* efektif yang setara dengan filter yang lebih besar namun dengan jumlah parameter lebih efisien sekaligus menambah non-linearitas.

Dalam implementasi penelitian ini, VGG19 tersedia sebagai model *pre-trained* pada dataset ImageNet yang dilatih pada lebih dari satu juta citra untuk mengklasifikasikan 1.000 kategori. Lapisan *fully connected* bawaan VGG19 digantikan dengan *custom classifier head* yang disesuaikan dengan sembilan kelas target penelitian, mencakup lapisan GlobalAveragePooling2D, dua lapisan Dense dengan regularisasi L2 dan *dropout*, serta lapisan output Dense(9) dengan fungsi aktivasi *softmax* [17].

2.4. Transfer Learning

Transfer learning adalah pendekatan yang memanfaatkan pengetahuan dari model yang telah dilatih pada domain sumber untuk meningkatkan kinerja pada domain target, terutama ketika data target terbatas [18]. Dalam penelitian ini, *transfer learning* diterapkan melalui dua strategi. Pertama, *feature extraction*, di mana seluruh lapisan konvolusi VGG19 dibekukan dan hanya lapisan klasifikasi akhir yang dilatih. Kedua, *fine-tuning*, di mana lapisan konvolusi terluar (*block5_conv4*) dibuka dan dilatih ulang menggunakan *learning rate* yang lebih kecil agar model lebih adaptif terhadap karakteristik domain target tanpa merusak bobot *pre-trained* [18]. Pendekatan ini selaras dengan temuan survei komprehensif tahun 2024 yang menekankan *transfer learning* sebagai strategi penting untuk meningkatkan adaptasi model pada variasi data lapangan [19].

2.5. Evaluasi Model

Evaluasi model dilakukan menggunakan *confusion matrix* sebagai instrumen dasar untuk memetakan perbandingan antara prediksi model dengan label aktual [20]. Dari *confusion matrix* tersebut, empat metrik statistik diturunkan: akurasi, presisi, *recall*, dan *F1-score* [21]. Akurasi mengukur proporsi prediksi benar dari seluruh prediksi sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Presisi mengukur ketepatan prediksi positif (3), *recall* mengukur kemampuan model mendeteksi

seluruh data positif aktual (4), dan *F1-score* merupakan rata-rata harmonik dari presisi dan *recall* (5) [21].

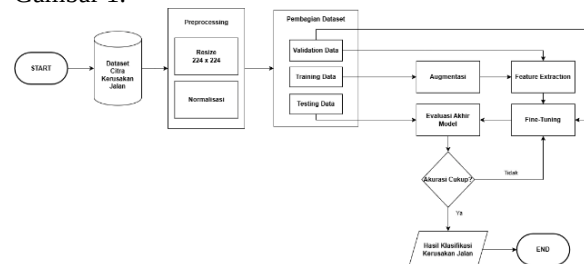
$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1-Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (5)$$

3. METODE PENELITIAN

Bagian ini menjelaskan metode yang digunakan untuk membangun model klasifikasi kerusakan jalan berbasis VGG19, mulai dari penyusunan dataset hingga implementasi antarmuka inferensi. Alur proses klasifikasi secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Alur Proses Klasifikasi Kerusakan Jalan Berbasis VGG19

3.1. Dataset Citra Kerusakan Jalan

Dataset yang digunakan berupa foto permukaan jalan rusak di Kota Samarinda yang merepresentasikan kondisi nyata ruas jalan pada berbagai lingkungan. Data dikumpulkan dari tiga sumber utama, yaitu platform pelaporan Jalan Peduli DPUPR Kota Samarinda, dokumentasi lapangan menggunakan kamera smartphone dengan resolusi minimal 12 MP, serta arsip dokumentasi kerusakan jalan dari DPUPR.

Seluruh citra mentah melalui proses kurasi dengan kriteria: resolusi minimal 224×224 piksel, objek kerusakan terlihat jelas dan menutupi sedikitnya 30% area citra, tidak buram akibat guncangan kamera, dan hanya memuat satu jenis kerusakan dominan agar label kelas tidak rancu. Setelah kurasi diperoleh 1.731 citra yang diklasifikasikan ke dalam sembilan kelas, yaitu kombinasi tiga jenis kerusakan (lubang, retak, dan permukaan tidak rata) dengan tiga tingkat keparahan (ringan, sedang, dan berat) sebagaimana dirangkum pada Tabel 1.

Tabel 1. Skema Klasifikasi Kerusakan Jalan

No	Jenis Kerusakan	Ringan	Sedang	Berat
1	Lubang	Lubang Ringan	Lubang Sedang	Lubang Berat
2	Retak	Retak Ringan	Retak Sedang	Retak Berat
3	Permukaan Tidak Rata	PTR Ringan	PTR Sedang	PTR Berat



Gambar 2. Contoh Citra Kerusakan Jalan

3.2. Preprocessing Citra

Preprocessing dilakukan untuk menyamakan karakteristik data sebelum masuk ke model VGG19. Setiap citra diubah ukurannya menjadi 224×224 piksel sesuai spesifikasi masukan VGG19 dan agar struktur spasial fitur tetap konsisten. Selanjutnya, nilai piksel dinormalisasi ke rentang 0-1 dengan membagi setiap kanal RGB dengan 255, sehingga distribusi nilai menjadi lebih stabil dan membantu proses optimisasi selama pelatihan.

3.3. Pembagian Dataset

Dataset yang telah diproses kemudian dibagi menjadi tiga subset independen menggunakan metode stratified random sampling dengan rasio 70:15:15 untuk data latih, validasi, dan uji. Pendekatan stratifikasi memastikan proporsi masing-masing kelas tetap representatif di setiap subset, sehingga distribusi kelas pada data latih, validasi, dan uji tetap seimbang relatif terhadap dataset asal; ringkasan rasionya ditunjukkan pada Tabel 2.

Tabel 2. Rasio Pembagian Dataset Latih, Validasi, dan Uji

Kelas	Latih	Validasi	Uji	Total
Lubang_Berat	70	15	16	101
Lubang_Ringan	175	37	38	250
Lubang_Sedang	209	44	46	299
PTR_Berat	62	13	14	89
PTR_Ringan	16	3	5	24
PTR_Sedang	45	9	11	65
Retak_Berat	85	18	19	122
Retak_Ringan	366	78	80	524
Retak_Sedang	179	38	40	257
Total	1.207	255	269	1.731

3.4. Augmentasi Data

Distribusi awal dataset sangat tidak seimbang, dengan jumlah citra per kelas berkisar antara 16 hingga 366 citra sehingga rasio antara kelas terbanyak dan terkecil mencapai sekitar 21,8:1. Untuk mengurangi ketidakseimbangan ini, dilakukan augmentasi data secara dinamis pada subset latih hingga setiap kelas mencapai target sekitar 800 citra, sehingga total data latih setelah augmentasi menjadi 7.196 citra.

Teknik augmentasi yang digunakan meliputi rotasi acak hingga 15 derajat, *flip* horizontal, pergeseran posisi hingga 10% pada sumbu horizontal dan vertikal, serta pengaturan kecerahan pada rentang 0,8–1,2. Data validasi dan uji tidak diaugmentasi agar tetap merepresentasikan kondisi lapangan yang sebenarnya.

3.5. Pelatihan Model VGG19

Model dibangun menggunakan arsitektur VGG19 yang diinisialisasi dengan bobot pra-latih ImageNet melalui Keras API. Lapisan *fully connected* bawaan VGG19 dihilangkan dan diganti dengan classifier head khusus yang terdiri atas lapisan GlobalAveragePooling2D, lapisan Dense 512 dengan regularisasi L2 dan dropout 0,5, lapisan Dense 256 dengan regularisasi L2 dan dropout 0,3, serta lapisan output Dense 9 dengan fungsi aktivasi *softmax* untuk memetakan ke sembilan kelas kerusakan. Ringkasan konfigurasi arsitektur ditunjukkan pada Tabel 3.

Pelatihan dilakukan dalam dua skenario berurutan. Pada skenario pertama (*feature extraction*), seluruh lapisan konvolusi VGG19 dibekukan dan hanya *classifier head* yang dilatih menggunakan optimizer Adam dengan learning rate $1e-3$, batch size 32, dan maksimum 50 epoch, disertai mekanisme EarlyStopping dengan patience 10 epoch. Pada skenario kedua (*fine-tuning*), lapisan konvolusi block5_conv4 dibuka dan dilatih kembali bersama classifier head menggunakan learning rate lebih kecil ($1e-4$), dengan EarlyStopping ber-patience 5 epoch dan ReduceLROnPlateau yang menurunkan learning rate sebesar faktor 0,5 jika loss validasi tidak membaik selama 3 *epoch*. Seluruh eksperimen dijalankan pada lingkungan Google Colaboratory dengan GPU NVIDIA Tesla T4 menggunakan Python dan TensorFlow.

Tabel 3. Konfigurasi Arsitektur Model VGG19

No	Layer	Type	Output Shape	Parameter
1	vgg19	Functional (Base Model)	(None, 7, 7, 512)	20.024.384
2	global_average_pooling2d	GlobalAveragePooling2D	(None, 512)	0
3	dense	Dense (512, ReLU, L2)	(None, 512)	262.656
4	dropout	Dropout (0,5)	(None, 512)	0
5	dense_1	Dense (256, ReLU, L2)	(None, 256)	131.328
6	dropout_1	Dropout (0,3)	(None, 256)	0
7	dense_2	Dense (9, Softmax)	(None, 9)	2.313
Total Parameter				20.420.681
Trainable Parameters				396.297
Non-trainable Parameters				20.024.384

3.6. Evaluasi Model

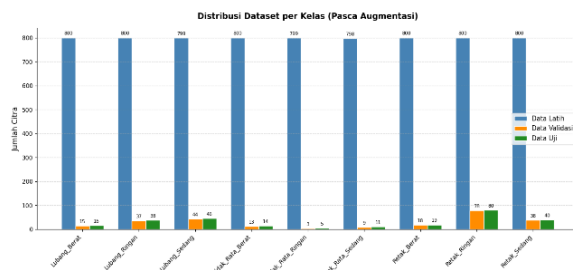
Evaluasi model dilakukan menggunakan confusion matrix yang memetakan hubungan antara label prediksi dan label sebenarnya pada data uji. Dari confusion matrix tersebut dihitung metrik akurasi, presisi, recall, dan F1-score baik secara per kelas maupun macro-average untuk menilai performa keseluruhan model pada sembilan kelas kerusakan.

Rumus masing-masing metrik dijelaskan pada tinjauan pustaka, sehingga pada bagian metode ini fokus diberikan pada cara penerapannya, yaitu dengan menggunakan nilai True Positive, True Negative, False Positive, dan False Negative yang diperoleh dari confusion matrix multi-kelas.

4. HASIL DAN PEMBAHASAN

4.1 Arsitektur Model dan Pelatihan

Setelah proses pengumpulan dan kurasi, diperoleh 1.731 citra kerusakan jalan dari Kota Samarinda. Augmentasi dinamis pada data latih menghasilkan 7.196 citra yang tersebar merata di seluruh sembilan kelas, dengan enam kelas mencapai tepat 800 citra dan tiga kelas mengalami selisih kurang dari 0,3% dari target (Lubang_Sedang 799 citra, PTR_Ringan 799 citra, PTR_Sedang 798 citra). Distribusi data validasi dan uji mencerminkan proporsi asli yang tidak seimbang, dengan kelas Retak_Ringan memiliki jumlah terbesar (78 citra validasi, 80 citra uji) dan kelas PTR_Ringan paling sedikit (3 citra validasi, 5 citra uji). Distribusi *dataset* pasca augmentasi ditunjukkan pada Gambar 3.

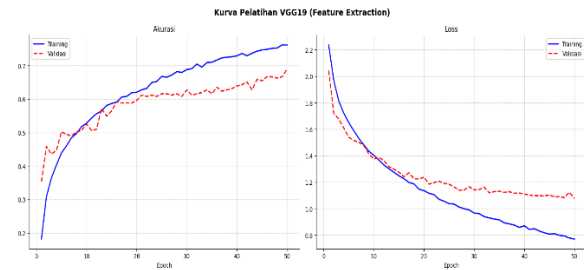


Gambar 3. Distribusi Dataset per Kelas Pasca Augmentasi

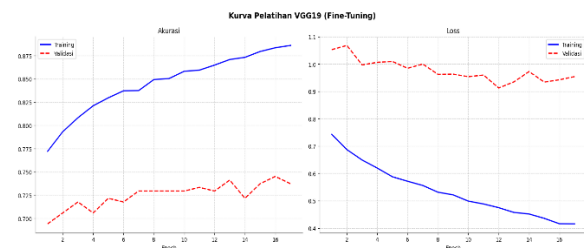
4.2 Hasil Pelatihan

Pada tahap *feature extraction*, kurva pelatihan menunjukkan akurasi data latih meningkat dari sekitar 19% pada *epoch* pertama hingga sekitar 75% pada *epoch* ke-50, sementara akurasi validasi meningkat dari sekitar 36% menjadi sekitar 69%. Pola ini mengindikasikan proses pelatihan yang stabil tanpa gejala *overfitting* yang signifikan. Pada tahap *fine-tuning*, pelatihan berjalan selama 17 *epoch* hingga *EarlyStopping* aktif, dengan akurasi data latih meningkat dari sekitar 77% menjadi sekitar 88% dan akurasi validasi dari sekitar 70% menjadi sekitar 74%. Kesenjangan antara *loss* latih yang terus menurun (0,75 menjadi 0,42) dan *loss* validasi yang stagnan pada rentang 0,95-1,05 merupakan konsekuensi yang

dapat diprediksi pada *dataset* berskala terbatas dengan distribusi kelas tidak seimbang [11]. Kurva pelatihan kedua skenario ditunjukkan pada Gambar 4.



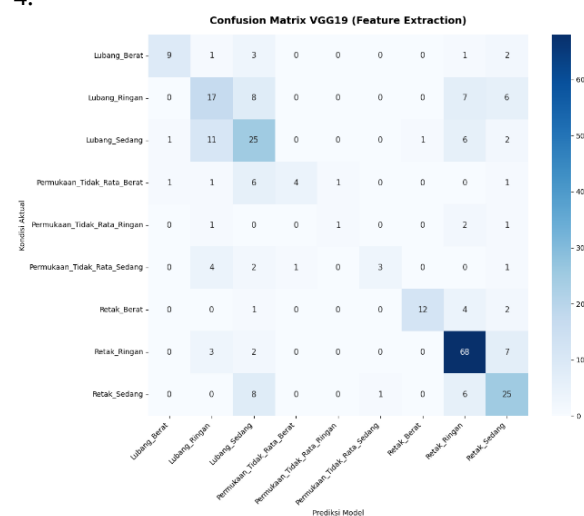
Gambar 4a. Kurva Pelatihan VGG19 Feature Extraction



Gambar 4b. Kurva Pelatihan VGG19 Fine-Tuning

4.3 Hasil Pengujian Feature Extraction

Pengujian dilakukan pada 269 citra data uji yang independen dari proses pelatihan. Hasil *confusion matrix* menunjukkan bahwa model *feature extraction* memperoleh performa terbaik pada kelas Retak_Ringan dengan 68 dari 80 citra berhasil diprediksi benar. Sebaliknya, kelompok Permukaan Tidak Rata (PTR) secara keseluruhan memiliki performa terendah: PTR_Ringan hanya benar 1 dari 5 citra dan PTR_Sedang benar 3 dari 11 citra. *Confusion matrix* ditunjukkan pada Gambar 5 dan hasil metrik evaluasi per kelas secara lengkap disajikan pada Tabel 4.



Gambar 5. Confusion Matrix VGG19 Feature Extraction

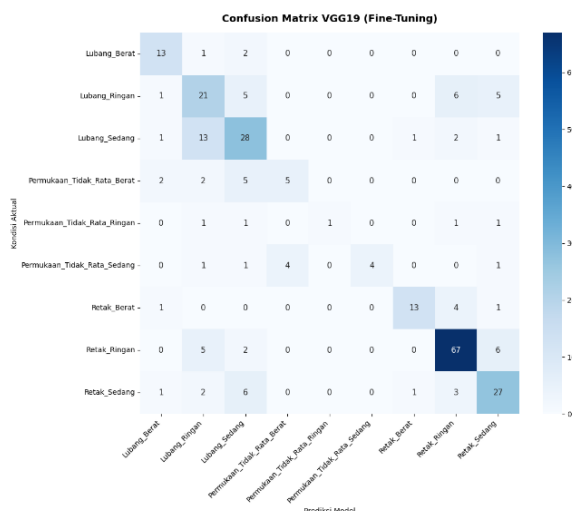
Tabel 4. Hasil Metrik Evaluasi per Kelas Feature Extraction

Kelas	Precision	Recall	F1-Score	Support
Lubang_Berat	0,8182	0,5625	0,6667	16
Lubang_Ringan	0,4474	0,4474	0,4474	38
Lubang_Sedang	0,4545	0,5435	0,4950	46
PTR_Berat	0,8000	0,2857	0,4211	14
PTR_Ringan	0,5000	0,2000	0,2857	5
PTR_Sedang	0,7500	0,2727	0,4000	11
Retak_Berat	0,9231	0,6316	0,7500	19
Retak_Ringan	0,7234	0,8500	0,7816	80
Retak_Sedang	0,5319	0,6250	0,5747	40
Macro Avg	0,6609	0,4909	0,5358	269
Weighted Avg	0,6306	0,6097	0,6020	269
Akurasi			0,6097	269

Model *feature extraction* memperoleh akurasi keseluruhan 60,97% dengan *macro-average F1-score* 0,5358. Kelas Retak_Berat mencatat *precision* tertinggi sebesar 0,9231, sedangkan kelas PTR_Ringan memperoleh nilai terendah pada seluruh metrik dengan *F1-score* hanya 0,2857.

4.4 Hasil Pengujian Fine-Tuning

Skenario *fine-tuning* menunjukkan peningkatan kemampuan prediksi yang lebih merata dibandingkan *feature extraction*. Kelas Lubang_Berat meningkat dari 9 menjadi 13 prediksi benar. Kelas Retak_Ringan tetap menjadi kelas dengan prediksi terbanyak dengan 67 prediksi benar dari 80 citra, meski terdapat 5 citra yang kini salah diklasifikasikan ke Lubang_Ringan yang merupakan kesalahan lintas jenis paling dominan pada skenario ini. Kelompok PTR masih menjadi tantangan utama. *Confusion matrix* dan metrik evaluasi lengkap ditunjukkan pada Gambar 6 dan Tabel 5.



Gambar 6. Confusion Matrix VGG19 Fine-Tuning

Tabel 5. Hasil Metrik Evaluasi per Kelas Fine-Tuning

Kelas	Precision	Recall	F1-Score	Support
Lubang_Berat	0,6842	0,8125	0,7429	16
Lubang_Ringan	0,4565	0,5526	0,5000	38
Lubang_Sedang	0,5600	0,6087	0,5833	46
PTR_Berat	0,5556	0,3571	0,4348	14
PTR_Ringan	1,0000	0,2000	0,3333	5
PTR_Sedang	1,0000	0,3636	0,5333	11
Retak_Berat	0,8667	0,6842	0,7647	19
Retak_Ringan	0,8072	0,8375	0,8221	80
Retak_Sedang	0,6429	0,6750	0,6585	40
Macro Avg	0,7303	0,5657	0,5970	269
Weighted Avg	0,6862	0,6654	0,6616	269
Akurasi			0,6654	269

Akurasi keseluruhan skenario *fine-tuning* adalah 66,54% dengan *macro-average F1-score* 0,5970. Kelas Retak_Ringan mencatat *F1-score* tertinggi sebesar 0,8221, sementara kelas PTR_Ringan dan PTR_Sedang mencatat *precision* sempurna 1,0000 dengan *recall* yang masih rendah, mencerminkan jumlah data uji yang sangat terbatas pada kelas-kelas tersebut.

4.5 Perbandingan Performa Kedua Skenario

Perbandingan performa keseluruhan kedua skenario secara ringkas disajikan pada Tabel 6.

Tabel 6. Perbandingan Performa Feature Extraction dan Fine-Tuning

Metrik	Feature Extraction	Fine-Tuning	Selisih
Akurasi	60,97%	66,54%	+5,57%
Macro Precision	0,6609	0,7303	+0,0694
Macro Recall	0,4909	0,5657	+0,0748
Macro F1-Score	0,5358	0,5970	+0,0612

Model *fine-tuning* secara konsisten mengungguli model *feature extraction* pada seluruh metrik evaluasi. Peningkatan akurasi sebesar 5,57% menunjukkan bahwa adaptasi bobot lapisan konvolusi *block5_conv4* terhadap karakteristik visual *dataset* lokal memberikan kontribusi signifikan terhadap kemampuan generalisasi model. Peningkatan *macro precision* dari 0,6609 menjadi 0,7303, *macro recall* dari 0,4909 menjadi 0,5657, dan *macro F1-score* dari 0,5358 menjadi 0,5970 mengonfirmasi bahwa model menjadi lebih adaptif terhadap pola kerusakan jalan setelah sebagian bobot lapisan akhir dilatih kembali.

4.6 Analisis Kesalahan Klasifikasi

Pola kesalahan yang paling dominan terjadi antartingkat keparahan dalam satu jenis kerusakan yang sama, khususnya pada kelompok Lubang. Pada skenario *fine-tuning*, sebanyak 5 citra Retak_Ringan salah diklasifikasikan sebagai Lubang_Ringan, yang merupakan kesalahan lintas jenis paling signifikan. Kesalahan lintas jenis lebih kritis secara operasional dibandingkan kesalahan antartingkat keparahan, karena perbedaan jenis kerusakan mengimplikasikan

metode perbaikan yang berbeda. Hal ini disebabkan oleh ambiguitas fitur visual antara kedua kelas pada kondisi pencahayaan atau sudut pengambilan tertentu.

Performa rendah pada kelompok Permukaan Tidak Rata (PTR) merupakan temuan yang paling menonjol pada kedua skenario. Pada skenario *fine-tuning*, ketiga kelas PTR hanya memperoleh *F1-score* 0,3333 (PTR_Ringan), 0,5333 (PTR_Sedang), dan 0,4348 (PTR_Berat), jauh di bawah kelompok Retak yang berkisar antara 0,6585 hingga 0,8221. Pola metriknya juga khas: *precision* PTR_Ringan dan PTR_Sedang mencapai 1,0000, tetapi *recall*-nya hanya 0,2000 dan 0,3636. Artinya, model hampir tidak pernah keliru ketika memutuskan sebuah citra sebagai PTR, namun sangat jarang berani mengambil keputusan tersebut; sebagian besar citra PTR justru dialihkan ke kelas mayoritas, terutama kelompok Retak dan Lubang. Rendahnya performa PTR karena itu bukan semata-mata cerminan ketidakmampuan model mengenali pola PTR, melainkan kecenderungan model bersikap terlalu konservatif terhadap kelas tersebut. Empat faktor berikut menjadi penyebabnya.

Faktor pertama adalah ketimpangan data yang ekstrem pada kelompok PTR. Ketiga kelas PTR hanya menyumbang 178 dari 1.731 citra atau sekitar 10,3% dari keseluruhan *dataset*, dengan PTR_Ringan sebagai kelas terkecil yang hanya memiliki 24 citra dan 16 di antaranya berada pada subset latih. Augmentasi dinamis memang menaikkan jumlah citra latih PTR_Ringan hingga 799 citra, namun seluruhnya diturunkan dari 16 citra sumber sehingga setiap citra direplikasi sekitar 50 kali. Augmentasi geometris dan fotometris hanya menghasilkan variasi rotasi, pergeseran, pencerminan, dan kecerahan dari pola yang sama, dan tidak menambah keragaman pola kerusakan yang sesungguhnya [23]. Akibatnya, keseimbangan jumlah citra latih antarkelas tercapai secara kuantitatif, tetapi keragaman informasi visual pada kelas PTR tetap jauh lebih rendah dibandingkan kelas lain, sehingga model cenderung menghafal karakteristik spesifik citra sumber alih-alih mempelajari pola PTR yang dapat digeneralisasi.

Faktor kedua bersumber dari karakteristik visual kerusakan PTR itu sendiri. Lubang dan retak memiliki penanda visual lokal berfrekuensi tinggi berupa tepi yang tegas, diskontinuitas tekstur, dan kontras gelap-terang yang kuat, sehingga sangat sesuai dengan cara kerja filter konvolusi 3×3 pada VGG19 yang memang dirancang menangkap pola tepi dan tekstur lokal [17]. Sebaliknya, permukaan tidak rata merupakan deformasi geometris berskala global berupa gelombang, penurunan, atau alur memanjang yang penanda utamanya adalah perubahan elevasi permukaan secara berangsur. Penanda semacam ini bersifat berfrekuensi rendah, bergantung pada sudut pandang dan arah pencahayaan, serta sebagian besar informasinya hilang ketika citra diubah ukurannya menjadi 224×224 piksel. Pada citra tunggal dua dimensi tanpa informasi kedalaman maupun objek acuan skala, permukaan yang bergelombang ringan

secara visual hampir tidak dapat dibedakan dari permukaan aspal normal yang bertekstur kasar, sehingga model kehilangan sinyal pembeda yang konsisten.

Faktor ketiga berkaitan dengan konsistensi pelabelan. Tingkat keparahan pada lubang dan retak masih dapat diperkirakan dari citra melalui luas bukaan atau lebar dan kerapatan celah, sedangkan tingkat keparahan PTR ditentukan oleh besarnya perbedaan elevasi yang tidak terukur langsung dari foto. Proses pelabelan kelas PTR karena itu lebih bergantung pada penilaian subjektif anotator dan berpotensi menghasilkan label yang kurang konsisten, terutama pada batas antara PTR_Ringan dan PTR_Sedang. Label yang tidak konsisten menyulitkan model membentuk batas keputusan yang tajam antarkelas PTR.

Faktor keempat adalah keterbatasan pengujian itu sendiri. Kelas PTR_Ringan hanya diwakili 5 citra uji dan PTR_Sedang 11 citra uji, sehingga resolusi metrik pada kelas-kelas tersebut sangat kasar: satu citra saja mengubah nilai *recall* PTR_Ringan sebesar 0,2000 dan *recall* PTR_Sedang sekitar 0,0909. Nilai *precision* 1,0000 pada kedua kelas tersebut juga perlu dibaca secara hati-hati karena masing-masing hanya dihitung dari 1 dan 4 prediksi positif. Dengan ukuran sampel sekecil ini, metrik per kelas pada kelompok PTR memiliki selang kepercayaan yang sangat lebar dan belum dapat diperlakukan sebagai estimasi performa yang stabil [20]. Dengan demikian, angka rendah pada kelompok PTR mencerminkan gabungan antara keterbatasan data, karakteristik kerusakan yang sulit direpresentasikan pada citra dua dimensi, ketidakkonsistenan pelabelan, dan ketidakpastian statistik akibat kecilnya data uji, bukan semata-mata kelemahan arsitektur yang digunakan.

4.7 Efisiensi Waktu Inferensi dan Implikasi Penerapan

Pengujian pada antarmuka Streamlit menunjukkan sistem menghasilkan prediksi dalam waktu rata-rata 3.479,40 milidetik ($\pm 3,48$ detik) per citra, mencakup proses *preprocessing*, *forward-pass* model, dan *rendering* hasil. Berdasarkan konfirmasi langsung dengan petugas DPUPR Kota Samarinda, proses validasi manual satu laporan foto kerusakan jalan memerlukan waktu sekitar 3 sampai 5 menit per laporan. Dengan demikian, model ini memberikan percepatan yang signifikan dibandingkan proses manual pada skala pelaporan massal.

Meski demikian, dengan akurasi 66,54%, penerapan model saat ini lebih tepat difungsikan sebagai alat bantu *decision support* bagi petugas daripada sebagai pengganti validasi manual sepenuhnya. Model dapat digunakan untuk membantu proses penyaringan dan prioritas awal laporan: laporan dengan nilai *confidence score* tinggi dapat menjadi kandidat untuk diperiksa lebih cepat, sedangkan laporan dengan keyakinan rendah atau kerusakan berat tetap memerlukan pemeriksaan manual. Pola

pemanfaatan ini selaras dengan prinsip *human-in-the-loop*, yaitu sistem kecerdasan buatan tetap melibatkan manusia sebagai pengambil keputusan akhir dalam layanan publik [7].

4.8 Keterkaitan dengan Penelitian Terdahulu

Perbandingan dengan penelitian terdahulu perlu dilakukan secara proporsional. Tang et al. [24] menggunakan VGG19 pada 2 hingga 4 kelas dengan *dataset* publik berskala besar dan mencapai akurasi 86,2%, sedangkan Wang et al. [14] mencapai 88,68% pada *dataset* RDD2020. Perbedaan angka tersebut bukan disebabkan oleh inferioritas metode, melainkan oleh perbedaan kompleksitas tugas. Penelitian ini

mengklasifikasikan 9 kelas secara simultan pada *dataset* lokal berskala terbatas dengan distribusi kelas yang sangat tidak seimbang. Dalam konteks tersebut, akurasi 66,54% merupakan hasil *proof-of-concept* yang valid dan memberikan fondasi yang dapat dikembangkan lebih lanjut. Augmentasi dinamis yang diterapkan terbukti efektif mengatasi ketidakseimbangan kelas yang ekstrem [22], dan pemanfaatan bobot *pre-trained* ImageNet melalui *transfer learning* memberikan keuntungan signifikan pada *dataset* yang terbatas [17], [18].

Ringkasan perbandingan kuantitatif antara hasil penelitian ini dan penelitian terdahulu disajikan pada Tabel 7.

Tabel 7. Perbandingan Hasil Penelitian dengan Penelitian Terdahulu

Penelitian	Metode	Cakupan Tugas	Performa Terbaik
Utaminigrum et al. [17]	GLCM + GA + ELM	Deteksi lubang	Akurasi 88,65%
Rachman et al. [18]	Faster R-CNN (MobileNetV2)	Deteksi kerusakan	mAP 79,7%
Vishwakarma & Vennelakanti [19]	Faster R-CNN (ResNet-50)	Deteksi kerusakan (data global)	Mean F1-score 0,542
Tang et al. [24]	VGG-19 + transfer learning	Jenis kerusakan (2-4 kelas)	Akurasi 86,2%
Wang et al. [14]	Improved VGG-19	Jenis kerusakan	Akurasi 88,68%
Penelitian ini	VGG19 + transfer learning	Jenis + keparahan, 9 kelas (data lokal 1.731 citra)	Akurasi 66,54%; macro F1-score 0,5970

Tabel 7 memperlihatkan bahwa seluruh penelitian terdahulu yang melaporkan akurasi di atas 85% bekerja pada ruang keputusan yang jauh lebih sempit, yaitu mendeteksi keberadaan satu jenis kerusakan atau mengklasifikasikan jenis kerusakan saja tanpa menyertakan tingkat keparahan. Penelitian ini membedakan sembilan kelas yang mencakup jenis sekaligus tiga tingkat keparahan di dalam setiap jenis. Sebagai gambaran, peluang menebak benar secara acak pada tugas empat kelas adalah 25%, sedangkan pada sembilan kelas hanya 11,1%. Sebagaimana ditunjukkan pada analisis kesalahan sebelumnya, sebagian besar kesalahan model justru terjadi antartingkat keparahan dalam jenis kerusakan yang sama, yaitu perbedaan yang tidak pernah diuji pada penelitian-penelitian pembandingan tersebut.

Perbandingan yang lebih setara dapat dilakukan terhadap penelitian yang melaporkan metrik dengan bobot setara antarkelas pada kondisi kelas tidak seimbang. Vishwakarma & Vennelakanti [19] memperoleh *mean F1-score* 0,542 pada *dataset* Road Damage berskala global, sedangkan penelitian ini memperoleh *macro-average F1-score* 0,5970 pada sembilan kelas dengan *dataset* lokal yang jauh lebih kecil. Pada metrik yang sama-sama tidak memihak kelas mayoritas, hasil penelitian ini berada pada rentang yang sebanding meskipun tugas klasifikasinya lebih kompleks dan jumlah datanya lebih terbatas.

Terhadap penelitian yang menggunakan arsitektur dasar yang sama, yaitu Tang et al. [24] dan Wang et al. [14], penelitian ini berbeda pada dua hal. Pertama, kedua penelitian tersebut menggunakan *dataset* publik berskala besar dengan distribusi kelas yang relatif seimbang, sementara penelitian ini

menggunakan citra lapangan dengan rasio kelas terbesar terhadap terkecil mencapai 21,8:1. Kedua, keluaran penelitian ini dirancang untuk langsung dipakai pada alur validasi laporan masyarakat, sehingga label yang dihasilkan harus memuat informasi tingkat keparahan yang menjadi dasar penentuan prioritas penanganan. Dengan mempertimbangkan kedua perbedaan tersebut, selisih akurasi sekitar 20% terhadap kedua penelitian itu lebih tepat dibaca sebagai konsekuensi perbedaan kompleksitas tugas dan kualitas *dataset* daripada sebagai indikasi kelemahan arsitektur, sekaligus menegaskan bahwa peningkatan performa berikutnya lebih ditentukan oleh perluasan dan penyeimbangan *dataset* daripada oleh penggantian arsitektur semata.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil membangun model klasifikasi jenis dan tingkat kerusakan jalan berbasis arsitektur VGG19 dengan pendekatan *transfer learning* pada 9 kelas yang dikembangkan dari *dataset* lokal Kota Samarinda sebanyak 1.731 citra. Skenario *fine-tuning* pada lapisan *block5_conv4* menghasilkan performa terbaik dengan akurasi 66,54% dan *macro-average F1-score* 0,5970, melampaui skenario *feature extraction* yang mencapai 60,97%, di mana kelas Retak_Ringan memperoleh *F1-score* tertinggi sebesar 0,8221. Waktu inferensi rata-rata 3.479 milidetik per citra jauh lebih cepat dibandingkan validasi manual petugas selama 3 sampai 5 menit, sehingga model berpotensi difungsikan sebagai alat bantu *decision support* pra-seleksi laporan pada sistem *Jalan Peduli*.

Pengujian pada penelitian ini memiliki sejumlah keterbatasan yang menjadi dasar penyusunan saran.

Pertama, evaluasi hanya dilakukan satu kali pada 269 citra uji dengan distribusi kelas yang sangat timpang, di mana kelas PTR_Ringan hanya diwakili 5 citra dan PTR_Sedang 11 citra, sehingga metrik per kelas pada kelompok minoritas belum stabil secara statistik. Kedua, penyeimbangan kelas hanya dilakukan melalui augmentasi geometris dan fotometris yang tidak menambah keragaman pola kerusakan yang sesungguhnya. Ketiga, kelompok PTR memperoleh performa terendah karena penandanya berupa perubahan elevasi permukaan yang sulit direpresentasikan pada citra tunggal dua dimensi berukuran 224×224 piksel. Keempat, pengujian hanya membandingkan dua skenario *transfer learning* pada satu arsitektur tanpa pembandingan arsitektur lain pada protokol yang sama. Kelima, model belum diintegrasikan dan diuji pada alur kerja nyata sistem *Jalan Peduli*.

Berdasarkan keterbatasan tersebut, penelitian selanjutnya disarankan untuk: (1) menambah jumlah citra pada kelas minoritas, khususnya ketiga kelas PTR, hingga mendekati jumlah kelas mayoritas, disertai protokol pengambilan citra yang baku mencakup sudut pengambilan, jarak, dan objek acuan skala agar tingkat keparahan dapat dilabeli secara lebih konsisten; (2) menerapkan *k-fold stratified cross-validation* atau memperbesar porsi data uji sehingga metrik per kelas pada kelompok minoritas dapat dilaporkan beserta selang kepercayaannya dan tidak lagi bergantung pada sedikit citra uji; (3) melengkapi augmentasi dengan strategi yang bekerja pada tingkat fungsi objektif, seperti *class weighting* atau *focal loss*, maupun augmentasi generatif, agar penyeimbangan kelas tidak hanya bersifat kuantitatif; (4) mengeksplorasi masukan yang mampu merepresentasikan deformasi permukaan berskala global, misalnya penambahan informasi kedalaman, citra multi-sudut, atau resolusi masukan yang lebih besar, mengingat kelemahan utama terjadi pada kelas PTR yang penandanya bersifat geometris; (5) membandingkan VGG19 dengan arsitektur lain seperti ResNet, EfficientNet, atau model berbasis *transformer* pada *dataset* dan protokol pengujian yang sama untuk memastikan pilihan arsitektur telah optimal; serta (6) mengintegrasikan model ke dalam sistem *Jalan Peduli* dengan ambang *confidence score* dan mekanisme *human-in-the-loop*, kemudian mengukur dampaknya terhadap waktu dan akurasi validasi petugas pada kondisi operasional yang sebenarnya.

DAFTAR PUSTAKA

- [1] S. Biswas and R. Jain, "The Road to Safety - Examining the Nexus between Road Infrastructure and Crime in Rural India," *Journal of Quantitative Economics*, pp. 1–24, 2025, doi: 10.1007/S40953-025-00466-6.
- [2] W. Shaw, G. Wang, G. K. M. Ampofo, C. L. Befeke, and D. Botei, "An analysis of the symmetric and asymmetric relationship between road traffic casualties and economic development of Ghana," *Journal of Transport Geography*, vol. 124, p. 104174, 2025, doi: 10.1016/J.JTRANGE.2025.104174.
- [3] A. Kurniawan, A. Patriadi, and S. Sajiyo, "Analysis of the Correspondence Between the Type of Road Damage and the Budget Costs Incurred in Handling on Provincial Roads in the Madura Region," *International Journal of Mechanical, Electrical and Civil Engineering*, vol. 2, no. 1, pp. 149–156, 2025, doi: 10.61132/IJMECIE.V211.140.
- [4] R. S. Falasyfa and D. Avianto, "Perancangan Aplikasi Layanan Pengaduan Kerusakan Jalan Berbasis Android," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 5, no. 1, pp. 944–953, 2024, doi: 10.35870/jimik.v5i1.569.
- [5] L. Probo Sumirat, D. Cahyono, and A. Ali Akbar, "Geographic Information System for Reporting Damage and Road Improvement in the National Road Implementation Center VIII Based on Web and Android," *Jurnal Sistem Informasi dan Bisnis Cerdas (SIBC)*, vol. 14, no. 1, 2021.
- [6] I. A. A. Angreni, "Klasifikasi Kerusakan Jalan: Retak, Retak Kulit Buaya dan Lubang Berdasarkan Citra Digital," *Jurnal Ilmiah Desain & Konstruksi*, vol. 23, no. 2, pp. 253–260, 2024, doi: 10.35760/DK.2024.V23I2.11389.
- [7] L. G. Denaro and R. Lim, "Analisis Deteksi Kerusakan pada Jalan Aspal menggunakan Deep Learning untuk Mendukung Efisiensi Biaya dan Waktu dalam Pemantauan Berkelanjutan," *Jurnal Dimensi Insinyur Profesional*, vol. 3, no. 1, pp. 16–25, 2025, doi: 10.9744/JDIP.3.1.16-25.
- [8] F. I. Rahman, L. Lukman, and H. Hildayati, "Sistem Klasifikasi Kerusakan Jalan Metode Machine Learning dengan Algoritma K-Means dan Random Forest," *Arus Jurnal Sains dan Teknologi*, vol. 3, no. 1, pp. 116–126, 2025, doi: 10.57250/AJST.V3I1.1212.
- [9] C. Zhang, E. Nateghinia, L. F. Miranda-Moreno, and L. Sun, "Pavement distress detection using convolutional neural network (CNN): A case study in Montreal, Canada," *International Journal of Transportation Science and Technology*, vol. 11, no. 2, pp. 298–309, 2022, doi: 10.1016/J.IJTST.2021.04.008.
- [10] H. Assemblali, S. Bouhsissin, and N. Sael, "Computer Vision-Based Detection and Classification of Road Obstacles: Systematic Literature Review," *IEEE Access*, vol. 13, pp. 128603–128638, 2025, doi: 10.1109/ACCESS.2025.3588668.
- [11] U. Baroudi, A. Bahamid, Y. Elalfy, and Z. Al Alami, "Enhancing Pothole Detection and Characterization: Integrated Segmentation and Depth Estimation in Road Anomaly Systems," 2025.
- [12] B. Hugo, A. Y. Zakiyyah, and K. A. Minor, "EfficientDet Model for Accurate Detection and Classification of Road Damages in Indonesia,"

- Journal of Machine Learning and Deep Learning, vol. 2, no. 1, pp. 11–17, 2025, doi: 10.64820/AEPJMLDL.21.11.17.62025.
- [13] B. Sasmito, B. H. Setiadji, and R. Isnanto, "Deteksi Kerusakan Jalan Menggunakan Pengolahan Citra Deep Learning di Kota Semarang," *TEKNIK*, vol. 44, no. 1, pp. 7–14, 2023, doi: 10.14710/teknik.v44i1.51908.
- [14] J. Wang, K. Wang, and K. Li, "Research on road damage recognition and classification based on improved VGG-19," *Mathematical Models in Engineering*, vol. 9, no. 4, pp. 115–129, 2023, doi: 10.21595/mme.2023.23455.
- [15] J. Liang, Q. Zhang, and X. Gu, "Small-sample data-driven lightweight convolutional neural network for asphalt pavement defect identification," *Case Studies in Construction Materials*, vol. 21, p. e03643, 2024, doi: 10.1016/J.CSCM.2024.E03643.
- [16] D. I. Mulyana, I. Wahyudi et al., "Deteksi Kerusakan Jalan Berdasarkan Citra Digital Menggunakan Convolutional Neural Network (CNN)," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi (JIMIK)*, vol. 6, no. 1, 2025.
- [17] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556, 2015. [Online]. Available: <http://arxiv.org/abs/1409.1556>.
- [18] F. Yu, X. Xiu, and Y. Li, "A Survey on Deep Transfer Learning and Beyond," *Mathematics*, vol. 10, no. 19, 2022, doi: 10.3390/math10193619.
- [19] P. Yan et al., "A Comprehensive Survey of Deep Transfer Learning for Anomaly Detection in Industrial Time Series: Methods, Applications, and Directions," *IEEE Access*, vol. 12, pp. 3768–3789, 2024, doi: 10.1109/ACCESS.2023.3349132.
- [20] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-56706-x.
- [21] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, 2024, doi: 10.53555/ajbr.v27i4s.4345.
- [22] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artificial Intelligence Review*, vol. 55, no. 3, pp. 2351–2377, 2022, doi: 10.1007/s10462-021-10066-4.
- [23] W. Sapitri, Y. N. Kunang, I. Z. Yadi, and Mahmud, "The Impact of Data Augmentation Techniques on the Recognition of Script Images in Deep Learning Models," *Jurnal Online Informatika*, vol. 8, no. 2, pp. 169–176, 2023, doi: 10.15575/join.v8i2.1073.
- [24] Y. Tang, K. Li, and K. Wang, "Research on intelligent detection of pavement damage based on CNN," *Mathematical Models in Engineering*, vol. 8, no. 4, pp. 98–107, 2022, doi: 10.21595/mme.2022.22918.
- [25] B. Kulambayev et al., "Real-Time Road Damage Detection System on Deep Learning Based Image Analysis," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 9, 2024.