Jurnal KomtekInfo
https://jkomtekinfo.org/ojs
2024 Vol. 11 No. 4 Hal: 464-469 p-ISSN: 2356-0010, e-ISSN: 2502-8758

Evaluasi Pergeseran Domain Zero-Shot dan Latensi Klasifikasi Pipeline Hibrida Fitur Terseleksi Penyakit Daun Padi

Agung Ramadhanu[1]✉, Rely Maryani[2]

[1]Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK, Jl. Raya Lubuk Begalung, 25221, Padang, Indonesia

[2]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK, Jl. Raya Lubuk Begalung, 25221, Padang, Indonesia

✉ agung_ramadhanu@upiyptk.ac.id

Abstract

Most image-based rice leaf disease classification models are evaluated using datasets captured under controlled laboratory environments, leaving their operational behaviour on real-world agricultural field images largely unverified. This study reports an empirical zero-shot domain transfer evaluation and classification latency profiling of a hybrid feature-fusion pipeline (configuration E5). The architecture combines deep representations extracted from EfficientNet-B0 on segmented leaf images (1,280 dimensions) with 73 mask-based handcrafted features, standardized via z-score and filtered to the top 128 features using the ANOVA F-score before classification by a Radial Basis Function Support Vector Machine (SVM-RBF, $C = 5.0$). Trained once on 4,410 deduplicated public images spanning five classes, the frozen model was directly applied without retraining or fine-tuning to 80 locally collected field images covering two mapped disease classes (Bacterial Blight and Brown Spot, 40 samples each). Results reveal a severe performance degradation: accuracy fell from 99.64% ± 0.18% in cross-validated in-domain evaluation to 18.75% (15 of 80 images; 95% Wilson score confidence interval: 11.7–28.7%), while Macro-F1 dropped from 99.47% ± 0.28% to 13.64%. Perturbation experiments adding Gaussian noise across four intensity levels ($\sigma = 0$, 0.05, 0.10, and 0.20) shifted the out-of-distribution accuracy marginally from about 18.75% to about 17.5%; because the baseline accuracy was already near the five-class random-guess floor, this flat trajectory reflects a floor effect rather than model robustness. Isolated profiling of the classification step alone—a single SVM prediction on a precomputed 128-feature vector on CPU—took 0.78 ± 0.54 ms (P95: 1.69 ms; 1,278.5 FPS) over 100 repetitions, and 0.70 ± 0.21 ms (P95: 1.24 ms; 1,421.0 FPS) in a repeat verification. Because this figure excludes leaf and lesion segmentation, handcrafted feature extraction, and the EfficientNet-B0 forward pass, it does not represent end-to-end latency. Underlying causes of the domain transfer failure are systematically examined as research hypotheses.

Keywords: Domain Shift, Out-of-Distribution, Zero-Shot Transfer, Rice Leaf Disease, Inference Latency

Abstrak

Mayoritas model klasifikasi penyakit daun padi berbasis pengolahan citra dievaluasi menggunakan himpunan data yang diakuisisi dalam kondisi laboratorium terkendali, sedangkan perilakunya ketika berhadapan langsung dengan citra lapangan jarang diukur secara objektif. Penelitian ini melaporkan pengujian transfer *zero-shot* lintas domain serta pemrofilan latensi tahap klasifikasi pada *pipeline* fusi fitur hibrida (konfigurasi E5). Arsitektur yang diuji menggabungkan fitur representasi dalam dari jaringan EfficientNet-B0 pada citra daun tersegmentasi (1.280 dimensi) dengan 73 fitur rekayasa manual (*handcrafted features*) berbasis masker segmentasi, yang selanjutnya distandardisasi dan diseleksi menjadi 128 fitur terbaik menggunakan skor F-ANOVA sebelum diklasifikasikan dengan *Support Vector Machine* berkernel *Radial Basis Function* (SVM-RBF, $C = 5{,}0$). Model dilatih satu kali pada 4.410 citra publik terdeduplikasi dari lima kelas, kemudian diterapkan secara langsung tanpa pelatihan ulang (*fine-tuning*) pada 80 citra lapangan yang dikumpulkan secara lokal dari dua kelas penyakit (*Bacterial Blight* dan *Brown Spot*, masing-masing 40 citra). Hasil pengujian menunjukkan penurunan kinerja yang tajam: akurasi merosot dari 99,64% ± 0,18% pada evaluasi *in-domain* tervalidasi silang menjadi 18,75% (15 dari 80 citra benar; interval kepercayaan skor Wilson 95%: 11,7–28,7%), dan Macro-F1 turun dari 99,47% ± 0,28% menjadi 13,64%. Penambahan derau Gaussian pada empat tingkat intensitas ($\sigma = 0$; 0,05; 0,10; dan 0,20) hanya mengubah akurasi *out-of-distribution* dari sekitar 18,75% menjadi sekitar 17,5%; mengingat akurasi awal sudah berada di dekat batas bawah tebakan acak, kurva mendatar ini menunjukkan fenomena *floor effect* dan bukan ketahanan model terhadap derau. Tahap klasifikasi saja, yaitu satu pemanggilan prediksi SVM pada vektor 128 fitur yang telah dihitung sebelumnya pada CPU, membutuhkan waktu 0,78 ± 0,54 ms (P95: 1,69 ms; 1.278,5 FPS) dalam 100 pengulangan, serta 0,70 ± 0,21 ms (P95: 1,24 ms; 1.421,0 FPS) pada pengukuran ulang. Angka ini tidak mencakup proses segmentasi daun dan lesi, ekstraksi fitur rekayasa manual, maupun *forward pass* EfficientNet-B0, sehingga bukan merupakan latensi menyeluruh (*end-to-end*). Berbagai faktor penyebab kegagalan transfer domain dianalisis secara mendalam sebagai hipotesis evaluasi.

Kata kunci: Pergeseran Domain, Out-of-Distribution, Transfer Zero-Shot, Penyakit Daun Padi, Latensi Inferensi

KomtekInfo is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

1. Pendahuluan

Klasifikasi penyakit tanaman berbasis pengolahan citra digital dan pembelajaran mesin umumnya dilaporkan menggunakan himpunan data uji yang diakuisisi dalam lingkungan laboratorium yang terkendali. Dalam skenario pengujian standar tersebut, citra daun biasanya dipotret satu per satu di atas latar belakang polos dengan pencahayaan merata, sehingga model dengan mudah mencapai tingkat akurasi di atas 98% hingga 99%. Namun, ketika model yang sama diuji langsung pada citra yang diambil dalam kondisi lingkungan yang berbeda dari data pelatihannya, kinerjanya kerap mengalami kemerosotan tajam. Temuan klasik oleh Mohanty dkk. [1] memperlihatkan bahwa meskipun arsitektur jaringan saraf konvolusional dalam (*deep convolutional neural network*) mampu meraih akurasi lebih dari 99% pada evaluasi internal dataset PlantVillage, akurasinya jatuh hingga

Diterima: 22-10-2024 | Revisi: 24-11-2024 | Diterbitkan: 31-12-2024 | doi: 10.35134/komtekinfo.v11i4.723

tersisa sedikit di atas 31% saat diuji pada citra tanaman dari sumber daring yang memiliki kondisi pencahayaan dan latar belakang berbeda. Tinjauan komprehensif oleh Barbedo [2] mengidentifikasi keanekaragaman latar belakang alami, fluktuasi pencahayaan matahari, sudut pengambilan gambar, serta variasi fisik gejala lesi sebagai tantangan teknis utama dalam identifikasi penyakit tanaman berbasis citra spektrum cahaya tampak (*visible range images*). Sejalan dengan hal tersebut, penelitian Ahmad dkk. [3] mengkaji secara khusus persoalan generalisasi model pembelajaran mendalam untuk identifikasi penyakit tanaman saat berpindah dari kondisi terkendali menuju kondisi lapangan (*field conditions*).

Penurunan performa secara drastis ketika model meninggalkan lingkungan pelatihannya bukanlah fenomena yang eksklusif terjadi pada patologi tanaman. Tolok ukur (*benchmark*) pergeseran distribusi dunia nyata berskala besar seperti WILDS yang disusun oleh Koh dkk. [4] mendemonstrasikan bahwa kinerja model pembelajaran mesin modern dapat merosot secara signifikan ketika distribusi probabilitas data uji berbeda dari distribusi data latih. Secara teoretis, penurunan kinerja ini berakar pada pergeseran distribusi antara domain sumber (*source domain*) tempat model mempelajari batas keputusan dan domain target (*target domain*) tempat model dioperasikan [5]. Terminologi pergeseran himpunan data (*dataset shift*) beserta taksonomi jenis-jenisnya—mulai dari pergeseran kovariat (*covariate shift*), pergeseran probabilitas prior (*prior probability shift*), hingga pergeseran konsep (*concept shift*)—telah dirangkum secara sistematis oleh Moreno-Torres dkk. [6]. Berbagai pendekatan metodologis untuk melatih model agar mampu melakukan generalisasi ke domain baru yang belum pernah terlihat sebelumnya (*domain generalization*) juga telah ditinjau secara luas oleh Zhou dkk. [7]. Prinsip fundamental yang perlu ditekankan dari literatur tersebut adalah bahwa ketika terjadi pergeseran domain yang nyata, jaminan performa statistik yang diperoleh dari prosedur validasi silang pada domain sumber tidak lagi berlaku pada domain target. Selain pergeseran domain alami, evaluasi ketahanan (*robustness*) model terhadap gangguan kualitas sinyal citra juga lazim dilakukan melalui penyuntikan korupsi buatan seperti derau (*noise*) [8]. Walaupun demikian, pengujian perturbasi derau semacam itu hanya menghasilkan informasi diagnostik yang bermakna apabila model memiliki kinerja awal (*baseline accuracy*) yang layak pada domain yang sedang diuji.

Aspek krusial lainnya yang sering kali kurang mendapatkan ketelitian metodologis dalam literatur klasifikasi penyakit tanaman adalah pelaporan biaya komputasi dan latensi inferensi. Dalam banyak publikasi yang mengusulkan sistem deteksi untuk perangkat pertanian cerdas, angka latensi eksekusi atau kecepatan *frames per second* (FPS) kerap dilaporkan tanpa menyebutkan secara spesifik bagian *pipeline* mana yang sebenarnya diukur. Akibatnya, waktu komputasi yang sangat singkat pada tahap pengklasifikasi akhir sering disalahartikan oleh pembaca sebagai kecepatan sistem secara menyeluruh (*end-to-end latency*). Dehghani dkk. [9] melalui kajian tentang *efficiency misnomer* menunjukkan bahwa mengandalkan satu indikator biaya komputasi secara terisolasi dapat memberikan gambaran efisiensi yang tidak lengkap dan berpotensi menyesatkan, sehingga beberapa indikator beserta batasan cakupannya perlu dilaporkan secara bersama-sama. Berangkat dari kritik metodologis tersebut, makalah ini memisahkan secara tegas komponen *pipeline* yang diukur dari komponen hulu yang tidak diukur.

Berdasarkan permasalahan di atas, penelitian ini dirancang untuk menjawab tiga pertanyaan penelitian utama:

(RQ1) Seberapa besar penurunan kinerja klasifikasi ketika *pipeline* fusi fitur hibrida yang dilatih pada citra laboratorium diterapkan secara *zero-shot* pada citra daun padi lapangan yang dikumpulkan secara lokal?

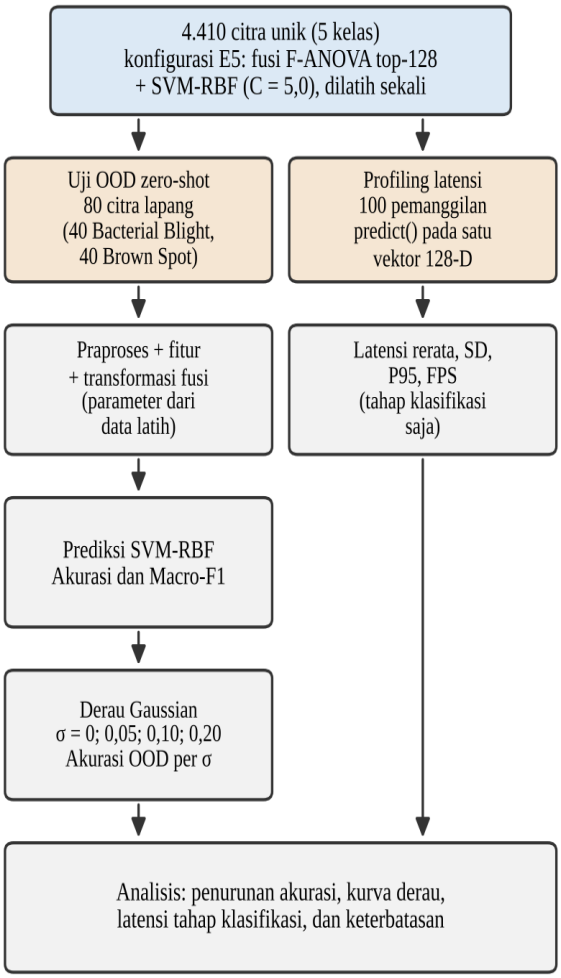
(RQ2) Bagaimana akurasi model pada citra lapangan tersebut berubah terhadap penambahan gangguan derau Gaussian pada berbagai tingkat intensitas, dan bagaimana menafsirkan kurva responsnya secara tepat?

(RQ3) Berapa latensi komputasi murni pada tahap klasifikasi menggunakan CPU, dan bagaimana batasan penafsirannya terhadap kelayakan implementasi pada perangkat tepi (*edge devices*)?

Kontribusi utama dari penelitian ini adalah pelaporan hasil negatif (*negative results*) secara transparan pada pengujian transfer domain *zero-shot* beserta analisis batas tafsir statistiknya, serta pengukuran profil latensi tahap klasifikasi dengan cakupan yang dinyatakan secara eksplisit.

2. Metodologi Penelitian

Alur tahapan penelitian yang dijalankan dalam studi ini diilustrasikan secara menyeluruh pada Gambar 1. Satu model klasifikasi dengan konfigurasi terbaik dilatih satu kali pada himpunan data laboratorium publik, kemudian parameter dan bobotnya dibekukan untuk digunakan pada dua cabang pengujian terpisah: (1) evaluasi transfer *zero-shot* pada citra lapangan beserta uji ketahanan terhadap derau Gaussian, dan (2) pemrofilan latensi komputasi pada tahap klasifikasi.



Gambar 1. Alur Penelitian

2.1 Model dan Domain Pelatihan

Arsitektur model yang dievaluasi menggunakan konfigurasi usulan E5 yang dikembangkan pada penelitian pendamping [12]. *Pipeline* hibrida ini menggabungkan dua cabang representasi visual yang saling melengkapi. Cabang pertama mengekstraksi fitur pembelajaran mendalam (*deep features*) menggunakan jaringan konvolusional EfficientNet-B0 [10] pada citra daun padi yang telah tersegmentasi, menghasilkan vektor fitur berdimensi 1.280. Cabang kedua mengekstraksi 73 fitur rekayasa manual (*handcrafted features*) berbasis masker segmentasi yang merepresentasikan deskriptor warna, tekstur, serta geometri morfologi daun dan lesi penyakit. Penggabungan (*concatenation*) kedua cabang tersebut menghasilkan vektor fitur gabungan berdimensi D = 1.353.

Sebelum memasuki tahap seleksi dan klasifikasi, setiap fitur ke-j pada vektor gabungan distandardisasi menggunakan transformasi z-score berdasarkan nilai rerata μ_j dan simpangan baku s_j yang dihitung dari data pelatihan:

$$z_{i,j} = \frac{x_{i,j} - \mu_j}{s_j} \qquad (1)$$

Selanjutnya, diterapkan seleksi fitur univariat berbasis analisis ragam satu arah (skor F-ANOVA) untuk menyaring fitur yang memiliki daya pisah antar-kelas paling tinggi. Untuk fitur ke-j yang dievaluasi pada K kelas dengan total N sampel pelatihan, rasio statistik F_j didefinisikan sebagai perbandingan antara varians antar-kelas (*between-class variance*) dan varians dalam-kelas (*within-class variance*):

$$F_j = \frac{\sum_{k=1}^{K} n_k \left(\bar{z}_{k,j} - \bar{z}_{\cdot,j}\right)^2 / (K-1)}{\sum_{k=1}^{K} \sum_{i=1}^{n_k} \left(z_{i,k,j} - \bar{z}_{k,j}\right)^2 / (N-K)} \qquad (2)$$

dengan n_k menyatakan jumlah sampel pada kelas ke-k, $\bar{z}_{k,j}$ adalah rerata fitur ke-j pada kelas ke-k, dan $\bar{z}_{\cdot,j}$ adalah rerata keseluruhan fitur ke-j. Dari 1.353 kandidat fitur, dipilih 128 fitur dengan nilai skor F_j tertinggi.

Vektor 128 fitur terpilih ($\mathbf{u} \in \mathbb{R}^{128}$) kemudian diklasifikasikan menggunakan algoritma *Support Vector Machine* dengan kernel *Radial Basis Function* (SVM-RBF) [11] pada parameter penalti regularisasi C = 5,0. Fungsi kernel RBF antara dua vektor fitur tereduksi **u** dan **v** dirumuskan sebagai:

$$K(\mathbf{u}, \mathbf{v}) = \exp(-\gamma \|\mathbf{u} - \mathbf{v}\|_2^2) \qquad (3)$$

Seluruh prosedur praproses, pembuatan masker segmentasi, dan ekstraksi fitur mengikuti protokol pada penelitian pendamping [12]. Untuk pembentukan model uji pada studi ini, konfigurasi E5 dilatih satu kali pada seluruh 4.410 citra unik (telah melalui proses deduplikasi) dari *Rice Leaf Disease Detection Dataset* di Kaggle [13] yang mencakup lima kelas, tanpa pemisahan data uji pada langkah pelatihan akhir ini. Sebagai tolok ukur pembanding *in-domain*, digunakan hasil evaluasi validasi silang berstrata (*stratified cross-validation*) 5-lipatan × 3 *seed* acak (total 15 lipatan uji) dari penelitian pendamping [12] untuk konfigurasi E5 yang sama, yaitu akurasi sebesar 99,64% ± 0,18% dan Macro-F1 sebesar 99,47% ± 0,28%.

2.2 Domain Uji Lapang

Domain uji lapangan (*out-of-distribution*/OOD) dibangun dari koleksi awal sebanyak 120 citra daun padi yang dikumpulkan secara lokal di area persawahan. Dari jumlah tersebut, sebanyak 80 citra yang terdiri atas kelas *Bacterial Blight* (40 citra) dan *Brown Spot* (40 citra) digunakan dalam pengujian karena label penyakitnya terpetakan secara langsung ke dalam ruang kelas pelatihan model, sedangkan 40 citra lainnya tidak dievaluasi. Selama pengujian berlangsung, bobot dan parameter model dibekukan sepenuhnya (*frozen*) dan tidak dilakukan proses pelatihan ulang maupun *fine-tuning* terhadap citra lapangan.

Setiap citra uji lapangan diproses melewati tahapan praproses, segmentasi masker, ekstraksi fitur, serta transformasi fusi yang identik dengan alur pelatihan. Guna menghindari kebocoran informasi (*data leakage*) dari domain target, parameter standardisasi (μ_j dan s_j pada Persamaan 1) serta indeks ke-128 fitur terpilih hasil skor F-ANOVA diambil sepenuhnya dari parameter data latih. Kinerja prediksi dievaluasi menggunakan metrik Akurasi dan Macro-F1 dari pustaka *scikit-learn*. Implementasi Macro-F1 tersebut menghitung rata-rata tak berbobot dari skor F1 atas seluruh label kelas unik L yang muncul baik pada label sebenarnya (*ground truth*) maupun pada keluaran prediksi model:

$$\text{Macro-F1} = \frac{1}{|L|} \sum_{c \in L} \frac{2\,\text{TP}_c}{2\,\text{TP}_c + \text{FP}_c + \text{FN}_c} \qquad (4)$$

Apabila model memprediksi kelas pelatihan di luar dua kelas uji lapangan, maka kelas yang hanya muncul pada prediksi tersebut tetap masuk ke dalam himpunan L dengan nilai $TP_c = 0$ sehingga menghasilkan skor F1 bernilai nol. Mengingat ukuran sampel uji lapangan yang terbatas ($n = 80$), ketidakpastian estimasi proporsi akurasi ($\hat{p}$) dihitung menggunakan interval kepercayaan skor Wilson [14] pada tingkat kepercayaan 95% ($z = 1{,}96$):

$$\mathrm{CI}_{\mathrm{Wilson}} = \frac{\hat{p} + \frac{z^2}{2n} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}} \qquad (5)$$

2.3 Uji Derau Gaussian

Untuk menguji respons prediksi model pada domain lapangan terhadap gangguan sinyal, ditambahkan derau Gaussian aditif $\varepsilon \sim N(0, \sigma^2)$ dengan empat variasi nilai simpangan baku secara bertahap, yaitu $\sigma = 0$; 0,05; 0,10; dan 0,20. Pada setiap tingkatan nilai σ, akurasi klasifikasi pada ke-80 citra uji lapangan dihitung ulang untuk memperoleh kurva respons terhadap intensitas gangguan.

2.4 Pengukuran Latensi

Latensi eksekusi tahap klasifikasi diukur secara spesifik dengan memanggil fungsi *predict()* dari model SVM-RBF terlatih pada satu vektor berdimensi 128 fitur yang telah dihitung sebelumnya (*precomputed feature vector*) sebanyak $R = 100$ kali pengulangan pada CPU. Hasil pengukuran waktu eksekusi $\{t_1, t_2, \ldots, t_{100}\}$ dilaporkan sebagai nilai rerata ($\bar{t}$) ± simpangan baku (s_t), persentil ke-95 (P95), serta laju *throughput* teoritis dalam satuan *frames per second* (FPS = $1.000/\bar{t}$ dengan $\bar{t}$ dalam milidetik). Pengukuran ini secara eksplisit hanya mencakup tahap klasifikasi akhir; proses segmentasi daun dan lesi, ekstraksi 73 fitur rekayasa manual, *forward pass* jaringan EfficientNet-B0, serta transformasi matriks fusi tidak diikutsertakan dalam perhitungan waktu tersebut.

3. Hasil dan Pembahasan

3.1 Kinerja pada Citra Lapang

Tabel 1 menyajikan perbandingan langsung antara kinerja model pada evaluasi *in-domain* dan hasil pengujian transfer *zero-shot* pada citra lapangan. Akurasi klasifikasi mengalami penurunan tajam dari 99,64% menjadi 18,75%, yang berarti hanya 15 dari 80 citra lapangan yang berhasil diprediksi dengan benar, atau terjadi kemerosotan sebesar 80,89 poin persentase. Berdasarkan Persamaan (5), interval kepercayaan Wilson 95% untuk akurasi lapangan berada pada rentang 11,7% hingga 28,7%. Sejalan dengan jatuhnya akurasi, skor Macro-F1 turut merosot dari 99,47% menjadi 13,64%.

Tabel 1. Kinerja In-Domain dan Zero-Shot Lapang

Domain	*N*	Akurasi (%)	Macro-F1 (%)
In-domain (validasi silang)	4.410	99,64 ± 0,18	99,47 ± 0,28
Lapang (*zero-shot*)	**80**	**18,75**	**13,64**

Keterangan: baris *in-domain* adalah rerata ± SD pada 15 lipatan uji; baris lapang adalah satu evaluasi pada 80 citra.

Terdapat dua aspek statistik penting yang perlu diperhatikan dalam menginterpretasikan angka penurunan ini. Pertama, capaian akurasi sebesar 18,75% berada jauh di bawah ambang 50% yang secara teoretis akan dicapai melalui tebakan acak seragam di antara dua kelas uji, dan justru berada sangat dekat dengan angka 20% yang merupakan probabilitas tebakan acak di antara lima kelas yang dikenal oleh model saat pelatihan. Indikasi ini sejalan dengan kecenderungan model yang sering memprediksi kelas penyakit di luar dua kelas uji yang tersedia pada sampel lapangan, meskipun distribusi prediksi per kelas dan matriks konfusi (*confusion matrix*) tidak dilaporkan pada eksperimen ini sehingga proporsi pastinya belum dapat diverifikasi secara langsung. Kedua, nilai Macro-F1 (13,64%) yang terpaut cukup jauh di bawah nilai akurasi (18,75%) sepenuhnya konsisten dengan mekanisme perhitungan pada Persamaan (4) yang menyertakan label kelas ekstra yang hanya muncul pada keluaran prediksi dengan skor F1 bernilai nol.

Akar penyebab terjadinya kegagalan transfer domain ini belum didiagnosis melalui uji ablasi terpisah pada penelitian ini, namun dapat dirumuskan ke dalam lima hipotesis teknis yang saling berkaitan:

(a) Perbedaan Iluminasi dan Pantulan Permukaan. Kondisi pencahayaan alami di lahan sawah sangat bervariasi akibat terik matahari langsung, bayangan antar-tajuk tanaman, serta pantulan spekular pada permukaan daun padi yang lembap, yang secara langsung mengubah distribusi intensitas piksel dibandingkan kondisi pencahayaan merata di laboratorium.

(b) Kerumitan Latar Belakang Alami. Berbeda dengan citra pelatihan yang memiliki latar belakang polos terkendali, citra lapangan memuat elemen latar belakang yang kompleks seperti tanah, genangan air sawah, batang tanaman lain, dan gulma yang memiliki kemiripan spektrum warna hijau dengan daun padi.

(c) Gejala Kerusakan Daun Non-Patogenik. Daun padi yang tumbuh di lingkungan terbuka sering kali memiliki artefak kerusakan fisik non-patogenik—seperti bekas gigitan hama, luka mekanis akibat gesekan angin, debu, atau penuaan ujung daun—yang menyerupai pola tekstur bercak penyakit.

(d) Kegagalan Segmentasi Daun dan Lesi. Arsitektur hibrida E5 memiliki ketergantungan struktural yang tinggi pada tahap pembuatan masker segmentasi di bagian hulu. Baik cabang ekstraksi fitur EfficientNet-B0 (yang menerima citra daun tersegmentasi) maupun cabang 73 fitur rekayasa manual sama-sama bergantung sepenuhnya pada keakuratan masker tersebut. Algoritma segmentasi yang dirancang berdasarkan kontras latar laboratorium berisiko tinggi gagal memisahkan objek daun dari latar belakang lapangan; terlebih lagi, pemeriksaan visual pada sebagian citra laboratorium menunjukkan bahwa masker yang dihasilkan terkadang tidak sepenuhnya andal bahkan pada domain pelatihan itu sendiri.

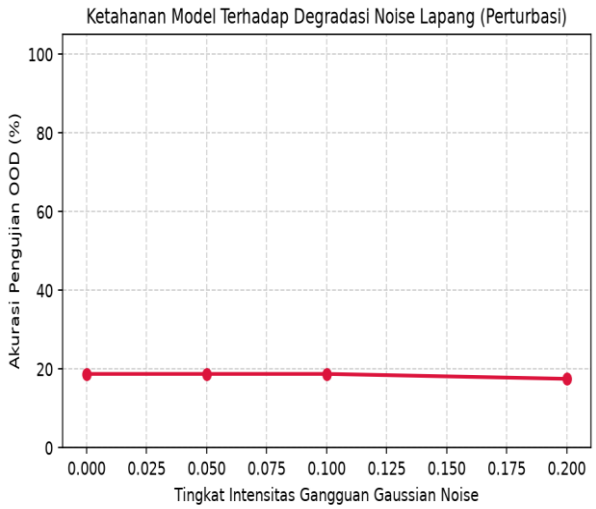
(e) Ketidaksesuaian Ruang Label. Ketidakselarasan antara ruang label pelatihan (lima kelas) dan ruang label pengujian lapangan (dua kelas) memperbesar peluang jatuhnya prediksi ke tiga kelas lain ketika distribusi fitur mengalami pergeseran kovariat.

Karena belum dilakukan analisis eksperimental atas citra lapangan yang memisahkan kontribusi kelima hipotesis di atas, langkah diagnostik yang diperlukan pada studi lanjutan meliputi inspeksi visual terhadap masker hasil segmentasi pada citra lapangan, penyusunan matriks

konfusi lengkap, pengujian tiap cabang fitur secara terpisah (skenario E1 hingga E3) pada data lapangan, serta perbandingan sebaran distribusi fitur antara domain sumber dan domain target.

3.2 Derau Gaussian

Gambar 2 memperlihatkan respons akurasi klasifikasi pada citra lapangan terhadap peningkatan simpangan baku derau Gaussian (σ). Berdasarkan pembacaan kurva pengujian, akurasi OOD bertahan mendatar di kisaran 18,75% pada tingkat gangguan σ = 0; 0,05; dan 0,10, kemudian bergeser tipis ke kisaran 17,5% pada tingkat gangguan tertinggi σ = 0,20.



Gambar 2. Akurasi pada citra lapang terhadap tingkat derau Gaussian

Kurva respons yang tampak hampir datar ini tidak dapat ditafsirkan sebagai bukti ketahanan (*robustness*) model terhadap gangguan derau. Sebaliknya, kemendataran kurva tersebut disebabkan oleh fenomena efek dasar (*floor effect*): karena akurasi awal model pada σ = 0 sudah merosot mendekati tingkat tebakan acak lima kelas (≈20%), gangguan derau tidak lagi memiliki ruang yang cukup untuk menurunkan kinerja model secara berarti. Pengujian ketahanan derau yang informatif mensyaratkan penyertaan kurva pengujian yang sama pada data *in-domain* (di mana akurasi awal berada pada 99,64%) sebagai pembanding, yang tidak dijalankan pada penelitian ini. Di samping itu, apabila perturbasi derau Gaussian tersebut disuntikkan pada ruang fitur yang telah distandardisasi (memiliki simpangan baku satuan bernilai 1,0), maka penambahan simpangan baku σ = 0,20 tergolong relatif kecil terhadap varians fitur sehingga dampak perubahan prediksi yang kecil memang secara matematis diharapkan.

3.3 Latensi Tahap Klasifikasi

Tabel 2 merangkum hasil pemrofilan latensi komputasi untuk tahap klasifikasi pada CPU selama 100 kali pemanggilan berulang. Eksekusi fungsi prediksi SVM-RBF pada satu vektor berdimensi 128 fitur membutuhkan waktu rerata 0,78 ms dengan simpangan baku 0,54 ms, batas persentil ke-95 (P95) sebesar 1,69 ms, serta *throughput* teoritis sebesar 1.278,5 FPS. Rasio simpangan baku yang cukup besar terhadap nilai reratanya (0,54 ms dibanding 0,78 ms) mencerminkan adanya fluktuasi waktu eksekusi antar-pemanggilan pada sistem operasi CPU. Saat dilakukan sesi pengukuran ulang untuk menguji keterulangan (*repeatability*), diperoleh waktu eksekusi sebesar 0,70 ± 0,21 ms, P95 sebesar 1,24 ms, dan *throughput* 1.421,0 FPS. Perbedaan hasil antar-sesi ini menunjukkan bahwa angka latensi tahap klasifikasi bervariasi antar-pengukuran dan sebaiknya dibaca pada tingkat orde besaran sekitar 1 ms per citra, bukan sebagai nilai presisi mutlak hingga dua angka desimal.

Tabel 2. Latensi Tahap Klasifikasi pada CPU (100 pemanggilan)

Tahap diukur	ms/citra	P95 (ms)	FPS
SVM-RBF *predict*, 128 fitur (Pengukuran 1)	0,78 ± 0,54	1,69	1.278,5
SVM-RBF *predict*, 128 fitur (Pengukuran ulang)	0,70 ± 0,21	1,24	1.421,0

Keterangan: tidak termasuk segmentasi, ekstraksi fitur, forward pass EfficientNet-B0, dan transformasi fusi.

Temuan pada Tabel 2 ini sama sekali tidak mendukung klaim bahwa sistem deteksi penyakit daun padi secara utuh dapat berjalan pada kecepatan lebih dari 1.000 FPS. Dalam alur inferensi menyeluruh, proses pembuatan masker segmentasi daun dan lesi, ekstraksi 73 fitur rekayasa manual, serta operasi *forward pass* jaringan konvolusional EfficientNet-B0 tidak diukur pada eksperimen ini dan hampir pasti membutuhkan waktu komputasi yang jauh lebih lama daripada sekadar mengevaluasi fungsi kernel SVM pada vektor 128 angka. Karena beban waktu dari tahapan-tahapan hulu tersebut belum diketahui, maka kelayakan penerapan *pipeline* hibrida ini pada perangkat tepi (*edge devices*) yang memiliki keterbatasan sumber daya—seperti ponsel pintar atau komputer papan tunggal pada drone pertanian—belum dapat dinilai dari hasil ini. Sebagai pembanding metodologis mengenai cakupan pengukuran yang tepat, arsitektur jaringan ringan yang dirancang khusus untuk perangkat bergerak seperti MobileNetV3 [15] dievaluasi kelayakannya dengan mengukur latensi jaringan secara utuh (*end-to-end*) langsung pada CPU ponsel.

3.4 Keterbatasan

Agar temuan dalam penelitian ini dapat ditafsirkan secara proporsional, tujuh keterbatasan metodologis berikut perlu dicatat secara eksplisit:

(1) Evaluasi *out-of-distribution* hanya menggunakan satu himpunan uji lapangan berukuran 80 citra yang mencakup dua kelas penyakit dari lima kelas pelatihan, serta dijalankan dalam satu kali evaluasi tanpa pengulangan penarikan sampel.

(2) Pengujian pada citra lapangan hanya menerapkan konfigurasi akhir (E5), sehingga kontribusi kegagalan dari masing-masing cabang fitur (cabang *deep features* EfficientNet-B0 maupun cabang 73 fitur rekayasa manual) belum dapat dipisahkan secara terukur.

(3) Ketiadaan pencatatan matriks konfusi (*confusion matrix*) menyebabkan pola distribusi kesalahan prediksi antar-kelas tidak diketahui secara rinci.

(4) Faktor penyebab kegagalan transfer domain belum didiagnosis secara eksperimental dan masih berstatus sebagai hipotesis yang memerlukan verifikasi lanjutan.

(5) Pengujian perturbasi derau Gaussian tidak dilengkapi dengan kurva pembanding pada data uji *in-domain*, serta titik penyuntikan derau di sepanjang *pipeline* belum terdokumentasi secara rinci.

(6) Pengukuran latensi hanya mencakup fungsi prediksi tahap akhir pada satu mesin CPU yang spesifikasi perangkat kerasnya belum dilaporkan, serta menunjukkan fluktuasi waktu antar-sesi pengukuran.

(7) Prosedur teknis perolehan citra di lahan sawah serta protokol verifikasi pelabelan oleh pakar agronomi belum terdokumentasi secara formal.

4. Kesimpulan

Evaluasi empiris pada penelitian ini menunjukkan bahwa *pipeline* fusi fitur hibrida (EfficientNet-B0 + 73 fitur rekayasa manual berbasis masker + seleksi 128 fitur F-ANOVA + SVM-RBF) yang mampu mencapai akurasi 99,64% pada validasi silang *in-domain* hanya menghasilkan akurasi 18,75% (interval Wilson 95%: 11,7–28,7%) dan Macro-F1 13,64% ketika diterapkan secara *zero-shot* pada 80 citra daun padi lapangan. Hasil ini membuktikan bahwa performa tinggi pada kondisi laboratorium tidak dapat dibawa begitu saja ke kondisi lapangan tanpa adaptasi distribusi. Selanjutnya, kurva uji derau Gaussian yang mendatar di kisaran 18,75% hingga 17,5% tidak bersifat informatif untuk mengklaim ketahanan model karena akurasi awal pada domain lapangan sudah terperangkap di dekat batas bawah tebakan acak (*floor effect*). Di sisi komputasi, tahap prediksi SVM-RBF memang hanya membutuhkan waktu sekitar 0,78 ms per citra (≈1.278,5 FPS), namun angka tersebut bukan merupakan latensi sistem secara menyeluruh (*end-to-end*) sehingga kelayakan operasionalnya pada perangkat tepi belum terbukti.

Penelitian lanjutan disarankan untuk menempuh empat agenda utama: (1) mendiagnosis akar kegagalan transfer domain melalui pencatatan matriks konfusi lengkap, visualisasi masker segmentasi pada citra lapangan, dan studi ablasi per cabang fitur; (2) mengumpulkan himpunan citra lapangan dengan ukuran sampel dan cakupan kelas yang lebih luas melalui protokol akuisisi yang terdokumentasi, sembari mengulang uji derau Gaussian pada data *in-domain* sebagai pembanding; (3) mengukur latensi inferensi menyeluruh (*end-to-end*) mulai dari praproses hingga prediksi akhir pada CPU dan perangkat tepi yang spesifikasinya dilaporkan secara transparan; serta (4) menguji metode adaptasi domain (*domain adaptation*), seperti pelatihan adversarial domain (*Domain-Adversarial Training of Neural Networks*/DANN) [16], guna menyelaraskan representasi fitur antara citra laboratorium dan citra lapangan.

Daftar Rujukan

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Front. Plant Sci.*, vol. 7, Art. no. 1419, 2016, doi: 10.3389/fpls.2016.01419.

[2] J. G. A. Barbedo, "A review on the main challenges in automatic plant disease identification based on visible range images," *Biosyst. Eng.*, vol. 144, pp. 52–60, 2016, doi: 10.1016/j.biosystemseng.2016.01.017.

[3] A. Ahmad, A. El Gamal, and D. Saraswat, "Toward generalization of deep learning-based plant disease identification under controlled and field conditions," *IEEE Access*, vol. 11, pp. 9042–9057, 2023, doi: 10.1109/ACCESS.2023.3240100.

[4] P. W. Koh *et al.*, "WILDS: A benchmark of in-the-wild distribution shifts," in *Proc. 38th Int. Conf. Mach. Learn. (ICML)*, PMLR 139, 2021, pp. 5637–5664.

[5] S. Ben-David *et al.*, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, no. 1–2, pp. 151–175, 2010, doi: 10.1007/s10994-009-5152-4.

[6] J. G. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N. V. Chawla, and F. Herrera, "A unifying view on dataset shift in classification," *Pattern Recognit.*, vol. 45, no. 1, pp. 521–530, 2012, doi: 10.1016/j.patcog.2011.06.019.

[7] K. Zhou *et al.*, "Domain generalization: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, pp. 1–20, 2022, doi: 10.1109/TPAMI.2022.3195549.

[8] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019, arXiv:1903.12261.

[9] M. Dehghani, A. Arnab, L. Beyer, A. Vaswani, and Y. Tay, "The efficiency misnomer," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2022, arXiv:2110.12894.

[10] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, PMLR 97, 2019, pp. 6105–6114.

[11] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[12] Fahruliazmi, I., Saputra, D., Nurdiani, S., & Sidauruk, J. (2026). Aspect-Based Sentiment Analysis Ulasan VGA RTX 5060 Berbasis LLM dan Naive Bayes. *Jurnal Informatika dan Teknologi Pendidikan*, *6*(1), 65-76..

[13] "Rice Leaf Disease Detection Dataset," Kaggle, pengguna shams069. [Online]. Available: https://www.kaggle.com/datasets/shams069/rice-leaf-disease-detection-dataset

[14] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval estimation for a binomial proportion," *Stat. Sci.*, vol. 16, no. 2, pp. 101–133, 2001, doi: 10.1214/ss/1009213286.

[15] A. Howard *et al.*, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[16] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 59, pp. 1–35, 2016.