Sistemasi: Jurnal Sistem Informasi
Volume 15, Nomor 11, 2026: 4581-4594
ISSN:2302-8149
e-ISSN:2540-9719

Perbandingan *ResNet-50* dan *Self-Supervised Contrastive Learning* untuk Klasifikasi Citra Daun Kopi Sehat dan Berkarat

Comparison of ResNet-50 and Self-Supervised Contrastive Learning for Healthy and Rust-Affected Coffee Leaf Classification

[1]Acce Venio Hasugian*, [2]Chaerul Anwar
[1,2]Program Studi Sistem Informasi, Fakultas Teknologi dan Desain, Universitas Pembangunan Jaya
[1,2]Jl. Boulevard UPJ, Bintaro Jaya, Kota Tangerang Selatan, Banten, Indonesia
e-mail: acce.veniohasugian@student.upj.ac.id,chaerulanwar@upj.ac.id*

(***received*:** 12 October 2026, ***revised*:** 5 November 2026, ***accepted*:** 6 November 2026)

Abstrak

Klasifikasi citra dapat membantu pengenalan karat daun kopi, tetapi kinerja model pada data uji internal belum tentu sama ketika dievaluasi menggunakan citra yang dikumpulkan secara terpisah. Penelitian ini membandingkan ResNet-50 dengan bobot awal ImageNet dan dua konfigurasi pembelajaran kontrastif berbasis SimCLR untuk mengklasifikasikan citra daun kopi ke dalam kelas *Healthy* dan *Rust*. Dataset sekunder setelah pembersihan terdiri atas 1.338 citra, yang dibagi menjadi 938 citra latih, 207 citra validasi, dan 193 citra uji internal. Evaluasi tambahan menggunakan 106 citra primer dengan label berdasarkan penilaian visual peneliti. Pada uji internal, ResNet-50 memperoleh akurasi 96,89%, sedangkan SimCLR konfigurasi pertama dan kedua masing-masing memperoleh 60,10% dan 91,71%. Pada citra primer, ResNet-50 mengklasifikasikan 94 dari 106 citra sesuai label, menghasilkan akurasi 88,68%, *precision* kelas *Rust* 95,65%, *recall* 81,48%, dan F-1 score 88,00%. Sepuluh citra berlabel *Rust* diprediksi sebagai *Healthy*. Penelitian ini juga melaporkan evaluasi YOLOv5 untuk tugas deteksi objek secara terpisah serta prototipe Gradio untuk menampilkan prediksi ResNet-50. Hasil menunjukkan bahwa ResNet-50 memberikan kinerja terbaik pada uji internal di antara konfigurasi yang dibandingkan, tetapi mengalami penurunan akurasi sebesar 8,21 poin persentase pada citra primer. Penelitian ini terbatas pada klasifikasi berdasarkan label *Healthy* dan *Rust* serta belum membuktikan kemampuan deteksi infeksi fase laten. Label citra primer belum diverifikasi secara independen oleh ahli tanaman.

Kata kunci: karat daun kopi, klasifikasi citra, ResNet-50, self-supervised contrastive learning, transfer learning.

Abstract

Image classification can support the identification of coffee leaf rust; however, model performance on internal test data may not necessarily generalize to independently collected images. This study compares ResNet-50 initialized with ImageNet weights against two SimCLR-based contrastive learning configurations for classifying coffee leaf images into two categories: Healthy and Rust. The cleaned secondary dataset consisted of 1,338 images, divided into 938 training images, 207 validation images, and 193 internal test images. An additional evaluation was conducted using 106 primary images labeled based on the researcher's visual assessment. On the internal test set, ResNet-50 achieved an accuracy of 96.89%, while the first and second SimCLR configurations achieved 60.10% and 91.71%, respectively. On the primary image dataset, ResNet-50 correctly classified 94 out of 106 images, achieving an accuracy of 88.68%, a Rust-class precision of 95.65%, a recall of 81.48%, and an F1-score of 88.00%. Ten images labeled as Rust were misclassified as Healthy. This study also reports a separate evaluation of YOLOv5 for object detection and presents a Gradio prototype for displaying ResNet-50 predictions. The results indicate that ResNet-50 achieved the highest performance on the internal test set among the evaluated configurations but experienced an accuracy decrease of 8.21 percentage points on the primary image dataset. This study is limited to classification into the Healthy and Rust categories and does not establish the model's ability to detect

 ISSN:2302-8149
e-ISSN:2540-9719

latent-stage infections. Furthermore, the labels assigned to the primary images have not been independently verified by plant pathology experts.
Keywords: *coffee leaf rust, image classification, ResNet-50, self-supervised contrastive learning, transfer learning.*

1 Pendahuluan

Karat daun merupakan penyakit yang relevan bagi tanaman kopi. Penelitian pada klon kopi Robusta di Lampung, misalnya menilai intensitas penyakit karat daun sebagai bagian dari pengamatan kesehatan tanaman [1]. Pemeriksaan kondisi daun secara visual membutuhkan perhatian pada gejala yang tampak, sehingga klasifikasi citra dapat digunakan untuk membantu mengenali daun sehat dan daun yang menunjukkan karat.

Berbagai pendekatan telah digunakan dalam pengenalan penyakit daun kopi. Novtahaning *et al*. mengembangkan pendekatan gabungan model pembelajaran mendalam [2], sedangkan Hitimana *et al*. meneliti pengenalan penyakit pada dataset citra kopi Arabika yang mereka susun [3]. Putra *et al.* menerapkan ResNet-50 pada citra daun kopi [4], dan Bath *et al*. membandingkan beberapa pendekatan pembelajaran mendalam untuk mengidentifikasi penyakit daun kopi [5]. Adelaja dan Pranggono turut mengkaji penerapan model untuk identifikasi penyakit daun kopi [6].

Penerapan model di luar eksperimen pelatihan juga menjadi perhatian penelitian terdahulu. Yamashita dan Leite mengembangkan klasifikasi penyakit kopi pada perangkat komputasi tepi [7]. Untuk tugas yang berbeda dari klasifikasi satu citra, Araaf *et al*. meneliti deteksi karat daun kopi [8], Silva E. Silva *et al* menggabungkan pelokalan dan klasifikasi penyakit [9], sedangkan Fragoso *et al* mengevaluasi model YOLO untuk mendeteksi penyakit dan hama pada daun kopi [10], Perbedaan tugas tersebut perlu diperhatikan karena nilai evaluasi deteksi objek tidak dapat langsung disamakan dengan akurasi klasifikasi citra.

Salah satu keterbatasan pengenalan penyakit berbasis citra adalah bahwa label kondisi visual tidak secara langsung memastikan status infeksi sebelum gejala terlihat. Dalam penelitian ini, label *Healthy* dan *Rust* digunakan untuk klasifikasi citra sesuai kategori yang tersedia. Penelitian tidak menggunakan data dengan status infeksi fase laten yang telah diverifikasi. Oleh karena itu, kemampuan mendeteksi infeksi fase laten tidak menjadi hasil yang diklaim dalam penelitian ini.

Self-Supervised Contrastive Learning memungkinkan model mempelajari representasi citra tanpa menggunakan label pada tahap prapelatihan. Salah satu pendekatannya adalah SimCLR, yang mempelajari representasi melalui dua tampilan hasil augmentasi dari citra yang sama [11] . Dalam penelitian ini, SimCLR digunakan dalam dua konfigurasi pelatihan dan dibandingkan dengan ResNet-50 yang menggunakan bobot awal ImageNet. Perbandingan tersebut dilakukan pada pembagian data uji internal yang sama agar hasil masing-masing konfigurasi dapat dinilai pada kelompok citra yang sama.

Kesenjangan yang dikaji dalam penelitian ini adalah perbedaan kinerja antara model dasar ResNet-50 dan dua konfigurasi SimCLR pada dataset daun kopi yang digunakan, serta perubahan kinerja ketika model dievaluasi menggunakan citra primer yang dikumpulkan secara terpisah. Posisi penelitian ini terletak pada penyajian perbandingan konfigurasi dalam satu rancangan evaluasi, disertai analisis kesalahan kelas *Rust* pada citra primer. Kedua konfigurasi SimCLR berbeda dalam beberapa pengaturan pelatihan sehingga hasilnya ditafsirkan sebagai perbandingan konfigurasi secara keseluruhan, bukan sebagai pengujian pengaruh satu parameter secara terpisah.

Berdasarkan kesenjangan tersebut, penelitian ini mengajukan tiga pertanyaan penelitian: (1) bagaimana perbandingan kinerja ResNet-50 dan dua konfigurasi SimCLR pada data uji internal; (2) bagaimana perubahan kinerja ResNet-50 ketika dievaluasi menggunakan citra primer; dan (3) bagaimana pola kesalahan klasifikasi kelas *Rust* pada citra primer berdasarkan *confusion matrix*, *precision*, *recall*, dan F1-score?

Penelitian ini bertujuan menjawab pertanyaan tersebut melalui evaluasi klasifikasi pada data uji internal dan evaluasi tambahan pada citra primer. Kontribusi utamanya adalah perbandingan hasil tiga konfigurasi model pada data uji internal yang sama serta analisis perubahan kinerja dan kesalahan kelas Rust pada citra primer. Evaluasi citra primer mencakup ResNet-50 dan SimCLR konfigurasi pertama, sedangkan SimCLR konfigurasi kedua belum dievaluasi pada kelompok citra primer yang

 ISSN:2302-8149
e-ISSN:2540-9719

sama. Penelitian juga menyajikan evaluasi YOLOv5 sebagai tugas deteksi objek yang terpisah serta prototipe Gradio untuk menampilkan prediksi ResNet-50.

2 Tinjauan Literatur

Dataset RoCoLe menyediakan citra daun kopi Robusta berlabel untuk penelitian pengolahan citra dan pengenalan penyakit tanaman [12]. Penelitian ini menggunakan kumpulan citra sekunder yang telah disiapkan untuk klasifikasi biner *Healthy* dan *Rust*, kemudian membersihkannya sebelum pembagian data. Sumber *dataset* perlu dijelaskan karena komposisi, label, dan pembagian data menentukan ruang lingkup hasil yang dapat disimpulkan.

ResNet memperkenalkan hubungan residual untuk melatih jaringan yang dalam [13]. Dalam penelitian penyakit daun kopi, pendekatan berbasis ResNet-50 telah digunakan oleh Putra et al. [4], sementara Mansouri et al. mengkaji klasifikasi penyakit daun kopi menggunakan transfer learning [14]. Chavarro et al. juga menyertakan ResNet-50 dalam kajian klasifikasi karat daun kopi dan interpretasi model [15]. Penelitian tersebut mendukung penggunaan ResNet-50 sebagai model pembanding, tetapi nilai akurasinya tidak dibandingkan langsung dengan penelitian ini karena data dan rancangan pengujiannya berbeda.

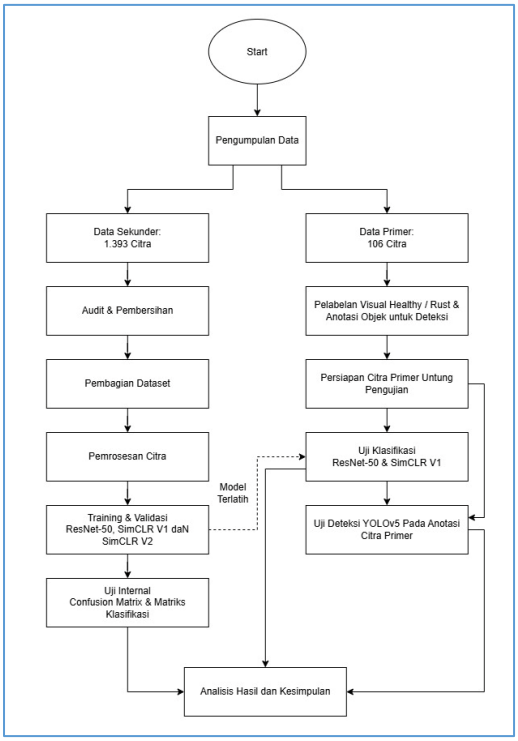
SimCLR mempelajari representasi citra dengan mendekatkan representasi dua tampilan hasil augmentasi dari citra yang sama [11]. Penelitian Wang et al. mengeksplorasi pembelajaran mandiri untuk pengenalan penyakit daun tanaman [16]. Huan et al. menguji kerangka pembelajaran mandiri pada citra penyakit tanaman dari sumber yang berbeda [17], sedangkan Shah et al. mengembangkan PlantCLR untuk prapelatihan kontrastif dalam pengenalan penyakit tanaman [18]. Penelitian-penelitian tersebut memberi dasar untuk menguji pendekatan kontrastif, tetapi tidak menjamin hasil yang sama pada dataset daun kopi dalam penelitian ini.

Kajian pustaka memperlihatkan tiga kebutuhan yang saling berkaitan. Pertama, perbandingan model klasifikasi perlu dilakukan pada pembagian data yang jelas. Kedua, evaluasi pada citra yang dikumpulkan secara terpisah diperlukan untuk melihat perbedaan kinerja dari uji internal. Ketiga klasifikasi citra dan deteksi objek harus dilaporkan dengan metrik yang sesuai dengan tugas masing masing. Penelitian ini berfokus pada ketiga kebutuhan tersebut, dengan tetap membatasi kesimpulan pada label dan Kumpulan citra yang benar benar diuji.

3 Metode Penelitian

Alur data, penelitian, dan evaluasi diperlihatkan pada Gambar 1. Citra sekunder digunakan untuk mengembangkan dan menguji model klasifikasi. Citra primer dikumpulkan terpisah untuk evaluasi tambahan. YOLOv5 ditangani sebagai tugas deteksi objek.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 1 Alur penelitian klasifikasi dan deteksi karat daun kopi

3.1 Data Penelitian

Penelitian menggunakan kumpulan citra daun kopi sekunder untuk klasifikasi biner *Healthy* dan *Rust* yang bersumber dari RoCoLe [12]. Kumpulan awal yang diproses terdiri atas 1.393 citra. Pembersihan duplikat dilakukan berdasarkan kesamaan nilai hash SHA-256 yang tercatat pada *manifest*. Sebanyak 55 citra duplikat dalam kelas yang sama dikeluarkan dengan mempertahankan satu citra perwakilan dari setiap kelompok duplikat. Setelah pembersihan, diperoleh 1.338 citra yang terdiri atas 759 citra *Healthy* dan 579 citra *Rust*. Angka 1.393 merujuk pada kumpulan awal yang diproses dalam penelitian ini, bukan jumlah keseluruhan citra pada publikasi dataset RoCoLe.

Pembagian data dilakukan dengan mempertahankan keutuhan kelompok berdasarkan group_id yang tercatat pada *manifest*. Sebanyak 384 kelompok dialokasikan ke dalam subset latih, validasi, dan uji internal menggunakan seed 42, dengan target proporsi masing-masing 70%, 15%, dan 15%. Seluruh citra dengan group_id yang sama ditempatkan dalam satu subset. Hasil pembagian menghasilkan 938 citra latih, 207 citra validasi, dan 193 citra uji internal. Proporsi aktual jumlah citra mengikuti hasil alokasi kelompok sehingga tidak harus sama persis dengan target proporsi pembagian. Rincian jumlah citra setiap kelas disajikan pada Tabel 1.

Pemeriksaan setelah pembagian dilakukan untuk mengidentifikasi tumpang tindih group_id dan hash SHA-256 antarsubset. Hasil pemeriksaan menunjukkan tidak terdapat group_id maupun hash SHA-256 yang muncul pada lebih dari satu subset. Temuan ini menunjukkan bahwa tidak ditemukan tumpang tindih berdasarkan identitas kelompok yang tercatat dan duplikat identik berdasarkan hash. Pemeriksaan tersebut belum memastikan ketiadaan seluruh bentuk kemiripan visual antarcitra.

Tabel 1 Pembagian data sekunder

Kelompok Data	Healthy	Rust	Jumlah
Training	532	406	938
Validation	118	89	207
Internal Test	109	84	193
Total	759	579	1.338

Selain itu, digunakan 106 citra primer yang dikumpulkan secara terpisah: 52 citra berlabel *Healthy* dan 54 citra berlabel *Rust.* Label ditetapkan melalui penelaian visual peneliti. Citra primer

 ISSN:2302-8149
e-ISSN:2540-9719

tidak digunakan dalam pelatihan model klasifikasi, dan labelnya belum divalidasi secara independen oleh ahli tanaman.

3.2 Prapemrosesan dan Pelatihan Model

Pelatihan dan evaluasi model dilakukan menggunakan PyTorch melalui Google Colab dengan GPU Tesla T4. Citra dibaca dalam format RGB dengan pemetaan kelas healthy = 0 dan rust = 1. Pada tahap pelatihan klasifikasi, augmentasi meliputi *RandomResizedCrop* berukuran 224 × 224 piksel dengan rentang skala luas 0,8–1,0, pembalikan horizontal acak, rotasi acak hingga ±15°, serta *ColorJitter* dengan *parameter brightness*, *contrast*, dan *saturation* masing-masing 0,2. Citra kemudian diubah menjadi tensor dan dinormalisasi menggunakan mean ImageNet [0,485; 0,456; 0,406] dan standar deviasi [0,229; 0,224; 0,225].

Pada validasi dan pengujian klasifikasi, citra diproses menggunakan *Resize*(256), dilanjutkan pemotongan tengah berukuran 224 × 224 piksel, konversi menjadi tensor, dan normalisasi yang sama. Augmentasi acak pelatihan tidak diterapkan pada tahap ini. *Batch size* untuk pelatihan dan evaluasi klasifikasi adalah 32. Istilah pengujian pada citra penuh merujuk pada penggunaan citra tanpa pemotongan berdasarkan kotak anotasi, tetapi citra tetap melewati pemotongan tengah pada tahap prapemrosesan model.

Model dasar menggunakan ResNet-50 dengan bobot awal ImageNet. Lapisan klasifikasi terakhir diganti dengan lapisan linear yang menghasilkan dua keluaran untuk kelas *Healthy* dan *Rust*. Seluruh parameter model dilatih selama 10 epoch menggunakan fungsi *loss CrossEntropyLoss* dan *optimizer* Adam dengan *learning rate* awal 0,0001. *Scheduler ReduceLROnPlateau* digunakan berdasarkan *loss* validasi, dengan *patience* 2 dan faktor penurunan *learning rate* 0,5. Bobot untuk pengujian dipilih berdasarkan akurasi validasi tertinggi, yang diperoleh pada epoch keenam. Data uji internal tidak digunakan untuk memilih bobot model.

Kedua konfigurasi SimCLR menggunakan *backbone* ResNet-50 sebagai *encoder* dan *projection head* berupa lapisan linear 2048→512, aktivasi ReLU, serta lapisan linear 512→128. Prapelatihan menggunakan citra pada subset latih tanpa memanfaatkan label kelas. Setiap citra menghasilkan dua tampilan melalui augmentasi acak yang diterapkan secara terpisah. Fungsi *loss* yang digunakan adalah NT-Xent dengan *temperature* 0,5. *Batch size* prapelatihan adalah 64, dengan batch terakhir yang tidak lengkap diabaikan. Kedua konfigurasi menggunakan *optimizer* Adam dengan *weight decay* 0,0001 dan *scheduler CosineAnnealingLR*. Bobot prapelatihan dipilih berdasarkan *loss* prapelatihan terendah.

SimCLR konfigurasi pertama menggunakan bobot awal acak dan menjalani prapelatihan selama 20 epoch dengan *learning rate* awal 0,0003. Augmentasinya meliputi *RandomResizedCrop* berukuran 224 × 224 piksel dengan rentang skala luas 0,2–1,0, pembalikan horizontal acak, ColorJitter dengan *brightness*, *contrast*, dan *saturation* masing-masing 0,4 serta *hue* 0,1 yang diterapkan dengan probabilitas 0,8, dan konversi *grayscale* dengan probabilitas 0,2. *Gaussian blur* diterapkan dengan probabilitas 0,5 menggunakan radius acak antara 0,1 dan 2,0. Selanjutnya, citra diubah menjadi tensor dan dinormalisasi menggunakan parameter ImageNet.

SimCLR konfigurasi kedua menggunakan bobot awal ImageNet dan menjalani prapelatihan selama 100 epoch dengan learning rate awal 0,0001. Augmentasinya meliputi *RandomResizedCrop* berukuran 224 × 224 piksel dengan rentang skala luas 0,5–1,0, pembalikan horizontal acak, ColorJitter dengan *brightness*, *contrast*, dan *saturation* masing-masing 0,2 serta *hue* 0,05 yang diterapkan dengan probabilitas 0,5, dan konversi *grayscale* dengan probabilitas 0,1. *Gaussian blur* menggunakan kernel berukuran 23 dan sigma antara 0,1 dan 2,0, dengan probabilitas penerapan 0,3. Konversi tensor dan normalisasi menggunakan parameter ImageNet yang sama.

Setelah prapelatihan, projection head tidak digunakan untuk klasifikasi. Keluaran *encoder* berdimensi 2048 dihubungkan dengan lapisan linear dua kelas. Pada kedua konfigurasi, tahap awal pelatihan klasifikasi dilakukan selama 10 epoch dengan parameter *encoder* dibekukan dan hanya parameter lapisan klasifikasi yang dioptimalkan. Tahap ini menggunakan *CrossEntropyLoss* dan *optimizer* Adam dengan *learning rate* awal 0,001 serta *weight decay* 0,0001. Bobot dipilih berdasarkan akurasi validasi tertinggi. Pada konfigurasi kedua, *encoder* juga ditempatkan dalam mode evaluasi selama tahap ini agar statistik *BatchNorm* tidak berubah, sedangkan kode konfigurasi pertama tetap menempatkan model dalam mode pelatihan.

 ISSN:2302-8149
e-ISSN:2540-9719

Tahap berikutnya adalah fine-tuning seluruh parameter encoder dan lapisan klasifikasi menggunakan CrossEntropyLoss dan optimizer Adam. Learning rate awal encoder adalah 0,00001, sedangkan learning rate awal lapisan klasifikasi adalah 0,0001, dengan weight decay 0,0001. Pelatihan kepala klasifikasi dan fine-tuning menggunakan scheduler ReduceLROnPlateau berdasarkan loss validasi, dengan patience 2 dan faktor 0,5.

Fine-tuning konfigurasi pertama dilakukan selama 10 epoch dengan pemilihan bobot berdasarkan akurasi validasi tertinggi. *Fine-tuning* konfigurasi kedua ditetapkan maksimal 15 epoch dengan *early stopping* apabila *loss* validasi tidak membaik selama lima epoch berturut-turut. Pada eksperimen yang dilaporkan, pelatihan konfigurasi kedua berhenti pada epoch ke-12 dan bobot dengan *loss* validasi terendah diperoleh pada epoch ketujuh. Pengujian internal dilakukan setelah bobot masing-masing model dipilih menggunakan data validasi.

Kedua konfigurasi SimCLR berbeda dalam bobot awal, jumlah epoch prapelatihan, *learning rate*, augmentasi, dan beberapa prosedur pelatihan klasifikasi. Oleh karena itu, perbandingan hasil ditafsirkan sebagai perbandingan konfigurasi secara keseluruhan. Penelitian ini tidak mengisolasi pengaruh satu pengaturan tertentu terhadap perbedaan kinerja model.

3.3 Evaluasi Kinerja dan Klasifikasi

Evaluasi klasifikasi menggunakan *confusion matrix*, akurasi, *precision*, *recall*, dan F1-score. Kelas *Rust* ditetapkan sebagai kelas positif. *True positive* (TP) adalah citra *Rust* yang diprediksi *Rust*; *true negative* (TN) adalah citra *Healthy* yang diprediksi *Healthy*; *false positive* (FP) adalah citra *Healthy* yang diprediksi *Rust*; dan *false negative* (FN) adalah citra *Rust* yang diprediksi *Healthy*.

Akurasi dihitung menggunakan Persamaan (1):

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\% \quad (1)$$

Precision kelas *Rust* dihitung menggunakan Persamaan (2):

$$\text{Precision}_{Rust} = \frac{TP}{TP+FP}\ x\ 100\% \quad (2)$$

Recall kelas *Rust* dihitung menggunakan Persamaan (3):

$$\text{Recall}_{Rust} = \frac{TP}{TP+FN}\ x\ 100\% \quad (3)$$

F1-score kelas *Rust* dihitung menggunakan Persamaan (4):

$$\text{F1} - \text{score}_{Rust} = \frac{2\ x\ \text{Precision}_{Rust}\ x\ \text{Recall}_{Rust}}{\text{Precision}_{Rust} + \text{Recall}_{Rust}} \quad (4)$$

Pada persamaan (4), *precision*, dan *recall* dimasukkan pada skala yang sama. Jika keduanya dimasukkan sebagai persentase, hasil F1-score juga berupa persentase dan tidak dikalikan 100% sekali lagi. Untuk membandingkan model pada uji internal, penelitian ini turut melaporkan F1-score berbobot dan ROC-AUC dari hasil evaluasi model.

3.4 Evaluasi Deteksi dan Prototipe

YOLOv5s digunakan untuk tugas deteksi objek dua kelas, yaitu *Healthy* dan *Rust*. *Dataset* deteksi disiapkan secara terpisah dari dataset klasifikasi menggunakan citra dan anotasi RoCoLe versi DatasetNinja. Anotasi poligon dikonversi menjadi kotak pembatas dengan mengambil koordinat minimum dan maksimum poligon, kemudian dinormalisasi ke format YOLO. Anotasi *healthy* dipetakan menjadi kelas *Healthy*, sedangkan rust_level_1 sampai rust_level_4 digabungkan menjadi kelas *Rust*. Anotasi *unhealthy* dan red_spider_mite tidak digunakan sebagai kelas target.

 ISSN:2302-8149
e-ISSN:2540-9719

Setelah pembersihan, dataset deteksi terdiri atas 1.334 citra, dengan 755 citra *Healthy* dan 579 citra *Rust*. Pembagian dilakukan secara acak dengan stratifikasi kelas menggunakan random_state 42, menghasilkan 933 citra latih, 200 citra validasi, dan 201 citra uji internal. Pembagian *dataset* deteksi ini berbeda dari pembagian dataset klasifikasi dan tidak menggunakan prosedur pengelompokan group_id yang dijelaskan pada bagian 3.1.

Pelatihan YOLOv5s menggunakan bobot awal yolov5s.pt, ukuran masukan 640 piksel, *batch size* 16, dan 100 epoch, dengan *seed* 42 serta *patience early stopping* 20. *Optimizer* yang digunakan adalah SGD dengan *learning rate awal* 0,01, *momentum* 0,937, dan *weight decay* 0,0005. *Warmup* diterapkan selama tiga epoch. Pengaturan augmentasi meliputi perubahan warna HSV dengan *parameter hue* 0,015, *saturation* 0,7, dan *value* 0,4; translasi 0,1; perubahan skala 0,5; pembalikan horizontal dengan probabilitas 0,5; serta *mosaic* dengan probabilitas 1,0. *Mixup*, pembalikan vertikal, dan rotasi tidak diaktifkan pada konfigurasi yang tersimpan.

Evaluasi tambahan deteksi dilaporkan pada 106 citra primer dengan 106 objek beranotasi, yang terdiri atas 52 objek *Healthy* dan 54 objek *Rust*. Pemetaan kelas pada konfigurasi evaluasi adalah *Healthy* = 0 dan *Rust* = 1. Metrik yang dilaporkan meliputi *precision*, *recall*, mAP@0,5, dan mAP@0,5:0,95. Metrik tersebut menilai tugas deteksi objek sehingga dilaporkan secara terpisah dan tidak disamakan dengan akurasi klasifikasi ResNet-50.

Untuk memastikan prosedur evaluasi terdokumentasi, pengujian ulang dilakukan menggunakan bobot hasil pelatihan YOLOv5s, yaitu best.pt, pada 106 citra primer. Evaluasi menggunakan ukuran masukan 640 piksel, *batch size* 16, *confidence threshold* awal 0,001, dan IoU *threshold* untuk *non-maximum suppression* sebesar 0,60. Pengujian menghasilkan nilai *precision* 0,744, *recall* 0,225, mAP@0,5 sebesar 0,431, dan mAP@0,5:0,95 sebesar 0,208. Nilai tersebut sama dengan ringkasan hasil evaluasi yang telah dilaporkan pada penelitian.

Prototipe antarmuka dibuat menggunakan *Gradio* di Google Colab. Pengguna dapat memasukkan citra dan melihat keluaran prediksi ResNet-50. Pengujian prototipe yang dilaporkan di sini terbatas pada fungsi inferensi tersebut. Aplikasi belum dievaluasi sebagai layanan yang di-*hosting* secara permanen.

4 Hasil dan Pembahasan

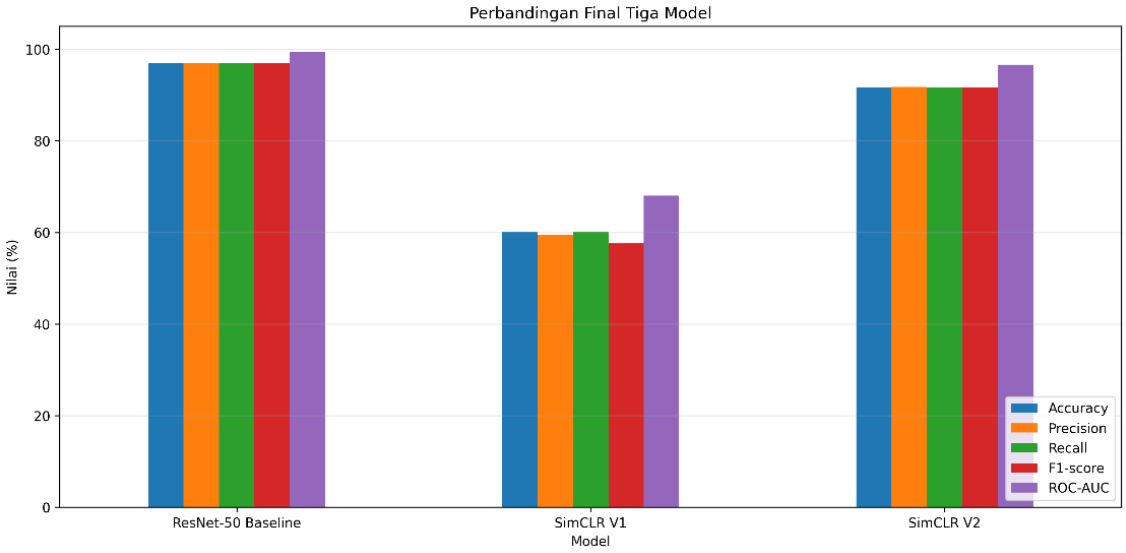
4.1 Perbandingan Model pada Uji Internal

Tabel 2 dan gambar 2 menunjukkan hasil tiga model pada 193 citra uji internal. ResNet-50 memberikan akurasi dan F1-score tertinggi, diikuti SimCLR konfigurasi kedua dan pertama. Pada gambar 2, *precision*, *recall*, dan F1-score merupakan agregat berbobot menurut jumlah sampel setiap kelas, berbeda dari metrik khusus kelas *Rust* yang dibahas berikutnya.

Tabel 2 Hasil klasifikasi pada uji internal

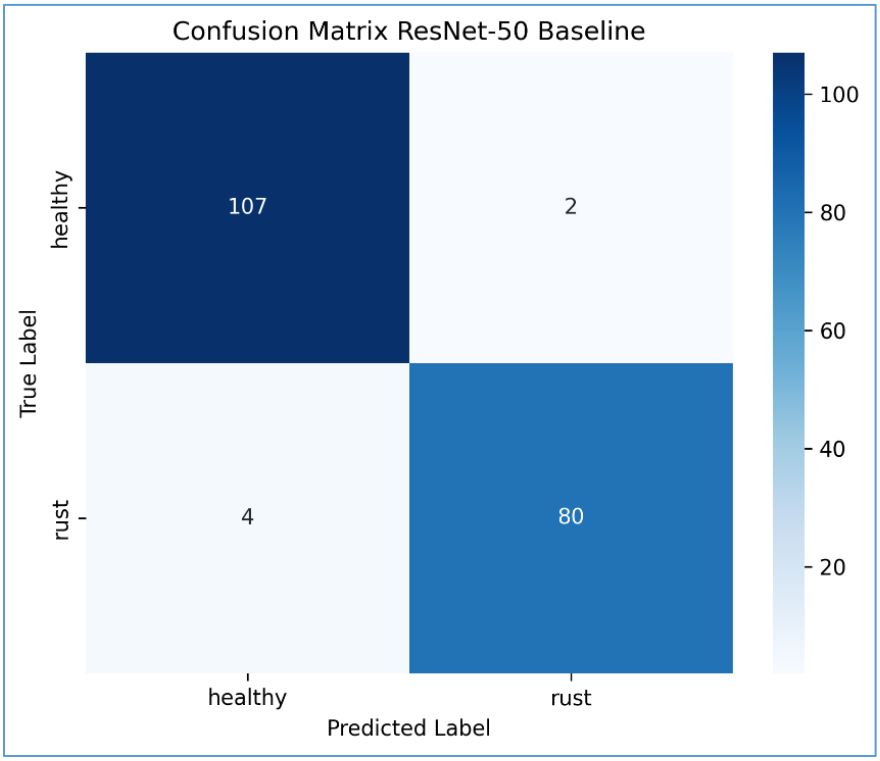
Model	Prediksi Benar	Akurasi	F-1 score berbobot	ROC-AUC
ResNet-50	187 dari 193	96,89%	0,9689	0,9938
SimCLR Konfigurasi 1	116 dari 193	60,10%	0,5761	0,6803
SimCLR Konfigurasi 2	177 dari 193	91,71%	0,9168	0,9664

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 2 Perbandingan kinerja tiga model pada uji internal

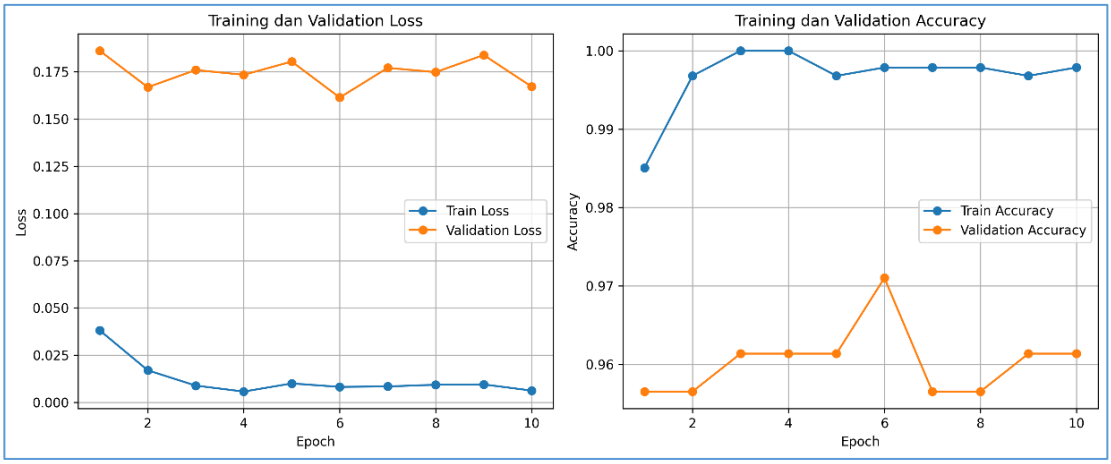
Gambar 3 menunjukkan bahwa ResNet-50 memprediksi benar 107 dari 109 citra *Healthy* dan 80 dari 84 citra *Rust*. Terdapat dua FP dan empat FN. *Precision* kelas *Rust* Adalah 97,56%, *recall* 95,24% dan F1-score 96,39%. SimCLR konfigurasi kedua lebih baik daripada konfigurasi pertama pada uji internal, tetapi perbedaan bobot awal, lama prapelatihan, dan augmentasi tidak memungkinkan satu faktor tunggal dinyatakan sebagai penyebab.



Gambar 3 *Confusion matrix* ResNet-50 pada 193 uji internal

Riwayat pelatihan ResNet-50 pada Gambar 4 menunjukkan akurasi validasi tertinggi pada *epoch* keenam dari sepuluh *epoch*. Nilai akurasi latih hampir mencapai 100% sedangkan akurasi validasi berada sekitar 95,7% - 97,1%. Pemilihan bobot pada *epoch* keenam mengikuti hasil validasi, bukan hasil uji internal. Selisih antara kurva latih dan validasi perlu diperhatikan ketika hasil model ditafsirkan.

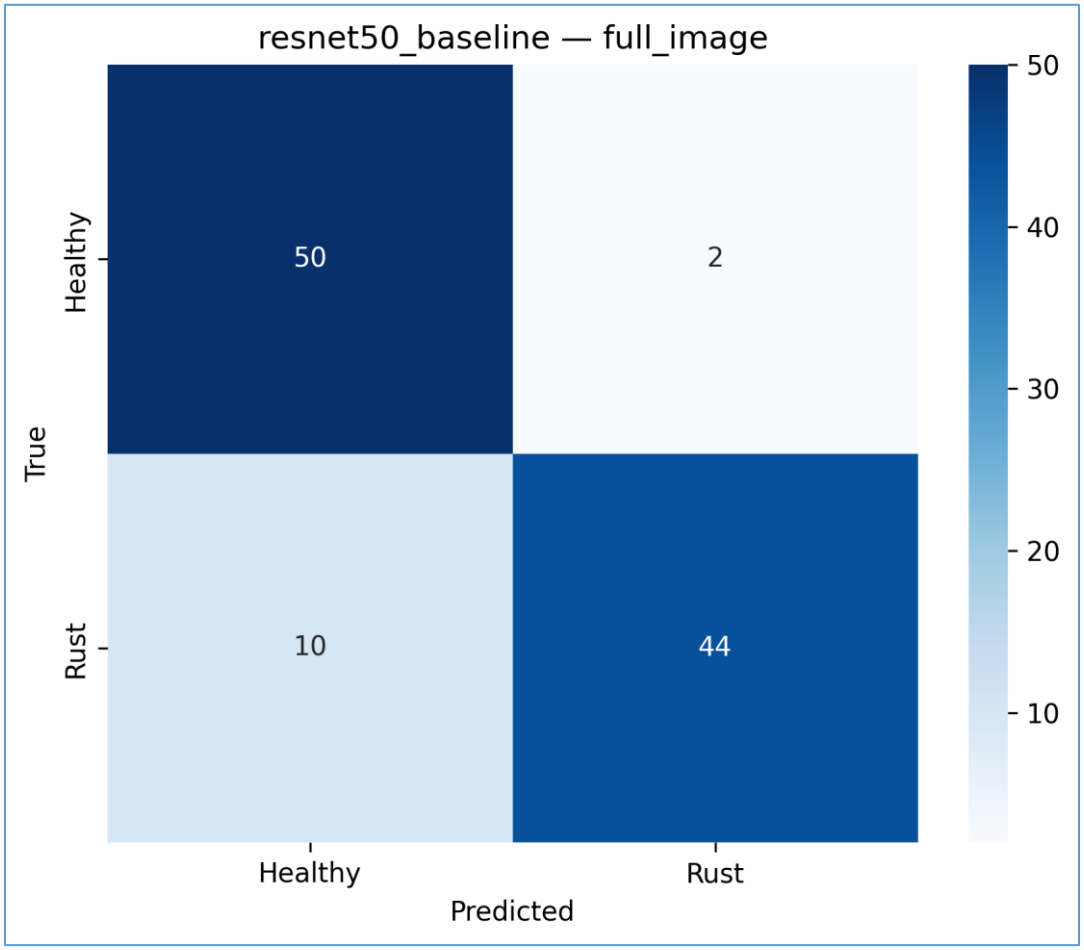
 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 4 Riwayat Loss dan Akurasi Pelatihan ResNet-50

4.2 Evaluasi ResNet-50 pada Citra Primer

Gambar 5 menunjukkan hasil ResNet-50 pada 106 citra primer. Sebanyak 50 citra *Healthy* dan 44 citra *Rust* diprediksi sesuai label visual peneliti. Sepuluh citra berlabel *Rust* diprediksi *Healthy* dan dua citra berlabel *Healthy* diprediksi *Rust*.



Gambar 5 Confusion matrix ResNet-50 pada 106 citra primer

Dengan kelas *Rust* sebagai kelas positif, hasil evaluasi pada 106 citra primer menghasilkan TP = 44, TN = 50, FP = 2, dan FN = 10. Berdasarkan Persamaan (1)–(4), ResNet-50 memperoleh akurasi

88,68%, *precision* kelas *Rust* 95,65%, *recall* 81,48%, dan F1-score 88,00%. Akurasi ini lebih rendah 8,21 poin persentase dibandingkan akurasi uji internal sebesar 96,89%.

Penurunan tersebut menunjukkan bahwa kinerja pada data uji internal belum sepenuhnya menggambarkan kinerja model pada citra yang dikumpulkan secara terpisah. Perbedaan pencahayaan, latar belakang, sudut pengambilan, jarak kamera, dan kejelasan gejala merupakan kemungkinan faktor yang perlu diperiksa. Pemotongan tengah pada tahap prapemrosesan juga berpotensi mengurangi bagian daun atau gejala yang terlihat oleh model apabila bagian tersebut berada di luar area potongan. Namun, penelitian ini belum melakukan analisis terkontrol terhadap setiap faktor sehingga penyebab penurunan kinerja belum dapat dipastikan.

Sebanyak sepuluh dari 54 citra berlabel *Rust* diprediksi sebagai *Healthy*, atau sekitar 18,52% dari seluruh citra berlabel *Rust*. Kesalahan ini menunjukkan bahwa sebagian citra yang diberi label berkarat masih terlewat oleh model. Apabila hasil prediksi digunakan sebagai bantuan penyaringan kondisi daun, kesalahan tersebut dapat menyebabkan citra yang memerlukan pemeriksaan lebih lanjut ditampilkan sebagai *Healthy*. Oleh karena itu, *precision* yang tinggi perlu dibaca bersama *recall* dan jumlah *false negative*, bukan sebagai jaminan bahwa seluruh citra *Rust* dapat dikenali.

Interpretasi hasil tersebut dibatasi oleh label citra primer yang ditetapkan melalui penilaian visual peneliti dan belum diverifikasi secara independen oleh ahli tanaman. Nilai evaluasi menggambarkan kesesuaian prediksi terhadap label tersebut, bukan pembuktian diagnosis biologis maupun kemampuan mendeteksi infeksi fase laten.

SimCLR konfigurasi pertama memperoleh akurasi 54,72% pada evaluasi citra primer yang dilaporkan. SimCLR konfigurasi kedua belum dievaluasi pada kelompok 106 citra primer yang sama. Oleh karena itu, penelitian ini belum dapat menyimpulkan perbandingan ketiga konfigurasi pada citra primer.

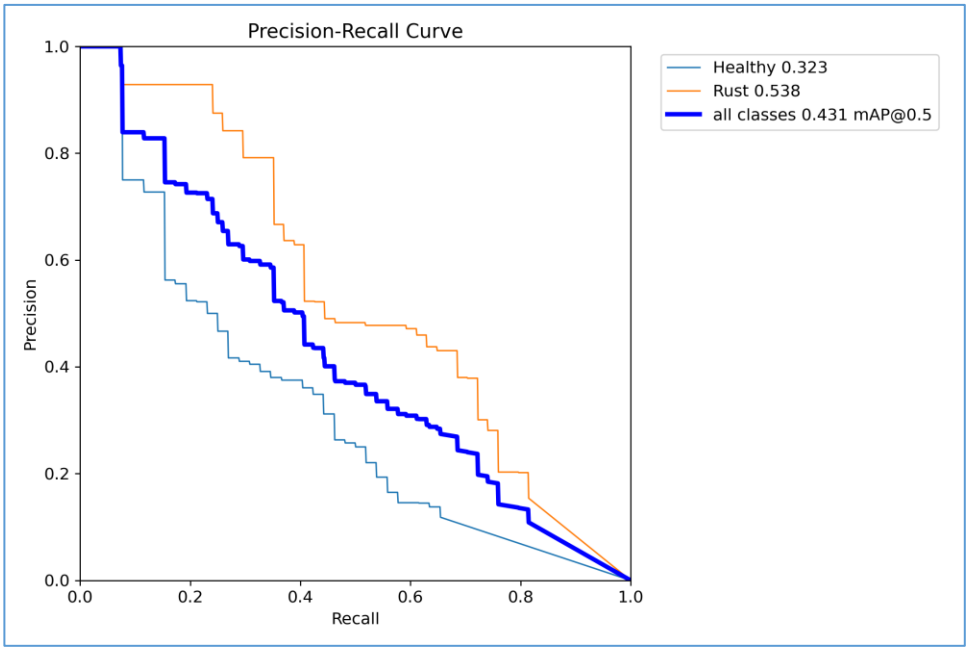
4.3 Evaluasi YOLOv5

Ringkasan hasil evaluasi YOLOV5 pada 106 citra primer mencatat *precision* 0,744, *recall* 0,225, mAP@0,5 sebesar 0,431, dan mAP@0,5:0,95, sebesar 0,208. *Recall* deteksi yang rendah menunjukkan bahwa sejumlah objek beranotasi belum terdeteksi menurut aturan pencocokan pada evaluasi tersebut. Proyek ini juga berisi run evaluasi YOLOv5 lainnya dengan hasil yang berbeda. Angka yang dilaporkan merupakan hasil evaluasi ulang YOLOv5s pada 106 citra primer menggunakan pengaturan yang dijelaskan pada bagian 3.4. Nilai mAP deteksi tidak dibandingkan langsung dengan akurasi klasifikasi citra ResNet-50.

Kurva *precision-recall* pada Gambar 6 berasal dari evaluasi ulang YOLOv5s pada 106 citra primer. Nilai AP@0,5 masing-masing adalah 0,323 untuk objek *Healthy* dan 0,538 untuk *Rust* dengan rerata 0,431. Kurva ini menggambarkan pertukaran *precision* dan *recall* pada ambang prediksi berbeda. Angka tersebut merupakan metrik deteksi, bukan peluang benar klasifikasi per citra.

Recall deteksi sebesar 0,225 menunjukkan keterbatasan model dalam menemukan objek beranotasi pada pengaturan evaluasi yang digunakan. Hasil tersebut perlu ditafsirkan bersama ambang prediksi dan aturan pencocokan deteksi yang dilaporkan pada metode. Nilai mAP@0,5 sebesar 0,431 merupakan ukuran kinerja deteksi objek dan tidak dapat ditafsirkan sebagai akurasi klasifikasi citra sebesar 43,1%. Oleh karena itu, hasil YOLOv5 dilaporkan secara terpisah dari evaluasi klasifikasi ResNet-50.

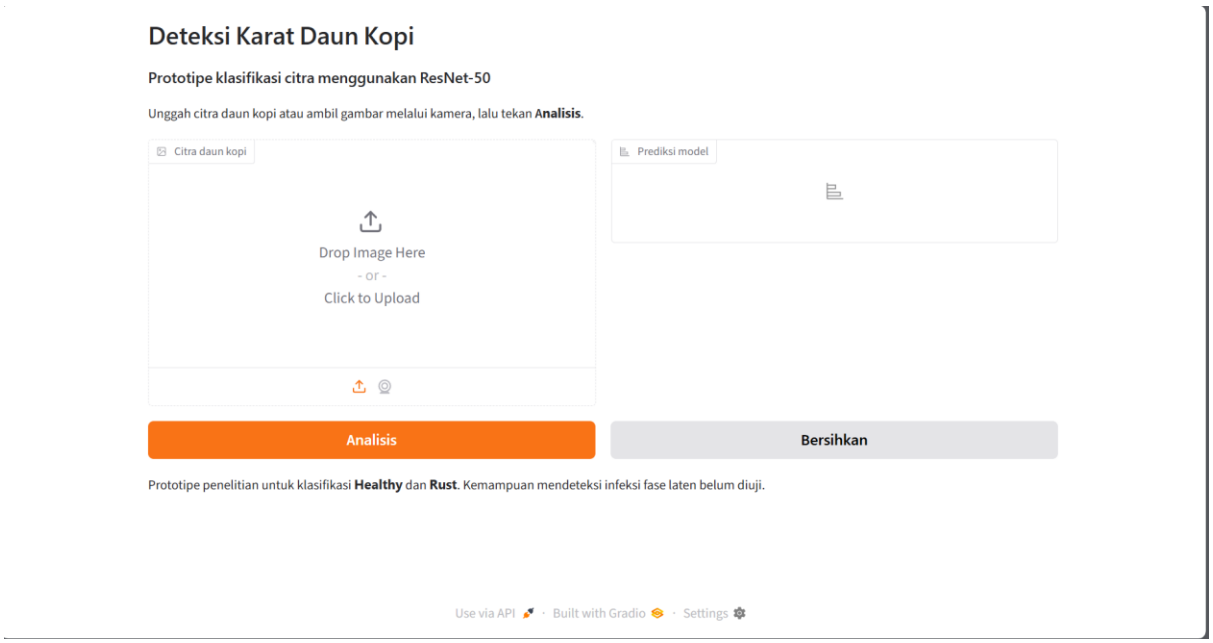
 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 6 Kurva *precision-recall* deteksi YOLOv5 pada 106 citra primer

4.4 Prototipe Aplikasi Web Berbasis Gradio

Prototipe aplikasi web dikembangkan menggunakan *Gradio* untuk mengintegrasikan model ResNet-50 ke dalam antarmuka klasifikasi citra daun kopi. Pengguna memasukkan citra daun kopi, kemudian menekan tombol **Analisis** untuk menjalankan prediksi. Aplikasi menampilkan hasil klasifikasi ke dalam kelas *Healthy* atau *Rust*. Tombol **Bersihkan** digunakan untuk menghapus citra masukan dan hasil sebelumnya. Tampilan awal prototipe ditunjukkan pada Gambar 7.

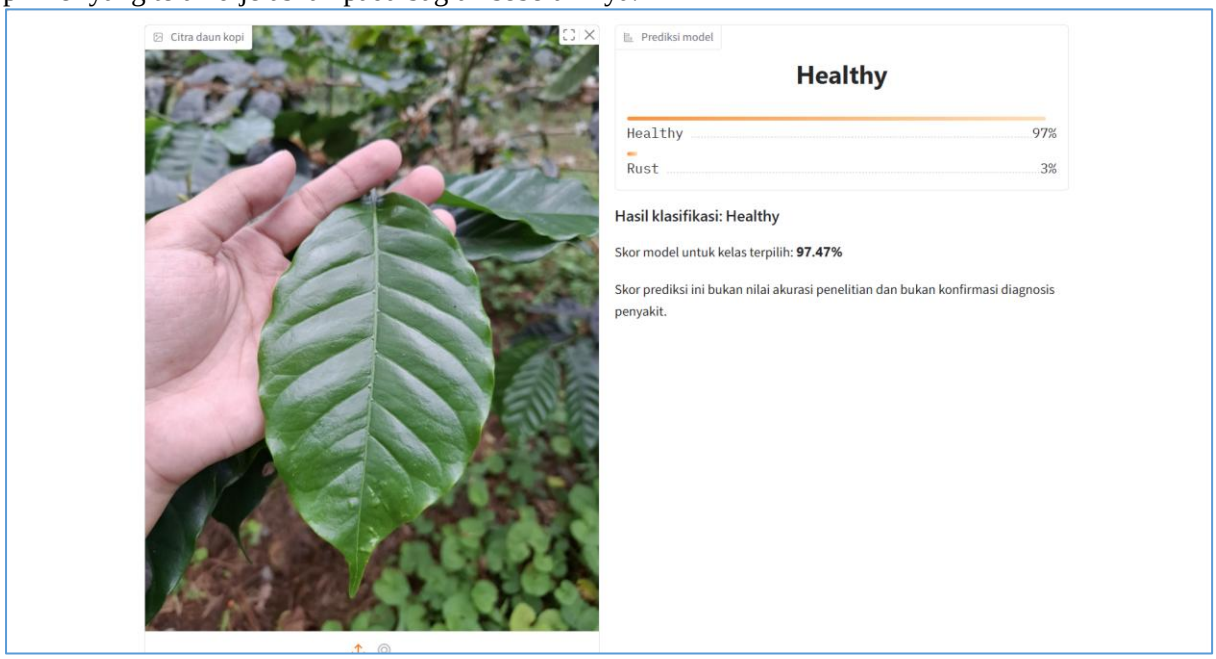


Gambar 7 Tampilan awal prototipe aplikasi *web* berbasis gradio

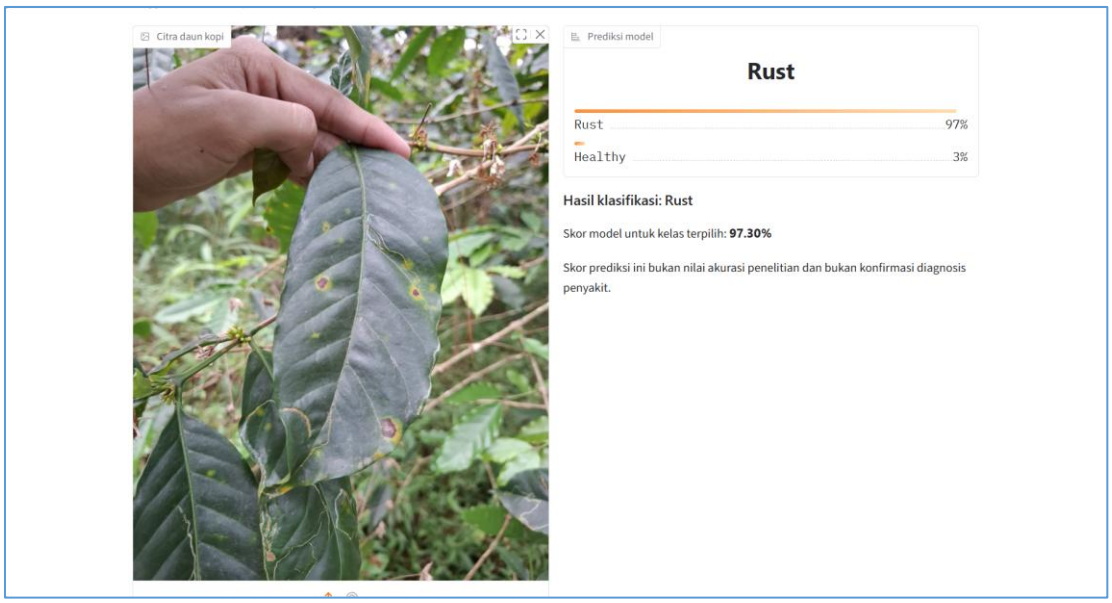
Contoh keluaran aplikasi ditunjukkan pada Gambar 8 dan Gambar 9. Setelah pengguna memasukkan citra dan proses analisis dijalankan, aplikasi menampilkan prediksi model untuk citra tersebut. Gambar 8 menunjukkan contoh keluaran prediksi *Healthy*, sedangkan Gambar 9 menunjukkan contoh keluaran dengan prediksi *Rust*. Kedua contoh tersebut merupakan demonstrasi

 ISSN:2302-8149
e-ISSN:2540-9719

fungsi aplikasi. Penilaian kinerja model tetap mengacu pada hasil pengujian data internal dan citra primer yang telah dijelaskan pada bagian sebelumnya.



Gambar 8 Contoh keluaran prototipe gradio prediksi healthy



Gambar 9 Contoh keluaran prototipe prediksi rust

Prototipe memungkinkan pengguna untuk menjalankan klasifikasi citra melalui antarmuka web tanpa menjalankan kode inferensi secara langsung. Pada tahap penelitian ini, aplikasi dijalankan melalui Google Colab menggunakan bobot ResNet-50 yang telah dilatih. Evaluasi waktu respons, kemudahan penggunaan, dan pengoperasian melalui layanan hosting permanen belum dilakukan. Oleh karena itu, hasil yang dilaporkan pada bagian ini terbatas pada demonstrasi fungsi prototipe dalam menerima citra dan menampilkan hasil prediksi.

4.5 Batasan Penelitian

Klasifikasi penelitian ini hanya mencakup dua kelas visual, *Healthy* dan *Rust*. Model tidak diuji untuk menentukan tingkat keparahan maupun fase laten infeksi. Label pada 106 citra primer berasal dari penilaian visual peneliti, bukan diagnosis ahli tanaman. Prapemrosesan dengan pemotongan

 ISSN:2302-8149
e-ISSN:2540-9719

tengah juga dapat mengubah bagian citra yang dilihat model. Selain itu, SimCLR konfigurasi kedua belum dievaluasi pada citra primer. Oleh karena itu, hasil penelitian berlaku untuk data, label, dan prosedur pengujian yang dilaporkan. Pengujian tambahan diperlukan sebelum menarik kesimpulan tentang penggunaan yang lebih luas.

5 Kesimpulan

ResNet-50 dengan bobot awal ImageNet memberikan kinerja terbaik pada uji internal di antara konfigurasi yang dibandingkan, dengan akurasi 96,89%, sedangkan SimCLR konfigurasi pertama dan kedua masing-masing memperoleh 60,10% dan 91,71%. Pada 106 citra primer, akurasi ResNet-50 menurun menjadi 88,68%, dengan precision kelas Rust 95,65%, recall 81,48%, dan F1-score 88,00%. Sepuluh citra berlabel Rust diprediksi sebagai Healthy, sehingga evaluasi perlu mempertimbangkan recall dan kesalahan per kelas selain akurasi. Evaluasi YOLOv5 disajikan secara terpisah untuk tugas deteksi objek, sementara prototipe Gradio memungkinkan penyajian prediksi ResNet-50 melalui antarmuka web. Penelitian ini belum membuktikan kemampuan mendeteksi infeksi fase laten karena tidak menggunakan data dengan status infeksi laten yang terverifikasi, dan label citra primer belum divalidasi secara independen oleh ahli tanaman. Penelitian selanjutnya perlu memvalidasi label bersama ahli tanaman, memperluas variasi citra primer, mengevaluasi SimCLR konfigurasi kedua pada kelompok citra primer yang sama, serta menguji kinerja dan kemudahan penggunaan aplikasi secara lebih menyeluruh.

Referensi

[1] Harmiansyah, D. Diptaningsari, N. Wardani, Meidaliyantisyah, R. Mawardi, and J. Hendra, "*Intensity of Leaf Rust Disease on Four Robusta Coffee Clones in Natar, South Lampung*," in *IOP Conference Series: Earth and Environmental Science*, South Lampung: Institute of Physics, Dec. 2023, pp. 1–5. DOI: 10.1088/1755-1315/1230/1/012097.

[2] D. Novtahaning, H. A. Shah, and J. M. Kang, "*Deep Learning Ensemble-based Automated and High-Performing Recognition of Coffee Leaf Disease*," *Agriculture (Switzerland)*, Vol. 12, No. 11, pp. 1–16, Nov. 2022, DOI: 10.3390/agriculture12111909.

[3] E. Hitimana *et al.*, "*An Intelligent System-based Coffee Plant Leaf Disease Recognition using Deep Learning Techniques on Rwandan Arabica Dataset*," *Technologies (Basel).*, Vol. 11, No. 5, Oct. 2023, DOI: 10.3390/technologies11050116.

[4] Y. Panca Putra, A. Susanto, and W. Novrian, "Penerapan Jaringan Syaraf Konvolusional untuk Deteksi Penyakit Tanaman Kopi berdasarkan Citra Daun," *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, Vol. 5, No. 3, pp. 507–513, Dec. 2025, DOI: 10.55382/jurnalpustakaai.v5i3.1123.

[5] N. Baht, E. Dominguez, and S. Aljumaili, "*Detection of Coffee Leaf Disease and Identification using Deep Learning*," *PeerJ Comput. SCI.*, Vol. 11, pp. 1–25, Oct. 2025, DOI: 10.7717/peerj-cs.3172.

[6] O. Adelaja and B. Pranggono, "*Leveraging Deep Learning for Real-Time Coffee Leaf Disease Identification*," *AgriEngineering*, Vol. 7, No. 1, Jan. 2025, DOI: 10.3390/agriengineering7010013.

[7] J. V. Y. Bordin Yamashita and J. P. R. R. Leite, "*Coffee Disease Classification at the Edge using Deep Learning*," *Smart Agricultural Technology*, Vol. 4, Aug. 2023, DOI: 10.1016/j.atech.2023.100183.

[8] R. T. Araaf, A. Minn, and T. Ahamed, "*Coffee Leaf Rust Disease Detection and Implementation of an Edge Device for Pruning Infected Leaves via Deep Learning Algorithms*," *Sensors*, Vol. 24, No. 24, pp. 1–20, Dec. 2024, DOI: 10.3390/s24248018.

[9] C. E. S. E. Silva, J. B. Fragoso, T. Paixao, A. B. Alvarez, and F. Palomino-Quispe, "*A Low Computational Cost Deep Learning Approach for Localization and Classification of Diseases and Pests in Coffee Leaves*," *IEEE Access*, Vol. 13, pp. 71943–71964, 2025, DOI: 10.1109/ACCESS.2025.3562832.

[10] J. Fragoso *et al.*, "*Coffee-Leaf Diseases and Pests Detection based on YOLO Models*," *Applied Sciences (Switzerland)*, Vol. 15, No. 9, May 2025, DOI: 10.3390/app15095040.

[11] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "*A Simple Framework for Contrastive Learning of Visual Representations*," pp. 1597–1607, Jul. 2020, [Online]. Available: http://arxiv.org/abs/2002.05709

[12] J. Parraga-Alava, K. Cusme, A. Elica Loor, and E. Santander, "*RoCoLe: A Robusta Coffee Leaf Images Dataset for Evaluation of Machine Learning based Methods in Plant Diseases Recognition*," 2019, DOI: 10.17632/c5yvn32dzg.2.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "*Deep Residual Learning for Image Recognition*." [Online]. Available: http://image-net.org/challenges/LSVRC/2015/

[14] N. Mansouri, H. Guesmi, and A. Alkhalil, "*A Deep Learning Model for Detection and Classification of Coffee-Leaf Diseases using the Transfer-Learning Technique*," *International Journal of Advances in Intelligent Informatics*, Vol. 10, No. 3, pp. 379–396, Aug. 2024, DOI: 10.26555/ijain.v10i3.1573.

[15] A. Chavarro, D. Renza, and E. Moya-Albor, "*ConvNext as a Basis for Interpretability in Coffee Leaf Rust Classification*," *Mathematics*, Vol. 12, No. 17, Sep. 2024, DOI: 10.3390/math12172668.

[16] Y. Wang *et al.*, "*Classification of Plant Leaf Disease Recognition based on Self-Supervised Learning*," *Agronomy*, Vol. 14, No. 3, Mar. 2024, DOI: 10.3390/agronomy14030500.

[17] X. Huan, B. Chen, and H. Zhou, "*A Unified Self-Supervised Framework for Plant Disease Detection on Laboratory and In-Field Images*," *Electronics (Switzerland)*, Vol. 14, No. 17, Sep. 2025, DOI: 10.3390/electronics14173410.

[18] S. S. A. Shah *et al.*, "*PlantCLR: Contrastive Self-Supervised Pretraining for Generalizable Plant Disease Detection*," *SCI. Rep.*, Vol. 16, No. 1, Dec. 2026, DOI: 10.1038/s41598-026-45684-x.