

PENANGANAN KETIDAKSEIMBANGAN KELAS PADA KLASIFIKASI PENYAKIT DAUN KENTANG MENGGUNAKAN DENSENET121: STUDI PERBANDINGAN CROSS ENTROPY DAN FOCAL LOSS

Daviq Ardli Wishesha, Henni Endah Wahanani, Andreas Nugroho Sihananto

Informatika, Fakultas Ilmu Komputer, UPN "VETERAN" Jawa Timur

henniendah.if@upnjatim.ac.id

ABSTRAK

Tanaman kentang merupakan komoditas pertanian strategis yang produktivitasnya kerap terganggu oleh berbagai penyakit daun, namun identifikasi secara manual masih bergantung pada keahlian ahli agronomi yang terbatas dan membutuhkan waktu lama. Permasalahan utama yang dihadapi adalah ketidakseimbangan kelas (class imbalance) pada dataset penyakit daun kentang di lingkungan tidak terkontrol, di mana kelas Nematode hanya memiliki 68 sampel (2,2%) berbanding kelas Fungi dengan 748 sampel (24,3%), sehingga model cenderung bias terhadap kelas mayoritas dan gagal mendeteksi penyakit langka yang justru krusial. Penelitian ini bertujuan membandingkan efektivitas fungsi loss Categorical Cross Entropy dan Focal Loss dalam menangani ketidakseimbangan kelas pada klasifikasi penyakit daun kentang menggunakan arsitektur DenseNet121 dengan transfer learning dari ImageNet. Dataset terdiri dari 3.076 gambar dalam 7 kelas dengan rasio pembagian 80:20, dilatih menggunakan optimizer Adam, augmentasi data pada subset training, serta mekanisme EarlyStopping dan ReduceLROnPlateau. Hasil evaluasi menunjukkan model Cross Entropy mencapai overall accuracy 68,63% dengan macro F1-Score 0,68, sedangkan Focal Loss mencapai 64,38% dengan macro F1-Score 0,66. Meskipun Cross Entropy unggul secara keseluruhan, Focal Loss terbukti lebih efektif pada kelas minoritas Nematode dengan F1-Score 0,75 berbanding 0,67 dan peningkatan Recall sebesar 15,4 poin, membuktikan bahwa pemilihan fungsi loss perlu disesuaikan dengan prioritas deteksi pada dataset yang tidak seimbang.

Kata kunci : *DenseNet121, Klasifikasi Penyakit Daun Kentang, Transfer Learning, Focal Loss, Categorical Cross Entropy, Deep Learning*

1. PENDAHULUAN

Sektor pertanian kentang menghadapi ancaman serius dari berbagai patogen penyakit daun yang secara langsung berdampak pada penurunan produktivitas dan ketahanan pangan global [1]. Deteksi otomatis penyakit daun kentang memiliki peran yang sangat krusial dalam dunia agrikultur nyata untuk mencegah penyebaran patogen dan mengamankan tingkat produksi pangan global [1]. Namun, implementasi teknologi di lapangan masih sangat bergantung pada inspeksi visual langsung oleh para ahli agronomi sebuah metode yang sangat memakan waktu, padat karya, dan rentan terhadap subjektivitas. Oleh karena itu, pengembangan sistem klasifikasi penyakit daun kentang berbasis kecerdasan buatan menjadi kebutuhan yang mendesak untuk memungkinkan intervensi pencegahan dini, pemantauan waktu nyata, serta meminimalisasi kerugian hasil panen secara masif [12].

Permasalahan fundamental yang dihadapi dalam pengembangan sistem deteksi berbasis Deep Learning untuk penyakit daun kentang adalah kondisi dataset alami yang tidak pernah terdistribusi secara merata, atau yang umumnya dikenal sebagai ketidakseimbangan kelas (class imbalance) [2]. Tantangan ini terlihat jelas pada dataset penyakit daun kentang yang mencakup 7 kelas, di mana terdapat disparitas yang ekstrem antara kelas yang sangat mendominasi seperti kelas Fungi (748 sampel), dan kelas yang sangat minoritas seperti kelas Nematode

(68 sampel). Dalam kondisi ketidakseimbangan data yang tinggi ini, model rentan mengalami bias karena algoritma akan cenderung menghafal pola dari kelas mayoritas dan mengabaikan kelas minoritas yang kurang terwakili dalam proses pelatihan [4]. Akibatnya, penggunaan metrik performa klasifikasi standar seperti akurasi global menjadi sangat menipu (deceptive), karena model seolah-olah menunjukkan tingkat keberhasilan yang tinggi secara keseluruhan padahal pada kenyataannya gagal mendeteksi sampel minoritas yang sering kali merupakan anomali atau penyakit yang krusial [5].

Untuk menyelesaikan permasalahan tersebut, penelitian ini mengusulkan dua pendekatan solusi yang diterapkan secara terintegrasi. Pertama, pada tingkat arsitektur, DenseNet-121 dipilih sebagai backbone ekstraksi fitur karena menawarkan keunggulan teoretis yang solid melalui konektivitas padat yang mampu mendorong penggunaan ulang fitur (feature reuse), meningkatkan efisiensi parameter, serta memitigasi masalah degradasi gradien (vanishing-gradient) [6], dikombinasikan dengan strategi transfer learning dari ImageNet untuk memaksimalkan kemampuan ekstraksi fitur pada kondisi keterbatasan data. Kedua, pada tingkat algoritma (algorithm-level approach), Focal Loss diusulkan sebagai pengganti fungsi Categorical Cross-Entropy (CCE) standar yang secara inheren memberikan bobot penalti sama rata untuk setiap sampel, sehingga menyebabkan gradien pembelajaran

didominasi oleh kelas mayoritas dan model menjadi malas mengenali kelas minoritas [8]. Focal Loss mengatasi kelemahan ini dengan secara dinamis menurunkan bobot penalti pada sampel kelas mayoritas yang sudah mudah ditebak (easy examples), sekaligus memaksa model untuk memusatkan atensinya pada sampel kelas minoritas yang sulit diidentifikasi (hard examples) [9]. Pendekatan CNN dengan optimasi Adam juga terbukti mampu meningkatkan performa klasifikasi citra dalam kondisi keterbatasan data, di mana penggunaan optimizer yang tepat terbukti lebih berpengaruh signifikan terhadap akurasi model dibandingkan rekayasa hyperparameter lainnya [15].

Sejumlah penelitian terdahulu telah menerapkan arsitektur CNN dan DenseNet untuk klasifikasi penyakit tanaman dengan hasil yang menjanjikan, seperti Ahmad & Alamsyah [7] yang mencapai akurasi 99,56% menggunakan DenseNet201 dan MobileNetV2, serta Nawaz et al. [1] yang mencapai 99,41% melalui kerangka kerja PotatoGuardNet berbasis Inception-ResNet-V2. Namun demikian, dari tinjauan studi sebelumnya, belum ada satupun kajian yang membandingkan secara komprehensif dan spesifik efektivitas intervensi tingkat algoritma menggunakan Focal Loss melawan Cross Entropy konvensional untuk menangani imbalance dataset ekstrem, khususnya pada pemetaan dataset tujuh kelas penyakit daun kentang. Kesenjangan penelitian (research gap) inilah yang menjadi landasan dan motivasi utama penelitian ini.

Bertolak dari landasan tersebut, penelitian ini bertujuan untuk melakukan komparasi langsung terhadap performa arsitektur DenseNet-121 yang dilatih menggunakan Cross-Entropy konvensional berbanding dengan fungsi Focal Loss pada dataset penyakit daun kentang 7 kelas berjumlah 3.076 gambar di lingkungan tidak terkontrol. Untuk menghindari bias dari evaluasi data yang timpang, pengukuran kinerja dititikberatkan pada metrik yang terbukti lebih kebal terhadap ketidakseimbangan kelas, yakni Precision, Recall, dan F1-Score untuk setiap kelas secara individual [10], dengan perhatian khusus pada kelas Nematode sebagai representasi kelas minoritas yang paling ekstrem. Harapannya, penelitian ini dapat memberikan kontribusi empiris yang menjembatani keterbatasan performa klasifikasi minoritas tanpa harus merombak dimensi ruang data aslinya, sekaligus menjadi panduan praktis bagi peneliti dalam memilih fungsi loss yang tepat sesuai karakteristik distribusi dataset penyakit tanaman di kondisi nyata.

2. TINJAUAN PUSTAKA

2.1. Penyakit Daun Kentang

Patologi daun kentang melibatkan berbagai infeksi yang secara visual dapat direpresentasikan melalui 7 (tujuh) kelas kondisi utama pada dataset agrikultur, yaitu: Bakteri, Fungi, Sehat, *Nematode*, Hama (*Pest*), *Phytophthora*, dan Virus. Setiap penyakit

atau agen patogen ini menghasilkan perubahan fisiologis dan tekstur yang kompleks pada struktur daun, yang sering kali sulit dibedakan dengan mata telanjang jika masih berada pada tahap awal infeksi [11]. Keberadaan penyakit tanaman ini merupakan salah satu ancaman paling signifikan terhadap ketahanan pangan global dan ekonomi agrikultur, yang dapat menyebabkan hilangnya persentase hasil panen secara masif [1]. Secara tradisional, diagnosis penyakit bergantung pada inspeksi visual secara langsung di lapangan oleh para ahli, sebuah metode yang sangat memakan waktu, padat karya, dan rentan terhadap subjektivitas. Oleh karena itu, mendeteksi patologi penyakit secara otomatis menggunakan sistem cerdas Computer Vision memiliki urgensi yang sangat krusial, karena memungkinkan intervensi pencegahan dini, pemantauan waktu nyata (*real-time*), dan meminimalisasi penggunaan pestisida yang berlebihan demi pertanian yang presisi dan berkelanjutan [12].

2.2. DenseNet121

Arsitektur *Densely Connected Convolutional Network (DenseNet)* secara radikal memodifikasi cara kerja jaringan saraf tiruan konvensional dengan mengimplementasikan prinsip konektivitas padat (*dense connectivity*) [6]. Dalam mekanisme *dense block* pada *DenseNet-121*, setiap lapisan dalam jaringan terhubung secara langsung dengan seluruh lapisan-lapisan berikutnya, sehingga memaksimalkan aliran informasi di seluruh jaringan. Struktur arsitektural ini secara teoretis mencegah terjadinya masalah degradasi gradien (*vanishing-gradient problem*) dengan memberikan jalur propagasi gradien langsung dari fungsi kerugian (*loss function*) ke lapisan-lapisan paling awal [6]. Keunggulan utama lainnya dari *DenseNet-121* untuk tugas klasifikasi citra adalah tingkat efisiensi parameternya yang tinggi serta dorongan kuat terhadap penggunaan ulang fitur (*feature reuse*), di mana peta fitur (*feature maps*) fundamental dari resolusi awal jaringan dipelihara tanpa modifikasi destruktif hingga mencapai lapisan klasifikasi, sehingga tidak kehilangan informasi patologis yang subtil [3].

2.3. Transfer Learning

Transfer Learning merupakan teknik pembelajaran mesin dengan mengeksploitasi arsitektur model jaringan yang sebelumnya telah dilatih (*pre-trained model*) pada dataset citra generik berskala masif, seperti *ImageNet* [9]. Dalam penerapannya untuk mengenali penyakit daun kentang, metode ini umumnya melibatkan skenario pembekuan (*freezing*) pada layer-layer awal jaringan, yang bertujuan untuk mempertahankan kemampuan model dalam mengekstraksi fitur visual generik seperti tekstur, tepi, dan bentuk dasar. Setelah itu, peneliti akan melakukan pelatihan ulang (*fine-tuning*) dengan membuka pembekuan pada layer-layer akhir, khususnya difokuskan pada 50 layer terakhir jaringan. Pelatihan ulang pada lapisan dalam jaringan ini memaksa

arsitektur untuk memfokuskan pembelajarannya pada ekstraksi fitur tingkat tinggi (*high-level features*) yang sangat spesifik terhadap patologi penyakit daun kentang, yang pada akhirnya meningkatkan keandalan model dan mengurangi waktu komputasi pelatihan [13].

2.4. Categorical Cross Entropy

Fungsi *Categorical Cross Entropy* (CCE) merupakan mekanisme fungsi kerugian (*loss function*) standar yang paling umum diaplikasikan dalam mengoptimalkan pembaruan bobot pada klasifikasi multikelas [14]. Secara matematis, CCE dapat diekspresikan dengan rumus berikut:

$$L_{CE} = -\frac{1}{N} \sum_{c=1}^C \sum_{i=1}^N y_{ic} \log(\hat{y}_{ic}) \quad (1)$$

Dalam formulasi tersebut, N merepresentasikan jumlah total sampel observasi, C merepresentasikan jumlah kelas secara keseluruhan, y_{ic} merupakan nilai kebenaran atau label (0 atau 1) yang mengindikasikan apakah sampel i benar merupakan kelas c , dan $\hat{y}_{ic}$ merupakan probabilitas prediksi yang dihasilkan oleh model untuk sampel tersebut [14].

Kelemahan utama dari fungsi CCE secara mendalam terlihat ketika fungsi ini dihadapkan pada dataset yang tidak seimbang (*imbalanced dataset*). CCE secara alami memberikan kontribusi nilai kerugian yang sama rata (*equal weight*) untuk semua sampel [3]. Pada distribusi data yang timpang, algoritma ini akan terus menerus mengakumulasi dan meminimalkan gradien kerugian dari kelas mayoritas karena jumlah sampelnya yang melimpah. Hal ini menyebabkan arah pembelajaran model secara keseluruhan akan didominasi oleh kelas mayoritas, yang pada akhirnya membuat model tersebut gagal belajar dan tidak mampu mengenali fitur krusial dari sampel kelas minoritas [3].

2.5. Focal Loss

Guna mengatasi keterbatasan fungsi objektif standar, *Focal Loss* diperkenalkan sebagai solusi murni pada tingkat algoritma (*algorithm-level approach*) yang berfungsi untuk menangani permasalahan ketidakseimbangan kelas [9]. Fungsi ini mengaplikasikan modifikasi matematis dari Cross Entropy dengan menambahkan faktor pembobot modulasi, yang secara lengkap dirumuskan sebagai berikut:

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (2)$$

Pada perumusan ini, p_t adalah probabilitas klasifikasi yang diprediksi oleh model untuk target kelas yang benar. Parameter α_t bertindak sebagai balancing factor yang mengontrol keseimbangan bobot kepatuhan antara kelas positif dan negatif. Sementara itu, parameter γ merupakan focusing parameter yang mengatur tingkat kecepatan

pelemahan fungsi kerugian berdasarkan probabilitas tebakan [9][3].

Fungsi kerugian ini bekerja dengan mengecilkan pengaruh dari faktor $(1-p_t)^\gamma$ ketika nilai prediksi p_t mendekati angka 1. Ini berarti Focal Loss secara dinamis akan sangat menurunkan bobot kontribusi nilai kerugian dari sampel kelas mayoritas yang umumnya mudah ditebak (*easy examples*) [9]. Akibatnya, reduksi bobot ini akan secara aktif memaksa jaringan saraf dalam untuk tidak lagi peduli pada kelas mayoritas yang sudah dikuasai, melainkan memusatkan seluruh fokus serta atensi pembelajarannya pada sampel kelas minoritas yang sering kali lebih sulit diklasifikasikan (*hard examples*) tanpa perlu merekayasa atau mendistorsi jumlah data input [3].

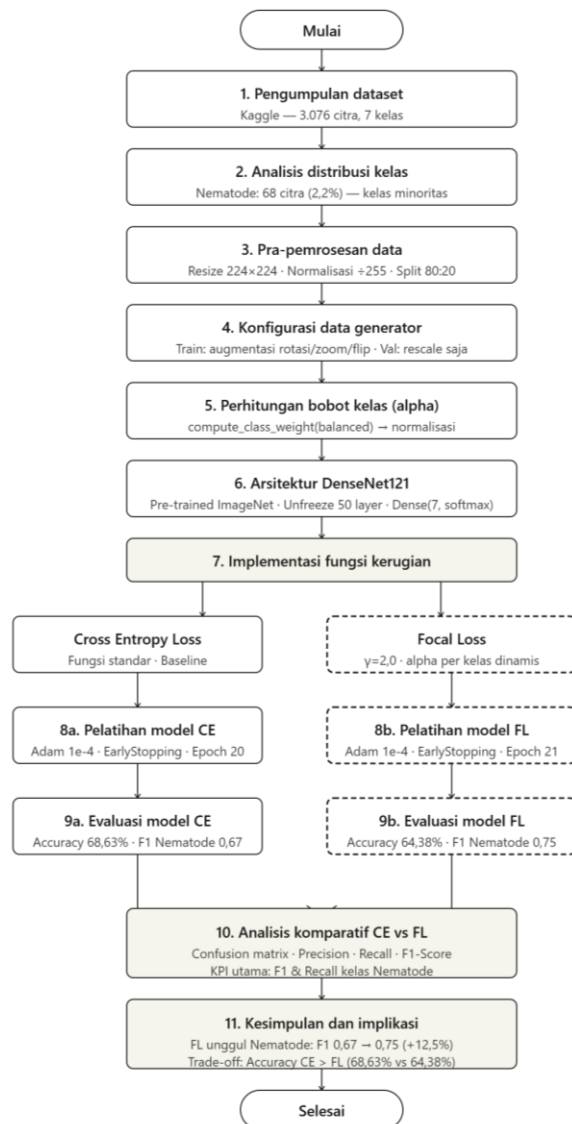
2.6. Penelitian Terkait

Sejumlah penelitian terbaru telah menerapkan jaringan saraf dalam dengan arsitektur CNN atau DenseNet untuk menyelesaikan permasalahan klasifikasi penyakit pada tanaman. Studi pada [7] mengembangkan model ensemble learning pada skala fitur (*feature-level fusion*) menggunakan kombinasi arsitektur DenseNet201 dan MobileNetV2 untuk mengklasifikasikan kondisi daun kentang, dan mencapai performa yang kuat dengan akurasi 99,56% menggunakan categorical cross-entropy. Di sisi lain, [1] memperkenalkan kerangka kerja PotatoGuardNet yang memanfaatkan Inception-ResNet-V2 sebagai arsitektur dasar (*base network*) pada Faster-RCNN guna mendeteksi sekaligus melokalisasi bercak penyakit pada daun kentang. Penggabungan region proposal network dengan ekstraksi fitur dalam tersebut menghasilkan akurasi yang menjanjikan sebesar 99,41%. Pada penelitian lainnya, [14] menangani masalah imbalanced data ekstrem pada deteksi penyakit daun tomat dengan memodifikasi Generative Adversarial Network (GAN) beserta penambahan pendekatan weighted loss function untuk kemudian diekstraksi ke arsitektur jaringan seperti DenseNet121 dan VGG19, dan mencapai pengenalan hingga 99,95%.

Walaupun berbagai metode mutakhir telah mendemonstrasikan efikasi arsitektur dalam seperti DenseNet dan strategi data augmentation tingkat lanjut (seperti GAN) guna meningkatkan akurasi model, masih terdapat sebuah kesenjangan penelitian (*research gap*) yang mendesak. Dari tinjauan studi sebelumnya, belum ada satupun kajian yang membandingkan secara komprehensif dan spesifik efektivitas intervensi tingkat algoritma menggunakan *Focal Loss* melawan *Cross Entropy* konvensional untuk menangani imbalance dataset ekstrem (seperti ketimpangan dominasi antara Fungsi dan kelas Nematode) khususnya pada pemetaan dataset tujuh kelas penyakit daun kentang. Pembuktian empiris ini menjadi sangat penting karena dapat menjembatani keterbatasan performa klasifikasi minoritas tanpa harus merombak dimensi ruang data aslinya.

3. METODE PENELITIAN

3.1. Alur Penelitian



Gambar 1. Flowchart diagram alur penelitian

Gambar 1 mengilustrasikan alur metodologi penelitian secara keseluruhan yang terdiri dari 11 tahapan. Penelitian diawali dengan pengumpulan dataset sebanyak 3.076 citra daun kentang dari Kaggle dalam 7 kelas, dilanjutkan dengan analisis distribusi kelas yang mengidentifikasi Nematode sebagai kelas minoritas dengan hanya 68 citra (2,2%). Setelah pra-pemrosesan berupa resize 224×224 piksel, normalisasi, dan pembagian data 80:20, dilakukan konfigurasi data generator dengan augmentasi hanya pada data training.

Bobot kelas (alpha) kemudian dihitung secara otomatis menggunakan fungsi `compute_class_weight (balanced)` yang dinormalisasi untuk digunakan sebagai parameter Focal Loss. Arsitektur DenseNet121 dengan bobot pretrained ImageNet dibangun dengan 50 layer terakhir yang dilatih ulang (fine-tuning) dan classification head Dense(7,

softmax). Pada tahap ke-7, alur bercabang menjadi dua jalur paralel garis utuh untuk model Cross Entropy (sebagai baseline) dan garis putus-putus untuk model Focal Loss yang masing-masing menjalani proses pelatihan dan evaluasi secara independen sebelum keduanya digabungkan kembali pada tahap analisis komparatif (tahap 10) menggunakan confusion matrix, Precision, Recall, dan F1-Score dengan fokus utama pada kelas Nematode, dan diakhiri dengan penarikan kesimpulan mengenai trade-off antara kedua fungsi loss.

3.2. Dataset

Dataset yang digunakan adalah Potato Leaf Disease Dataset in Uncontrolled Environment berisi 3.076 gambar dalam 7 kelas. Distribusi kelas ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Kelas Dataset Penyakit Daun Kentang

Kelas	Jumlah	Persentase
Bacteria	569	18,5%
Fungi	748	24,3%
Healthy	201	6,5%
Nematode	68	2,2%
Pest	611	19,9%
Phytophthora	347	11,3%
Virus	532	17,3%
Total	3.076	100%

Dari Tabel 1, kelas Nematode memiliki sampel paling sedikit (68 gambar, 2,2%) dibandingkan Fungi (748 gambar, 24,3%). Ketidakseimbangan ini menjadi motivasi penggunaan Focal Loss.

3.3. Preprocessing dan Augmentasi Data

Seluruh gambar diubah ukurannya menjadi 224×224 piksel dan dinormalisasi dengan membagi nilai pixel dengan 255. Augmentasi (rotasi 20°, zoom 20%, horizontal flip) diterapkan hanya pada data training. Dataset dibagi dengan rasio 80:20 menghasilkan 2.464 gambar training dan 612 gambar validasi.

3.4. Arsitektur Model

DenseNet121 diinisialisasi dengan bobot pretrained ImageNet. Seluruh layer dibekukan kecuali 50 layer terakhir yang dilatih ulang. Classification head terdiri dari Global Average Pooling 2D dan Dense layer dengan 7 neuron keluaran (softmax).

3.5. Penanganan Ketidakseimbangan Kelas

Alpha per kelas pada Focal Loss dihitung dengan `class_weight.compute_class_weight (mode balanced)` kemudian dinormalisasi (total alpha = 1). Nilai gamma = 2,0 digunakan sesuai rekomendasi Sun et al. [8]. Kelas Nematode mendapat bobot alpha tertinggi karena jumlah sampelnya paling sedikit.

3.6. Konfigurasi Training

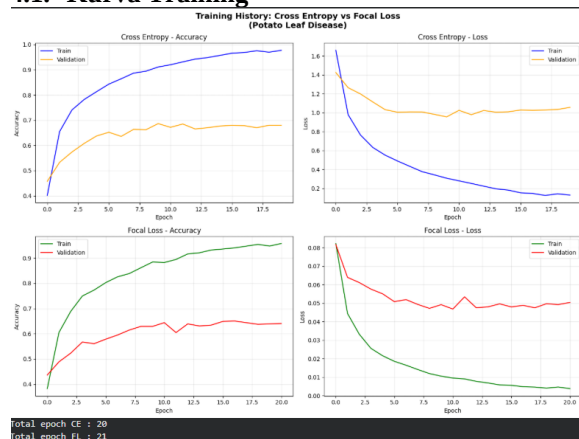
Kedua model dilatih dengan konfigurasi identik: optimizer Adam ($\text{lr} = 1 \times 10^{-4}$), maks 50 epoch, batch size 32. Callback: EarlyStopping (patience=10) dan ReduceLROnPlateau (factor=0,3, patience=5, min_lr= 1×10^{-7}). Model CE berhenti pada epoch 20, model FL pada epoch 21.

3.7. Evaluasi

Evaluasi menggunakan Accuracy, Precision, Recall, dan F1-Score per kelas. Perhatian khusus diberikan pada kelas Nematode untuk mengukur efektivitas Focal Loss terhadap kelas minoritas.

4. HASIL DAN PEMBAHASAN

4.1. Kurva Training



Gambar 2. Kurva Training History: Cross Entropy vs Focal Loss

Gambar 2 menampilkan kurva training accuracy, validation accuracy, training loss, dan validation loss untuk kedua model selama proses pelatihan. Model Cross Entropy berhenti pada epoch ke-20 dan model Focal Loss pada epoch ke-21, keduanya dihentikan oleh mekanisme EarlyStopping dengan patience = 10.

Model Cross Entropy menunjukkan peningkatan training accuracy yang konsisten dari $\sim 0,40$ pada epoch awal hingga mendekati 1,0 pada epoch terakhir. Sebaliknya, validation accuracy hanya meningkat hingga $\sim 0,68$ – $0,70$ dan cenderung stagnan sejak epoch ke-10. Pola yang sama terlihat pada kurva loss: training loss turun signifikan dari $\sim 1,6$ menjadi $\sim 0,15$, sementara validation loss turun hingga titik tertentu kemudian stagnan di kisaran $\sim 0,9$ – $1,0$. Celah (gap) yang lebar antara kurva training dan validation pada kedua metrik ini merupakan indikasi kuat terjadinya overfitting, di mana model terlalu menyesuaikan diri terhadap data latih namun gagal melakukan generalisasi dengan baik pada data yang belum pernah dilihat.

Model Focal Loss menunjukkan pola serupa namun dengan karakteristik yang berbeda. Training accuracy meningkat dari $\sim 0,40$ hingga $\sim 0,95$, sementara validation accuracy konvergen di $\sim 0,64$ – $0,66$. Pada kurva loss, nilai training loss turun dari $\sim 0,08$ menuju mendekati 0,001 skala yang jauh lebih

kecil dibandingkan Cross Entropy karena Focal Loss secara desain menekan bobot loss pada sampel yang sudah mudah diklasifikasikan. Validation loss turun hingga $\sim 0,045$ – $0,050$ lalu stagnan. Gap antara training dan validation juga mengindikasikan overfitting, meskipun tren validation loss-nya relatif lebih stabil dibandingkan Cross Entropy.

Fenomena overfitting pada kedua model ini kemungkinan disebabkan oleh keterbatasan jumlah data, khususnya pada kelas minoritas seperti Nematode (68 sampel). Meskipun augmentasi data telah diterapkan, variasi sampel pada kelas langka tetap terbatas sehingga model cenderung menghafalkan pola data latih. Mekanisme ReduceLROnPlateau berhasil memperlambat divergensi, namun tidak sepenuhnya mampu mengatasi overfitting tanpa penambahan data yang lebih representatif..

4.2. Perbandingan Evaluasi Per Kelas

Tabel 2 menyajikan hasil evaluasi lengkap kedua model. Secara keseluruhan, model CE mencapai overall accuracy 68,63% dan macro F1-Score 0,68, sedangkan model FL mencapai 64,38% dan 0,66.

Tabel 2. Perbandingan Hasil Evaluasi Per Kelas

Kelas	Model	Precision	Recall	F1-Score
Bacteria	CE	0,9383	0,6726	0,7835
	FL	0,9241	0,6460	0,7604
Fungi	CE	0,6628	0,7651	0,7103
	FL	0,6471	0,6644	0,6556
Healthy	CE	0,4217	0,8750	0,5691
	FL	0,4146	0,8500	0,5574
Nematode*	CE	0,8750	0,5385	0,6667
	FL	0,8182	0,6923	0,7500
Pest	CE	0,6535	0,5410	0,5919
	FL	0,5253	0,4262	0,4706
Phytophthora	CE	0,7879	0,7536	0,7704
	FL	0,7067	0,7681	0,7361
Virus	CE	0,6931	0,6604	0,6763
	FL	0,6549	0,6981	0,6758
Overall Acc	CE	—	—	0,6863
	FL	—	—	0,6438

Kelas minoritas (hanya 68 sampel). Baris FL pada Nematode menunjukkan peningkatan F1-Score (0,75 vs 0,67) dibandingkan CE.

4.3. Analisis Kelas Minoritas: Nematode

Kelas Nematode merupakan kelas dengan jumlah sampel paling sedikit dalam dataset, yakni hanya 68 gambar (2,2% dari total 3.076 gambar). Ketimpangan ini menjadikan Nematode sebagai kelas yang paling rentan terhadap bias model akibat ketidakseimbangan kelas (class imbalance). Pada data validasi, kelas ini hanya diwakili oleh 13 sampel, sehingga setiap kesalahan prediksi memiliki dampak yang besar terhadap nilai metrik.

Hasil evaluasi menunjukkan perbedaan yang cukup signifikan antara kedua model pada kelas ini. Model Cross Entropy menghasilkan Precision 0,875

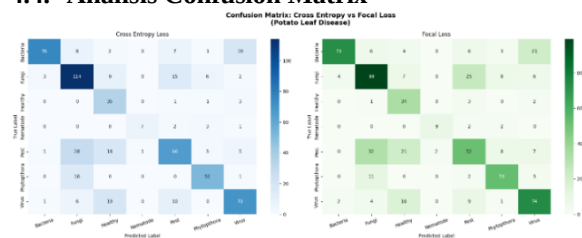
namun Recall hanya 0,5385, menghasilkan F1-Score sebesar 0,6667. Nilai Recall yang rendah ini berarti model CE hanya mampu mendeteksi dengan benar sekitar 7 dari 13 sampel Nematode pada data validasi 6 sampel lainnya salah diklasifikasikan sebagai kelas lain. Precision yang tinggi (0,875) menunjukkan bahwa ketika model CE memprediksi suatu gambar sebagai Nematode, prediksinya cenderung benar, namun model ini terlalu "ragu-ragu" untuk memberikan label Nematode sehingga banyak sampel positif yang terlewat (false negative tinggi).

Model Focal Loss memberikan hasil yang berbeda. Meskipun Precision sedikit turun menjadi 0,8182, Recall meningkat signifikan menjadi 0,6923, sehingga F1-Score naik menjadi 0,75 peningkatan sebesar 0,083 poin atau sekitar 12,5% relatif terhadap CE. Secara absolut, model FL berhasil mendeteksi sekitar 9 dari 13 sampel Nematode dengan benar, dua sampel lebih banyak dibandingkan model CE.

Peningkatan ini merupakan dampak langsung dari mekanisme pembobotan alpha pada Focal Loss. Karena nilai alpha untuk kelas Nematode dihitung secara otomatis berbanding terbalik dengan frekuensinya dalam dataset, kelas ini mendapat bobot alpha tertinggi di antara semua kelas. Akibatnya, setiap kesalahan prediksi pada sampel Nematode memberikan kontribusi loss yang jauh lebih besar selama pelatihan, memaksa model untuk lebih "memperhatikan" kelas ini. Selain itu, faktor $(1 - p_t)^\gamma$ dengan $\gamma = 2,0$ turut memperkuat efek ini dengan memberikan penalti lebih besar pada sampel Nematode yang diprediksi dengan keyakinan rendah.

Penurunan Precision dari 0,875 menjadi 0,8182 pada model FL adalah konsekuensi yang wajar dari peningkatan Recall model menjadi lebih sensitif terhadap kelas Nematode sehingga lebih berani memberikan label Nematode, yang pada gilirannya meningkatkan kemungkinan false positive. Dalam konteks diagnosis penyakit tanaman, trade-off ini umumnya dapat diterima karena biaya dari false negative (gagal mendeteksi penyakit yang sebenarnya ada) jauh lebih merugikan secara agronomis dibandingkan false positive (mengira daun sehat sebagai penyakit).

4.4. Analisis Confusion Matrix



Gambar 3. Confusion Matrix: Cross Entropy vs Focal Loss

Gambar 3 menampilkan confusion matrix kedua model. Pada model CE, Bacteria memiliki true positive tertinggi (76/113). Kelas Pest menjadi kelas terlemah pada kedua model (CE F1: 0,59; FL F1: 0,47)

karena kemiripan visual dengan kelas Phytophthora dan Virus. Kelas Healthy menunjukkan precision rendah pada kedua model akibat ketidakseimbangan dan kesamaan visual dengan kondisi daun terserang awal.

4.5. Pembahasan

Hasil evaluasi secara keseluruhan menunjukkan bahwa model Cross Entropy mengungguli Focal Loss dalam dua metrik utama, yaitu overall accuracy (68,63% vs 64,38%) dan macro F1-Score (0,68 vs 0,66). Namun demikian, interpretasi terhadap perbedaan ini tidak dapat dilakukan secara dangkal, mengingat karakteristik dataset yang sangat tidak seimbang antar kelas. Perbedaan overall accuracy sebesar 4,25 poin dan macro F1-Score sebesar 0,02 poin dapat sangat dipengaruhi oleh dominasi kelas mayoritas seperti Fungi (748 sampel) dan Pest (611 sampel) terhadap perhitungan rata-rata, sehingga diperlukan analisis yang lebih menyeluruh per kelas untuk memperoleh kesimpulan yang adil.

4.6. Performa pada Kelas Mayoritas

Pada kelas-kelas dengan jumlah sampel besar, Cross Entropy secara konsisten mengungguli Focal Loss. Kelas Bacteria mencatat F1-Score 0,7835 (CE) vs 0,7604 (FL), Fungi 0,7103 vs 0,6556, dan Pest 0,5919 vs 0,4706. Pola ini dapat dijelaskan oleh sifat dasar Cross Entropy yang memberikan bobot loss yang proporsional terhadap frekuensi kemunculan sampel. Karena kelas mayoritas mendominasi jumlah data latih, model CE secara natural lebih teroptimasi untuk mengenali pola pada kelas-kelas tersebut.

Sebaliknya, mekanisme down-weighting pada Focal Loss di mana sampel yang sudah mudah diklasifikasikan mendapat kontribusi loss yang ditekan melalui faktor $(1 - p_t)^\gamma$ justru menjadi bumerang pada kelas mayoritas. Sampel-sampel dari kelas Pest dan Fungi yang umumnya mudah dikenali oleh model mendapat bobot loss lebih kecil, sehingga model FL "kurang belajar" dari kelas-kelas ini dibandingkan model CE. Penurunan F1-Score yang paling mencolok terjadi pada kelas Pest (-0,1213), mengkonfirmasi bahwa kelas dengan representasi besar namun karakteristik visual yang beririsan dengan kelas lain paling terdampak oleh mekanisme Focal Loss.

4.7. Performa pada Kelas Minoritas

Situasi berbalik ketika analisis difokuskan pada kelas dengan sampel sedikit. Kelas Nematode kelas paling langka dengan hanya 68 sampel (2,2%) menunjukkan keunggulan FL yang signifikan: F1-Score 0,75 vs 0,67, dengan peningkatan Recall dari 0,5385 menjadi 0,6923. Kelas Virus juga menunjukkan keunggulan FL pada Recall (0,6981 vs 0,6604) meskipun F1-Score akhirnya setara. Hal ini konsisten dengan desain Focal Loss yang memberikan alpha lebih tinggi pada kelas minoritas, memaksa model untuk lebih sensitif terhadap sampel yang jarang muncul selama pelatihan.

Perlu digarisbawahi bahwa peningkatan Recall pada kelas Nematode diiringi penurunan Precision yang moderat ($0,8750 \rightarrow 0,8182$). Ini adalah trade-off yang inheren: dengan menjadi lebih sensitif terhadap Nematode, model FL juga menjadi lebih "berani" memberikan label Nematode pada sampel yang ambigu, sehingga tingkat false positive sedikit meningkat. Dalam konteks aplikasi di lapangan, trade-off ini umumnya dapat diterima karena kerugian akibat gagal mendeteksi penyakit (false negative) jauh lebih besar secara ekonomi dan agronomis dibandingkan salah mendiagnosis daun sehat sebagai terserang Nematode.

4.8. Fenomena Overfitting dan Implikasinya

Kedua model mengalami overfitting yang terlihat jelas dari kurva training (Gambar 1), di mana training accuracy mendekati 1,0 namun validation accuracy stagnan jauh di bawahnya. Fenomena ini secara langsung membatasi potensi generalisasi kedua model. Overfitting yang lebih parah pada model CE ditunjukkan oleh gap yang lebih lebar antara training loss ($\sim 0,15$) dan validation loss ($\sim 1,0$) turut berkontribusi pada penurunan kemampuannya mengenali kelas minoritas, karena model terlalu "menghafal" pola kelas mayoritas yang dominan dalam data latih.

Model FL, meskipun juga mengalami overfitting, menunjukkan validation loss yang lebih stabil (stagnan di $\sim 0,05$) berkat mekanisme pembobotan yang memaksa model untuk tidak mengabaikan kelas langka. Ini menjadi salah satu keunggulan implisit Focal Loss pada dataset tidak seimbang: model tetap "mengingat" kelas minoritas meskipun frekuensi kemunculannya sangat rendah.

4.9. Implikasi Pemilihan Fungsi Loss

Berdasarkan analisis di atas, pemilihan antara Cross Entropy dan Focal Loss sebaiknya disesuaikan dengan prioritas aplikasi. Jika sistem ditujukan untuk skrining awal penyakit kentang secara umum di mana akurasi keseluruhan menjadi tolok ukur utama, Cross Entropy merupakan pilihan yang lebih tepat. Namun jika sistem diutamakan untuk mendeteksi penyakit langka seperti infeksi Nematode yang justru sering kali membutuhkan penanganan paling segera Focal Loss memberikan keunggulan yang tidak dapat diabaikan.

Untuk penelitian selanjutnya, penggunaan Focal Loss dapat dikombinasikan dengan teknik oversampling seperti SMOTE atau augmentasi berbasis generative model (misalnya GAN) khusus untuk kelas minoritas, sehingga keunggulan Focal Loss pada kelas langka dapat dipertahankan sekaligus mengurangi dampak negatifnya pada kelas mayoritas.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan sistem klasifikasi penyakit daun kentang menggunakan arsitektur DenseNet121 dengan transfer learning dari ImageNet pada dataset 7 kelas (Bacteria, Fungi,

Healthy, Nematode, Pest, Phytophthora, dan Virus) berjumlah 3.076 gambar yang diambil di lingkungan tidak terkontrol.

Perbandingan antara Categorical Cross Entropy dan Focal Loss menghasilkan temuan yang tidak bersifat absolut: Cross Entropy unggul secara keseluruhan dengan overall accuracy 68,63% dan macro F1-Score 0,68, namun keunggulan ini didominasi oleh performa superior pada kelas mayoritas seperti Bacteria, Fungi, dan Phytophthora. Sebaliknya, Focal Loss terbukti lebih efektif dalam menangani ketidakseimbangan kelas, khususnya pada kelas Nematode yang hanya memiliki 68 sampel F1-Score meningkat dari 0,67 menjadi 0,75 dengan lonjakan Recall sebesar 15,4 poin, yang dalam konteks diagnosis penyakit tanaman memiliki nilai praktis tinggi karena kegagalan mendeteksi penyakit langka berdampak lebih besar secara agronomis dibandingkan kesalahan pada kelas mayoritas.

Dengan demikian, pemilihan fungsi loss sebaiknya disesuaikan dengan prioritas sistem: Cross Entropy untuk optimasi akurasi umum, dan Focal Loss untuk sistem yang mengutamakan sensitivitas terhadap kelas minoritas. Untuk penelitian selanjutnya, disarankan mengeksplorasi kombinasi Focal Loss dengan teknik oversampling seperti SMOTE atau augmentasi berbasis generative model pada kelas langka, penerapan strategi fine-tuning bertahap dengan jumlah layer yang lebih bervariasi, serta pengujian pada arsitektur backbone alternatif seperti EfficientNet atau MobileNetV3 guna memperoleh keseimbangan yang lebih baik antara akurasi keseluruhan dan sensitivitas terhadap kelas dengan representasi data yang sangat terbatas.

DAFTAR PUSTAKA

- [1] M. Nawaz, A. Javed, and A. K. J. Saudagar, "PotatoGuardNet: a refined deep learning framework for potato leaf disease detection," *Front. Plant Sci.*, vol. 17, 2026, doi: 10.3389/fpls.2026.1720276.
- [2] L. Bi and G. Hu, "Improving Image-Based Plant Disease Classification With Generative Adversarial Network Under Limited Training Set," *Front. Plant Sci.*, vol. 11, Dec. 2020, doi: 10.3389/fpls.2020.583438.
- [3] M. Saini and S. Susan, "Tackling class imbalance in computer vision: a contemporary review," *Artif. Intell. Rev.*, vol. 56, pp. 1279–1335, Oct. 2023, doi: 10.1007/s10462-023-10557-6.
- [4] R. Lumbantoruan, N. Rajagukguk, A. U. Lubis, M. Claudia, and H. Simanjuntak, "Two-step convolutional neural network classification of plant disease," *IAES International Journal of Artificial Intelligence*, vol. 14, no. 1, pp. 584–591, Feb. 2025, doi: 10.11591/ijai.v14.i1.pp584-591.
- [5] Y. Fan, M. Yu, L. Shen, J. Ma, Z. Zeng, and H. Wang, "Robust plant disease segmentation in complex field environments: an in-depth analysis

- and validation with STAR-Net,” *Front. Plant Sci.*, vol. 16, 2026, doi: 10.3389/fpls.2025.1706072.
- [6] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” Jan. 2018, [Online]. Available: <http://arxiv.org/abs/1608.06993>
- [7] B. Ahmad, “Journal of Information System Exploration and Research Ensemble Learning-based Potato Leaf Disease Classification Using DenseNet201 and MobileNetV2 Article Info,” *Journal of Information System Exploration and Research*, vol. 4, no. 1, pp. 1–8, 2026, doi: 10.52465/joiser.v4i1.597.
- [8] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, “Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation,” Nov. 2021, [Online]. Available: <http://arxiv.org/abs/2102.04525>
- [9] L. S. Shabadi and L. M. Mathews, “A comprehensive survey of deep learning methods for Capsicum (bell pepper) leaf disease detection,” Nov. 06, 2025, *Horizon e-Publishing Group*. doi: 10.14719/pst.10384.
- [10] T. D. Salka, M. B. Hanafi, S. M. S. A. A. Rahman, D. B. M. Zulperi, and Z. Omar, “Plant leaf disease detection and classification using convolution neural networks model: a review,” *Artif. Intell. Rev.*, vol. 58, no. 10, Oct. 2025, doi: 10.1007/s10462-025-11234-6.
- [11] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” *Front. Plant Sci.*, vol. 7, no. September, Sep. 2016, doi: 10.3389/fpls.2016.01419.
- [12] A. K. Mohammed, Y. A. Nayyef, and H. N. Tuama, “Real-Time Plant Disease Detection by AI,” *European Journal of Ecology, Biology and Agriculture*, vol. 2, no. 4, pp. 52–69, Jul. 2025, doi: 10.59324/ejeba.2025.2(4).05.
- [13] S. M. Hassan, A. K. Maji, M. Jasiński, Z. Leonowicz, and E. Jasińska, “Identification of plant-leaf diseases using cnn and transfer-learning approach,” *Electronics (Switzerland)*, vol. 10, no. 12, Jun. 2021, doi: 10.3390/electronics10121388.
- [14] C. Ravi and S. Gaddadevara Matt, “IC-CGAN: Imbalanced class-conditional generative adversarial network with weighted loss function,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 2, p. 1632, Apr. 2025, doi: 10.11591/ijece.v15i2.pp1632-1646.
- [15] N. E. Imandayanti, H. E. Wahanani, and A. M. Rizki, “Klasifikasi Jenis Wayang menggunakan Convolutional Neural Network (CNN) dan Optimasi Adaptive Moment Estimation (ADAM),” *KERNEL: Jurnal Riset Inovasi Bidang Informatika dan Pendidikan Informatika*, vol. 5, no. 2, pp. 88–96, Dec. 2024, doi: 10.31284/j.kernel.2024.v5i2.6862