

Article history:
Received: May 28, 2025
Published: Dec 7, 2025

KLASIFIKASI PENYAKIT PADA TANAMAN DAUN SINGKONG MENGUNAKAN VISION TRANSFORMER

¹⁾Luthfi Rakan Nabila ²⁾ Aditya Dwi Putro Wicaksono

^{1,2)} Fakultas Informatika, Universitas Telkom

email: luthfrakan@student.telkomuniversity.ac.id, adityaw@telkomuniversity.ac.id

Abstract

This research develops a cassava leaf disease classification model using Vision Transformer (ViT) to identify four types of diseases and healthy leaves. With a dataset from Kaggle (3,000 images/class), the TinyViT model was tested through parameter variations to achieve optimal performance. Results showed that the combination of SGD, 50 epochs, and batch size 32 gave the highest validation accuracy (83.16%), outperforming Adam/AdamW. Despite overfitting (100% training accuracy), the model showed good generalization with 81% precision and recall. These findings confirm the potential of ViT in plant disease detection, while highlighting the need to address overfitting through further regularization. Future research can explore dataset expansion and fine-tuning for accuracy improvement.

Keywords: Cassava Leaf Disease, Image Classification, Stochastic Gradient Descent, Vision Transformer.

Abstrak

Penelitian ini mengembangkan model klasifikasi penyakit daun singkong menggunakan Vision Transformer (ViT) untuk mengidentifikasi empat jenis penyakit dan daun sehat. Dengan *dataset* dari Kaggle (3.000 gambar/kelas), model TinyViT diuji melalui variasi parameter untuk mencapai performa optimal. Hasil menunjukkan kombinasi SGD, 50 epoch, dan batch size 32 memberikan akurasi validasi tertinggi (83.16%), mengungguli Adam/AdamW. Meski mengalami *overfitting* (akurasi pelatihan 100%), model menunjukkan generalisasi yang baik dengan nilai *precision* dan *recall* 81%. Temuan ini memperlihatkan potensi ViT dalam tugas deteksi penyakit tanaman, sekaligus menyoroti perlunya penanganan *overfitting* melalui regularisasi lebih lanjut. Penelitian selanjutnya dapat mengeksplorasi perluasan dataset dan *fine-tuning* untuk peningkatan akurasi.

Kata Kunci: Klasifikasi Gambar, Penyakit Daun Singkong, Stochastic Gradient Descent, Vision Transformer

1. INTRODUCTION

Indonesia terletak di posisi strategis dengan iklim tropis yang hangat dan lembap, sehingga cocok untuk tumbuh beragam jenis tumbuhan. Di Kalimantan, hampir setengah dari total wilayahnya dimanfaatkan sebagai lahan perkebunan, sedangkan sekitar 50 juta hektar areanya terdiri dari hutan. Kondisi ini membuat Kalimantan menjadi salah satu wilayah penghasil oksigen terpenting di dunia. [1] Iklim ini membuat Indonesia cocok menjadi negara dengan focus usaha pertanian. Salah satu komoditas yang sering dibudidayakan dalam pertanian Indonesia adalah tanaman singkong. Tanaman singkong telah banyak diproses menjadi aneka produk olahan yang memberikan nilai ekonomi lebih besar. [2]. Singkong (*Manihot esculenta* Crantz) adalah

tanaman asli kawasan Amazon Selatan, Brasil, yang kini menjadi sumber pangan utama bagi lebih dari 800 juta orang. Sebagai komoditas pertanian strategis, singkong memiliki ketahanan tinggi terhadap kekeringan sehingga dijuluki "tanaman penyelamat saat krisis pangan". Produksi globalnya melonjak dari 124 juta ton (1980) menjadi 278 juta ton (2018), menempatkannya sebagai salah satu komoditas pangan. [3].

Permintaan singkong di Indonesia menunjukkan tren peningkatan tahunan, didorong oleh berkembangnya industri kecil berbasis olahan singkong. Kondisi ini menjadikan pasokan bahan baku singkong sebagai faktor krusial dalam menopang keberlangsungan usaha mikro yang menjadi tulang punggung perekonomian banyak masyarakat. [4].

Laporan Global Hunger Index (GHI) 2022 menempatkan Indonesia pada peringkat ke-77 dari 127 negara, dengan skor 16.9 yang mengindikasikan tingkat kelaparan moderat. Meskipun berbagai upaya penguatan ketahanan pangan telah dilakukan, isu aksesibilitas, distribusi, dan kualitas gizi tetap menjadi tantangan struktural. [5]. Fakta yang cukup ironis adalah posisi Indonesia sebagai negara dengan tingkat kelaparan tertinggi kedua di Asia Tenggara, hanya di bawah Timor Leste, padahal potensi agrarisnya sangat besar. Kondisi ini memunculkan paradoks: sebagai negara yang mengandalkan sektor pertanian dengan jumlah petani signifikan, Indonesia seharusnya mampu mencapai kemandirian pangan yang lebih baik.

Tak berbeda dengan komoditas pertanian lainnya, tanaman singkong memiliki kerentanan terhadap infeksi penyakit dan serangan organisme pengganggu tumbuhan. [6]. Penyakit yang umumnya mengganggu tanaman ini adalah *Cassava Bacterial Blight*, *Cassava Brown Streak*, *Cassava Green Mottle* dan *penyakit Cassava Mosaic Disease*. Sebagai komoditas vital, singkong perlu perlindungan dari penyakit. Namun, deteksi penyakitnya sulit karena gejalanya yang mirip, pemeriksaan lab mahal dan lama, serta fasilitas terbatas. Maka dari itu diperlukan metode identifikasi yang lebih cepat dan murah. [2].

Salah satu bukti pesatnya perkembangan teknologi saat ini adalah adanya AI atau kecerdasan buatan. Implementasi pada umumnya adalah penerapan suatu algoritma pada masalah khusus.

2. LITERATURE REVIEW

Deteksi manual penyakit daun singkong kurang efisien karena memakan banyak waktu, tenaga, dan biaya - terutama di lahan luas dengan tanaman pada yang membutuhkan pemeriksaan teliti. Karena itu, dibutuhkan sistem otomatis yang fleksibel untuk klasifikasi penyakit secara akurat.

Vision Transformer (ViT)

Implementasi pada umumnya adalah penerapan suatu algoritma pada masalah khusus. Salah satu algoritma yang bisa diterapkan adalah Vision Transformer. Vision Transformer (ViT), arsitektur berbasis Transformer untuk pemrosesan citra sudah terbukti efektif untuk analisis gambar termasuk salah satunya yaitu pemantauan

tanaman. (Figo, Yudistira, and Widodo 2023). Implementasi Vision Transformer (ViT) dalam pertanian memungkinkan deteksi dini penyakit tanaman singkong, yang mana berguna untuk meningkatkan produktivitas.

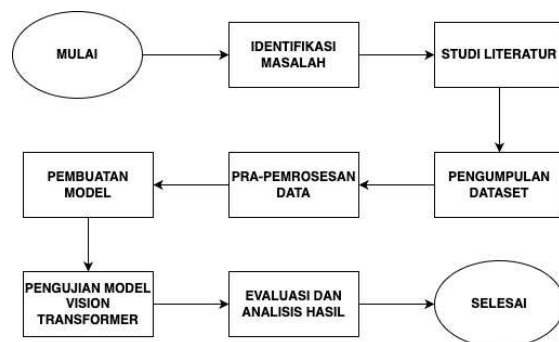
Penelitian Terdahulu

Terdapat beberapa penelitian yang mendukung penelitian ini. Di antaranya yaitu penelitian [8] yang di mana mereka membandingkan ViT dan CNN untuk klasifikasi 4 tingkat kematangan pisang. Hasilnya, ViT-L/16 mencapai akurasi 91.61% (lebih tinggi dari CNN 74.49%), dengan ViT unggul dalam generalisasi sementara CNN lebih efisien secara komputasi. Studi menggunakan *dataset* primer pisang Cavendish dan Ambon.

Kemudian penelitian yang dilakukan oleh [9] di mana mereka mengimplementasikan Vision Transformer (ViT) untuk klasifikasi gender menggunakan *dataset* AFAD (165.432 gambar wajah). Hasilnya menunjukkan akurasi hingga 98,43% pada resolusi 224×224 piksel dengan 28 patch. Meski belum mengungguli metode terkini, penelitian ini memberikan dasar berharga untuk pengembangan sistem klasifikasi gender yang lebih akurat di masa depan.

3. Method

Penelitian ini dilakukan melalui beberapa tahapan metodologis yang sistematis. Untuk mempermudah pemahaman, alur penelitian tersebut disajikan dalam bentuk diagram berikut:



Gambar 1. Diagram Alir Penelitian

Rumusan Masalah

Tahap pertama yang dilakukan penulis dalam melakukan penelitian ini adalah mengidentifikasi rumusan masalah yang terjadi pada industri pertanian di Indonesia, yaitu terkait penyakit yang menyerang tanaman singkong serta Solusi terhadap efisiensi waktu dan tenaga yang bisa

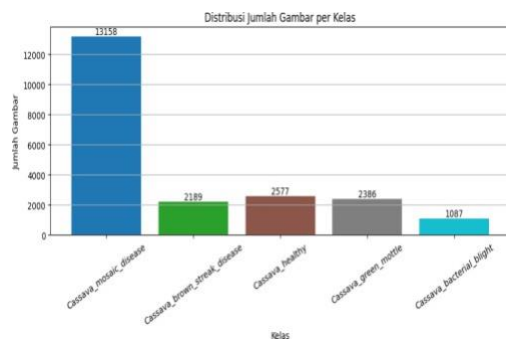
diterapkan dalam mendiagnosis penyakit pada tanaman singkong. Penulis kemudian menetapkan tujuan dan manfaat yang bisa didapat dari penelitian ini.

Kajian Literatur

Lalu selanjutnya adalah Kajian literatur yang di dalamnya meliputi Vision Transformer, teknik klasifikasi citra, dan patologi singkong, yang menjadi dasar pengembangan sistem diagnosis penyakit daun singkong. Studi terhadap penelitian serupa juga memberikan referensi metode evaluasi yang relevan.

Pengumpulan Data

Tahap berikutnya adalah pengumpulan dataset. Data yang berhasil dikumpulkan penulis pada penelitian ini merupakan jenis *dataset* sekunder. Penulis mengumpulkan data dari situs Kaggle dengan judul *Cassava Leaf Disease Classification*. Seluruhnya terbagi ke dalam 5 kelas, yaitu : *Cassava Bacterial Blight*, *Cassava Brown Streak Disease*, *Cassava Green Mottle*, *Cassava Mosaic Disease*, dan *Healthy*.



Gambar 2. Distribusi Data

Berikut ini adalah tampilan data citra dari *dataset* yang digunakan pada penelitian ini:



Gambar 3. Penyakit Cassava Bacterial Blight



Gambar 4. Penyakit Cassava Brown Streak



Gambar 5. Penyakit Cassava Green Mottle



Gambar 6. Penyakit Cassava Mosaic



Gambar 7. Daun Singkong yang sehat

Pra-pemrosesan Data

Tahap berikutnya adalah pra-pemrosesan data. Tahap pra-pemrosesan diawali dengan membagi data menjadi 70:15:15. Penulis memutuskan untuk menggunakan 3000 data per kelas dengan pertimbangan keterbatasan komputasi namun tetap ingin mendapatkan performa yang andal. Sehingga didapatkan data gambar untuk *training* sebesar 10500 gambar, lalu sebanyak 2250 untuk validasi dan 2250 untuk *testing*.

Keudian dilakukan *resizing* citra ke ukuran 224×224 piksel untuk memenuhi kebutuhan input ViT. Selanjutnya dilakukan augmentasi data melalui rotasi dan zooming guna meningkatkan variasi *dataset* dan mencegah *overfitting*. Kedua langkah ini krusial untuk optimalisasi performa model.

Pembuatan Model

Kemudian dilakukan pembuatan model. Model dibangun menggunakan arsitektur Vision Transformer (ViT) yang memproses citra melalui pembagian *patch*, transformasi *embedding*, dan mekanisme *multi-head attention*. Adapun versi Vision Transformer yang digunakan adalah versi Tiny Vision Transformer (TinyViT) yang mana memiliki perbedaan utama dengan versi base yaitu di bagian jumlah parameter yang dimiliki di mana TinyViT memiliki 21 juta parameter sementara itu versi *base* ViT memiliki 86 juta parameter. Pelatihan dilakukan dengan *optimizer* Adam dan *loss function categorical cross-entropy* melalui beberapa epoch untuk optimalisasi model.

Pengujian Model

Lalu setelah pembuatan model, maka dilakukan pengujian model. Proses pengujian model Vision Transformer dilakukan dengan mengevaluasi berbagai kombinasi parameter untuk mendapatkan performa optimal dalam mengklasifikasikan penyakit daun singkong. Penelitian menguji tiga parameter utama secara sistematis. Pertama, variasi jumlah epoch dengan dua opsi berbeda (50 dan 70 epoch) untuk melihat pengaruh durasi pelatihan terhadap akurasi model. Kedua, pengujian batch size dalam tiga ukuran (8, 16, dan 32) yang mempengaruhi kecepatan pelatihan dan stabilitas model, di mana *batch size* kecil cenderung lebih cepat konvergen tetapi membutuhkan komputasi intensif, sedangkan *batch size* besar memberikan estimasi gradien lebih stabil namun memerlukan sumber daya memori lebih besar. Terakhir, dilakukan perbandingan tiga jenis *optimizer* (Adam, AdamW, dan SGD) untuk menganalisis pengaruh masing-masing algoritma optimasi terhadap kemampuan klasifikasi model. Keseluruhan pengujian ini bertujuan untuk menemukan konfigurasi parameter terbaik yang dapat memaksimalkan kinerja model Vision Transformer dalam tugas identifikasi penyakit tanaman singkong. Berikut adalah tabel yang berisikan detail konfigurasi parameter yang diujikan terhadap model.

Parameter	Variasi
Epoch	50
	70
Batch Size	8
	16
	32
Optimizer	Adam
	AdamW
	SGD

Tabel 1. Detail Parameter

Evaluasi Model

Kemudian tahap selanjutnya adalah evaluasi model. Proses evaluasi kinerja model Vision Transformer dalam mendeteksi penyakit daun singkong dilakukan melalui analisis komprehensif terhadap berbagai metrik performa. Tahap pengujian menggunakan *dataset* baru berisi citra daun singkong yang belum pernah diproses sebelumnya, dengan confusion matrix sebagai alat evaluasi utama. Hasil pengujian kemudian dianalisis melalui beberapa indikator kunci meliputi akurasi keseluruhan, presisi, *recall*, serta *F1-score* yang memberikan gambaran menyeluruh tentang kemampuan model dalam mengidentifikasi berbagai jenis penyakit. *Confusion matrix* khususnya berperan penting dalam mengevaluasi ketepatan klasifikasi untuk setiap kategori penyakit, sekaligus mengidentifikasi potensi kesalahan klasifikasi antar kelas. Analisis multidimensi ini memungkinkan penilaian objektif terhadap efektivitas model dalam tugas diagnostik tanaman.

4. PENGUJIAN

Parameter yang pertama diuji adalah berdasarkan jumlah epoch di mana penulis menguji model dengan jumlah *epoch* 50 dan 70. Berikut adalah hasilnya :

Epoch	Training Accuracy (%)	Validation Accuracy (%)	Precision (%)	Recall (%)
50	98.71	72.04	74	70
70	99.00	75.16	76	75

Tabel 2. Hasil Pengujian Epoch

Di sini penulis mendapatkan hasil yang tercantum dalam tabel di atas. Berdasarkan hasil evaluasi model, terlihat bahwa performa model mengalami peningkatan seiring dengan penambahan jumlah *epoch* dari 50 ke 70. Pada epoch 50, model mencapai akurasi pelatihan yang sangat tinggi sebesar 98.71%, namun akurasi validasi hanya 72.04%. Nilai presisi 74% dan recall 70% pada epoch ini mengindikasikan bahwa model sudah cukup baik dalam meminimalkan *false positive* dan

false negative, meskipun masih ada ruang untuk perbaikan.

Ketika epoch ditingkatkan menjadi 70, terjadi peningkatan yang signifikan pada semua metrik evaluasi. Akurasi validasi meningkat menjadi 75.16%, disertai dengan peningkatan presisi menjadi 76% dan *recall* menjadi 75%. Peningkatan ini menunjukkan bahwa model menjadi lebih baik dalam menggeneralisasi pola ke data yang belum pernah dilihat sebelumnya.

Berikutnya penulis melakukan pengujian dengan parameter *optimizer*. Berikut adalah hasil yang didapatkan :

Optimizer	Training Accuracy (%)	Validation Accuracy (%)	Precision (%)	Recall (%)
Adam	98.90	74.62	74	73
AdamW	99.00	75.16	76	75
SGD	100.00	79.60	79	78

Tabel 3. Hasil Pengujian Optimizer

Berdasarkan hasil evaluasi ketiga *optimizer* yang diuji, dapat diamati bahwa SGD menunjukkan performa terbaik dibandingkan Adam dan AdamW. Meskipun ketiga *optimizer* mencapai akurasi pelatihan yang sangat tinggi (98.90% untuk Adam, 99.00% untuk AdamW, dan sempurna 100% untuk SGD), namun yang paling menarik adalah performa pada data validasi. SGD berhasil mencapai akurasi validasi tertinggi sebesar 79.60%, jauh melampaui Adam (74.62%) dan AdamW (75.16%).

Hasil ini juga konsisten dengan metrik evaluasi lainnya. SGD mencatat nilai presisi 79% dan *recall* 78%, yang lebih unggul dibandingkan Adam (presisi 74%, recall 73%) dan AdamW (presisi 76%, recall 75%). Hal ini menunjukkan bahwa model dengan *optimizer* SGD tidak hanya lebih akurat, tetapi juga lebih seimbang dalam meminimalkan kesalahan klasifikasi baik *false positive* maupun *false negative*.

Di sini, penulis memutuskan untuk menguji Kembali variasi *epoch* 50 yang sebelumnya memiliki akurasi yang lebih rendah dari *epoch*

70 namun kali ini sambil digunakan *optimizer* SGD. Berikut hasilnya:

Epoch	Training Accuracy (%)	Validation Accuracy (%)	Precision (%)	Recall (%)
50	100.00	83.16	81	81
70	100.00	79.60	79	78

Tabel 4. Hasil Pengujian Ulang Epoch + Optimizer SGD

Hasil eksperimen menunjukkan perkembangan menarik ketika *optimizer* SGD diaplikasikan pada model dengan 50 *epoch*. Terjadi peningkatan dimana akurasi validasi melonjak dari 79.60% menjadi 83.16% - suatu peningkatan hampir 4 poin persentase. Fenomena ini membalikkan temuan sebelumnya dimana *epoch* lebih tinggi selalu memberikan hasil lebih baik.

Penulis kemudian melakukan pengujian batch size dengan 3 opsi yaitu 8, 16, dan 32 terhadap model dengan 50 *epoch* dan *optimizer* SGD dan berikut adalah hasilnya:

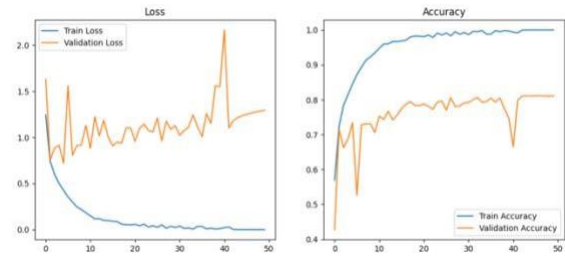
Batch Size	Training Accuracy (%)	Validation Accuracy (%)	Precision (%)	Recall (%)
8	100.00	81.07	81	81
16	100.00	75.96	77	77
32	100.00	83.16	81	81

Tabel 5. Hasil Pengujian Batch Size

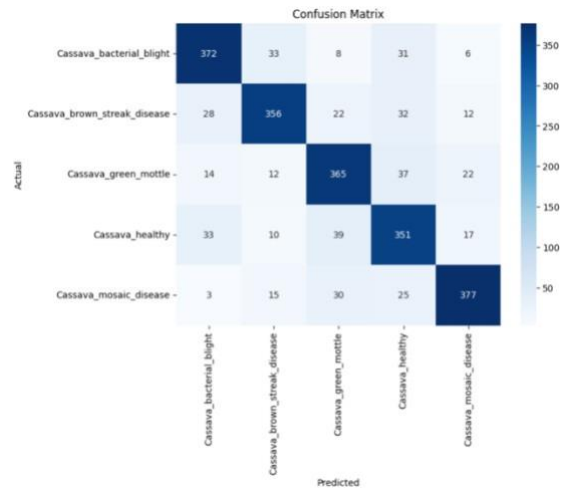
Hasilnya menunjukkan bahwa semua *batch size* menghasilkan *training accuracy* sempurna (100%), mengindikasikan *overfitting*. Namun, pada *validation accuracy*, batch size 32 memberikan performa terbaik (83.16%), diikuti oleh *batch size* 8 (81.07%) dan 16 (75.96%). *Precision* dan *recall* untuk *batch size* 8 dan 32 sama-sama mencapai 81%, sementara *batch size* 16 sedikit lebih rendah (77%). Dengan demikian, *batch size* 32 merupakan pilihan terbaik dalam eksperimen ini karena menghasilkan validasi akurasi tertinggi meskipun *overfitting* masih terjadi.

Berikut adalah gambar grafik *loss* dan akurasi beserta gambar *confusion matrix* dari tiap-tiap model dengan *optimizer* SGD setelah yang diujikan dengan variasi *batch size*:

a. Batch size 8

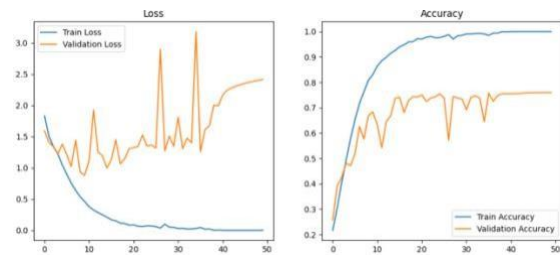


Gambar 8. Grafik Akurasi dan Loss Batch Size 8



Gambar 9. Confusion Matrix Batch Size 8

b. Batch Size 16

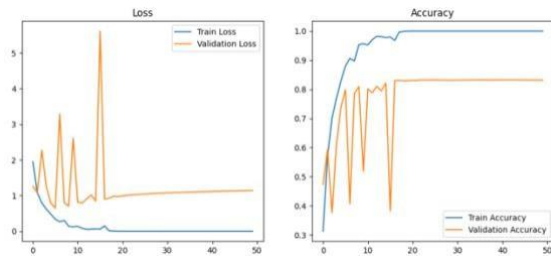


Gambar 10. Grafik Akurasi dan Loss Batch Size 16



Gambar 11. Confusion Matrix Batch Size 16

c. Batch Size 32



Gambar 12. Grafik Akurasi dan Loss Batch Size 32



Gambar 13. Confusion Matrix Batch Size 32

Kesimpulan

Berdasarkan serangkaian eksperimen klasifikasi penyakit pada daun singkong menggunakan

Vision Transformer, dapat disimpulkan bahwa kombinasi parameter *optimizer* SGD dengan 50 *epoch* dan *batch size* 32 menghasilkan performa terbaik. Model ini mencapai akurasi validasi tertinggi (83.16%) dibandingkan konfigurasi lain, meskipun menunjukkan gejala *overfitting* (*training accuracy* 100%). Peningkatan jumlah *epoch* dari 50 ke 70 semula dianggap dapat memperbaiki kinerja model, tetapi justru penggunaan SGD pada 50 *epoch* terbukti lebih efektif dengan lonjakan akurasi validasi hampir 4%.

Optimizer SGD secara konsisten mengungguli Adam dan AdamW dalam hal akurasi validasi (79.60% vs 75%), presisi (79% vs 75%), dan recall (78% vs 74%), menunjukkan kemampuannya yang lebih baik dalam generalisasi. Sementara itu, evaluasi *batch size* mengonfirmasi bahwa *batch size* 32 paling optimal, dengan akurasi validasi 83.16% dan metrik presisi-*recall* seimbang (81%). Dengan demikian, konfigurasi terbaik yang direkomendasikan adalah ViT dengan SGD, 50 *epoch*, dan *batch size* 32, sebagai *baseline* untuk pengembangan model selanjutnya.

Saran

1. Penambahan jumlah *dataset* yang digunakan serta augmentasi data untuk membantu mengurangi *overfitting*.
2. Penggunaan tenaga komputasi yang lebih tinggi agar proses *training data* dapat dilakukan dengan variasi *epoch* dan *batch size* yang lebih tinggi.
3. Melakukan *fine tuning* agar bisa mendapatkan model dengan performa yang paling baik.

5. REFERENSI

[1] N. Kharistian Gulo *et al.*, “Sistem Pakar Mendiagnosa Penyakit Pada Tumbuhan Ketela dengan Menggunakan Metode Teorema Bayes,” *Jurnal CyberTech*, vol. 4, no. 4, 2021, [Online]. Available: <https://ojs.trigunadharma.ac.id/>

- [2] A. Lianardo, N. Kumalasari, and C. Pratiwi, "Klasifikasi Gejala Penyakit Daun Pada Tanaman Singkong Berbasis Vision Menggunakan Metode CNN Dengan Arsitektur Mobilenet." [Online]. Available: www.kaggle.com
- [3] P. S. Hareesh, T. R. Resmi, M. N. Sheela, and T. Makesh Kumar, "Cassava mosaic disease in South and Southeast Asia: current status and prospects," 2023, *Frontiers Media SA*. doi: 10.3389/fsufs.2023.1086660.
- [4] G. Putra Bintang, D. Andreana, J. Yanto, A. Marisa, D. Putri Ayu, and S. Kumala Dewi, "Revitalisasi Sumber Pangan Nabati dan Hewani Pascapandemi dalam Mendukung Pertanian Lahan Suboptimal secara Berkelanjutan Inventarisasi Spesies Hama pada Tanaman Ubi Kayu (*Manihot esculante* Crantz) di Desa Tanjung Pering, Kabupaten Ogan Ilir, Sumatera Selatan Inventory of Pest Species on Cassava (*Manihot esculante* Crantz) in Tanjung Pering Village, Ogan Ilir Regency, South Sumatra," 2022.
- [5] K. von Grebmer *et al.*, "Global Hunger Index How Gender Justice Can Advance Climate Resilience And Zero Hunger." [Online]. Available: www.globalhungerindex.org
- [6] M. Faturrachman and I. Yustiana, "Disease Detection System in Cassava Leaves Using Android-Based Deep Learning and Tensorflow."
- [7] J. A. Figo, N. Yudistira, and A. W. Widodo, "Deteksi Covid-19 dari Citra X-ray menggunakan Vision Transformer," 2023. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] A. Pangestu, B. Purnama, and R. Risnandar, "Vision Transformer untuk Klasifikasi Kematangan Pisang," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 75–84, Feb. 2024, doi: 10.25126/jtiik.20241117389.
- [9] G. Gingin Tahyudin, M. D. Sulistiyo, M. Arzaki, and E. Rachmawati, "International Journal on Informatics Visualization journal homepage : www.joiv.org/index.php/joiv International Journal on Informatics Visualization Classifying Gender Based on Face Images Using Vision Transformer." [Online]. Available: www.joiv.org/index.php/joiv