

KLASIFIKASI TANAMAN OBAT TRADISIONAL BERBASIS CITRA BUAH DAN DAUN

Nurul Kusumawardani¹⁾, Chyquitha Danuputri²⁾, Darniati³⁾, Muhammad Faisal⁴⁾, Muhyiddin A.M Hayat⁵⁾, Muhammad Syafaat S.Kuba⁶⁾, Desi Anggreani⁷⁾

1. Program Studi Informatika, Universitas Muhammadiyah Makassar
email: 105841101821@student.unismuh.ac.id
2. Program Studi Informatika, Universitas Muhammadiyah Makassar
email: chyquithadanuputri@unismuh.ac.id
3. Program Studi Informatika, Universitas Muhammadiyah Makassar
email: darniati@unismuh.ac.id
4. Program Studi Informatika, Universitas Muhammadiyah Makassar
email: muhfaisal@unismuh.ac.id
5. Program Studi Informatika, Universitas Muhammadiyah Makassar
email: muhyiddin@unismuh.ac.id
6. Program Studi Informatika, Universitas Muhammadiyah Makassar
email: syafaat_skuba@unismuh.ac.id
7. Program Studi Informatika, Universitas Muhammadiyah Makassar
email: desianggreani@unismuh.ac.id

Abstract

Indonesia is a megabiodiversity country with extensive use of traditional medicinal plants; however, plant identification in natural environments remains largely manual and error-prone. Recent advances in deep learning, particularly Vision Transformer (ViT), provide a promising solution by effectively capturing global spatial features for image classification. This study applies a ViT-Base/16 model to automatically classify fruit and leaf images of Indonesian medicinal plants. The dataset comprises 1,000 field-collected images from Galung Village, West Sulawesi, covering 20 classes (10 medicinal and 10 non-medicinal plants). The model was fine-tuned using the AdamW optimizer with a learning rate of 2×10^{-5} and trained for 30 epochs with cosine annealing. The proposed approach achieved high performance, with 99.33% accuracy, 99.41% precision, 99.33% recall, and a 99.33% F1-score, while binary classification between medicinal and non-medicinal plants reached 100% accuracy. The system was deployed as a Flask-based web application, demonstrating reliable functionality and practical response times. Overall, the results confirm the effectiveness of Vision Transformer for medicinal plant classification under natural conditions and highlight its potential to support digital documentation, education, and the preservation of local ethnobotanical knowledge.

Keywords: Deep Learning, Ethnobotany, Image Classification, Traditional Medicinal Plants, Vision Transformer

A. PENDAHULUAN

Indonesia merupakan negara megabiodiversitas dengan kekayaan tanaman obat yang telah dimanfaatkan secara turun-temurun dalam sistem pengobatan tradisional. Badan Riset dan Inovasi Nasional (BRIN) mencatat terdapat sekitar 2.850 spesies tanaman obat yang telah teridentifikasi secara ilmiah, dengan lebih dari 22.000 ramuan obat tradisional yang diwariskan oleh Masyarakat [1]. Namun, pengetahuan etnobotani ini menghadapi ancaman serius akibat modernisasi dan minimnya dokumentasi ilmiah yang sistematis. Penelitian etnobotani di Leuwiliang, Bogor berhasil mengidentifikasi 101 spesies tanaman obat dari 44 famili yang digunakan masyarakat setempat [2], sementara penelitian lain mendokumentasikan pengetahuan suku Rejang tentang 38 spesies tanaman obat untuk berbagai keperluan pengobatan [3].

Perkembangan teknologi *deep learning* menawarkan solusi potensial untuk mengatasi keterbatasan metode konvensional. *Convolutional Neural Network* (CNN) telah banyak digunakan untuk klasifikasi tanaman dengan hasil yang menjanjikan [4]. Namun, arsitektur *Vision Transformer* (ViT) menawarkan keunggulan signifikan melalui mekanisme *self-attention* yang mampu menangkap hubungan spasial jangka panjang antar elemen gambar [5]. Kemampuan ini menjadikan ViT sangat cocok untuk menangani klasifikasi tanaman obat yang memiliki variasi bentuk, warna, dan tekstur yang kompleks.

Efektivitas *Vision Transformer* dalam klasifikasi tanaman telah dibuktikan melalui berbagai penelitian. PlantXViT yang dikembangkan oleh Thakur et al. menghasilkan akurasi tinggi dalam klasifikasi penyakit tanaman sekaligus memberikan interpretabilitas visual yang baik [6]. Penelitian lain menunjukkan

kemampuan ViT dalam mengenali hubungan spasial kompleks pada citra botani [7], serta ketahanan terhadap berbagai gangguan visual seperti *noise*, rotasi, dan perubahan skala yang merupakan tantangan umum dalam pengolahan citra tanaman di lingkungan alami [8]. Meskipun demikian, aplikasi *Vision Transformer* untuk klasifikasi tanaman obat tradisional Indonesia yang memiliki keragaman morfologi khas masih sangat terbatas.

Ketersediaan dataset berkualitas merupakan tantangan utama dalam penelitian ini. Dataset IndoHerb telah menghimpun lebih dari 10.000 gambar dari 100 spesies tanaman obat [9], namun belum mencakup informasi kontekstual seperti nama lokal, manfaat spesifik, dan cara penggunaan tradisional. Keterbatasan ini mendorong pendekatan pengumpulan data mandiri yang melibatkan masyarakat lokal, selaras dengan studi etnobotani yang merekam pengetahuan lokal melalui metode partisipatif [10]. Pendekatan *citizen science* terbukti efektif dalam menyediakan data keanekaragaman hayati yang beragam dan relevan dengan kondisi lokal [11].

Untuk mengatasi keterbatasan jumlah data, teknik augmentasi citra menjadi strategi penting. Penelitian menunjukkan bahwa *preprocessing* seperti penghapusan latar belakang dapat meningkatkan performa klasifikasi [12]. Sebagai pengembangan lebih lanjut dari strategi augmentasi, pendekatan berbasis kelas di mana setiap kelas tanaman diberikan teknik augmentasi yang disesuaikan dengan karakteristik visualnya telah terbukti efektif dalam meningkatkan akurasi klasifikasi citra tanaman [13].

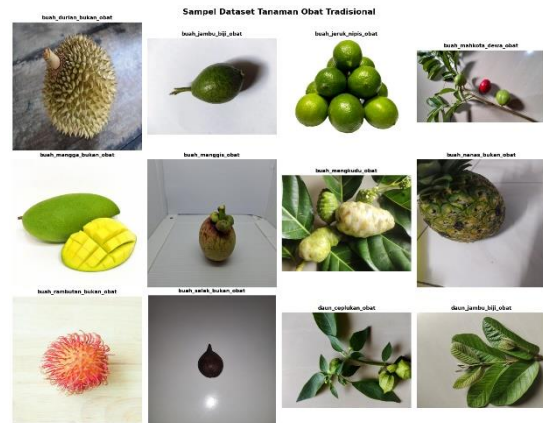
Penelitian ini bertujuan mengimplementasikan *Vision Transformer* untuk klasifikasi tanaman obat tradisional Indonesia berbasis citra buah dan daun. Fokus utama adalah menguji performa ViT dalam

membedakan tanaman obat dari tanaman bukan obat secara otomatis, dengan evaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Sistem yang dikembangkan juga menyajikan informasi etnobotani seperti nama ilmiah, manfaat, dan cara penggunaan sebagai sarana edukatif.

B. METODE PENELITIAN

Pengumpulan Data

Pengumpulan data dilakukan di lingkungan alami seperti pekarangan rumah, kebun tanaman obat (Taman Obat Keluarga/TOGA), dan area sekitar pemukiman di wilayah Desa Galung, Kecamatan Tapalang, Kabupaten Mamuju, Sulawesi Barat. Pendekatan ini dipilih untuk memastikan bahwa data yang dikumpulkan merepresentasikan kondisi nyata dari tanaman obat tropis Indonesia, yang memiliki keragaman morfologi tinggi dan latar belakang alami yang kompleks. Dataset penelitian ini terdiri dari 1.000 citra yang terbagi ke dalam 20 kelas, dengan setiap kelas berisi 50 citra yang mewakili 10 tanaman obat dan 10 tanaman bukan obat. Seluruh citra dikumpulkan melalui pengambilan langsung di lingkungan alami menggunakan *smartphone* beresolusi 16 MP yang dilengkapi fitur *autofocus* dan *auto white balance* untuk menjaga kestabilan kualitas visual. Metode pengumpulan data berbasis observasi lapangan ini mengikuti pendekatan dokumentasi etnobotani yang menekankan pengambilan citra langsung dari habitat alami tanaman, sebagaimana diterapkan pada studi sebelumnya di Indonesia [14], [15]. Untuk mempertahankan distribusi kelas yang seimbang, dataset dibagi secara stratifikasi menjadi 70% data latih, 15% data validasi, dan 15% data uji.



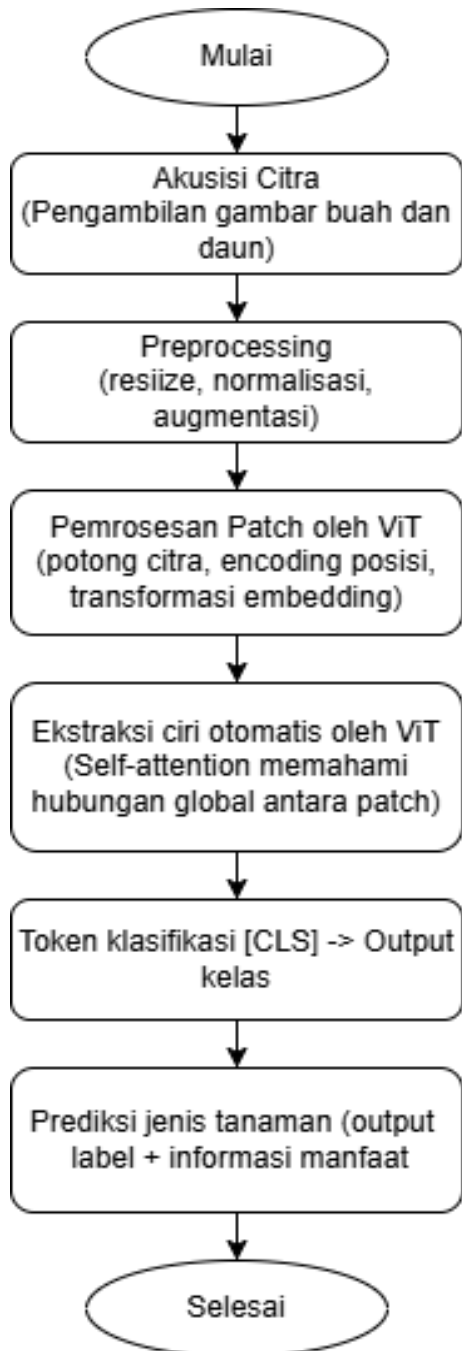
Gambar 1. Dataset tanaman obat

Gambar tersebut menampilkan sampel dataset yang digunakan dalam penelitian ini, yang mencerminkan variasi visual dari citra buah dan daun tanaman obat maupun bukan obat yang dikumpulkan di lingkungan alami. Representasi ini menunjukkan keragaman tekstur, warna, dan kondisi pencahayaan yang menjadi karakteristik utama dataset lapangan, sehingga memberikan gambaran yang jelas mengenai kompleksitas data yang diolah pada tahap pemodelan.

Tahapan Penelitian

Penelitian ini dilakukan melalui serangkaian tahap yang mencakup studi literatur untuk menentukan landasan teoretis, pengumpulan citra tanaman obat dan bukan obat di lingkungan alami, serta prapemrosesan data melalui *resize*, normalisasi, dan augmentasi. Dataset kemudian dibagi secara stratifikasi menjadi data latih, validasi, dan uji sebelum digunakan pada pelatihan model *Vision Transformer* yang memproses citra melalui *patch embedding* dan *Transformer Encoder*. Kinerja model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan *confusion matrix* [16]. Model terbaik selanjutnya diintegrasikan ke dalam sistem identifikasi berbasis web dan diuji

menggunakan metode *blackbox* untuk memastikan fungsionalitas, ketepatan klasifikasi, dan konsistensi waktu respons. Gambar 2 merangkum keseluruhan alur penelitian tersebut, mulai dari akuisisi data hingga implementasi dan evaluasi sistem.



Gambar 2. Flowchart Proses Penelitian

Flowchart ini menggambarkan alur proses penelitian mulai dari pengumpulan data, prapemrosesan, pelatihan model *Vision Transformer*, hingga implementasi dan pengujian sistem.

Arsitektur Model Vision Transformer (ViT)

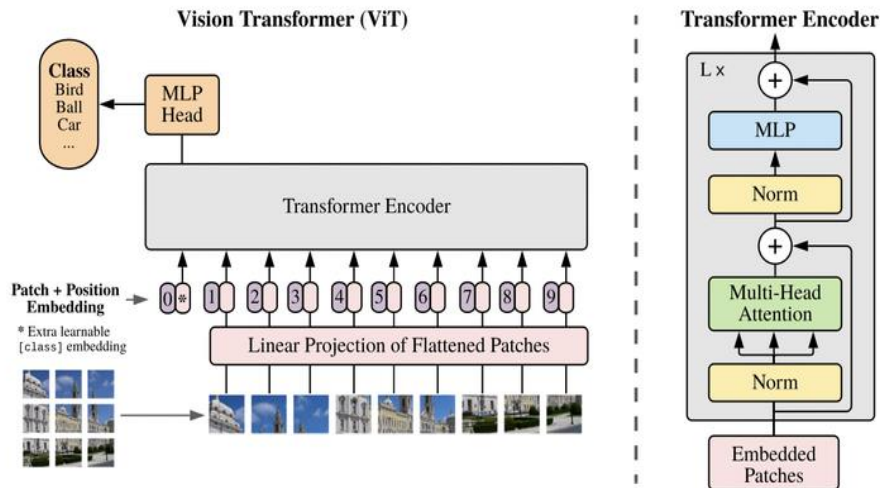
Model yang digunakan pada penelitian ini adalah *Vision Transformer (ViT)*, yang memproses citra dengan membaginya menjadi patch berukuran 16×16 piksel. Setiap patch kemudian di-flatten dan diproyeksikan ke dalam patch embedding berdimensi 768. Untuk mempertahankan informasi posisi setiap patch, positional encoding ditambahkan pada embedding sebelum seluruh patch diproses oleh *Transformer Encoder*. Struktur *encoder* terdiri dari *Multi-Head Self-Attention (MHSA)* dan *Feed-Forward Network (FFN)* yang tersusun dalam 12 blok *encoder*, di mana setiap blok memiliki mekanisme perhatian berlapis untuk menangkap hubungan antar *patch* secara global [17]. Token khusus [CLS] disisipkan pada urutan input dan digunakan sebagai representasi global citra. Setelah melalui seluruh blok *encoder*, vektor [CLS] digunakan sebagai basis untuk tahap klasifikasi menggunakan *fully connected layer* [18]. Konfigurasi lengkap arsitektur ViT yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Konfigurasi Arsitektur Vision Transformer

Parameter	Nilai
Jumlah Patch	196
Dimensi Embedding	768
Jumlah Attention Heads	12
Dimensi per Head	64
Dimensi FFN Hidden Layer	3.072
Total Parameter	85.814.036
Jumlah Kelas Output	20

Arsitektur lengkap *Vision Transformer* yang digunakan dalam penelitian ini dapat dilihat pada Gambar 3. Diagram tersebut memperlihatkan alur

pemrosesan citra mulai dari pembagian *patch* hingga tahap klasifikasi menggunakan token [CLS].



Gambar 3. Vision Transformer Sumber: (Dosovitskiy et al., 2021)

Metrik Evaluasi

Dalam penelitian ini, evaluasi performa model dilakukan menggunakan empat metrik utama, yaitu akurasi, *precision*, *recall*, dan *F1-score*. Metrik-metrik ini umum digunakan dalam tugas klasifikasi citra dan memberikan gambaran menyeluruh mengenai kemampuan model dalam mengenali dan membedakan kelas dengan tepat. Persamaan untuk masing-masing metrik ditunjukkan sebagai berikut:

- a. Akurasi mengukur proporsi prediksi yang benar terhadap seluruh data [20]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- b. Precision mengukur ketepatan model dalam memprediksi kelas positif:

$$Precision = \frac{TP}{TP + FP}$$

- c. *Recall* mengukur seberapa banyak data positif yang berhasil dikenali dengan benar oleh model:

$$Recall = \frac{TP}{TP + FN}$$

- d. *F1-Score* adalah rata-rata harmonis dari *precision* dan *recall*, yang digunakan ketika keseimbangan antara keduanya penting:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Berdasarkan metrik evaluasi tersebut, semakin tinggi nilai akurasi, precision, recall, dan F1-score maka semakin baik performa model dalam melakukan

klasifikasi. Nilai *precision* dan *recall* digunakan untuk melihat keseimbangan antara kemampuan model mendeteksi kelas positif dan tingkat ketepatannya, sedangkan *F1-score* memberikan gambaran menyeluruh terhadap stabilitas performa model. Dengan demikian, keempat metrik ini menjadi dasar untuk menentukan efektivitas model yang digunakan dalam penelitian.

Konfigurasi Pelatihan

Model dilatih selama 30 *epoch* menggunakan konfigurasi hyperparameter yang ditunjukkan pada Tabel 2. Parameter pelatihan ini digunakan untuk mengoptimalkan kinerja *Vision Transformer* pada dataset citra tanaman obat dan bukan obat.

Tabel 2. Konfigurasi Hyperparameter Pelatihan

Hyperparameter	Nilai
Jumlah Epoch	30
Batch Size	16
Learning Rate	2×10^{-5}
Optimizer	AdamW
Weight Decay	0.01
Scheduler	Cosine Annealing
Loss Function	CrossEntropyLoss

Konfigurasi hyperparameter tersebut digunakan secara konsisten selama seluruh proses pelatihan untuk memastikan stabilitas pembaruan bobot dan pencapaian konvergensi yang optimal. Pengaturan ini dipilih berdasarkan performa terbaik pada eksperimen awal serta rekomendasi studi sebelumnya mengenai pelatihan model *Vision Transformer*. Efektivitas konfigurasi pelatihan ini dapat dilihat dari hasil pelatihan yang disajikan pada bagian berikutnya.

C. HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Bagian ini menyajikan hasil pengolahan data serta pembahasan terkait performa model *Vision Transformer* (ViT) dalam mengklasifikasikan citra tanaman obat tradisional. Hasil penelitian diuraikan berdasarkan tahapan pada metode, dimulai dari gambaran dataset, proses pelatihan model, evaluasi kinerja, hingga analisis hasil klasifikasi.

Dataset

Dataset penelitian ini dikumpulkan secara langsung di Desa Galung, Kecamatan Tapalang, Kabupaten Mamuju, Sulawesi Barat menggunakan kamera *smartphone* beresolusi 16 MP dengan fitur *autofocus* dan *auto exposure*. Total dataset terdiri dari 1.000 citra yang terbagi secara seimbang ke dalam 20 kelas, meliputi 10 kelas tanaman obat (5 kelas buah dan 5 kelas daun) serta 10 kelas bukan tanaman obat (5 kelas buah dan 5 kelas daun mirip tanaman obat). Setiap kelas berisi 50 gambar untuk memastikan distribusi data yang seimbang dan mencegah bias model selama pelatihan. Komposisi lengkap dataset ditunjukkan pada Tabel 3 menunjukkan rincian lengkap jumlah citra pada masing-masing kelas.

Tabel 3. Distribusi Dataset Tanaman Obat dan Bukan Obat

No	Nama Kelas	Jumlah Citra	Format
1	buah_durian	50	jpg
2	buah_jambu_biji	50	jpg
3	buah_jeruk_nipis	50	jpg
4	buah_mahkota_dewa	50	jpg
5	buah_mangga	50	jpg
6	buah_manggis	50	jpg

7	buah_mengkudu	50	jpg
8	buah_nanas	50	jpg
9	buah_rambutan	50	jpg
10	buah_salak	50	jpg
11	daun_ceplukan	50	jpg
12	daun_jambu_biji	50	jpg
13	daun_kersen	50	jpg
14	daun_mengkudu	50	jpg
15	daun_pepaya	50	jpg
16	mirip_daun	50	jpg
17	mirip_daun_jambu_biji	50	jpg
18	mirip_daun_kersen	50	jpg
19	mirip_daun_mengkudu	50	jpg
20	mirip_daun_pepaya	50	jpg
Total Citra		1000	jpg

Pembagian Dataset

Dataset dibagi menggunakan metode stratifikasi untuk memastikan bahwa setiap kelas memiliki representasi yang seimbang pada data latih, validasi, dan uji. Pendekatan ini penting untuk mengurangi bias distribusi kelas dan menjaga konsistensi performa model selama proses pelatihan dan evaluasi. Proporsi pembagian dataset ditunjukkan pada Tabel 4.

Tabel 4. Distribusi Dataset

Split Data	Jumlah Citra	Persentase
Data Latih (Training)	700	70%
Data Validasi (Validation)	150	15%
Data Uji (Testing)	150	15%
Total	1.000	100%

Hasil Pelatihan

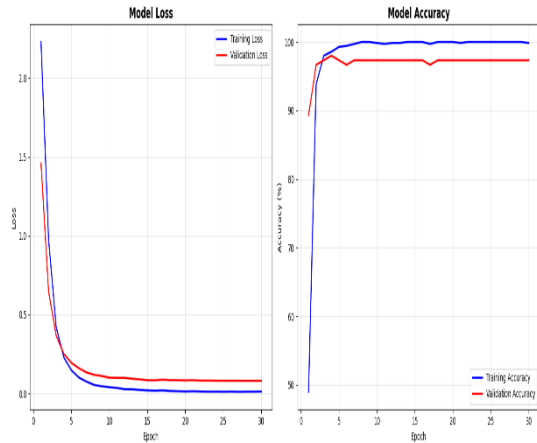
Proses pelatihan model menunjukkan bahwa Vision Transformer mampu mencapai konvergensi yang stabil dalam jumlah epoch pelatihan yang relatif singkat. Nilai training loss mengalami penurunan tajam pada fase awal pelatihan dan kemudian bergerak menuju nilai yang konstan, sedangkan validation loss mempertahankan pola yang konsisten tanpa lonjakan signifikan. Hal ini menunjukkan bahwa model tidak mengalami overfitting dan mampu belajar representasi visual yang baik dari data pelatihan. Ringkasan hasil pelatihan ditampilkan pada Tabel 5.

Tabel 5. Ringkasan Hasil Akhir Proses Pelatihan Model

Metrik Evaluasi	Nilai Akhir
Final Training Loss	0.0126
Final Training Accuracy	99.86%
Final Validation Loss	0.0806
Final Validation Accuracy	97.33%
Best Validation Accuracy	98.00%

Visualisasi *learning curve* yang ditampilkan pada Gambar 4 memberikan gambaran lebih jelas mengenai stabilitas proses pelatihan. Grafik tersebut memperlihatkan perbedaan yang konsisten dan relatif kecil antara *training*

loss dan *validation loss*, sekaligus menunjukkan peningkatan akurasi yang stabil pada kedua subset data. Pola ini mengindikasikan bahwa model berhasil belajar secara efektif tanpa menghadapi masalah *underfitting* maupun *overfitting*.



Gambar 4. Grafik Perubahan Nilai Loss dan Akurasi Model ViT Selama Pelatihan

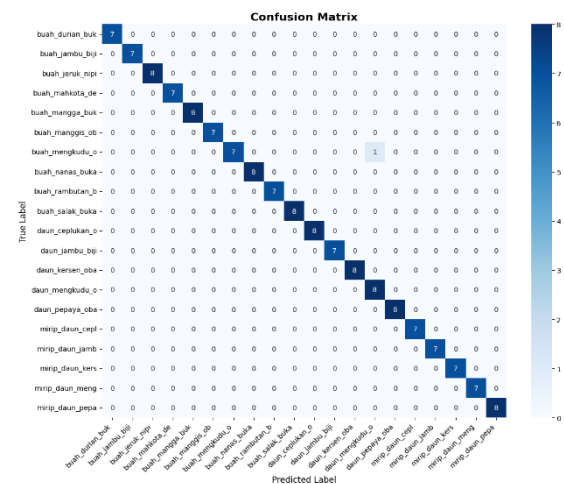
Pola *learning curve* pada Gambar 4 menunjukkan bahwa proses pelatihan berlangsung secara stabil, di mana model mampu mempertahankan keseimbangan antara penurunan *loss* dan peningkatan akurasi pada data latih maupun validasi. Konsistensi ini menegaskan bahwa konfigurasi hyperparameter dan strategi optimasi yang digunakan telah memberikan proses pembelajaran yang efisien bagi model *Vision Transformer*. Dengan model yang telah mencapai konvergensi tanpa indikasi *overfitting*, evaluasi kinerja pada data uji dapat dilakukan secara lebih objektif. Hasil evaluasi performa model pada data uji dijelaskan pada subbagian berikutnya.

Kinerja Model

1. Hasil Evaluasi

Evaluasi performa model *Vision Transformer* dilakukan pada *test set* yang terdiri dari 150 sampel gambar tanaman menggunakan empat metrik utama, yaitu

accuracy, *precision*, *recall*, dan *F1-score* yang telah dijelaskan pada 2.4. *Confusion matrix* untuk klasifikasi 20 kelas tanaman ditunjukkan pada Gambar 5, yang merupakan representasi visual hubungan antara label aktual dan label prediksi untuk setiap kelas dalam dataset.



Gambar 5. Confusion Matrix untuk Klasifikasi Multi-Class 20 Kelas Tanaman

Dari *confusion matrix* pada Gambar 5, dapat diamati bahwa sebagian besar prediksi model berada pada diagonal utama (warna biru gelap), yang mengindikasikan prediksi yang benar. Hanya terdapat satu nilai di luar diagonal utama pada posisi kelas *buah_mengkudu_obat* yang diklasifikasikan keliru sebagai *daun_mengkudu_obat* akibat kemiripan morfologi. Dengan demikian, dari 150 sampel *test set*, model berhasil memprediksi dengan benar sebanyak 149 sampel dengan tingkat kesalahan yang sangat minimal. Berdasarkan *confusion matrix* tersebut, perhitungan metrik evaluasi menggunakan persamaan:

- Accuracy:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= \frac{150 + 149}{150 + 149 + 1 + 1} \\ &= 0.9933 \end{aligned}$$

- Precision

$$\begin{aligned} \text{Precision}_{\text{weighted}} &= \frac{\sum_{i=1}^{20} (\text{Precision}_i \times \text{Support}_i)}{\text{Total Support}} \\ &= \frac{149.125}{150} = 0.9941 \end{aligned}$$

- Recall

$$\begin{aligned} \text{Recall}_{\text{buah_mengkudu_obat}} &= \frac{TP}{TP + FN} \\ &= \frac{7}{7 + 1} = 0.875 \end{aligned}$$

$$\begin{aligned} \text{Recall}_{\text{weighted}} &= \frac{(1.00 \times 142) + (0.875 \times 8)}{150} = \frac{149}{150} \\ &= 0.9933 \end{aligned}$$

- F1-Score

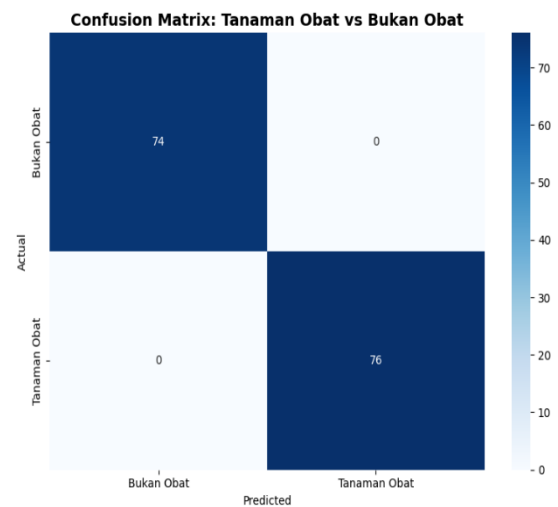
$$\begin{aligned} \text{F1-Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= 2 \times \frac{0.9941 \times 0.9933}{0.9941 + 0.9933} \\ &= 0.9933 \end{aligned}$$

Hasil evaluasi lengkap dari keempat metrik klasifikasi multi-class dapat dilihat pada Tabel 6 berikut:

Tabel 6. Hasil Evaluasi Kinerja Model Vision Transformer

Metrik	Nilai	Persentase
Accuracy	0.9933	99.33%
Precision	0.9941	99.41%
Recall	0.9933	99.33%
F1-Score	0.9933	99.33%

Selain evaluasi *multi-class*, model juga dievaluasi untuk klasifikasi *binary* dalam membedakan tanaman obat (10 kelas) dan bukan tanaman obat (10 kelas). *Confusion matrix binary* ditunjukkan pada Gambar 6, yang menggambarkan performa model dalam membedakan kedua kategori utama tersebut.



Gambar 6. Confusion Matrix untuk Klasifikasi Binary Tanaman Obat vs Bukan Tanaman Obat

Dari *confusion matrix* pada Gambar 6, terlihat bahwa semua prediksi berada pada diagonal utama dengan nilai 74 untuk kategori bukan obat dan 76 untuk kategori tanaman obat, tanpa adanya kesalahan klasifikasi. Berdasarkan komponen *confusion matrix* dengan $TP = 76$, $TN = 74$, $FP = 0$, dan $FN = 0$, perhitungan metrik evaluasi binary menghasilkan:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= \frac{76 + 74}{76 + 74 + 0 + 0} = \frac{150}{150} \\ &= 1.00 \end{aligned}$$

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{76}{76 + 0} = \frac{76}{76} = 1.00$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{76}{76 + 0} = \frac{76}{76} = 1.00$$

$$\begin{aligned} \text{F1-Score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= \frac{2 \times 1.00 \times 1.00}{1.00 + 1.00} = \frac{2.00}{2.00} \\ &= 1.00 \end{aligned}$$

Hasil evaluasi lengkap untuk klasifikasi binary dapat dilihat pada Tabel 7 berikut:

Tabel 7. Hasil Evaluasi Kinerja Model untuk Klasifikasi Binary

Metrik	Nilai	Persentase
Accuracy	1.0000	100.00%
Precision	1.0000	100.00%
Recall	1.0000	100.00%
F1-Score	1.0000	100.00%

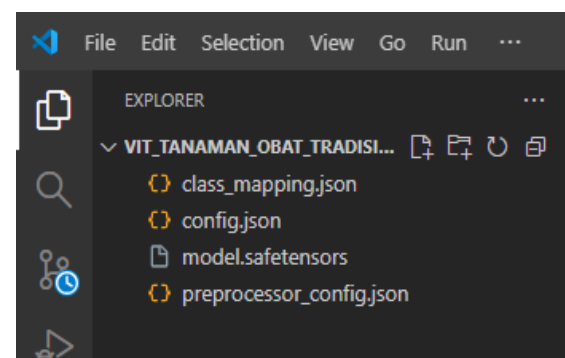
2. Analisis Hasil Evaluasi

Berdasarkan Tabel 6 dan Tabel 7, model *Vision Transformer* menunjukkan performa yang sangat baik dengan akurasi 99.33% pada klasifikasi *multi-class* dan akurasi sempurna 100% pada klasifikasi binary. Dari 150 sampel *test set*, hanya terdapat 1 kesalahan prediksi pada kelas *buah_mengkudu_obat* yang salah diklasifikasikan sebagai *daun_mengkudu_obat* akibat kemiripan morfologi visual antara buah dan daun dari spesies yang sama. Nilai *precision* 99.41%, *recall* 99.33%, dan *F1-score* 99.33% menunjukkan keseimbangan yang baik antara ketepatan prediksi dan kemampuan deteksi dengan tingkat kesalahan yang sangat minimal.

Akurasi sempurna 100% pada klasifikasi binary mengindikasikan bahwa model tidak pernah salah dalam membedakan kategori utama tanaman obat dan bukan tanaman obat, sehingga kesalahan yang terjadi hanya dalam membedakan subkelas dengan kemiripan visual tinggi dalam kategori yang sama. Performa tinggi ini diperoleh melalui mekanisme *self-attention* pada arsitektur *Vision Transformer* yang mampu menangkap relasi spasial antar bagian gambar secara global, sehingga lebih efektif dalam mengenali pola kompleks dan fitur diskriminatif yang membedakan setiap kelas tanaman. Hasil ini menunjukkan bahwa pendekatan yang diusulkan sangat efektif untuk tugas klasifikasi citra tanaman dengan kemampuan generalisasi yang baik pada data yang belum pernah dilihat sebelumnya.

Pengujian Sistem

Pengujian sistem dilakukan untuk memastikan bahwa aplikasi web klasifikasi tanaman dapat berfungsi dengan baik dan sesuai dengan kebutuhan pengguna. Pengujian dilakukan dalam dua aspek, yaitu pengujian fungsionalitas menggunakan metode *blackbox* testing dan pengujian kecepatan respon sistem. Struktur file hasil pelatihan model *Vision Transformer* yang digunakan dalam aplikasi ditunjukkan pada Gambar 7.

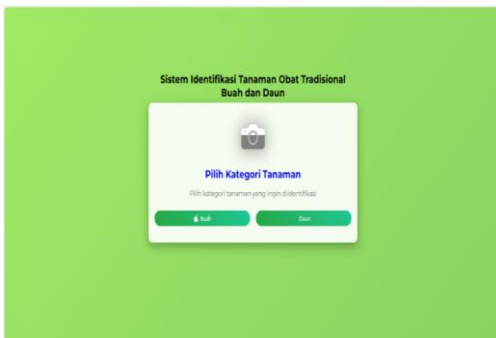


Gambar 7. Struktur File Hasil Pelatihan Model ViT

Gambar 7 menunjukkan struktur file yang digunakan dalam sistem, yang terdiri dari `class_mapping.json` untuk pemetaan kelas, `config.json` untuk konfigurasi model, `model.safetensors` untuk bobot model terlatih, dan `preprocessor_config.json` untuk konfigurasi *preprocessing* gambar. File-file ini diperlukan agar model dapat melakukan inferensi pada aplikasi web dengan benar.

1. Pengujian Fungsionalitas

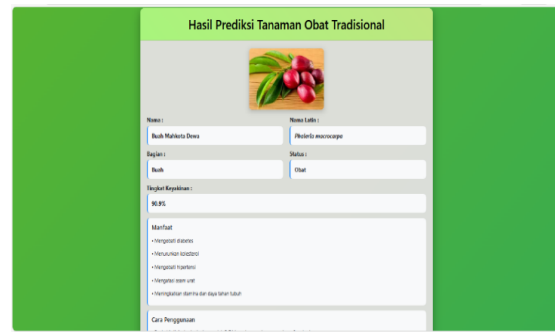
Pengujian fungsionalitas dilakukan menggunakan metode *blackbox testing* untuk memverifikasi bahwa setiap fitur pada aplikasi web dapat berjalan sesuai dengan yang diharapkan tanpa perlu mengetahui detail implementasi internal. Antarmuka utama aplikasi web ditunjukkan pada Gambar 8, yang menyediakan opsi untuk memilih kategori tanaman yang akan diidentifikasi.



Gambar 8. Halaman Utama Aplikasi Web

Pada Gambar 8, halaman utama menampilkan judul "Sistem Identifikasi Tanaman Obat Tradisional Buah dan Daun" dengan dua tombol pilihan kategori: Buah dan Daun. Pengguna dapat memilih salah satu kategori sebelum mengunggah gambar untuk diidentifikasi. Setelah pengguna mengunggah gambar dan sistem melakukan prediksi, hasilnya

ditampilkan pada halaman seperti yang ditunjukkan pada Gambar 9.



Gambar 9. Halaman Hasil Prediksi

Setelah pengguna mengunggah gambar dan sistem melakukan prediksi, hasilnya ditampilkan pada halaman seperti yang ditunjukkan pada Gambar 9. Gambar 9 menunjukkan halaman hasil prediksi yang menampilkan gambar tanaman yang diunggah, prediksi kelas Buah Mahkota Dewa (kategori tanaman obat) dengan tingkat akurasi 90,9%, serta informasi manfaat dan cara penggunaan tanaman.

Untuk memastikan seluruh fungsi aplikasi berjalan dengan baik, dilakukan pengujian *blackbox testing* terhadap fitur-fitur utama sistem. Hasil pengujian lengkap ditunjukkan pada Tabel 8.

Tabel 8. Hasil Pengujian Black-Box Testing Sistem

N o	Skenario Pengujian	Input	Output Diharapkan	Status
1	Identifikasi tanaman obat	Citra daun jambu biji	Nama benarStatus obat	Sesuai
2	Identifikasi tanaman bukan obat	Citra buah rambutan	Nama benarStatus bukan obat	Sesuai
3	Gambar tidak dalam dataset	Citra acak	Notifikasi error	Sesuai
4	Upload bukan gambar	File PDF/DOC X	Input ditolak	Sesuai

N o	Skenario Pengujian	Input	Output Diharapkan	Status
5	Navigasi ulang	Klik tombol kembali	Kembali ke upload	Sesuai

Berdasarkan Tabel 8, seluruh fungsi utama aplikasi telah berhasil diuji dan berfungsi sesuai harapan, meliputi identifikasi tanaman, validasi input, dan navigasi sistem.

2. Pengujian Kecepatan Respon Sistem

Pengujian kecepatan respons sistem dilakukan untuk mengukur waktu yang dibutuhkan sistem dalam memproses prediksi gambar, mulai dari saat pengguna mengunggah gambar hingga hasil prediksi ditampilkan. Pengujian dilakukan pada lima jenis citra dengan karakteristik yang berbeda, termasuk citra tanaman obat, bukan obat, citra tidak dalam dataset, citra berukuran besar, dan citra dengan kualitas rendah. Hasil pengujian kecepatan respon sistem ditunjukkan pada Tabel 9.

Tabel 9. Hasil Pengujian Kecepatan Respon Sistem

N o	Jenis Pengujian	Jenis Citra	Waktu Respon	Kate gori
1	Prediksi tanaman obat	Daun jambu biji	6.67 detik	Cepat
2	Prediksi tanaman bukan obat	Buah rambutan	5.79 detik	Cepat
3	Citra tidak dalam dataset	Daun acak	6.68 detik	Cepat
4	Citra resolusi besar	Gambar >5 MB	6.93 detik	Responsif
5	Citra buram/gelap	Daun gelap	7.50 detik	Responsif

R ata a- ra ta	-	-	6.71 detik	Cepat & Respon sif
----------------------------	---	---	---------------	--------------------------

Berdasarkan Tabel 9, sistem menunjukkan waktu respon rata-rata 6.71 detik dengan rentang 5.79 hingga 7.50 detik. Waktu tercepat dicapai pada prediksi buah rambutan (5.79 detik), sedangkan waktu terlama pada citra berkualitas rendah (7.50 detik). Seluruh pengujian menunjukkan kategori "Cepat" hingga "Responsif", yang mengindikasikan bahwa sistem memiliki performa yang baik dan dapat memberikan pengalaman pengguna yang responsif.

D. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa arsitektur *Vision Transformer* (ViT) mampu memberikan performa yang sangat baik dalam mengklasifikasikan citra tanaman obat tradisional Indonesia berbasis buah dan daun. Model yang dilatih menggunakan teknik *fine-tuning* dengan *optimizer* AdamW dan *cosine annealing scheduler* menghasilkan tingkat akurasi sebesar 99,33% dengan nilai *precision*, *recall*, dan *F1-score* masing-masing sebesar 99,41%, 99,33%, dan 99,33% pada data uji. Sistem yang dikembangkan juga berhasil membedakan tanaman obat dan bukan obat secara sempurna pada klasifikasi biner dengan akurasi 100%. Implementasi sistem berbasis web menunjukkan bahwa model dapat digunakan secara praktis dengan waktu respons rata-rata sebesar 6,71 detik. Hasil ini membuktikan bahwa *Vision Transformer* efektif digunakan sebagai solusi digital untuk membantu identifikasi tanaman obat tradisional serta mendukung pelestarian pengetahuan etnobotani

Indonesia. Untuk pengembangan selanjutnya, disarankan agar jumlah dan variasi dataset diperluas dengan melibatkan lebih banyak spesies tanaman obat dari berbagai daerah di Indonesia guna meningkatkan kemampuan generalisasi model. Selain itu, sistem dapat dikembangkan ke dalam bentuk aplikasi mobile agar dapat diakses lebih luas oleh masyarakat, serta ditambahkan fitur deteksi otomatis objek tanaman untuk meningkatkan kemudahan penggunaan di lingkungan lapangan.

E. REFERENSI

- [1] Badan Riset dan Inovasi Nasional, "BRIN catat ada 2.850 spesies tanaman obat di Indonesia," pp. 1–8, 2024.
- [2] M. P. Pratami, A. Anggraeni, and W. Sujarwo, "Ethnobotany of Medicinal Plants in Leuwiliang (Bogor), Indonesia," *Ethnobot. Res. Appl.*, vol. 27, pp. 1–41, 2024, doi: 10.32859/era.27.1.1-41.
- [3] T. N. Sitorban, E. Nursaadah, and A. Primairyani, "Keanekaragaman Hayati Tumbuhan Obat Tradisional dan Pemanfaatannya," *BIOEDUSAINS Jurnal Pendidik. Biol. dan Sains*, vol. 6, no. 2, pp. 531–544, 2023, doi: 10.31539/bioedusains.v6i2.7547.
- [4] S. Salsabila, A. Suharso, and P. Purwantoro, "Comparison of Deep Learning Architectures in Identifying Types of Medicinal Plant Leaf Images," *J. Appl. Informatics Comput.*, vol. 8, no. 1, pp. 39–46, 2024, doi: 10.30871/jaic.v8i1.6289.
- [5] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv Prepr.*, 2021, doi: 10.48550/arXiv.2010.11929.
- [6] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, "Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT," no. D1, 2022.
- [7] M. R. Tonmoy, N. Dey, S. Member, M. F. Mridha, and S. Member, "MobilePlantViT: A Mobile-friendly Hybrid ViT for Generalized Plant," pp. 1–8, 2025.
- [8] D. Zhou *et al.*, "Understanding The Robustness in Vision Transformers," *Proc. Mach. Learn. Res.*, vol. 162, pp. 27378–27394, 2022.
- [9] M. S. Ikrar Musyaffa, N. Yudistira, M. A. Rahman, A. H. Basori, A. B. Firdausiah Mansur, and J. Batoro, "IndoHerb: Indonesia medicinal plants recognition using transfer learning and deep learning," *Heliyon*, vol. 10, no. 23, p. e40606, 2024, doi: 10.1016/j.heliyon.2024.e40606.
- [10] A. P. Dewi *et al.*, "Ethnobotany of food, medicinal, construction and household utilities producing plants in Cikaniki, Gunung Halimun Salak National Park, Indonesia," *J. Mt. Sci.*, vol. 20, no. 1, pp. 163–181, 2023, doi: 10.1007/s11629-021-7108-5.
- [11] Y. Ferdi, "Data Augmentation through Background Removal for Apple Leaf Disease Classification Using the MobileNetV2 Model," 2024.
- [12] S. Saleem, M. I. Sharif, M. I. Sharif, M. Z. Sajid, and F. Marinello, "Comparison of Deep Learning Models for Multi-Crop Leaf

- Disease Detection with Enhanced Vegetative Feature Isolation and Definition of a New Hybrid Architecture,” *Agronomy*, vol. 14, no. 10, 2024, doi: 10.3390/agronomy14102230.
- [13] N. Siahaan *et al.*, “Ethnobotanical Study of Medicinal Plants Used To Treat Human Diseases In The Urban Forest Area Of Sangga Buana Cilandak Jakarta Selatan,” *J. Trop. Biodivers.*, vol. 2, no. 3, pp. 117–132, 2022.
- [14] S. J. David and D. Chitra, “Vision Transformer for Medicinal Leaf Image,” vol. 14, no. 1, pp. 45–53, 2024.
- [15] A. M. Akbar DB, M. Faisal, and M. AM Hayat, “IMPLEMENTASI HYBRID CNN, FACIAL LANDMARK DAN LIVENESS DETECTION PADA SISTEM ABSENSI WAJAH,” *J. Inform. Prog.*, vol. 17, no. 2, pp. 116–120, Sep. 2025, doi: 10.56708/progres.v17i2.483.
- [16] Sarina *et al.*, “KLASIFIKASI PENYAKIT TANAMAN NILAM BERDASARKAN CITRA DAUN MENGGUNAKAN GLCM DAN SVM,” *J. Inform. Prog.*, vol. 17, no. 2, pp. 12–22, Sep. 2025, doi: 10.56708/progres.v17i2.469.
- [17] S. Akbar *et al.*, “IMPLEMENTASI K-MEANS DAN ANALISIS SENTIMEN KRITIK SARAN BERBASIS NLP PADA DATA MONEV BBPSDMP KOMINFO MAKASSAR,” *J. Inform. Prog.*, vol. 17, no. 2, pp. 36–43, Sep. 2025, doi: 10.56708/progres.v17i2.465.
- [18] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jégou, “Training data-efficient image transformers & distillation through attention,” *Proc. Mach. Learn. Res.*, vol. 139, pp. 10347–10357, 2021.
- [19] A. Dosovitskiy *et al.*, “an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale,” *ICLR 2021 - 9th Int. Conf. Learn. Represent.*, 2021.
- [20] J. I. Mujidah *et al.*, “KLASIFIKASI TINGKAT KEMATANGAN LADA MENGGUNAKAN ENSEMBLE LEARNING BERDASARKAN CITRA WARNA KULIT,” *J. Inform. Prog.*, vol. 17, no. 2, pp. 1–11, Sep. 2025, doi: 10.56708/progres.v17i2.467.