

KLASIFIKASI BUNGA HERBAL MENGGUNAKAN MODEL INCEPTIONV3 DAN VISION TRANSFORMER

Diaz Prisheila Dharmawan, Eva Yulia Puspaningrum, Afina Lina Nurlaili

Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasion “Veteran” Jawa Timur
Jalan Rungkut Madya, Gunung Anyar, Kec. Gunung Anyar, Surabaya, Jawa timur 60294
evapuspaningrum.if@upnjatim.ac.id

ABSTRAK

Bunga merupakan bagian penting dari tumbuhan yang memiliki berbagai manfaat medis. Namun, beberapa jenis bunga memiliki kemiripan dalam bentuk dan warna, sehingga sulit diklasifikasikan secara manual. Kesulitan dalam membedakan spesies yang mirip menjadi tantangan utama dalam klasifikasi otomatis, sehingga diperlukan model yang dapat menangkap informasi visual lebih kompleks. Penelitian ini mengusulkan pendekatan InceptionV3 dan Vision Transformer untuk ekstraksi fitur mendalam, dengan tujuan mengembangkan model klasifikasi yang lebih akurat dalam membedakan bunga herbal. Metode penelitian terdiri dari enam tahap utama, yaitu akuisisi data, pengubahan ukuran citra, ekstraksi fitur menggunakan InceptionV3 dan Vision Transformer, pembagian data, klasifikasi dengan MLP, dan evaluasi model. Model diuji dengan berbagai konfigurasi parameter, termasuk data split 90:10, learning rate 0.0001, batch size 64, serta arsitektur hidden layer 1024–512–256. Hasil eksperimen menunjukkan bahwa konfigurasi terbaik menghasilkan akurasi 78.29% dengan loss 0.6162. Evaluasi lebih lanjut menunjukkan bahwa model memiliki precision, recall, dan F1-score sebesar 78.4%, dengan prediksi terbaik pada kelas Common Daisy dan Black-Eyed Susan, di mana tingkat prediksi benar mencapai lebih dari 84%. Penelitian ini berkontribusi dalam pengembangan metode klasifikasi bunga herbal dengan memanfaatkan model InceptionV3 dan Vision Transformer untuk ekstraksi fitur mendalam. Hasil yang diperoleh menunjukkan efektivitas pendekatan ini dalam meningkatkan akurasi klasifikasi bunga yang memiliki karakteristik visual serupa.

Kata kunci : Bunga, InceptionV3, Transfer Learning, Vision Transformer, dan Multilayer Perceptron (MLP).

1. PENDAHULUAN

Bunga merupakan struktur reproduksi pada tumbuhan yang terdiri dari kelopak, mahkota, benang sari, dan putik, yang berfungsi dalam proses penyerbukan dan pembuahan [1]. Bunga memiliki manfaat medis yang berbeda-beda, walaupun beberapa jenisnya hanya dapat dinikmati keindahannya saja. Bunga termasuk hasil dari tunas yang mencakup batang dan daun, dengan perubahan pada bentuk, warna, dan susunannya untuk mendukung reproduksi [2]. Setiap bunga memiliki karakteristik unik dalam bentuk, warna, dan aroma. Namun, beberapa bunga menunjukkan kemiripan dalam bentuk dan warna meskipun berasal dari jenis yang berbeda, sehingga menyulitkan proses klasifikasi. Dalam konteks klasifikasi gambar, variasi bentuk, warna, dan tekstur bunga menjadi objek menarik untuk penelitian pengenalan pola dan klasifikasi citra. Fitur seperti warna dan tekstur sering digunakan untuk mendefinisikan dan menggambarkan citra bunga [3]. Penelitian ini menggunakan jenis bunga seperti Black-Eyed Susan, Calendula, Common Daisy, Coreopsis, dan Sunflower.

Secara manual, klasifikasi bunga biasanya dilakukan oleh ahli botani atau pertanian, namun metode ini memiliki keterbatasan karena membutuhkan waktu lama dan jumlah pakar yang terbatas. Tidak semua petani atau masyarakat umum memiliki kemampuan untuk mengidentifikasi bunga dengan tepat. Oleh karena itu, diperlukan teknologi yang mampu mengenali karakteristik bunga

berdasarkan gambar. Klasifikasi jenis bunga sangat kompleks dan cukup penting karena ada bunga yang sangat mirip dengan fungsi yang berbeda-beda [4]. Klasifikasi bunga berbasis teknologi dapat membantu mengatasi permasalahan tersebut sekaligus meningkatkan pemahaman akan keragaman hayati serta pentingnya pelestarian ekosistem. Namun, tantangan seperti kualitas gambar yang kurang optimal, latar belakang yang tidak mendukung, dan adanya noise tetap menjadi kendala dalam proses klasifikasi bunga.

Seiring perkembangan teknologi, klasifikasi bunga kini dapat dibantu oleh kecerdasan buatan. Pendekatan ini diharapkan mampu mempercepat proses klasifikasi sekaligus meningkatkan akurasi. Salah satu penelitian menggunakan Transfer Learning dengan model CNN pra-latih seperti MobileNetV2, VGG16, dan InceptionV3 untuk klasifikasi bunga iris [5]. Hasil penelitian menunjukkan bahwa MobileNetV2 memiliki akurasi tertinggi sebesar 91%, diikuti oleh InceptionV3 sebesar 87%, dan VGG16 sebesar 84%. Penelitian ini memberikan hasil yang menjanjikan dalam klasifikasi bunga dan dapat menjadi referensi untuk penelitian lanjutan. Penelitian lainnya mengombinasikan Transfer Learning dengan Vision Transformer (ViT) untuk klasifikasi penyakit daun tanaman [6]. Dalam pendekatan ini, model pra-latih seperti ResNet50, AlexNet, VGG16, VGG19, dan InceptionV3 digunakan untuk mengekstraksi fitur awal, sementara ViT memanfaatkan mekanisme self-attention untuk menangkap interaksi global antar fitur.

Model Multi-Layer Perceptron (MLP) dengan dua lapisan tersembunyi (dimensi 1024 dan 512) kemudian digunakan untuk menentukan kelas gambar. Hasil terbaik dicapai oleh kombinasi VGG19 dan ViT, yang menghasilkan akurasi 98,81% pada dataset PlantVillage dan 99,86% pada dataset Wheat Rust. Penelitian ini menghasilkan nilai yang nyaris sempurna dan menunjukkan potensi besar dalam penerapan Vision Transformer untuk meningkatkan akurasi klasifikasi.

Penelitian terkait yang menerapkan *deep learning* untuk klasifikasi jenis bunga menunjukkan hasil yang memuaskan. Model Transfer Learning dan Vision Transformer diharapkan dapat meningkatkan akurasi dan efisiensi klasifikasi bunga herbal. Transfer Learning adalah metode yang memanfaatkan model CNN yang telah dilatih pada dataset besar untuk diterapkan pada dataset yang lebih kecil. Pendekatan ini terbukti lebih efektif dibandingkan membangun model CNN dari nol, karena dapat mengatasi keterbatasan dataset, mempercepat proses pelatihan, dan meningkatkan akurasi [7]. Salah satu model yang sering digunakan dalam klasifikasi citra adalah InceptionV3, dengan 42 lapisan yang mampu menangani objek bertekstur kompleks secara efisien. Dibandingkan dengan VGG16, VGG19, dan ResNet50, model ini menggunakan faktorisasi filter besar menjadi lebih kecil untuk mengurangi komputasi tanpa kehilangan kemampuan ekstraksi fitur. Didukung teknik regulasi seperti Dropout, InceptionV3 unggul dalam keseimbangan akurasi dan sensitivitas, menjadikannya pilihan optimal untuk klasifikasi citra kompleks [8].

Vision Transformer (ViT) adalah model yang menunjukkan bagaimana Transformer dapat sepenuhnya menggantikan operasi konvolusi dalam jaringan saraf untuk dataset gambar berskala besar. Dengan mekanisme *self-attention*, ViT mampu menangkap hubungan global antar fitur dalam gambar, menjadikannya alat yang sangat andal untuk klasifikasi visual [9]. Pendekatan ini lebih unggul dibandingkan metode machine learning seperti SVM, yang menunjukkan akurasi kurang memuaskan dalam klasifikasi bunga daisy dan sunflower [10]. Selain itu, ViT lebih unggul dibanding model tradisional khususnya CNN dikarenakan ViT mampu memodelkan ketergantungan jangka panjang antara elemen input. Sehingga, dapat menangkap hubungan kompleks dalam gambar yang mungkin terlewatkan oleh Convolutional Neural Networks (CNN). Fleksibilitas ini sangat bermanfaat untuk tugas-tugas yang memerlukan pemahaman konteks global suatu gambar, seperti deteksi objek, segmentasi gambar, dan pengenalan pola yang lebih kompleks. Dengan kemampuan ini, ViT dapat memberikan hasil yang baik dalam tugas visi komputer [11].

Berdasarkan latar belakang, penelitian ini bertujuan untuk mengembangkan model klasifikasi bunga herbal menggunakan kombinasi Transfer Learning dan Vision Transformer. Pendekatan ini

dipilih karena Transfer Learning memanfaatkan model pra-latih yang telah dilatih pada dataset besar (seperti ImageNet) dan diaplikasikan pada dataset yang lebih kecil. Sedangkan, Vision Transformer (ViT) dengan mekanisme *self-attention* mampu menangkap hubungan global antar fitur dalam gambar untuk membedakan bunga dalam bentuk, warna, atau tekstur. Kombinasi ini menggabungkan kemampuan Transfer Learning dalam mengekstraksi fitur awal yang relevan dengan ViT dalam memahami pola mendalam. Dengan metode ini, diharapkan hasil penelitian dapat memberikan kontribusi dalam pengenalan pola dan klasifikasi citra bunga yang memiliki kemiripan visual.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian mengenai klasifikasi bunga telah banyak dilakukan dengan berbagai pendekatan, baik *machine learning* maupun *deep learning*. Studi terdahulu menunjukkan bahwa Transfer Learning meningkatkan efisiensi waktu, sementara Vision Transformer (ViT) menangkap pola visual kompleks untuk meningkatkan akurasi. Evaluasi terhadap penelitian sebelumnya membantu memahami kelebihan dan keterbatasan yang menjadi dasar model klasifikasi bunga herbal.

Penelitian oleh Tabbakh dan Barpanda [6] mengusulkan kombinasi Transfer Learning dan Vision Transformer untuk klasifikasi penyakit tanaman. Model ini terdiri dari empat tahap utama: akuisisi data, augmentasi gambar, ekstraksi fitur, dan klasifikasi. Dataset PlantVillage dan wheat rust digunakan untuk pelatihan serta evaluasi. Ekstraksi fitur dilakukan dalam dua tahap, menggunakan Transfer Learning dengan model seperti ResNet50, AlexNet, VGG16, VGG19, dan Inception V3, lalu dilanjutkan dengan Vision Transformer untuk menangkap fitur yang lebih mendalam. Klasifikasi dilakukan menggunakan Multilayer Perceptron (MLP). Hasil eksperimen menunjukkan bahwa kombinasi VGG19 + Vision Transformer memberikan akurasi tertinggi, yaitu 98,81% pada PlantVillage dataset dan 99,86% pada wheat rust dataset, melebihi model Transfer Learning konvensional. Studi ini menegaskan efektivitas Vision Transformer dalam meningkatkan klasifikasi gambar daun tanaman yang memiliki kompleksitas visual tinggi.

Penelitian oleh Gramandha Wega Intyanto [12] mengembangkan model klasifikasi bunga menggunakan Convolutional Neural Network (CNN) dengan dua arsitektur, arsitektur yang dirancang sendiri dan VGG16. Dataset yang digunakan mencakup lima jenis bunga: daisy, dandelion, rose, sunflower, dan tulip, dengan total 4.317 gambar yang diambil dari Kaggle. Proses pre-processing meliputi resize gambar menjadi 150x150 piksel dan pembagian dataset dengan rasio 80% untuk pelatihan serta 20% untuk pengujian. Hasil eksperimen menunjukkan bahwa arsitektur CNN yang dirancang sendiri

memiliki akurasi 62%, sedangkan VGG16 mencapai 80%, menunjukkan performa yang lebih baik dalam klasifikasi bunga.

Penelitian oleh Millenia Mudita Chandra dan Yoannita [10] menggunakan metode Support Vector Machine (SVM) dengan fitur warna Hue Saturation Value (HSV) untuk klasifikasi bunga matahari dan daisy. Dataset yang digunakan berasal dari Kaggle, dengan gambar yang telah di-resize ke ukuran 320x240 piksel sebelum dilakukan ekstraksi fitur HSV. Hasil penelitian menunjukkan bahwa metode SVM menghasilkan akurasi 63,36% dalam membedakan kedua jenis bunga yang memiliki kemiripan bentuk dan warna.

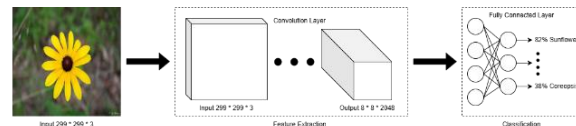
Penelitian oleh Han dan Guo [13] mengusulkan model otomatis untuk klasifikasi penyakit daun tanaman berkayu menggunakan Hierarchical Vision Transformer dan Transfer Learning. Model ini memanfaatkan multi-head attention untuk menangkap informasi kontekstual serta Multilayer Perceptron (MLP) sebagai klasifikator, dengan dataset yang terdiri dari 22 jenis penyakit daun tanaman berkayu yang diambil dari AI Challenger 2018. Untuk meningkatkan efisiensi, transfer learning diterapkan dengan pre-trained weight dari ImageNet. Hasil eksperimen menunjukkan bahwa model yang diusulkan mencapai akurasi lebih dari 85%, melebihi model berbasis CNN seperti AlexNet, GoogleNet, VGG, dan ResNet, serta model Vision Transformer konvensional. Dengan pendekatan hierarchical Vision Transformer, penelitian ini berhasil mengurangi kompleksitas komputasi tanpa mengorbankan akurasi, menunjukkan potensi besar dalam sistem otomatis deteksi penyakit tanaman dalam Precision Agriculture.

Penelitian sebelumnya umumnya menggunakan SVM dan CNN untuk klasifikasi gambar bunga, tetapi SVM kurang efektif dalam menangkap informasi global, sehingga kesulitan membedakan bunga dengan karakteristik yang serupa. Selain itu, masih terbatasnya penelitian yang mengintegrasikan Transfer Learning dan Vision Transformer untuk klasifikasi bunga. Sebagian besar dataset yang digunakan dalam studi terdahulu berfokus pada bunga umum, sedangkan penelitian ini menyoroti bunga herbal yang memiliki kemiripan visual.

2.2. Transfer Learning

Transfer learning merupakan metode yang telah dilatih pada suatu tugas digunakan kembali untuk membantu penyelesaian tugas yang berbeda. Metode ini sangat cocok dalam kasus di mana data penelitian yang digunakan untuk melatih model jumlahnya terbatas [7]. Transfer learning menjadi salah satu metode yang populer dalam pengembangan model, khususnya *deep learning*. Hal tersebut dikarenakan pelatihan model dari awal membutuhkan komputasi yang besar serta data dalam jumlah besar. *Transfer learning* bekerja dengan memanfaatkan pengetahuan dan fitur yang telah dipelajari oleh model pada dataset sumber untuk diterapkan pada dataset target. Pada

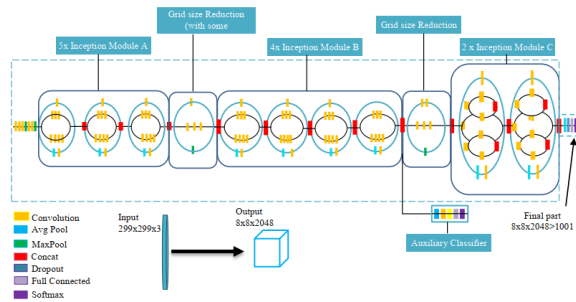
metode ini, model awal dilatih terlebih dahulu menggunakan kumpulan data besar, kemudian parameter model yang telah dilatih sebelumnya dibekukan dan ditransfer ke model target. Gambar 1, menunjukkan proses Transfer Learning dalam melatih model baru untuk mengklasifikasikan kumpulan data baru. Pendekatan ini mengatasi tantangan seperti keterbatasan dataset, penghematan sumber daya komputasi, kecepatan proses pelatihan, dan meningkatkan kinerja model. Saat ini banyak model CNN yang telah dilatih sebelumnya, seperti AlexNet, VGG, MobileNet, Inception, dan lainnya [14].



Gambar 1. Transfer learning

2.3. InceptionV3

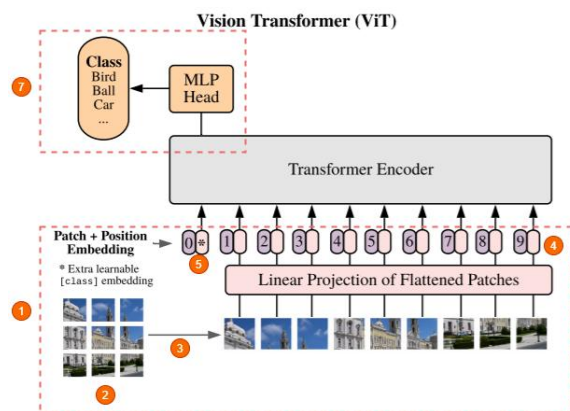
InceptionV3 adalah salah satu arsitektur Convolutional Neural Network (CNN) yang dikembangkan Google untuk tugas klasifikasi gambar dalam bidang visi komputer. InceptionV3 merupakan versi ketiga dari arsitektur InceptionV1 dan V2 yang diperkenalkan pada tahun 2015 [15]. InceptionV3 dirancang untuk meningkatkan kinerja dan efisiensi tugas klasifikasi gambar melalui pendekatan inovatif dalam desain arsitekturnya. Arsitektur InceptionV3 terdiri dari beberapa modul yang dirancang untuk menangkap fitur lokal dan global dalam gambar. Model ini menggunakan teknik faktorisasi kovolusi untuk mengurangi jumlah parameter tanpa mengorbankan akurasi. InceptionV3 memiliki beberapa keunggulan utama yang membuatnya menonjol dalam jaringan saraf konvolusional. Pertama, faktorisasi spasial konvolusi besar menjadi kecil, seperti 5x5, dipecah menjadi dua lapisan konvolusi 3x3, yang secara signifikan mengurangi kompleksitas komputasi. Kedua, penggunaan batch normalization membantu menstabilkan proses pelatihan dengan mengurangi perubahan distribusi input, sehingga mempercepat konvergensi dan meningkatkan kinerja model. Terakhir, InceptionV3 dilatih sebelumnya pada dataset besar seperti ImageNet, yang memungkinkan model ini dapat digunakan pada dataset yang lebih kecil dengan teknik fine-tuning, sehingga meningkatkan fleksibilitas dan efisiensi dalam berbagai aplikasi. Salah satu inovasi utama dari Inception V3 adalah kemampuannya untuk mengekstrak fitur dari berbagai skala. Dengan menggabungkan fitur dari beberapa saluran, model ini dapat menangkap informasi yang lebih beragam dan relevan dari data input, yang sangat penting dalam tugas-tugas seperti klasifikasi gambar dan deteksi objek [16]. Arsitektur InceptionV3 yang terdiri dari beberapa lapisan dapat dilihat pada gambar 2.



Gambar 2. InceptionV3

2.4. Vision Transformer

Vision Transformer (ViT), sebuah arsitektur berbasis Transformer yang awalnya dikembangkan untuk Natural Language Processing (NLP) dan kini banyak diaplikasikan pada tugas-tugas visi komputer [11]. ViT menggunakan mekanisme self-attention untuk menangkap hubungan global antar fitur dalam gambar, sehingga memberikan kinerja yang lebih baik dibandingkan CNN dalam beberapa tugas visual. Pada tugas klasifikasi citra ViT memecah gambar menjadi potongan kecil (patches) untuk diklasifikasi, memberikan performa yang kompetitif pada dataset besar seperti ImageNet. ViT memerlukan dataset besar untuk pelatihan agar dapat mencapai hasil terbaik, karena kurangnya inductive bias seperti pada CNN. Pada gambar 3, menunjukkan arsitektur atau alur kerja Vision Transformer.



Gambar 3. Vision Transformer

2.5. Patch Partitioning

Patch partitioning adalah proses pembagian gambar menjadi potongan-potongan kecil yang disebut patch, dengan ukuran yang tetap dan seragam. Pada tahap ini, gambar diinputkan dengan dimensi $H \times W \times C$ (di mana H adalah tinggi, W adalah lebar, dan C adalah jumlah saluran warna). Kemudian, gambar dibagi menjadi N petak persegi dengan bentuk (p, p, c) , di mana p adalah parameter yang telah ditetapkan sebelumnya, dalam urutan raster (kiri ke kanan, atas ke bawah). Patch ini biasanya memiliki dimensi seperti 16×16 atau 32×32 . Semakin banyak petak berarti semakin mudah melatih jaringan karena petak-petak itu sendiri semakin kecil. Jumlah total patch N , dapat dihitung dengan persamaan (1):

$$N = \frac{h * w}{p^2} \quad (1)$$

2.6. Linear Projection

Setelah dibagi menjadi beberapa patch, patch tersebut akan diratakan menjadi vektor satu dimensi. Tujuannya untuk mengubah representasi patch ke dimensi yang sesuai untuk pemrosesan oleh transformer. Setiap vektor patch dikalikan dengan bobot berat (w) dan bias (b). Perhitungan ini dituliskan dalam persamaan (2), di mana x merupakan inputan untuk linear projection.

$$z = x * w + b \quad (2)$$

Agar model dapat memahami posisi patch di dalam gambar, position embedding ditambahkan ke setiap patch embedding. Penambahan position embedding membantu model memahami struktur spasial gambar. Tensor ini menjadi input ke transformer encoder, proses position embedding dapat dilihat dalam persamaan (3) dan (4).

$$z' = z + PE \quad (3)$$

$$PE(p, 2i) = \sin\left(\frac{p}{10000^{\frac{2i}{D}}}\right) \quad (4)$$

Sebuah token khusus bernama [CLS] (classification token) ditambahkan di awal urutan patch. Setelah token ini ditambahkan, urutan patch menjadi $(n+1, d)$ dengan n adalah jumlah patch dan $+1$ tambahan untuk token [CLS]. Hasilnya (Z_0) memiliki bentuk $(n+1, d)$ menjadi input awal untuk transformer encoder, dituliskan dalam persamaan (5).

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{pos} \quad (5)$$

2.7. Transformer Encoder

Transformer encoder terdiri dari Multi-Head Attention (MHA), Multilayer Perceptron (MLP), dan lapisan normalisasi. Multi-Head Attention (MHA) yang berfungsi sebagai perhitungan hubungan antara semua patch untuk memahami bagaimana informasi antar-patch saling terkait. MHA sebagai mekanisme *self-attention* melibatkan proses perhitungan matriks berat untuk masing-masing *query* (q), *key* (k), dan *value* (v), dituliskan dalam persamaan (6).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_h}}\right) \quad (6)$$

Setelah didapatkan nilai *attention* (A), *self-attention* dihitung dengan mengalikan matriks *attention* (A) dengan V (*value*). Hasil *self-attention* akan digabungkan menjadi satu dan menghasilkan output akhir MHA(z) yang dituliskan dalam persamaan (7).

$$MHA(z) = [SA_1(z); SA_2(z); \dots; SA_h(z)] \quad (7)$$

2.8. MLP

Hasil dari MHA akan dilalui lapisan normalisasi untuk meningkatkan stabilitas selama pelatihan, perhitungan lapisan ini dituliskan dalam persamaan (8).

$$X_{normalisasi} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (8)$$

Setelah itu, hasil representasi akan digunakan untuk klasifikasi menggunakan Multilayer Perceptron (MLP). MLP berfungsi untuk memproses fitur yang dihasilkan oleh MHA dengan transformasi non-linear menggunakan jaringan saraf berlapis. Proses ini melibatkan *hidden layer* MLP, di mana token dilewatkan ke lapisan tersembunyi dengan fungsi aktivasi non-linear (ReLU) yang dituliskan dalam persamaan (9).

$$H = \text{ReLU}(X.W1 + b1) \quad (9)$$

2.9. Confusion Matrix

Confusion matrix adalah alat evaluasi yang digunakan untuk mengukur kinerja model klasifikasi dalam machine learning. Matriks ini menyajikan perbandingan antara prediksi model dan nilai aktual dari data dalam bentuk tabel [17]. Confusion matrix terdiri dari empat elemen utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Dari keempat elemen tersebut, beberapa metrik evaluasi dapat dihitung untuk menilai kinerja model, seperti akurasi, precision, recall, dan F1-score. Akurasi mengukur rasio prediksi yang benar terhadap total prediksi, sebagaimana dinyatakan dalam persamaan (10). Precision mengukur rasio prediksi positif yang benar terhadap total prediksi positif, sesuai dengan persamaan (11). Recall mengukur rasio prediksi positif yang benar terhadap total aktual positif, sebagaimana dinyatakan dalam persamaan (12). F1-score merupakan rata-rata harmonik antara precision dan recall, sebagaimana dinyatakan dalam persamaan (13).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

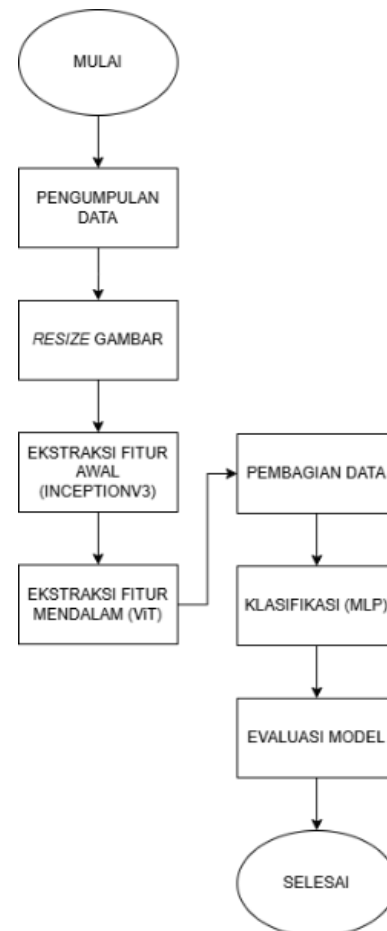
$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (12)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

3. METODE PENELITIAN

Dalam metode penelitian yang dikerjakan adapun tahapan untuk menyelesaikan masalah, berikut diagram alur metodologi penelitian ini:



Gambar 4. Alur metode penelitian

Penelitian ini mengusulkan klasifikasi bunga herbal dengan model ekstraksi fitur mendalam berdasarkan model Transfer Learning dan Vision Transformer, seperti pada gambar 4. Meliputi enam tahap utama, sebagai berikut:

- Pengumpulan Data:** Dataset Flower Images digunakan sebagai input utama untuk melatih dan menguji model. Terdapat lima jenis bunga yang berbeda dalam dataset Flower Images (Black Eyed Susan, Calendula, Common Daisy, Coreopsis, dan Sunflower).
- Resize Gambar:** Dataset Flower Images yang digunakan memiliki ukuran yang berbeda-beda, sehingga perlu mengubah ukuran gambar untuk memastikan semua gambar sama dan sesuai dengan model yang diusulkan.
- Ekstraksi Fitur:** Proses ini mengidentifikasi dan mengekstraksi informasi dari dataset, seperti tepi, sudut, dan tekstur yang digunakan sebagai input klasifikasi. Terdapat dua tahap untuk mengekstraksi fitur:
 - Tahap pertama: Model InceptionV3 berbasis pra-latih digunakan untuk ekstraksi fitur awal.
 - Tahap kedua: Vision Transformer digunakan untuk mengekstraksi fitur mendalam dari fitur awal.

- d. Pembagian Data: Setelah melewati ekstraksi fitur, dataset akan dibagi menjadi data latih dan uji untuk melatih dan menguji model.
- e. Klasifikasi: Data yang telah siap kemudian akan dilatih dan diuji pada Multilayer Preceptron (MLP) sebagai klasifikator.
- f. Evaluasi Model: Model akan dievaluasi setelah dilatih dan diuji menggunakan metrik akurasi, precision, recall, dan F1-score untuk mengukur kinerja model.

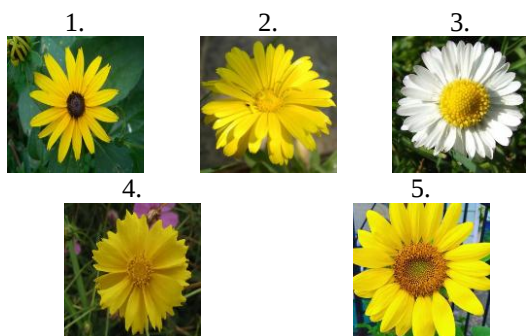
4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari situs Kaggle, yaitu dataset Flower Images. Dataset ini digunakan untuk mengevaluasi model yang diusulkan dalam mengekstraksi fitur mendalam dari daun.

Tabel 1. Dataset bunga herbal.

No.	Nama Kelas	Jumlah Data
1.	Black Eyed Susan	820
2.	Calendula	820
3.	Common Daisy	820
4.	Coreopsis	820
5.	Sunflower	820
Total:		4.100

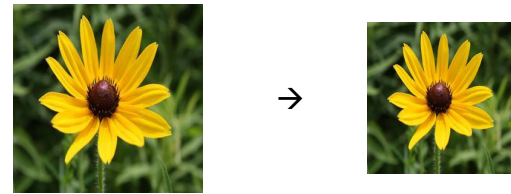


Gambar 5. Contoh gambar dataset

Tabel 1 dan gambar 5 menyajikan contoh dataset yang digunakan dalam penelitian ini, mencakup kelas bunga, jumlah gambar per kelas, serta contoh gambar bunga. Total dataset ini terdiri dari 4.100 gambar, dengan 820 gambar untuk setiap kelas.

4.2. Resize Gambar

Dalam dataset penelitian ini, terdapat perbedaan resolusi antara data per kelas, sehingga diperlukan penyesuaian ukuran agar model dapat mengenali pola visual secara konsisten. Jika proses resize tidak diterapkan, model akan menerima input gambar dengan ukuran bervariasi, sehingga dapat menyebabkan distorsi fitur dan performa klasifikasi yang kurang optimal. Semua gambar diubah ukurannya menjadi dimensi 299x299 piksel sesuai dengan ukuran input standar untuk ekstraksi fitur Transfer Learning model InceptionV3 yang dapat dilihat pada gambar 6.



Gambar 6. Hasil resize gambar

4.3. Ekstraksi Fitur

Dalam penelitian ini, ekstraksi fitur dilakukan dalam dua tahap, yaitu ekstraksi fitur awal menggunakan Transfer Learning dan ekstraksi fitur mendalam menggunakan Vision Transformer.

4.4. Ekstraksi fitur Transfer Learning

Ekstraksi fitur awal dilakukan menggunakan model InceptionV3, yang hanya digunakan untuk mengekstraksi fitur tanpa terlibat dalam proses pelatihan dan pengujian. Model ini digunakan hingga sebelum mencapai lapisan fully connected. InceptionV3 memiliki ukuran default untuk citra input sebesar 299*299*3, dengan hasil ekstraksi fitur akhir berukuran 8*8*2.048. Hasil ekstraksi fitur awal menggunakan Transfer Learning menghasilkan fitur dengan dengan ukuran 8*8*2.048, seperti pada gambar 7. Fitur yang telah diekstraksi kemudian digunakan sebagai input untuk tahap berikutnya, yaitu ekstraksi fitur mendalam menggunakan Vision Transformer (ViT).

```
Contoh fitur (X): (1, 8, 8, 2048)
[[[1.82806700e-03 0.00000000e+00 0.00000000e+00 ... 1.08434737e+00
  4.71200228e-01 1.44685733e+00]
 [0.00000000e+00 0.00000000e+00 0.00000000e+00 ... 1.10115266e+00
  3.84808093e-01 1.90036225e+00]
 [9.20824409e-02 0.00000000e+00 0.00000000e+00 ... 1.21272254e+00
  2.65876621e-01 2.46489930e+00]
 ...
```

Gambar 7. Hasil ekstraksi fitur awal

4.5. Ekstraksi Fitur Vision Transformer

Hasil ekstraksi fitur awal menggunakan Transfer Learning dengan model InceptionV3 berukuran 8*8*2.048 akan diproses lebih lanjut dalam ekstraksi fitur mendalam menggunakan Vision Transformer (ViT). Model ini akan memecah gambar menjadi patch-patch kecil, lalu meratakan setiap patch menjadi vektor satu dimensi menggunakan linear projection. Setelah itu, position embedding ditambahkan untuk memberikan informasi posisi setiap patch. Selanjutnya, transformer encoder digunakan untuk menghitung nilai self-attention, yang memungkinkan model memahami hubungan antar fitur dalam gambar sebelum akhirnya diklasifikasikan menggunakan Multilayer Perceptron (MLP). Hasil ekstraksi fitur mendalam menggunakan Vision Transformer menghasilkan fitur dengan dimensi 256, dapat dilihat pada gambar 8.

Beberapa data fitur mendalam (X):

```
[[ [ 2.8013315  1.7459958 -1.1988602 ... -1.0413414  0.05375116
    0.9364154 ]
  [ 2.4375832  1.9381729 -0.6341806 ... -1.2329541  0.21589257
    0.0530435 ]
  [ 2.5260072  1.409654  -1.0800761 ... -0.73330164  0.91704816
    0.80861545]
  [ 1.5619397  1.8538551  0.36586395 ... -0.55075854  0.2572539
    0.34815377]
  [ 1.072824  1.0860347 -0.6860157 ... -0.62413496 -0.23182991
    -0.24125746]]
```

Gambar 8. Hasil ekstraksi fitur mendalam

4.6. Pembagian Data

Membagi data menjadi data train dan test untuk pelatihan dan pengujian model, dengan tujuan memastikan bahwa model dapat mempelajari pola dari data train serta menguji akurasi prediksi menggunakan data test. Dalam penelitian ini menerapkan pembagian data dengan rasio 90:10. Rasio ini dipilih karena dampaknya terhadap kinerja model dalam proses pelatihan dan pengujian [18]. Sebanyak 3.690 gambar sebagai data pelatihan dan 410 gambar untuk data pengujian.

4.7. Klasifikasi Multilayer Perceptron (MLP)

Pengklasifikasi yang digunakan adalah Multilayer Perceptron (MLP). Setelah encoder Transformer menghasilkan fitur-fitur mendalam pada tahap sebelumnya, fitur-fitur tersebut dimasukkan ke dalam pengklasifikasi untuk menentukan kelas gambar yang sesuai. Vektor fitur mendalam digunakan sebagai input ke pengklasifikasi MLP, di mana lapisan input memiliki 256 neuron, yang sesuai dengan dimensi vektor fitur mendalam. Lapisan tersembunyi dengan dimensi yang ditentukan akan digunakan, dan lapisan terakhir terdiri dari 5 neuron, yang merepresentasikan jumlah kelas dalam dataset. Selain itu, lapisan Dropout ditambahkan di setiap lapisan tersembunyi untuk mencegah terjadinya *overfitting*.

4.8. Evaluasi Model

Semua percobaan dalam penelitian ini dilakukan pada laptop Windows 10 versi 22H2, dengan sistem operasi 64-bit dan prosesor Intel(R) Core(TM) i7 2.40GHz dengan RAM 8 GB. Python 3.12.6 digunakan sebagai bahasa pemrograman dan semua percobaan dilakukan pada Visual Studio Code.

Tabel 2. Parameter pengujian

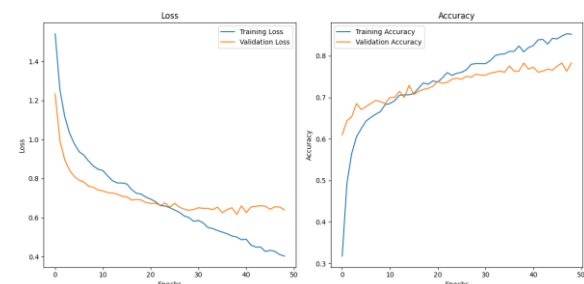
Parameter	Value
Pembagian Data	90:10
Learning Rate	0.0001
Batch Size	64
Hidden Layers	1024, 512, 256
Epoch	100
Optimizer	Adam

Setelah proses pelatihan model Multilayer Perceptron (MLP) berdasarkan parameter pengujian pada tabel 2, langkah selanjutnya adalah melakukan evaluasi terhadap kinerja model yang telah dibangun. Evaluasi ini dilakukan untuk menilai seberapa baik model dalam mengklasifikasikan data baru, dengan

menggunakan metrik-metrik evaluasi yang telah ditentukan sebelumnya.

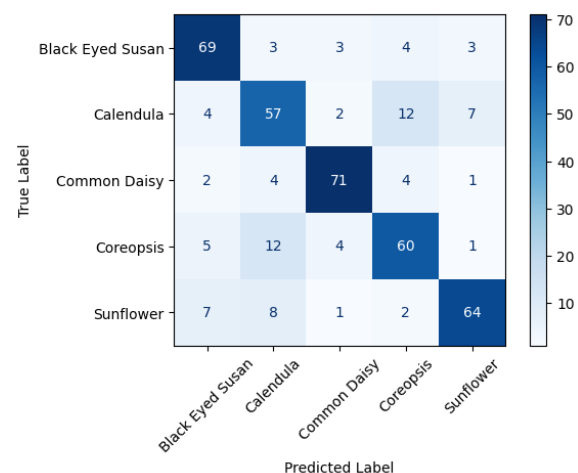
4.9. Hasil Pengujian

Berdasarkan skenario pengujian yang disajikan pada tabel 2, terdapat beberapa parameter utama yang diuji, yaitu split data, learning rate, batch size, arsitektur layer MLP, epoch, dan optimizer Adam. Dari pengujian yang dilakukan, hasil menunjukkan kombinasi parameter split data 90:10, *learning rate* 0.0001, *batch size* 64, struktur *hidden layers* 1024-512-256, *epoch* 100, dan optimizer Adam. Model ini mencapai akurasi sebesar 78,29% dengan nilai loss 0,6162. Adapun nilai precision, recall, dan F1-score yang diperoleh adalah 78%.



Gambar 9. Grafik accuracy dan loss hasil terbaik

Grafik pada gambar 9 menunjukkan perubahan loss dan akurasi model selama 49 epoch pelatihan. Model mengalami *early stopping* pada epoch ke-49 karena akurasi validasi tidak menunjukkan peningkatan signifikan. Performa terbaik dicapai pada epoch ke-39 dengan akurasi validasi 78,29% dan loss 0,6162. Setelahnya, loss validasi berfluktuasi tanpa perbaikan, mengindikasikan potensi *overfitting*, di mana model kurang mampu menggeneralisasi data baru. Oleh karena itu, bobot model dikembalikan ke kondisi terbaik pada epoch ke-39 untuk memastikan performa optimal.



Gambar 10. Confusion matrix hasil terbaik

Gambar 10 menunjukkan *confusion matrix* hasil evaluasi model klasifikasi untuk lima kelas bunga

herbal. Model berhasil mengklasifikasikan sebagian besar citra dengan benar, dengan Common Daisy sebagai kelas dengan prediksi terbaik, diikuti oleh Black Eyed Susan dan Sunflower. Kesalahan klasifikasi terutama terjadi antara Coreopsis dan Calendula. Secara keseluruhan, model menunjukkan kinerja klasifikasi yang stabil, meskipun masih terdapat peluang untuk meningkatkan akurasi pada kelas dengan karakteristik visual serupa.

Classification Report:				
	precision	recall	f1-score	support
0	0.79	0.84	0.82	82
1	0.68	0.70	0.69	82
2	0.88	0.87	0.87	82
3	0.73	0.73	0.73	82
4	0.84	0.78	0.81	82
accuracy			0.78	410
macro avg	0.78	0.78	0.78	410
weighted avg	0.78	0.78	0.78	410

Gambar 11. Classification report

Pada gambar 11, model klasifikasi mencapai akurasi keseluruhan 78% dari 410 data uji. Evaluasi metrik *precision*, *recall*, dan *F1-score* menunjukkan bahwa kelas Common Daisy memiliki performa terbaik (*precision* 0,88, *recall* 0,87, *F1-score* 0,87), sementara kelas Calendula masih memiliki ruang untuk peningkatan (*precision* 0,68, *recall* 0,70, *F1-score* 0,69). Kinerja model tetap stabil di seluruh kelas, dengan rata-rata *precision*, *recall*, dan *F1-score* sebesar 0,78, mengindikasikan kemampuan klasifikasi yang konsisten.

5. KESIMPULAN DAN SARAN

Berdasarkan tahapan yang telah dilakukan dalam penelitian ini, klasifikasi bunga herbal menggunakan ekstraksi fitur model InceptionV3 dan Vision Transformer. Penelitian ini berhasil mengembangkan model klasifikasi bunga herbal dengan kombinasi model InceptionV3 sebagai ekstraksi fitur awal dan Vision Transformer sebagai ekstraksi fitur mendalam. Serta, Multilayer Perceptron sebagai pengklasifikasi untuk menentukan kelas bunga herbal. Hasil menunjukkan bahwa kombinasi kedua metode ini menghasilkan akurasi tertinggi sebesar 78.29%, nilai *precision*, *recall*, dan *F1-score* yang seimbang di angka 78.4%. Dengan rata-rata prediksi benar tertinggi lebih dari 84% pada kelas Common Daisy dan Black-Eyed Susan. Meskipun model menunjukkan hasil yang baik, terdapat tantangan dalam membedakan beberapa kelas tertentu, seperti Calendula yang sering salah diprediksi sebagai Coreopsis. Hal ini menunjukkan bahwa karakteristik visual bunga herbal memiliki kemiripan yang dapat menjadi kendala dalam klasifikasi.

Berdasarkan hasil penelitian yang dilakukan, terdapat beberapa saran yang dapat dijadikan pertimbangan untuk penelitian selanjutnya. Pada penelitian selanjutnya, disarankan untuk memperbanyak dataset dan diperlukan dataset yang

lebih bersih dan representatif, dengan cara mengurangi gambar yang memiliki noise tinggi, menghindari latar belakang yang kompleks. Penelitian ini dapat menjadi dasar untuk penelitian lanjutan di bidang terkait, dengan fokus pada pengembangan model menjadi sebuah sistem atau aplikasi.

DAFTAR PUSTAKA

- [1] M. Chen and J. Z. Zhang, "Flowering and Flower Development in Plants," *Agronomy*, vol. 14, no. 2, pp. 16–19, 2024, doi: 10.3390/agronomy14020256.
- [2] I. Gilang Perwati, N. Suarna, and T. Suprapti, "Analisis Klasifikasi Gambar Bunga Lily Menggunakan Metode Convolutional Neural Network (Cnn) Dalam Pengolahan Citra," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, pp. 2908–2915, 2024, doi: 10.36040/jati.v8i3.9193.
- [3] R. Rafie, "Klasifikasi Bunga Menggunakan Naïve Bayes Berdasarkan Fitur Warna Dan Texture," *J. Sains Komput. Dan Teknol. ...*, vol. 4, no. 1, pp. 90–94, 2021, [Online]. Available: <http://journal.umpalangkaraya.ac.id/index.php/jsakti/article/view/3173>
- [4] L. Farokhah, "Implementasi K-Nearest Neighbor Untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna Rgb Implementation of K-Nearest Neighbor for Flower Classification With Extraction of Rgb Color Features," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1129–1136, 2020, doi: 10.25126/jtiik.202072608.
- [5] A. Sabri, "Transfer Learning Model CNN Pralatih Untuk Klasifikasi Bunga Iris Berbasis Citra," *Semin. Nas. Teknol. Inf. dan Komun. STI&K*, vol. 7, no. 1, pp. 2581–2327, 2023.
- [6] A. Tabbakh and S. S. Barpanda, "A Deep Features Extraction Model Based on the Transfer Learning Model and Vision Transformer 'TLMViT' for Plant Disease Classification," *IEEE Access*, vol. 11, no. May, pp. 45377–45392, 2023, doi: 10.1109/ACCESS.2023.3273317.
- [7] M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," *Technologies*, vol. 11, no. 2, pp. 1–14, 2023, doi: 10.3390/technologies11020040.
- [8] O. Iparraquirre-Villanueva, V. Guevara-Ponce, O. R. Paredes, F. Sierra-Liñan, J. Zapata-Paulini, and M. Cabanillas-Carbonell, "Convolutional Neural Networks with Transfer Learning for Pneumonia Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, pp. 544–551, 2022, doi: 10.14569/IJACSA.2022.0130963.
- [9] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in Vision: A Survey," *ACM Comput. Surv.*, vol. 54, no. 10, pp. 1–30, 2022, doi: 10.1145/3505244.
- [10] M. M. Chandra and Yoanita, "Klasifikasi Jenis Bunga Menggunakan Metode Svm Berdasarkan Citra Dengan Fitur Hsv," *J. Indones. Sos.*

- Teknol.*, vol. 4, no. 2, pp. 255–264, 2023.
- [11] K. Han *et al.*, “A Survey on Vision Transformer,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, 2023, doi: 10.1109/TPAMI.2022.3152247.
- [12] G. W. Intyanto, “Klasifikasi Citra Bunga dengan Menggunakan Deep Learning: CNN (Convolution Neural Network),” *J. Arus Elektro Indones.*, vol. 7, no. 3, p. 80, 2021, doi: 10.19184/jaei.v7i3.28141.
- [13] D. Han and C. Guo, “Automatic classification of ligneous leaf diseases via hierarchical vision transformer and transfer learning,” *Front. Plant Sci.*, vol. 14, no. January, pp. 1–12, 2023, doi: 10.3389/fpls.2023.1328952.
- [14] C. Narvekar and M. Rao, “Flower classification using CNN and transfer learning in CNN-Agriculture Perspective,” *Proc. 3rd Int. Conf. Intell. Sustain. Syst. ICISS 2020*, pp. 660–664, 2020, doi: 10.1109/ICISS49785.2020.9316030.
- [15] A. E. Minarno, L. Aripa, Y. Azhar, and Y. Munarko, “Classification of Malaria Cell Image Using Inception-V3 Architecture,” *Int. J. Informatics Vis.*, vol. 7, no. 2, pp. 273–278, 2023, doi: 10.30630/joiv.7.2.1301.
- [16] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, *A review of convolutional neural networks in computer vision*, vol. 57, no. 4. Springer Netherlands, 2024. doi: 10.1007/s10462-024-10721-6.
- [17] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.
- [18] H. Bichri, A. Chergui, and M. Hain, “Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, pp. 331–339, 2024, doi: 10.14569/IJACSA.2024.0150235.