PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 1, (2025) E-ISSN: 3063-5845

Perbandingan Convolutional Neural Network dan Vision Transformer Untuk Klasifikasi Penyakit Daun Pada Tomat

Suryatna Sacadibrata[1], Taufiqur Rahman[2], Sajarwo Anggai[3]
[1,2,3)] Program Studi Teknik Informatika S-2, Universitas Pamulang
Email: [1] sacadibrata@gmail.com, [2] taufiqur.rahman@msn.com, [3] sajarwo@gmail.com

Abstrak--Penyakit daun pada tanaman tomat, seperti early blight dan late blight, dapat menyebabkan penurunan hasil panen yang signifikan, mengancam keberlanjutan usaha pertanian dan ketahanan pangan global. Deteksi dini menjadi langkah penting untuk meminimalkan kerugian yang dialami petani. Penelitian ini bertujuan mengevaluasi kinerja dua model deep learning, yaitu ResNet-152 dan Vision Transformer (ViT), dalam mendeteksi dan mengklasifikasikan penyakit daun tomat. Dataset yang digunakan terdiri dari 10.000 citra daun tomat yang diklasifikasikan ke dalam 10 kategori penyakit, dengan proporsi data pelatihan dan validasi sebesar 80:20. Augmentasi data dilakukan untuk meningkatkan variasi dataset. Model dilatih dengan parameter yang dioptimalkan dan dievaluasi menggunakan metrik akurasi, loss, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Vision Transformer (ViT) lebih unggul dibandingkan ResNet-152, dengan akurasi 98,60% dan loss 0,0568, sedangkan ResNet-152 mencapai akurasi 96,09% dan loss 0,1683. Mekanisme self-attention pada ViT memungkinkan model menangkap pola visual kompleks, menghasilkan generalisasi lebih baik pada dataset dengan variasi tinggi. Sebaliknya, ResNet-152 menunjukkan performa solid tetapi lebih rentan terhadap kesalahan pada kategori penyakit dengan gejala visual yang serupa. Implikasi dari penelitian ini adalah bahwa Vision Transformer dapat diimplementasikan untuk mendeteksi penyakit daun tomat secara otomatis, membantu petani mengambil langkah pengendalian lebih dini, sehingga mengurangi kerugian hasil panen dan meningkatkan produktivitas. Penelitian selanjutnya disarankan untuk mengoptimalkan parameter Vision Transformer dan mengembangkan teknik augmentasi data yang lebih baik untuk meningkatkan akurasi pada kategori penyakit dengan gejala yang sulit dibedakan.

Kata Kunci--*Tomato leaf disease, Vision Transformer, ResNet-152, deep learning, classification, self-attention, residual learning, data augmentation, pattern recognition.*

I. PENDAHULUAN

Tanaman tomat (*Solanum lycopersicum*) adalah salah satu komoditas hortikultura yang memiliki peran penting dalam mendukung sektor pertanian serta ketahanan pangan secara global. Sebagai salah satu sayuran yang banyak dikonsumsi di seluruh dunia, tomat tidak hanya berperan dalam memenuhi kebutuhan pangan masyarakat, tetapi juga menjadi sumber nutrisi penting bagi kesehatan manusia. Tomat mengandung vitamin C, asam folat, serta senyawa antioksidan seperti likopen, yang diketahui memberikan berbagai manfaat kesehatan, termasuk membantu mencegah penyakit degeneratif dan meningkatkan sistem imunitas tubuh.

Permintaan pasar terhadap tomat terus mengalami peningkatan, baik untuk konsumsi langsung maupun sebagai bahan baku dalam berbagai produk olahan, seperti saus, pasta, dan jus. Namun, keberhasilan budidaya tomat sering kali terhambat oleh berbagai tantangan, terutama serangan penyakit tanaman. Beberapa penyakit utama yang menyerang daun tomat meliputi *bacterial spot*, *early blight*, *late blight*, *leaf mold*, *septoria leaf spot*, *spider mites*, *target spot*, *mosaic virus*, dan *yellow leaf curl virus*. Serangan penyakit-penyakit ini tidak hanya berdampak pada kualitas hasil panen, tetapi juga mengancam keberlanjutan usaha pertanian. Akibatnya, ketersediaan tomat di pasar dapat terganggu, yang berpotensi memicu lonjakan harga.

Untuk mengatasi masalah ini, teknologi berbasis kecerdasan buatan (*Artificial Intelligence*/AI) mulai diterapkan dalam bidang pertanian, khususnya untuk mendukung proses identifikasi dan klasifikasi penyakit pada tanaman. Teknologi ini bertujuan meningkatkan efisiensi serta akurasi diagnosis, sehingga petani dapat mengambil tindakan pengendalian dengan lebih cepat dan tepat. Dua pendekatan yang populer dalam klasifikasi gambar adalah *Convolutional Neural Network* (CNN) dengan arsitektur *ResNet* dan *Vision Transformer (ViT)*.

Arsitektur *Convolutional Neural Network* (CNN), seperti *ResNet*, memiliki berbagai varian, salah satunya adalah *ResNet-152*,

yang menggunakan 152 lapisan dalam jaringannya. *ResNet-152* memenangkan kompetisi *ILSVRC* pada tahun 2015 dengan tingkat kesalahan *top-5* sebesar 3,6%, setara dengan kemampuan manusia dalam melakukan klasifikasi. Arsitektur ini terdiri dari lima tahap utama proses konvolusi yang mencakup sejumlah blok konvolusi serta satu lapisan *fully connected*. *ResNet-152* telah terbukti sangat efektif dalam tugas-tugas analisis citra berkat kemampuannya mengenali pola visual melalui operasi konvolusi. Dengan memanfaatkan mekanisme *residual learning*, *ResNet-152* memungkinkan pelatihan model yang sangat dalam tanpa terkendala masalah *vanishing gradient*. Model ini mampu mengekstraksi fitur-fitur visual yang penting, sehingga sangat andal dalam membedakan berbagai jenis penyakit daun dengan tingkat akurasi yang tinggi.

Sebaliknya, *Vision Transformer (ViT)* adalah pendekatan inovatif yang memanfaatkan mekanisme *self-attention* untuk menganalisis hubungan spasial dalam data visual. Pendekatan ini memiliki kemampuan untuk memahami pola-pola kompleks dalam gambar dengan cara yang lebih fleksibel dibandingkan dengan CNN konvensional. Dengan menggunakan representasi gambar berbasis *patch*, *ViT* dapat mengolah gambar secara efisien, terutama pada dataset dengan tingkat variasi yang tinggi. *Vision Transformer* juga memiliki keunggulan dalam menangani data dengan resolusi tinggi tanpa memerlukan arsitektur tambahan yang rumit. Selain itu, pendekatan ini mampu mengurangi risiko hilangnya informasi penting selama proses ekstraksi fitur. *ViT* telah menunjukkan performa yang kompetitif dalam berbagai tugas klasifikasi citra, terutama pada dataset yang besar dan beragam.

Penelitian ini bertujuan untuk menganalisis kinerja model *Convolutional Neural Network* (CNN) *ResNet* dan *Vision Transformer* (*ViT*) dalam mengidentifikasi penyakit pada daun tomat. Dataset yang digunakan terdiri dari gambar daun tomat yang telah diklasifikasikan berdasarkan jenis penyakitnya. Untuk meningkatkan variasi data dan mengurangi risiko *overfitting*, dataset diproses menggunakan teknik augmentasi data. Kedua model dilatih dengan menerapkan parameter yang disesuaikan, seperti pengaturan *learning rate*, penerapan metode augmentasi, dan proses *fine-tuning* pada arsitektur masing-masing. Penelitian ini juga mengevaluasi efektivitas kedua model dalam memanfaatkan fitur visual untuk klasifikasi. Hasil pelatihan akan dibandingkan berdasarkan metrik seperti akurasi, *loss*, dan stabilitas model selama proses pelatihan. Analisis ini diharapkan dapat memberikan wawasan baru mengenai penggunaan model *deep learning* untuk klasifikasi penyakit tanaman.

Penggunaan teknologi *deep learning*, khususnya arsitektur *Convolutional Neural Network* (CNN), telah menjadi solusi canggih untuk diagnosis penyakit tanaman, termasuk pada daun tomat. CNN terbukti sangat efektif dalam menganalisis data visual yang kompleks, seperti gambar daun yang terinfeksi. Menurut Krichen (2023), kemampuan CNN dalam mengekstraksi fitur penting dengan tingkat akurasi yang tinggi menjadikannya alat yang andal untuk klasifikasi penyakit. Salah satu arsitektur CNN yang menonjol adalah *ResNet*, yang memanfaatkan mekanisme *residual learning* untuk mengatasi tantangan pelatihan jaringan yang dalam tanpa terkendala masalah *vanishing gradient*. Sebagai contoh, penerapan *ResNet-152* telah menunjukkan kinerja luar biasa dalam mengenali pola-pola visual pada berbagai dataset.

Selain CNN, pendekatan berbasis *Vision Transformer* (*ViT*) menawarkan metode inovatif dalam analisis gambar dengan memanfaatkan mekanisme *self-attention*. Pendekatan ini memungkinkan *ViT* untuk memahami hubungan spasial dalam data visual dengan lebih baik dibandingkan metode konvensional. Menurut Qin et al. (2023), *Vision Transformer* mampu menangani gambar dengan resolusi tinggi secara efisien, yang sangat berguna untuk dataset dengan tingkat variasi yang tinggi. *ViT* memanfaatkan representasi berbasis *patch* untuk mengenali pola-pola visual yang kompleks, memberikan keunggulan kompetitif dalam tugas-tugas klasifikasi penyakit tanaman. Pendekatan ini menjadi alternatif yang menarik untuk analisis citra, terutama pada kasus dengan data yang beragam dan kompleks.

Penelitian ini bertujuan untuk membandingkan performa model *ResNet152* dan *Vision Transformer* dalam mengklasifikasikan penyakit daun pada tanaman tomat. Studi ini akan menganalisis efektivitas dan efisiensi kedua model dalam mendeteksi dan mengklasifikasikan gambar daun yang terindikasi penyakit. Parameter utama yang akan dievaluasi meliputi akurasi deteksi, kecepatan pemrosesan, dan kompleksitas model. Akurasi deteksi mengukur kemampuan model dalam mengidentifikasi jenis penyakit dari kumpulan data gambar secara tepat. Penelitian ini juga akan mengkaji waktu pemrosesan dan kebutuhan komputasi masing-masing model untuk mendapatkan wawasan terkait efisiensi sumber daya.

Penelitian ini memberikan kontribusi signifikan terhadap pengembangan teknologi berbasis *deep learning* di sektor agrikultur, khususnya dalam klasifikasi penyakit tanaman. Menurut Naeem et al. (2020), penerapan teknologi kecerdasan buatan memungkinkan petani mengidentifikasi penyakit tanaman secara lebih cepat dan akurat. Dengan menggunakan model seperti *ResNet152* dan *Vision Transformer*, sistem deteksi otomatis dapat diimplementasikan untuk meningkatkan efisiensi pengendalian penyakit tanaman. Selain itu, teknologi ini diharapkan mampu mengurangi kerugian hasil panen akibat penyakit yang tidak terdeteksi. Hal ini juga mendukung ketahanan pangan global melalui sistem pertanian yang lebih modern dan adaptif.

Melalui penelitian ini, kedua model akan dibandingkan secara komprehensif untuk menentukan model mana yang paling optimal dalam mendukung pengendalian penyakit tanaman. Analisis akan mencakup evaluasi terhadap kemampuan masing-masing model dalam mengolah data dengan jumlah besar dan variasi gambar yang kompleks. Selain itu, waktu pemrosesan akan diuji dalam skenario praktis untuk menentukan kelayakan model dalam aplikasi lapangan. Model dengan kebutuhan komputasi dan memori yang lebih ringan juga akan diprioritaskan untuk memastikan keberlanjutan implementasi di daerah dengan sumber daya terbatas. Hal ini penting untuk mendorong penerapan teknologi ini di berbagai wilayah, termasuk pedesaan.

Hasil penelitian ini diharapkan tidak hanya memberikan wawasan baru tentang penerapan teknologi *deep learning*, tetapi juga

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 1, (2025) E-ISSN: 3063-5845

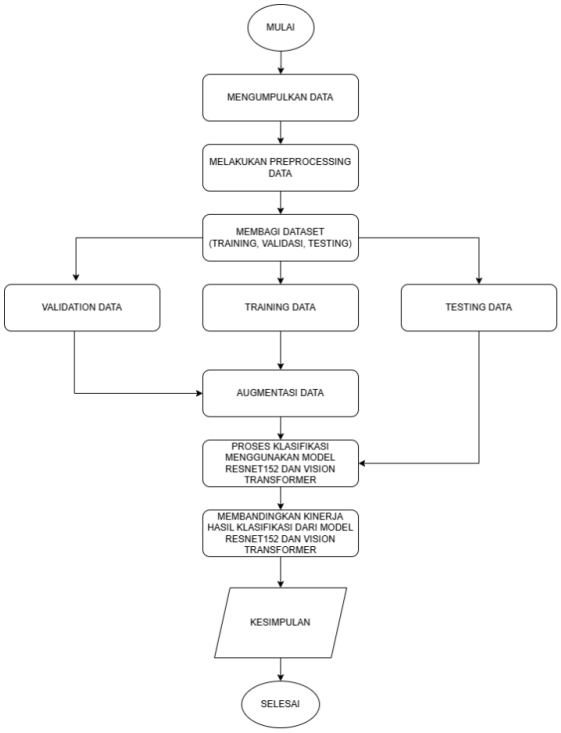
mendukung keberlanjutan sektor pertanian. Teknologi deteksi otomatis berbasis kecerdasan buatan ini dapat membantu petani mengidentifikasi penyakit tanaman dengan lebih cepat, akurat, dan efisien. Dengan demikian, petani dapat mengambil langkah pengendalian penyakit secara dini untuk meminimalkan kerugian hasil panen. Hal ini juga berkontribusi terhadap peningkatan produktivitas dan pengelolaan sumber daya yang lebih baik di sektor pertanian. Pada akhirnya, teknologi ini memiliki potensi besar untuk memperbaiki kualitas dan kuantitas hasil pertanian.

Penelitian terdahulu dilakukan oleh Dosovitskiy et al. (2020) memperkenalkan Vision Transformer (ViT) dan membandingkannya dengan CNN pada berbagai dataset, seperti ImageNet. Studi tersebut menunjukkan bahwa ViT unggul dalam memahami pola visual pada dataset besar dengan variasi tinggi, berkat mekanisme self-attention. Di sisi lain, CNN seperti ResNet lebih efisien pada dataset yang lebih kecil karena memanfaatkan convolutions yang terstruktur. Selanjutnya, Qin et al. (2023) mengevaluasi performa CNN (termasuk ResNet) dan ViT untuk tugas klasifikasi penyakit tanaman. Hasilnya menunjukkan bahwa ViT memberikan akurasi lebih tinggi pada dataset dengan resolusi tinggi dan variasi pencahayaan yang kompleks, sementara CNN masih memiliki keunggulan dalam efisiensi komputasi dan kecepatan inferensi pada perangkat dengan keterbatasan sumber daya.

Dengan landasan tersebut, penelitian ini dapat menjadi dasar pengembangan sistem deteksi otomatis yang lebih handal dan efisien untuk mendukung keberlanjutan sektor hortikultura. Sistem ini diharapkan mampu mengurangi dampak negatif penyakit tanaman terhadap hasil panen, meningkatkan efisiensi pengendalian, dan memberikan solusi yang lebih terjangkau bagi petani. Selain itu, implementasi model *deep learning* yang lebih efektif dapat membantu mempercepat transformasi digital di sektor agrikultur. Penelitian ini juga berpotensi mendukung ketahanan pangan global melalui penerapan teknologi modern untuk menghadapi tantangan penyakit tanaman. Pada akhirnya, penelitian ini memberikan kontribusi nyata terhadap pengembangan teknologi cerdas di bidang pertanian.

II. METODE PENELITIAN

Metodologi penelitian memiliki peran yang sangat penting dalam memastikan bahwa seluruh proses penelitian dan penulisan berjalan secara sistematis dan sesuai dengan fokus masalah yang sedang diteliti. Metodologi ini tidak hanya berfungsi sebagai panduan dalam pelaksanaan penelitian, tetapi juga sebagai kerangka kerja yang memungkinkan hasil yang diperoleh dapat dipertanggungjawabkan dan sesuai dengan tujuan yang ingin dicapai. Dalam penelitian ini, metode yang digunakan melibatkan beberapa langkah utama yang saling terkait, mulai dari pengumpulan data, *preprocessing data*, membagi dataset, augmentasi data, proses klasifikasi model *ResNet* dan *ViT*, membandingkan kinerja model, dan menarik kesimpulan. Setiap langkah dilakukan dengan cermat untuk memastikan data yang digunakan relevan dan berkualitas tinggi, model yang dirancang memiliki kinerja optimal, serta hasil analisis dapat memberikan wawasan yang signifikan terhadap permasalahan yang diteliti. Pendekatan ini tidak hanya memastikan keakuratan hasil, tetapi juga menjamin bahwa penelitian dapat memberikan kontribusi nyata di bidang yang dikaji.

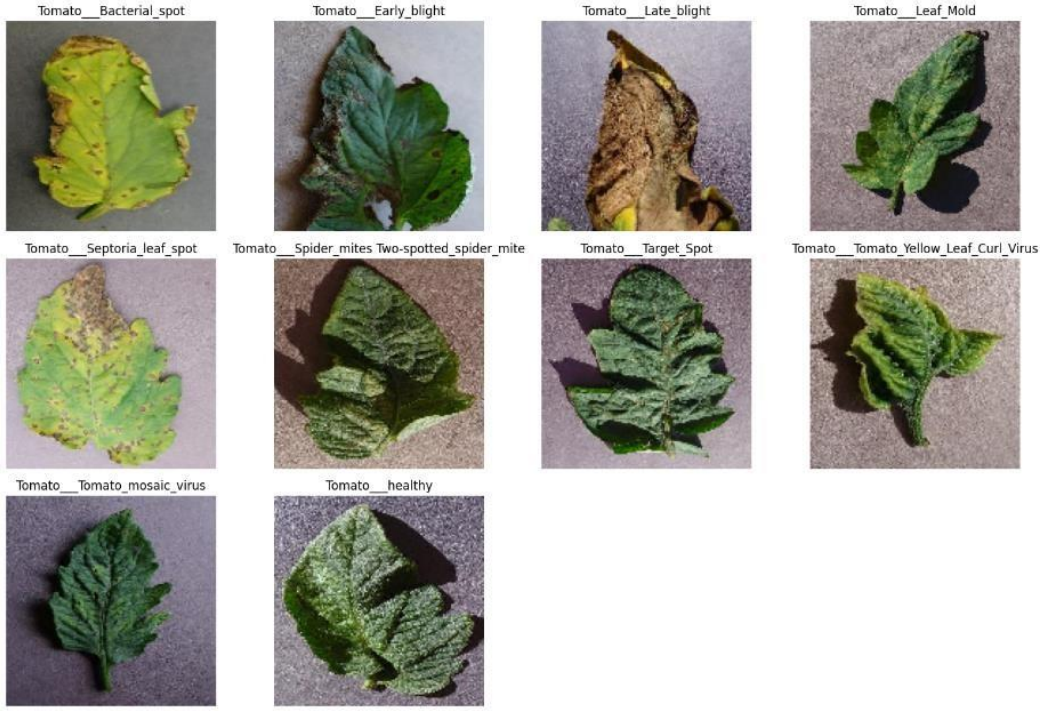


Gambar 1 Diagram Alir Penelitian

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 1, (2025) E-ISSN: 3063-5845

A. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan kumpulan citra digital yang berisi gambar daun tomat yang terinfeksi berbagai jenis penyakit. *Dataset* ini diambil dari sumber terpercaya, yaitu *Kaggle*, sebuah platform penyedia *dataset* publik yang telah diakui secara luas di kalangan peneliti dan praktisi data. Semua gambar dalam *dataset* diunduh dalam format standar .jpg untuk memastikan kompatibilitas dengan berbagai alat pemrosesan data. *Dataset* ini mencakup total 10.000 gambar, yang dibagi secara merata ke dalam 10 kelas berbeda. Dengan masing-masing kelas berisi setidaknya 1.000 gambar.



Gambar 2 Dataset Penyakit Daun Tomat

B. Preprocessing Data

Sebelum melanjutkan ke tahap pelatihan model, *preprocessing* data menjadi langkah krusial untuk memastikan kualitas data yang optimal.

C. Membagi Dataset

Dataset akan dibagi menjadi dua proses utama: pelatihan dan pengujian, dengan proporsi masing-masing 80% dan 20%. Pemisahan ini menggunakan parameter *validation_split*=0.2, yang secara otomatis memisahkan *dataset* menjadi 80% untuk pelatihan dan 20% untuk validasi. Rasio 80:20 dipilih sebagai titik awal yang dianggap optimal karena dapat secara signifikan meningkatkan akurasi model. Pada dasarnya, data pelatihan dirancang untuk mencakup sebagian besar variasi dalam *dataset*, sehingga model dapat belajar dengan lebih efektif.

D. Augmentasi Dataset

Augmentasi data adalah teknik untuk memperluas *dataset* dengan membuat variasi dari gambar yang ada melalui transformasi seperti *flipping*, *rotasi*, *zooming*, perubahan kecerahan, dan lainnya. Tujuannya adalah meningkatkan kemampuan generalisasi model dengan membuatnya lebih robust terhadap variasi data di dunia nyata.

E. Desain Arsitektur Model

Desain arsitektur model merupakan tahap penting dalam proses pengembangan jaringan neural yang dirancang untuk mempelajari dan menganalisis data secara efektif. Proses ini melibatkan pemilihan dan pengaturan elemen-elemen utama dari jaringan, seperti jumlah lapisan, jenis lapisan (konvolusi, *pooling*, atau *fully connected*), fungsi aktivasi, serta mekanisme regulasi seperti *dropout* atau *batch normalization*. Arsitektur yang dirancang dengan baik tidak hanya meningkatkan efisiensi pelatihan dengan mempercepat konvergensi model, tetapi juga mampu mengatasi tantangan seperti *overfitting*, yaitu ketika model terlalu menyesuaikan diri dengan data latih sehingga kurang mampu menangani data baru.

Selain itu, desain arsitektur yang optimal dapat meningkatkan kemampuan model dalam mengekstraksi fitur-fitur penting dari *dataset*, memungkinkan model untuk mengenali pola-pola yang kompleks dengan lebih baik. Sebagai contoh, dalam klasifikasi penyakit daun tomat, arsitektur seperti ResNet-152 atau *Vision Transformer* (ViT) dirancang untuk menangkap detail visual pada

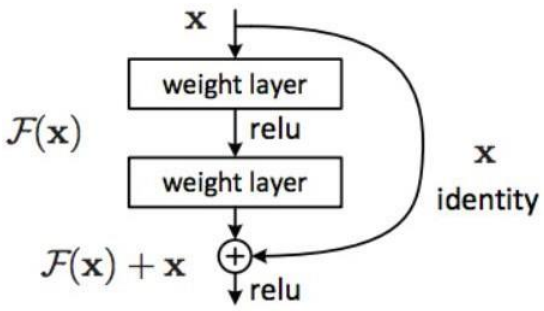
berbagai skala, sehingga model mampu membedakan jenis penyakit dengan lebih akurat. Penggunaan mekanisme seperti *residual learning* dalam *ResNet-152*, misalnya, membantu menjaga aliran informasi melalui jaringan yang sangat dalam, sehingga mengurangi risiko hilangnya informasi penting selama proses pelatihan.

Lebih lanjut, desain arsitektur model juga mencakup pemilihan parameter *hyperparameter*, seperti *learning rate, batch size,* dan *optimizer* yang digunakan, yang semuanya berkontribusi terhadap stabilitas dan kinerja model selama pelatihan. Penyesuaian *hyperparameter* ini dilakukan dengan cermat untuk memastikan bahwa model mampu mempelajari pola secara efektif tanpa terjebak dalam *local minimal* atau mengalami pelatihan yang terlalu lambat. Selain itu, penggunaan metode *fine-tuning* pada arsitektur *pretrained* sering kali diterapkan untuk mempercepat pelatihan model, terutama pada *dataset* yang berukuran relatif kecil.

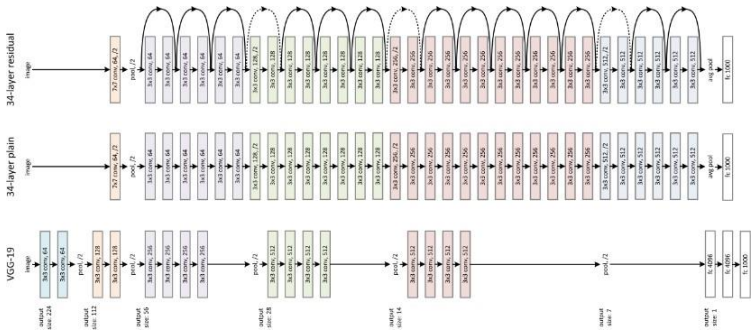
Dalam konteks aplikasi praktis, desain arsitektur model juga mempertimbangkan efisiensi komputasi dan keterbatasan perangkat keras yang digunakan. Model yang terlalu kompleks mungkin membutuhkan waktu pelatihan yang lama atau sumber daya yang besar, sehingga tidak praktis untuk diimplementasikan di lapangan. Oleh karena itu, desain yang baik harus mampu menyeimbangkan antara kompleksitas model, akurasi, dan efisiensi komputasi, memastikan bahwa model dapat diimplementasikan secara nyata tanpa mengorbankan kinerjanya.

1) ResNet – 152

ResNet (*Residual Network*) adalah arsitektur jaringan saraf yang dirancang untuk mengatasi masalah *vanishing gradient* yang sering terjadi pada jaringan yang sangat dalam. Konsep utama *ResNet* adalah penggunaan *residual block*, di mana jaringan mempelajari selisih antara *input* dan *output* yang diharapkan (residu), bukan langsung memprediksi *output*. Untuk mendukung hal ini, digunakan *skip connection* atau *shortcut connection*, yang memungkinkan *output* dari satu lapisan langsung dilewatkan ke lapisan berikutnya dengan melewati beberapa lapisan di antaranya. Pendekatan ini membantu jaringan tetap stabil meskipun kedalaman meningkat. Salah satu varian *ResNet* yang populer adalah *ResNet-152*, yang memiliki 152 lapisan dan dirancang untuk menangani *dataset* kompleks dengan memberikan performa tinggi tanpa mengorbankan efisiensi pelatihan. Dalam implementasinya, *ResNet* sering dimodifikasi dengan menambahkan lapisan seperti *reshape*, *flattened*, *dense*, *dropout*, dan *SoftMax* agar sesuai dengan tugas spesifik, seperti klasifikasi gambar. Selain itu, model ini sering digunakan sebagai *pre-trained model* pada *dataset* seperti *ImageNet* untuk mempercepat pelatihan dan meningkatkan akurasi karena sudah memahami fitur umum dalam data gambar. Keunggulan utama ResNet terletak pada *skip connection*, yang tidak hanya mengatasi lapisan yang mungkin mengganggu performa, tetapi juga memastikan gradien dapat mengalir lebih bebas hingga ke lapisan awal, sehingga masalah *vanishing gradient* dapat diminimalkan.



Gambar 3 Residual building block

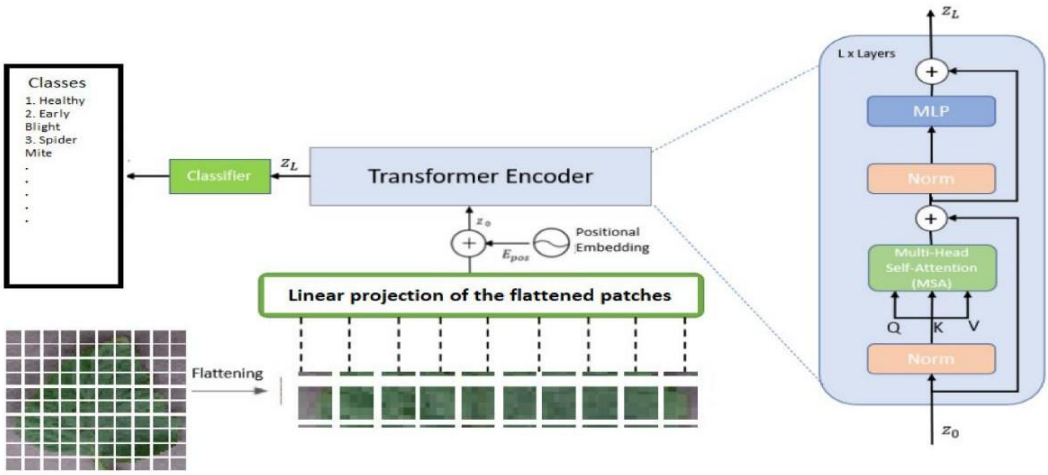


Gambar 4 Arsitektur Resnet-152

2) *Vision Transformer* (ViT)

Vision Transformer (ViT) menggunakan mekanisme perhatian (*attention mechanism*) untuk memodelkan hubungan antara piksel-piksel yang dekat dan yang jauh dalam sebuah gambar. Proses dimulai dengan membagi gambar input menjadi blok-blok kecil dengan ukuran tetap yang tidak tumpang tindih, kemudian meratakan (*flatten*) blok-blok tersebut. Selanjutnya, untuk mendukung mekanisme perhatian, gambar dibagi menjadi potongan-potongan kecil atau *patch*. Setiap *patch* kemudian diberikan *embedding* posisi (*position embedding*), yang selanjutnya dimasukkan ke dalam *encoder transformer*. Pada akhirnya, gambar diklasifikasikan menggunakan lapisan *Multi-Layer Perceptron* (MLP Head). Proses ini memiliki kesamaan dengan lapisan konvolusi (*convolution layer*) yang menggunakan *kernel*, di mana *output*nya berupa matriks 4D yang diindeks berdasarkan *batch*, serta tiga dimensi lainnya mewakili baris, kolom, dan kedalaman.

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 1, (2025) E-ISSN: 3063-5845



Gambar 5 Arsitektur *Vision Transformer (ViT)*

F. Pelatihan Model

Pada penelitian ini, pelatihan dilakukan dengan menggunakan dua arsitektur jaringan neural yang berbeda, yaitu *ResNet-152* dan *Vision Transformer* (ViT). Kedua arsitektur ini dipilih karena keunggulannya masing-masing dalam menangani data visual yang kompleks. *ResNet-152*, sebagai varian dari *ResNet* yang telah ditingkatkan, menggunakan mekanisme *residual learning* yang lebih maju, memungkinkan model untuk dilatih hingga kedalaman yang sangat besar tanpa menghadapi masalah *vanishing gradient*. Dengan 152 lapisan dalam jaringannya, ResNet-152 dirancang untuk mengekstraksi fitur visual dengan sangat detail, menjadikannya pilihan yang kuat untuk klasifikasi penyakit daun tomat. Pada pelatihan menggunakan *ResNet-152*, dilakukan penyesuaian *hyperparameter* seperti *learning rate*, *batch size*, dan metode augmentasi data untuk memaksimalkan akurasi dan stabilitas model.

Sebaliknya, *Vision Transformer* (ViT) menawarkan pendekatan baru yang menggunakan mekanisme *self-attention* untuk menganalisis gambar secara efisien. *Vision Transformer* (ViT) memecah gambar menjadi *patch-patch* kecil yang kemudian dianalisis untuk memahami pola spasial di seluruh gambar. Keunggulan *Vision Transformer* (ViT) terletak pada kemampuannya menangani gambar dengan resolusi tinggi dan variasi yang kompleks, menjadikannya sangat fleksibel dalam aplikasi klasifikasi citra. Pelatihan menggunakan *Vision Transformer (ViT)* melibatkan teknik *fine-tuning* dan augmentasi data untuk meningkatkan kemampuan model dalam mengenali berbagai jenis penyakit pada daun tomat.

Masing-masing arsitektur memiliki pendekatan unik dalam pengolahan data citra, yang diuraikan sebagai berikut:

1) Pelatihan Menggunakan ResNet-152

ResNet-152 adalah arsitektur *deep learning* berbasis *Convolutional Neural Networks* (CNN) yang menggunakan konsep *residual connections* untuk menangani masalah *vanishing gradient* pada jaringan yang sangat dalam. Berikut adalah langkah-langkah pelatihan model ResNet-152:

a) Inisialisasi Model

Model *ResNet-152* yang telah dilatih sebelumnya pada dataset *ImageNet* digunakan sebagai dasar. Bagian awal model di-*freeze* untuk mempertahankan fitur dasar, sementara lapisan-lapisan akhir diubah untuk menyesuaikan jumlah kelas *dataset* daun tomat.

b) Pengaturan *Hyperparameter*

(1) *Optimizer*: *Stochastic Gradient Descent* (SGD) dengan *learning rate* 0.001 dan momentum 0.9.

Stochastic Gradient Descent adalah metode optimasi klasik yang telah terbukti efektif untuk pelatihan jaringan saraf dalam. *Stochastic Gradient Descent* SGD memperbarui bobot model berdasarkan gradien dari subset data (mini-batch), yang memberikan efisiensi dan memungkinkan generalisasi yang baik.

Learning rate adalah parameter kunci yang mengontrol seberapa besar langkah pembaruan bobot pada setiap iterasi. Nilai kecil seperti 0.001 digunakan untuk memastikan konvergensi yang stabil dan menghindari osilasi atau *overshooting* (melampaui *minimum loss*).

Momentum menambahkan komponen kecepatan (*velocity*) ke pembaruan gradien, yang membantu model untuk melewati lembah datar (*flat regions*) dalam lanskap *loss* dan mempercepat konvergensi. Momentum 0.9 adalah nilai standar yang banyak digunakan karena memberikan keseimbangan antara stabilitas dan kecepatan pelatihan.

(2) *Loss Function*: *Categorical Crossentropy*.

Loss function ini dirancang untuk tugas klasifikasi multi-kelas di mana target adalah representasi *one-hot encoding*. *Categorical crossentropy* mengukur perbedaan antara distribusi prediksi model (probabilitas output) dan distribusi target sebenarnya, sehingga cocok untuk tugas klasifikasi dengan lebih dari dua kelas.

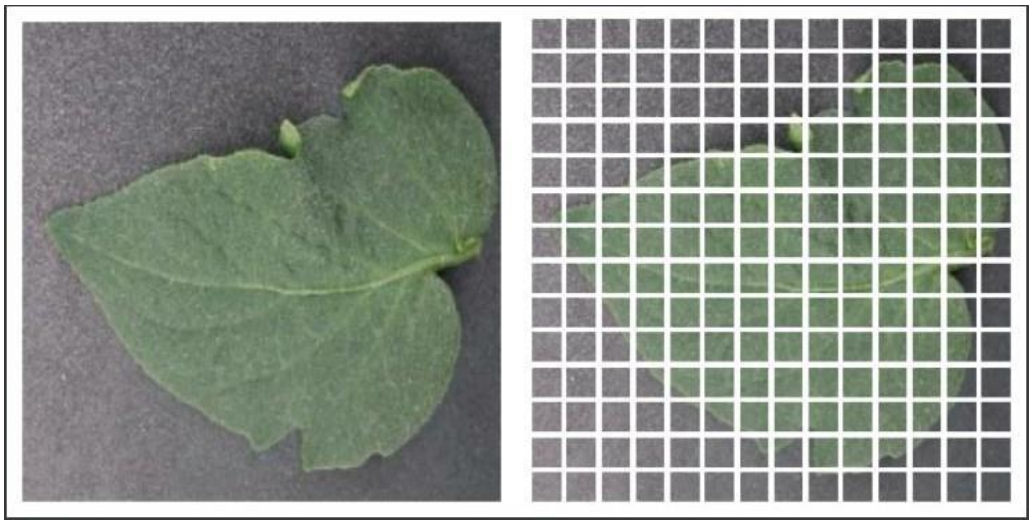
PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 1, (2025) E-ISSN: 3063-5845

(3) *Regularization*: *Dropout* dengan tingkat 0.3 pada beberapa lapisan *fully-connected* untuk mencegah *overfitting*.
Dropout adalah teknik regularisasi yang secara acak menonaktifkan sejumlah unit neuron selama pelatihan. Ini mencegah unit-unit tertentu menjadi terlalu dominan, sehingga mengurangi risiko *overfitting* dan meningkatkan kemampuan generalisasi model.
Tingkat *dropout* 0.3 berarti 30% neuron secara acak dinonaktifkan dalam setiap iterasi pelatihan. Angka ini dipilih untuk memberikan keseimbangan antara regularisasi yang cukup dan pemeliharaan informasi yang cukup untuk pembelajaran.
Lapisan *fully-connected* cenderung memiliki jumlah parameter yang sangat besar dibandingkan dengan lapisan *convolutional*, sehingga lebih rentan terhadap *overfitting*. *Dropout* membantu mengurangi risiko ini dengan membuat jaringan lebih *robust*.

c) Callback yang Digunakan
(1) *EarlyStopping* untuk menghentikan pelatihan jika validasi tidak meningkat dalam 3 epoch.
(2) *ModelCheckpoint* untuk menyimpan model terbaik selama pelatihan.
(3) *ReduceLROnPlateau* untuk menurunkan *learning rate* jika validasi *loss* stagnan.

d) Proses Pelatihan
Model dilatih selama 25 epoch dengan batch size 16 menggunakan generator data yang telah diaugmentasi. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk validasi.

2) Pelatihan Menggunakan *Vision Transformer (ViT)*
a) Inisialisasi Model
Model *Vision Transformer* dibuat menggunakan pendekatan custom dengan lapisan-lapisan berikut:
(1) *Patches* Layer: Membagi gambar menjadi *patch* kecil berukuran 16x16 piksel.



Gambar 6 Membagi gambar menjadi potongan-potongan kecil

(2) *Patch Encoder*: Memberikan *embedding* posisi pada setiap *patch* untuk mempertahankan informasi spasial. *Embedding* posisi ini membantu model mempertahankan informasi tentang urutan spasial gambar, yang hilang setelah gambar diubah menjadi *patch*.
(3) *Transformer Layers*: Terdiri dari delapan blok yang menggunakan *multi-head attention* dan *multi-layer perceptron* (MLP). Memproses *patch* yang telah diberi *encoding* menggunakan arsitektur *transformer*. Terdiri dari komponen *Multi-Head Attention* (MHA) dan Multi-Layer *Perceptron* (MLP).
(4) *Classification Head*: MLP dengan lapisan *dense* untuk klasifikasi akhir. Terdiri dari MLP dengan beberapa lapisan *dense* dan lapisan terakhir untuk menghasilkan *logits* (nilai sebelum *softmax*).

b) Pengaturan Hyperparameter
(1) AdamW (Adaptive Moment Estimation with Weight Decay): Digunakan pada ViT untuk mengoptimalkan pembaruan bobot. Optimizer ini sangat efektif untuk model besar karena menggabungkan keuntungan dari Adam dan regularisasi *weight decay*, yang membantu mencegah *overfitting*.
(2) *Loss Function: Sparse Categorical Crossentropy* untuk menangani data label integer. Fungsi *loss* ini optimal untuk menangani data multi-kelas dengan label integer. Membuat proses pelatihan lebih efisien karena tidak perlu transformasi tambahan pada label.

c) Callback yang Digunakan
ModelCheckpoint untuk menyimpan bobot model terbaik selama pelatihan

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 1, (2025) E-ISSN: 3063-5845

G. Evaluasi Model

Evaluasi model dalam penelitian ini dilakukan dengan menggunakan tiga metrik utama, yaitu *loss function*, akurasi, dan *confusion matrix*, untuk memastikan performa model secara komprehensif. *Loss function* digunakan untuk mengukur seberapa besar perbedaan antara prediksi model dengan label yang sebenarnya. Nilai *loss* yang lebih rendah menunjukkan bahwa model mampu menghasilkan prediksi yang lebih mendekati kebenaran. Selama proses pelatihan, nilai *loss* ini digunakan sebagai indikator untuk mengevaluasi dan menyempurnakan model melalui iterasi berulang, seperti penyesuaian parameter model. Selain itu, *monitoring loss* pada data validasi membantu mengidentifikasi potensi *overfitting* jika model mulai kehilangan kemampuan generalisasi meskipun nilai *loss* pada data latih terus menurun.

Metrik kedua adalah akurasi, yang mengukur persentase prediksi model yang benar terhadap total data yang diuji. Akurasi memberikan gambaran langsung tentang seberapa baik model dalam mengklasifikasikan gambar-gambar daun tomat ke dalam kelas-kelas penyakit yang benar. Namun, untuk mengevaluasi lebih dalam, digunakan juga *confusion matrix*, yang memberikan rincian distribusi prediksi model, seperti jumlah *true positive (TP), true negative (TN), false positive (FP),* dan *false negative (FN). Confusion matrix* sangat penting dalam mengidentifikasi pola kesalahan model, misalnya jika model sering salah mengklasifikasikan satu jenis penyakit sebagai jenis lain.

Berikut adalah penjelasan mendetail terkait metrik-metrik tersebut:

1) Loss of function and accuracy

Loss function bertugas membandingkan nilai prediksi model dengan nilai sebenarnya (*ground truth*) secara kontinu selama proses pelatihan. Hasil perbandingan tersebut digunakan untuk memperbarui bobot model guna meminimalkan kesalahan yang dihasilkan. Dalam penelitian ini, digunakan *cross-entropy loss* sebagai fungsi *loss*. Secara formal, fungsi ini dapat didefinisikan sebagai berikut:

$$L = \frac{1}{N}\sum_{i} L_i = -\frac{1}{N}\sum_{i}\sum_{c=1}^{m} y_{ic}\log(p_{ic}) \tag{1}$$

Dimana N adalah jumlah kategori, y_{ic} adalah nilai biner dari label sebenarnya (0 atau 1), dan p_{ic} adalah probabilitas prediksi bahwa kasus i termasuk dalam kategori c.

Empat karakteristik dasar digunakan untuk mendefinisikan akurasi precision, recall, dan F1 Score dan confusion matrix. Keempat karakteristik tersebut adalah *True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)*. Berikut adalah penjelasannya:

a) *True Positive* (TP): Jumlah kasus yang secara akurat menunjukkan keberadaan suatu lesi tertentu.
b) *True Negative* (TN): Jumlah kasus yang secara akurat menunjukkan tidak adanya suatu lesi tertentu.
c) *False Positive* (FP): Jumlah kasus yang secara keliru menunjukkan keberadaan suatu lesi tertentu.
d) *False Negative* (FN): Jumlah kasus yang secara keliru menunjukkan tidak adanya suatu lesi tertentu.

Secara formal, akurasi, *precision, recall*, dan F1 *Score* didefinisikan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

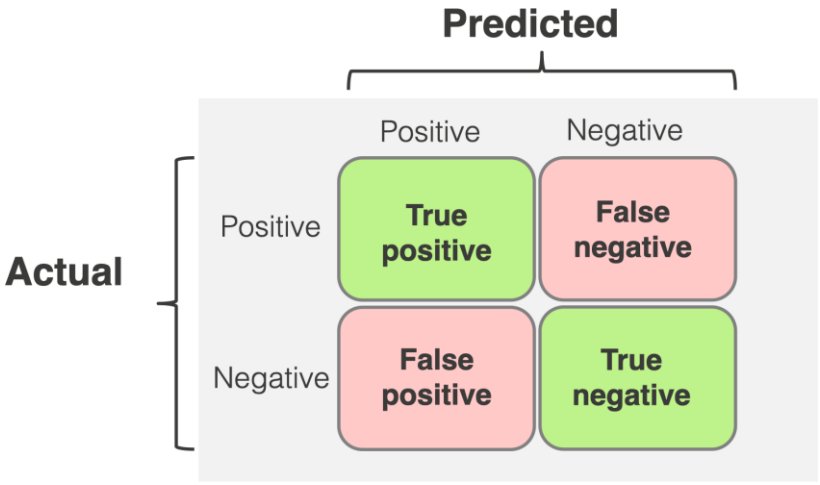
$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1\,Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

2) Confusion matrix

Confusion matrix memberikan gambaran umum tentang hasil prediksi dari masalah klasifikasi. Matriks ini mempermudah visualisasi hasil kategorisasi. Dalam *confusion matrix*, jumlah berbagai jenis penyakit daun pada tomat yang diklasifikasikan oleh model ditampilkan di bawah masing-masing kategori. Matriks ini memberikan demonstrasi yang baik tentang performa model dalam mengklasifikasikan penyakit daun pada tomat secara spesifik.

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 1, (2025) E-ISSN: 3063-5845



Gambar 7 *Confusion Matrix*

H. Visualisasi

Visualisasi merupakan langkah penting dalam memahami performa model dalam tugas klasifikasi penyakit daun pada tomat, karena memungkinkan analisis mendalam terhadap proses pelatihan dan evaluasi prediksi. Visualisasi ini merupakan elemen penting dalam evaluasi kinerja model klasifikasi, terutama untuk memastikan model bekerja sesuai dengan tujuan yang diharapkan. Gambar visualisasi yang dihasilkan menampilkan label aktual, label prediksi, dan tingkat kepercayaan model untuk setiap gambar daun tomat. Dengan adanya informasi ini, peneliti dapat secara langsung memeriksa akurasi prediksi model pada setiap kelas penyakit daun tomat, seperti *Tomato Mosaic Virus, Early Blight*, *Late Blight*, dan lainnya. Visualisasi juga membantu mengidentifikasi potensi kesalahan model dalam mengklasifikasikan penyakit tertentu, seperti prediksi yang salah atau tingkat kepercayaan yang rendah, sehingga dapat menjadi dasar untuk meningkatkan performa model di masa mendatang.

Selain untuk evaluasi kinerja, visualisasi ini juga memberikan wawasan tambahan mengenai pola atau fitur yang mungkin digunakan model untuk membuat keputusan klasifikasi. Misalnya, dalam kasus gambar dengan tingkat kepercayaan tinggi, model menunjukkan keyakinan yang kuat terhadap prediksinya, yang dapat mencerminkan kejelasan fitur pada gambar tersebut. Sebaliknya, pada gambar dengan tingkat kepercayaan rendah atau prediksi yang salah, visualisasi ini bisa menjadi dasar untuk menganalisis karakteristik gambar yang menyebabkan kebingungan model. Dengan memanfaatkan visualisasi seperti ini, peneliti dapat memperbaiki kualitas data atau menyempurnakan parameter model untuk mencapai hasil yang lebih optimal.

III. HASIL DAN PEMBAHASAN

Berikut adalah analisis efektivitas model *ResNet-152* dan *Vision Transformer (ViT)* dalam mengklasifikasikan penyakit daun pada tanaman tomat, dengan perhatian khusus pada kemampuan masing-masing model dalam menangkap dan memahami pola-pola visual yang kompleks. Evaluasi dilakukan menggunakan metrik utama seperti akurasi dan confusion matrix, yang memberikan wawasan mendalam mengenai kinerja model secara keseluruhan.

Akurasi digunakan untuk mengukur tingkat kesesuaian prediksi model dengan label sebenarnya, sehingga menjadi indikator utama dalam menilai performa model. Di sisi lain, *confusion matrix* memberikan detail yang lebih rinci tentang distribusi kesalahan klasifikasi, seperti jumlah prediksi benar dan salah di setiap kategori. Hal ini memungkinkan analisis mendalam terhadap kekuatan dan kelemahan masing-masing model, termasuk kemampuan mereka dalam menangani kelas-kelas yang sulit dibedakan.

Tabel 1 Hasil Perbandingan Model

Model	Optimizer	Learning Rate	Accuracy	loss
ResNet-152	*Stochastic Gradient Descent* (SGD)	0.001	0.9609	0.1683
Vision Transformer (ViT)	*AdamW (Adaptive Moment Estimation with Weight Decay)*	0.001	0.9860	0.0568

Hasil menunjukkan perbandingan kinerja dua model, *ResNet-152* dan *Vision Transformer (ViT),* dalam klasifikasi penyakit daun pada tomat. Evaluasi didasarkan pada beberapa metrik utama untuk memahami sejauh mana kemampuan masing-masing model dalam mengenali pola-pola visual yang kompleks. Kedua model menggunakan *batch size* sebesar 16 dan dilatih selama 25 *epoch*, memastikan bahwa kondisi pelatihan konsisten untuk perbandingan yang adil. *Optimizer* yang digunakan *ResNet-152* adalah *Stochastic Gradient Descent (SGD),* yang terkenal karena kesederhanaannya dan sering digunakan untuk model konvolusi. Lalu,

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 1, (2025) E-ISSN: 3063-5845

Vision Transformer menggunakan Adam dengan *weight decay*, yang memberikan stabilitas dan kontrol regularisasi yang lebih baik pada model berbasis *transformer*. Kedua model menggunakan *learning rate* sebesar 0.001, memastikan bahwa laju pembelajaran tetap konstan selama pelatihan.

Vision Transformer unggul dengan akurasi sebesar 98.60%, menunjukkan bahwa model ini lebih mampu mengenali pola visual pada *dataset*. Sedangkan, ResNet-152 menghasilkan akurasi sebesar 96.09%. ViT terbukti sangat efektif dalam mengenali pola visual kompleks, berkat mekanisme *self-attention* yang memungkinkannya memahami hubungan spasial pada gambar dengan lebih baik. Keunggulan ini menjadikan ViT sebagai model yang sangat andal untuk tugas klasifikasi penyakit tanaman, terutama pada *dataset* yang memiliki tingkat variasi tinggi. Di sisi lain, ResNet-152, meskipun memiliki akurasi yang sedikit lebih rendah, tetap menunjukkan kemampuan yang solid dalam mengidentifikasi pola visual melalui mekanisme *residual learning*-nya. Model ini berhasil menjaga nilai *loss* yang konsisten pada data validasi, menunjukkan kemampuannya untuk menggeneralisasi pola dari data pelatihan ke data baru. Namun, beberapa kelemahan terlihat pada beberapa kategori penyakit dengan karakteristik visual yang serupa, seperti *early blight* dan *late blight*, yang sering kali tumpang tindih dalam hasil prediksi.

Hasil perbandingan model menunjukkan bahwa *Vision Transformer (ViT)* memiliki nilai *loss* yang lebih rendah (0.0568) dibandingkan dengan *ResNet-152* (0.1683). Nilai *loss* ini mencerminkan tingkat kesalahan model dalam memprediksi kelas penyakit daun tomat pada data validasi. Pada *ResNet-152*, meskipun nilai *loss* cukup rendah, model ini masih menunjukkan beberapa kelemahan dalam membedakan pola-pola visual tertentu, terutama pada kelas penyakit yang memiliki karakteristik serupa. Mekanisme *residual learning* pada *ResNet-152* membantu mengurangi kesalahan, tetapi tidak seefektif pendekatan yang digunakan oleh ViT. Sebaliknya, ViT berhasil mencapai nilai *loss* yang lebih rendah, yang menunjukkan kemampuannya untuk meminimalkan kesalahan prediksi secara lebih efisien. Keunggulan ini sebagian besar disebabkan oleh mekanisme *self-attention* yang mampu memahami hubungan spasial kompleks dalam data citra. Dengan kemampuan tersebut, ViT tidak hanya menghasilkan nilai akurasi yang lebih tinggi, tetapi juga memiliki tingkat generalisasi yang lebih baik terhadap data validasi. Hal ini menjadikan ViT sebagai model yang lebih unggul dalam mengklasifikasikan penyakit daun tomat berdasarkan *dataset* yang digunakan dalam penelitian ini.

	precision	recall	f1-score
Tomato___Bacterial_spot	0.98	0.97	0.98
Tomato___Early_blight	0.93	0.98	0.96
Tomato___Late_blight	1.00	0.94	0.97
Tomato___Leaf_Mold	1.00	0.98	0.99
Tomato___Septoria_leaf_spot	1.00	0.92	0.95
Tomato___Spider_mites Two-spotted_spider_mite	0.99	0.84	0.91
Tomato___Target_Spot	0.91	0.98	0.94
Tomato___Tomato_Yellow_Leaf_Curl_Virus	1.00	0.99	0.99
Tomato___Tomato_mosaic_virus	1.00	1.00	1.00
Tomato___healthy	0.85	1.00	0.92

Gambar 8 Matrik Evaluasi Vision Transformer

	precision	recall	f1-score
Tomato___Bacterial_spot	0.99	0.85	0.91
Tomato___Early_blight	0.84	0.92	0.88
Tomato___Late_blight	0.92	0.99	0.95
Tomato___Leaf_Mold	0.82	0.96	0.88
Tomato___Septoria_leaf_spot	0.94	0.90	0.92
Tomato___Spider_mites Two-spotted_spider_mite	0.91	0.86	0.89
Tomato___Target_Spot	0.95	0.81	0.88
Tomato___Tomato_Yellow_Leaf_Curl_Virus	0.95	0.96	0.96
Tomato___Tomato_mosaic_virus	0.97	0.97	0.97
Tomato___healthy	0.92	0.96	0.94

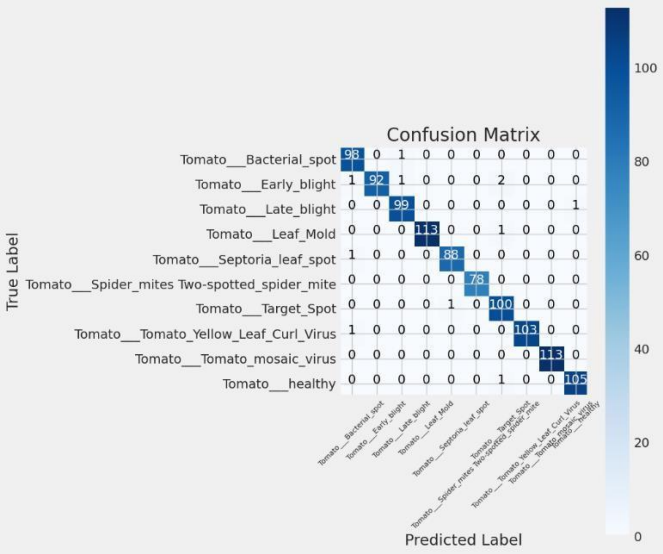
Gambar 9 Matrik Evaluasi ResNet152

Gambar 8 dan Gambar 9 menyajikan perbandingan hasil evaluasi performa model *Vision Transformer* dan *ResNet152* dalam klasifikasi penyakit daun tomat berdasarkan metrik *precision*, *recall*, dan *f1-score*. Masing-masing metrik ini digunakan untuk menilai tingkat akurasi deteksi terhadap 10 jenis penyakit dan kondisi sehat pada tanaman tomat. Vision Transformer unggul secara konsisten dengan rata-rata *precision*, *recall*, dan *f1-score* lebih tinggi dibandingkan *ResNet152*. Model *Vision Transformer* lebih efektif dalam mengidentifikasi penyakit dengan tingkat kesalahan lebih rendah, menjadikannya pilihan yang lebih unggul untuk klasifikasi penyakit daun tomat dalam dataset yang digunakan.

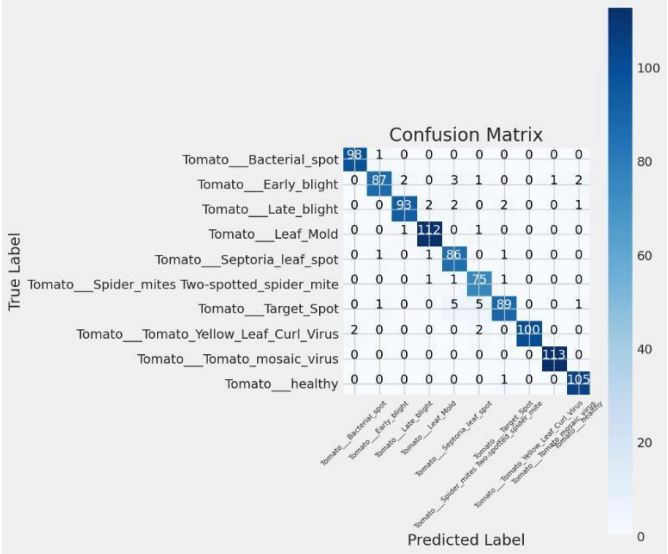
Pada Gambar 8, hasil evaluasi *Vision Transformer* menunjukkan performa yang konsisten tinggi dengan *F1-score* rata-rata mendekati 0.95 hingga 1.0 pada beberapa kelas, seperti *Tomato__Late_blight* dan *Tomato__Tomato_mosaic_virus*. Namun, ada sedikit penurunan pada kelas *Tomato__healthy* dengan *F1-score* sebesar 0.92, meskipun *recall* mencapai nilai sempurna (1.0). Hal ini mengindikasikan bahwa *Vision Transformer* memiliki kemampuan yang sangat baik dalam mendeteksi penyakit, tetapi sedikit kurang optimal pada data kategori tanaman sehat.

Sementara itu, Gambar 9 yang menampilkan evaluasi *ResNet152* menunjukkan performa yang juga baik tetapi sedikit lebih rendah dibandingkan *Vision Transformer* pada beberapa kelas, seperti *Tomato__Early_blight* dengan F1-score 0.88 dan *Tomato__Spider_mites_Two-spotted_spider_mite* dengan nilai yang sama. Meskipun demikian, model ini tetap menunjukkan kekuatan pada kelas-kelas tertentu, seperti *Tomato Tomato_mosaic_virus* dengan F1-score 0.97. Secara keseluruhan, ResNet152 tampak lebih bervariasi dalam kinerjanya dibandingkan *Vision Transformer*, dengan potensi yang lebih besar untuk peningkatan melalui pengoptimalan lebih lanjut.

PROKASDADIK: Prosiding Kecerdasan Artifisial, Sains Data, dan Pendidikan Masa Depan
Vol. 3, No. 1, (2025) E-ISSN: 3063-5845



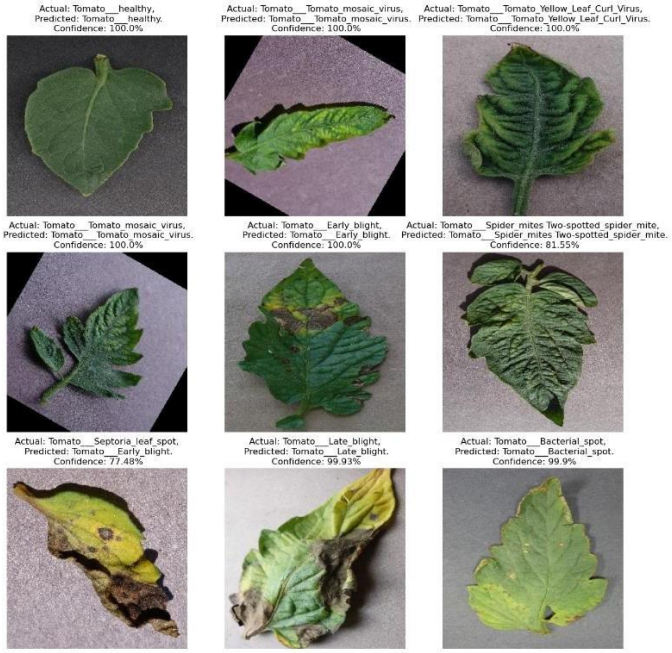
Gambar 10 Confusion Matrix Vision Transformer



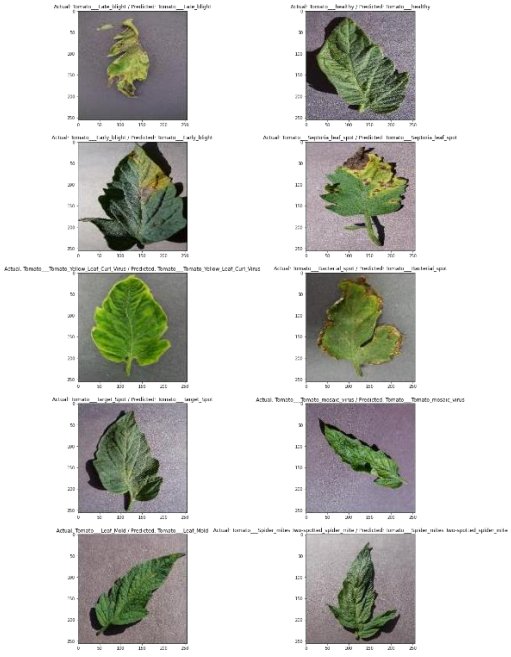
Gambar 11 Confusion Matrix ResNet-152

Berdasarkan *confusion matrix*, *Vision Transformer* (ViT) menunjukkan performa yang lebih unggul dibandingkan *ResNet-152* dalam klasifikasi penyakit daun tomat. ViT menghasilkan jumlah prediksi benar (*true positive*) yang lebih tinggi dengan tingkat kesalahan, seperti *false positive* dan *false negative*, yang lebih rendah dibandingkan *ResNet-152*. Mekanisme *self-attention* pada ViT memungkinkan model ini untuk menangkap hubungan spasial dan pola visual kompleks dengan lebih baik, sehingga mampu mengklasifikasikan kelas-kelas yang sering sulit dibedakan oleh *ResNet-152*, seperti *early blight* dan *late blight*. Sebaliknya, *ResNet-152* memiliki keterbatasan dalam membedakan pola-pola halus di antara kelas-kelas tersebut, meskipun mekanisme *residual learning* membantu mengurangi beberapa kesalahan prediksi.

ViT juga terbukti lebih stabil dalam menangani variasi pada data citra, seperti perbedaan pencahayaan dan latar belakang, sehingga lebih jarang menghasilkan kesalahan prediksi. Hal ini berbanding terbalik dengan *ResNet-152* yang cenderung kesulitan menghadapi variasi tersebut, ditunjukkan oleh jumlah *false negative* yang lebih tinggi pada beberapa kelas. Secara keseluruhan, ViT lebih andal untuk tugas klasifikasi penyakit daun tomat karena mampu meminimalkan kesalahan prediksi dan memberikan distribusi prediksi yang lebih akurat berdasarkan *confusion matrix* yang dihasilkan.



Gambar 12 Visualisasi Vision Transformer



Gambar 13 Visualisasi ResNet152

Pada Gambar 12, hasil visualisasi menggunakan model *Vision Transformer* menunjukkan kemampuan model dalam mengklasifikasikan berbagai kondisi daun tomat, baik yang sehat maupun yang terkena penyakit. Model ini berhasil mengidentifikasi beberapa kategori penyakit dengan tingkat kepercayaan (*confidence*) yang sangat tinggi, seperti *Tomato mosaic virus*, *Tomato Yellow Leaf Curl Virus*, dan daun tomat yang sehat dengan *confidence* mencapai 100%. Selain itu, penyakit seperti *Late blight* juga berhasil diklasifikasikan dengan *confidence* 99.93%, menandakan kemampuan *Vision Transformer* yang unggul dalam mendeteksi pola-pola kompleks.

Namun, terdapat kelemahan pada klasifikasi penyakit tertentu, seperti Septoria leaf spot, yang hanya mencapai tingkat kepercayaan 77.48%. Hal ini mengindikasikan bahwa *Vision Transformer* mengalami kesulitan dalam membedakan gejala penyakit ini dengan kategori lainnya, terutama jika ciri-ciri visualnya mirip dengan penyakit lain. Selain itu, penyakit *seperti Spider mites Two-spotted spider mite* juga memiliki tingkat *confidence* yang lebih rendah (81.55%), menunjukkan tantangan dalam mendeteksi penyakit yang melibatkan gejala visual tidak khas atau kompleks.

Sementara itu, Gambar 13 menampilkan visualisasi menggunakan model *ResNet152*, yang juga menunjukkan performa yang baik dalam mengklasifikasikan kondisi daun tomat. Visualisasi ini memberikan tata letak gambar yang lebih sistematis untuk memudahkan analisis. Namun, tidak seperti *Vision Transformer*, *confidence level* dari klasifikasi tidak disertakan secara eksplisit dalam visualisasi ini, sehingga tidak memungkinkan untuk membandingkan tingkat kepercayaan klasifikasi antar model secara langsung melalui gambar.

Secara keseluruhan, *Vision Transformer* menunjukkan keunggulan dalam klasifikasi penyakit daun tomat dengan tingkat kepercayaan yang tinggi untuk sebagian besar kategori, terutama dalam mendeteksi penyakit dengan ciri visual yang khas. Namun, tantangan tetap ada untuk penyakit dengan gejala yang sulit dibedakan. Di sisi lain, *ResNet152* menawarkan performa yang sebanding, meskipun dengan pendekatan visualisasi yang berbeda. Penelitian lebih lanjut dapat dilakukan untuk mengevaluasi performa kedua model ini dalam skenario klasifikasi yang lebih kompleks.

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa *Vision Transformer* (ViT) memiliki keunggulan signifikan dalam klasifikasi penyakit daun tomat dibandingkan *ResNet-152*. Dengan akurasi 98,60% dan *loss* 0,0568, ViT mengungguli *ResNet-152* yang memiliki akurasi 96,09% dan *loss* 0,1683. Mekanisme *self-attention* pada ViT memungkinkan model untuk memahami pola visual yang kompleks, menjadikannya lebih andal dalam menangani variasi data dan mendeteksi pola-pola sulit. ViT lebih stabil dalam mengidentifikasi penyakit yang sering sulit dibedakan, seperti *early blight* dan *late blight*, dengan tingkat kesalahan lebih rendah. Namun, model ini masih memiliki keterbatasan pada kategori penyakit dengan gejala kompleks, seperti *Septoria leaf spot* dan *Spider mites*. Sebaliknya, *ResNet-152* menunjukkan performa yang baik, tetapi lebih rentan terhadap kesalahan pada kelas-kelas dengan karakteristik visual yang serupa. Dalam aspek evaluasi metrik seperti *precision, recall, dan F1-score, Vision Transformer* konsisten menunjukkan nilai lebih tinggi dibandingkan ResNet-152. Hal ini mengindikasikan kemampuan ViT dalam memberikan prediksi yang lebih akurat dan generalisasi yang lebih baik terhadap data validasi. Mekanisme *self-attention* juga membantu ViT dalam menghadapi variasi pada data citra, seperti pencahayaan dan latar belakang. Hasil penelitian ini menegaskan bahwa *Vision Transformer* merupakan pilihan yang lebih efektif untuk klasifikasi penyakit daun tomat, khususnya pada dataset dengan tingkat variasi tinggi. Namun, untuk meningkatkan performa lebih lanjut, penelitian di masa depan dapat berfokus pada pengoptimalan parameter ViT atau pengembangan teknik augmentasi data yang lebih baik untuk mengatasi tantangan pada kategori penyakit dengan gejala kompleks.

DAFTAR PUSTAKA

[1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. Vol. 3, J. Peters, Ed. New York: McGraw-Hill (1964) 15–64.
[2] W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth (1993) 123–135.
[3] H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag (1985) Ch. 4.
[4] B. Smith, "An approach to graphs of linear forms (Unpublished work style)," belum dipublikasikan.
[5] E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, akan dipublikasikan.
[6] J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, didaftarkan untuk dipublikasikan.
[7] C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, komunikasi pribadi, (1995, May).
[8] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Studi elektron spektroskopi pada media optik-pembesar dan antarmuka substrat plastik (gaya jurnal terjemahan)," *IEEE Transl. J. Magn.Jpn.*, Vol. 2 (1987) 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan (1982) 301].
[9] M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science (1989).
[10] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility (Periodical style)," *IEEE Trans. Electron Devices*, Vol. ED-11 (1959, Jan.) 34– 39.
[11] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks,"
IEEE Trans. Neural Networks, Vol. 4 (1993, Jul.) 570–578.

[12] R. W. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, Vol. 44, No. 4 (1965, Apr.) 547–588.
[13] S. P. Bingulac, "On the compatibility of adaptive controllers (Published Conference Proceedings style)," in *Proc. 4th Annu. Allerton Conf. Circuits and Systems Theory*, New York (1994) 8–16.
[14] G. R. Faulhaber, "Design of service systems with priority reservation," in *Conf. Rec. 1995 IEEE Int. Conf. Communications*, 3–8.

[15] W. D. Doyle, "Magnetization reversal in films with biaxial anisotropy," in *1987 Proc. INTERMAG Conf.*, 2.2-1–2.2-6.
[16] G. W. Juette and L. E. Zeffanella, "Radio noise currents n short sections on bundle conductors (Presented Conference Paper style)," presented at the IEEE Summer power Meeting, Dallas, TX, Jun. 22–27 (1990) Paper 90 SM 690-0 PWRS.
[17] J. G. Kreifeldt, "An analysis of surface-detected EMG as an amplitude-modulated noise," presented at the 1989 Int. Conf. Medicine and Biological Engineering, Chicago, IL.
[18] J. Williams, "Narrow-band analyzer (Thesis or Dissertation style)," Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA (1993).
[19] N. Kawasaki, "Parametric study of thermal and chemical nonequilibrium nozzle flow," M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan (1993).
[20] J. P. Wilkinson, "Nonlinear resonant circuit devices (Patent style)," U.S. Patent 3 624 12, July 16, (1990).
[21] *IEEE Criteria for Class IE Electric Systems* (Standards style), IEEE Standard 308 (1969).
[22] *Letter Symbols for Quantities*, ANSI Standard Y10.5 (1968).
[23] R. E. Haskell and C. T. Case, "Transient signal propagation in lossless isotropic plasmas (Report style)," USAF Cambridge Res. Lab., Cambridge, MA Rep. ARCRL-66-234 (II) (1994), Vol. 2.
[24] E. E. Reber, R. L. Michell, and C. J. Carter, "Oxygen absorption in the Earth's atmosphere," Aerospace Corp., Los Angeles, CA, Tech. Rep. TR-0200 (420- 46)-3 (Nov. 1988).
[25] (Handbook style) *Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC (1985) 44–60.