

BAB II

TINJAUAN PUSTAKA

2.1 Merdeka Belajar Kampus Merdeka (MBKM) dan Proyek Independen

Program Merdeka Belajar Kampus Merdeka (MBKM) merupakan kebijakan strategis Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia yang bertujuan untuk memberikan kebebasan belajar bagi mahasiswa agar dapat mengembangkan kompetensi secara lebih holistik dan kontekstual sesuai dengan kebutuhan zaman. Salah satu bentuk implementasi dari kebijakan ini adalah Proyek Independen, yaitu kegiatan pembelajaran berbasis proyek (project-based learning) yang menempatkan mahasiswa sebagai pelaku utama dalam perencanaan, pelaksanaan, dan evaluasi terhadap suatu permasalahan nyata dengan pendekatan ilmiah dan terukur [1].

Secara konseptual, proyek independen sejalan dengan pendekatan konstruktivistik dalam pendidikan tinggi, di mana pengetahuan dibangun secara aktif oleh peserta didik melalui interaksi langsung dengan konteks dunia nyata. Kegiatan proyek dalam MBKM tidak hanya dirancang untuk memperdalam penguasaan materi akademik, tetapi juga untuk mengembangkan kemampuan berpikir kritis, keterampilan komunikasi ilmiah, literasi digital, kemampuan manajerial, dan etika profesi. Hal ini mendukung learning outcomes yang mengintegrasikan hard skills dan soft skills secara seimbang dalam kerangka outcome-based education [15], [16].

Menurut Buku Panduan MBKM Tahun 2024, proyek independen harus memenuhi tiga prinsip dasar, yaitu: (1) berbasis pada penyelesaian permasalahan nyata di masyarakat atau industri, (2) menghasilkan luaran konkret seperti prototipe, publikasi, atau solusi aplikatif, serta (3) dapat dikonversi menjadi satuan kredit semester (sks) berdasarkan capaian pembelajaran yang terukur. Kegiatan ini idealnya berlangsung selama satu semester penuh dengan total beban kerja ekuivalen 20 sks, dan dapat dilakukan secara individu maupun dalam kelompok lintas disiplin [1].

Dalam proyek ini, keterlibatan mahasiswa dalam skema lomba dan kompetisi ilmiah seperti Program Kreativitas Mahasiswa bidang Artikel Ilmiah (PKM-AI) menjadi bentuk aktualisasi konkret dari pelaksanaan proyek independen. Dengan mengusung topik klasifikasi penyakit daun kentang berbasis Vision Transformer (ViT), proyek ini dirancang tidak hanya sebagai latihan teknis dalam penerapan model deep learning, tetapi juga sebagai wahana untuk membangun kapasitas riset mahasiswa, berpijak pada metode ilmiah, dan menghasilkan luaran akademik berupa artikel ilmiah.

Proyek independen berperan sebagai jembatan antara pendidikan tinggi dan ekosistem inovasi. Mahasiswa dilatih untuk bersikap otonom, reflektif, dan solutif, yang merupakan karakter penting dalam membentuk lulusan adaptif dan kompetitif secara global. Dengan demikian, proyek ini menjadi bagian dari high impact learning strategy yang berfokus pada keterlibatan aktif mahasiswa dalam proses pembelajaran, sejalan dengan arah transformasi pendidikan tinggi berbasis student-centered learning dan experiential learning [17].

2.2 Pengolahan Citra Digital

Pengolahan citra digital merupakan disiplin ilmu yang mempelajari teknik-teknik untuk memanipulasi citra dalam bentuk representasi diskret agar informasi visual di dalamnya dapat diekstraksi, dianalisis, atau dimanfaatkan lebih lanjut. Dalam konteks pengembangan sistem klasifikasi penyakit tanaman berbasis citra, tahap pengolahan citra memiliki peran yang sangat fundamental untuk memastikan bahwa data masukan yang diterima oleh model pembelajaran mesin telah memenuhi standar kualitas dan keseragaman representasi.

Citra digital sendiri didefinisikan sebagai fungsi dua dimensi $f(x,y)$, di mana x dan y menyatakan koordinat spasial, dan f merepresentasikan intensitas atau tingkat keabuan pada titik tersebut [18]. Untuk memungkinkan pemrosesan secara komputasional, fungsi kontinu ini didiskretisasi melalui proses sampling dan kuantisasi, menghasilkan array dua dimensi yang terdiri atas elemen-elemen pixel [19].

Dalam praktiknya, data citra yang diperoleh dari lapangan, seperti daun tanaman kentang yang diambil dari lingkungan terbuka, cenderung mengandung berbagai jenis ketidakaturan, seperti variasi pencahayaan, noise, perbedaan latar belakang, serta orientasi objek yang tidak seragam. Oleh karena itu, sebelum digunakan sebagai input dalam model klasifikasi berbasis deep learning, citra tersebut harus melalui tahapan prapemrosesan (preprocessing) untuk meningkatkan kualitas visual dan memperkuat representasi fitur.

Beberapa teknik pengolahan citra yang umum digunakan dalam studi klasifikasi tanaman meliputi:

- Penyesuaian ukuran (resizing): dilakukan untuk menyeragamkan dimensi citra sesuai spesifikasi arsitektur model (misalnya 224×224 piksel untuk ViT).
- Normalisasi: transformasi nilai pixel ke dalam rentang tertentu (misalnya 0 hingga 1 atau -1 hingga 1) untuk mempercepat proses pelatihan dan menjaga stabilitas numerik.
- Augmentasi data: metode augmentasi citra seperti rotasi, flipping, zooming, serta penyesuaian brightness dan kontras digunakan untuk meningkatkan keberagaman data dan mengurangi risiko overfitting [20].
- Transformasi ke tensor: konversi dari array piksel ke tensor numerik tiga dimensi agar kompatibel dengan input layer pada jaringan neural.

Proses-proses ini tidak hanya berfungsi untuk memperbaiki kualitas visual citra, tetapi juga untuk menjaga konsistensi distribusi data di seluruh tahap pelatihan, validasi, dan pengujian. Terlebih lagi, dalam model berbasis Vision Transformer (ViT), yang tidak menggunakan lapisan konvolusi tetapi mengandalkan patch embedding, ketepatan struktur spasial dan konsistensi pencahayaan menjadi sangat penting untuk menjaga ketajaman representasi fitur global. Dengan pengolahan citra yang baik, model ViT akan mampu mengenali pola-pola visual dari berbagai

kondisi daun kentang secara lebih akurat dan efisien, meskipun data diambil dari lingkungan pertanian yang tidak terkontrol.

2.3 Transfer Learning

Transfer learning merupakan pendekatan dalam pembelajaran mesin di mana pengetahuan yang telah diperoleh dari suatu domain atau tugas tertentu dimanfaatkan kembali untuk meningkatkan kinerja model pada tugas lain yang berkaitan. Teknik ini menjadi sangat populer dalam domain deep learning, terutama ketika jumlah data latih pada tugas target terbatas, namun tersedia model pralatih dari domain sumber yang memiliki representasi fitur yang kaya [21], [22].

Secara umum, transfer learning bekerja dengan mengadaptasi model yang telah dilatih pada dataset berskala besar, seperti ImageNet, untuk digunakan dalam tugas klasifikasi baru dengan cara melakukan pelatihan ulang sebagian parameter (fine-tuning) atau melatih head classifier saja. Dalam konteks visi komputer, pendekatan ini terbukti mempercepat waktu pelatihan, mengurangi kebutuhan komputasi, dan menghasilkan akurasi yang kompetitif bahkan pada dataset dengan kondisi visual yang tidak terkontrol [23].

Menurut IBM, transfer learning memiliki tiga komponen utama, yaitu[21]:

- Domain Sumber (Source Domain): mencakup ruang fitur dan tugas pembelajaran awal tempat model dilatih pertama kali.
- Domain Target (Target Domain): merupakan ruang fitur atau tugas baru tempat model akan diterapkan.
- Fungsi Transfer: yaitu strategi bagaimana pengetahuan dari domain sumber ditransfer ke domain target, baik melalui parameter, arsitektur, maupun representasi fitur.

Beberapa pendekatan umum dalam transfer learning mencakup:

- Fine-tuning: penyesuaian ulang parameter pada seluruh atau sebagian jaringan neural, memungkinkan model belajar fitur spesifik dari data target.

- Feature Extraction: menggunakan lapisan awal dari model pralatih sebagai feature encoder dan hanya melatih classifier head pada data target.
- Domain Adaptation: digunakan ketika terdapat perbedaan distribusi data antara domain sumber dan domain target.

Studi oleh Tan et al. (2018) menunjukkan bahwa transfer learning pada model convolutional neural networks (CNN) meningkatkan akurasi klasifikasi daun tanaman hingga lebih dari 90%, bahkan ketika dataset target memiliki volume data yang terbatas [24]. Hal serupa juga ditemukan dalam studi klasifikasi penyakit tanaman oleh Ramcharan et al. (2019), di mana model pralatih seperti ResNet dan Inception menunjukkan kinerja tinggi saat ditransfer ke tugas klasifikasi daun pisang yang terserang penyakit [25].

Vision Transformer (ViT) sebagai arsitektur berbasis self-attention juga dapat memanfaatkan transfer learning secara efektif. Dosovitskiy et al. (2021) dalam studi awal ViT menunjukkan bahwa model ViT yang dilatih pada ImageNet-21k dapat memberikan hasil unggul pada berbagai tugas downstream seperti klasifikasi dan deteksi objek. Dalam konteks penelitian ini, ViT pralatih digunakan sebagai basis untuk mendeteksi tujuh kategori penyakit daun kentang pada citra dari lingkungan tidak terkontrol, memanfaatkan kekuatan generalisasi dari fitur global yang telah dipelajari sebelumnya [26].

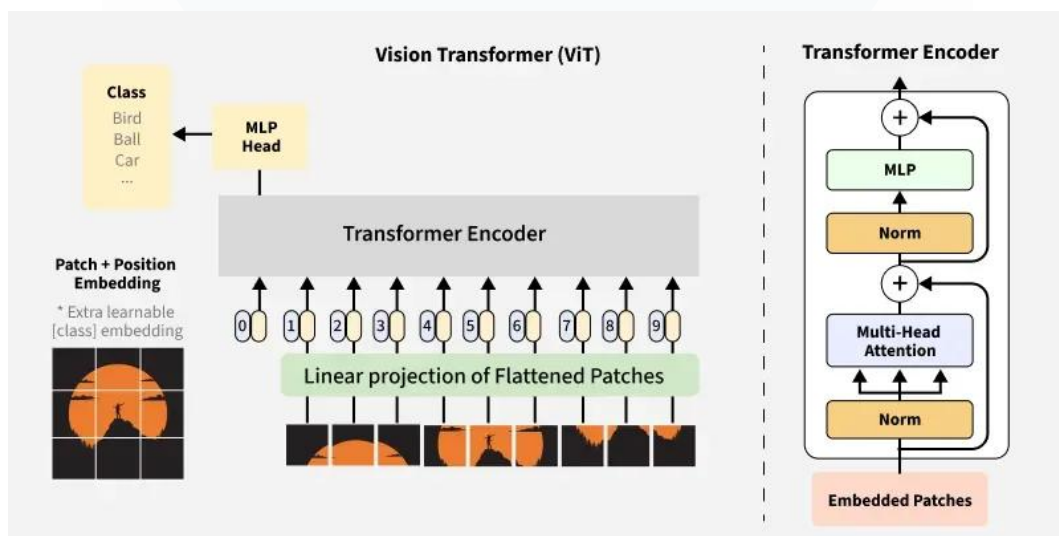
Dengan demikian, transfer learning tidak hanya memberikan efisiensi pelatihan, tetapi juga membuka peluang untuk penerapan model state-of-the-art pada domain yang memiliki keterbatasan data, seperti pertanian presisi. Strategi ini semakin relevan dalam skenario dunia nyata, di mana pengumpulan data besar seringkali memerlukan waktu dan biaya yang tinggi.

2.4 Vision Transformer (ViT)

Vision Transformer (ViT) merupakan pendekatan baru dalam bidang pengolahan citra digital berbasis deep learning yang mengadopsi arsitektur Transformer dari domain pemrosesan bahasa alami (Natural Language Processing, NLP) ke ranah visi komputer. Berbeda dengan pendekatan tradisional seperti

Convolutional Neural Networks (CNN), ViT tidak menggunakan operasi konvolusi untuk mengekstraksi fitur visual, melainkan mengandalkan perhatian global melalui mekanisme self-attention yang dapat menangkap hubungan antar piksel dalam skala luas secara paralel dan efisien [26].

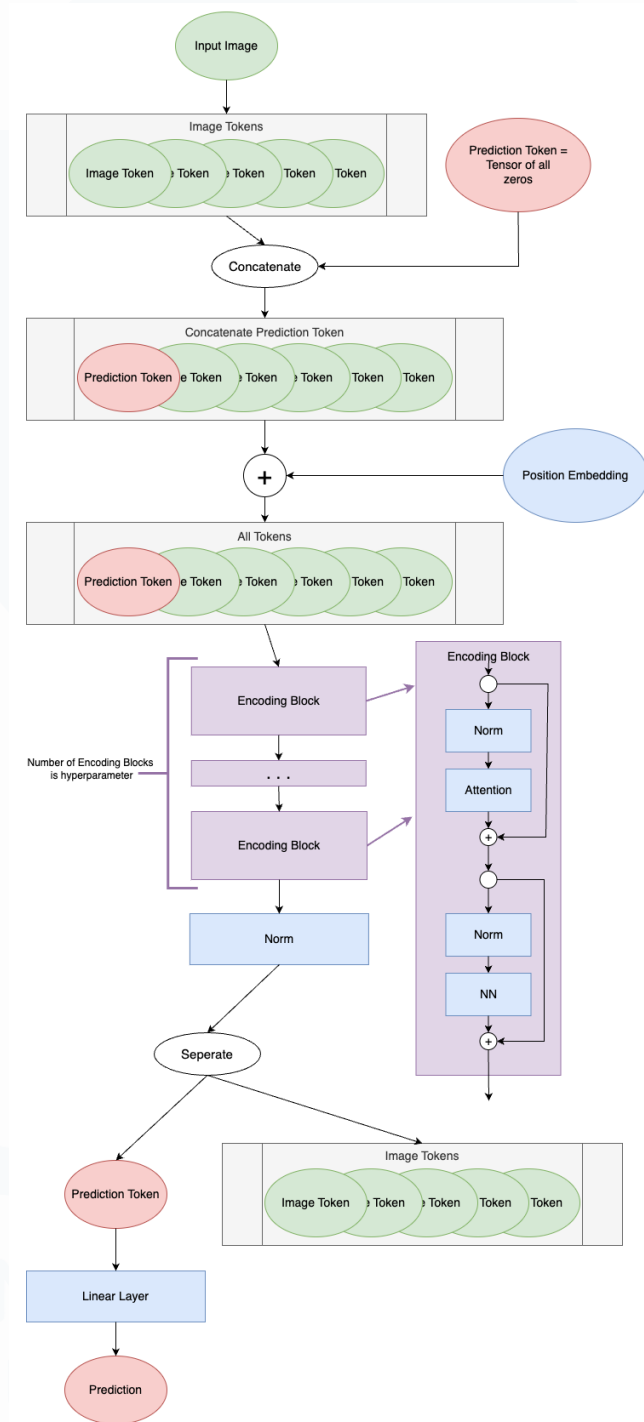
Dalam arsitektur ViT, citra masukan berukuran tetap (224×224) dibagi menjadi patch-patch kecil berukuran 16×16 piksel, menghasilkan total 196 patch. Setiap patch ini kemudian diubah menjadi vektor berdimensi tetap melalui proses linear projection. Hasil vektorisasi ini kemudian dikombinasikan dengan positional encoding untuk mempertahankan informasi spasial, dan dimasukkan sebagai input tokens ke dalam Transformer Encoder. Selain patch embedding, juga ditambahkan token khusus bernama class token yang berperan sebagai representasi akhir untuk proses klasifikasi. Arsitektur umum ini dapat dilihat pada Gambar 2.6 [27].



Gambar 2. 1 Arsitektur Vision Transformer [27]

Secara teknis, Transformer Encoder terdiri atas beberapa blok yang masing-masing memuat dua komponen utama: multi-head self-attention (MSA) dan feed-forward network (FFN). MSA memungkinkan model untuk memperhatikan berbagai bagian gambar secara bersamaan, sementara FFN memberikan kapasitas representasi non-linear. Setelah melalui beberapa lapisan encoder, class token yang telah mengalami pembelajaran dikirimkan ke Multilayer Perceptron (MLP) untuk

proses klasifikasi akhir. Diagram konseptual dari proses ini digambarkan pada Gambar 2.7 [28].



Gambar 2. 2 Diagram Model ViT [28]

Perbedaan fundamental antara ViT dan CNN terletak pada pendekatan pemrosesan spasial. CNN mengandalkan local receptive fields dan weight sharing untuk mengekstraksi fitur dari area gambar yang sempit dan bertumpuk, sedangkan ViT memproses gambar sebagai rangkaian token, sehingga dapat menangkap konteks global sejak awal proses pelatihan. Hal ini menjadikan ViT sangat efisien dalam memahami struktur citra yang kompleks, terutama ketika dilatih dengan dataset berukuran besar [27], [29].

Keunggulan ViT mencakup fleksibilitas terhadap variasi struktur input, kemampuan menangani konteks spasial yang luas, dan performa kompetitif dalam berbagai tugas klasifikasi gambar. Namun, ViT juga memiliki kekurangan, seperti kebutuhan akan data pelatihan dalam jumlah besar serta waktu komputasi yang relatif tinggi dibanding CNN konvensional. Oleh karena itu, pendekatan transfer learning dan pretraining pada dataset berskala besar seperti ImageNet-21k sering kali digunakan untuk meningkatkan efisiensi model [29].

Dengan mempertimbangkan kekuatan dan kelemahannya, ViT menjadi pilihan yang menjanjikan dalam konteks klasifikasi citra daun tanaman di lingkungan nyata yang penuh variasi visual. Model ini mampu menangkap fitur-fitur global yang tidak selalu dapat diakomodasi oleh CNN, menjadikannya relevan untuk diterapkan dalam proyek klasifikasi penyakit daun kentang.

2.5 Penelitian Terdahulu

Tinjauan pustaka ini disusun untuk menguraikan kerangka teoretis dan temuan empiris yang relevan sebagai landasan ilmiah dalam pelaksanaan Project Independent melalui skema Merdeka Belajar Kampus Merdeka (MBKM). Sesuai dengan Buku Panduan MBKM Tahun 2024, Project Independent merupakan aktivitas pembelajaran berbasis proyek yang menuntut mahasiswa untuk merancang, mengimplementasikan, dan merefleksikan solusi terhadap permasalahan nyata secara mandiri maupun kolaboratif, dengan menghasilkan luaran yang terukur [1]. Dalam konteks ini, mahasiswa tidak hanya berperan sebagai pelaksana kegiatan, tetapi juga sebagai perancang dan peneliti yang

bertanggung jawab terhadap validitas pendekatan ilmiah yang digunakan. Oleh karena itu, keberhasilan kegiatan ini sangat dipengaruhi oleh pemahaman terhadap literatur yang mendasari bidang permasalahan yang dikaji. Tinjauan ini mencakup pemetaan pendekatan metodologis dan arsitektur model dalam studi-studi terdahulu yang berkaitan dengan klasifikasi penyakit tanaman berbasis citra digital, yang menjadi fokus utama dari proyek yang diangkat dalam kompetisi PKM-AI.

Penelitian mengenai klasifikasi penyakit daun kentang berbasis citra digital telah mengalami perkembangan signifikan seiring dengan kemajuan teknologi deep learning. Salah satu metode yang paling banyak digunakan adalah Convolutional Neural Networks (CNN), yang telah terbukti efektif dalam mengenali pola visual dari citra tanaman.

Penelitian oleh Deep Learning Approaches for Crop Disease Classification (2024) menggunakan model DenseXnet dan menghasilkan akurasi sebesar 99,01% setelah menerapkan preprocessing dehazing untuk meningkatkan kualitas citra yang berkabut [30]. Hal ini menunjukkan bahwa teknik preprocessing yang tepat dapat berkontribusi besar terhadap akurasi model.

Dalam studi lain, Assessing the Performance of Domain-Specific Models for Plant Leaf Disease Classification (2024), sebanyak 23 arsitektur CNN dibandingkan dengan menggunakan 18 dataset publik. DenseNet, EfficientNet, dan ConvNeXt terbukti lebih unggul dibandingkan arsitektur lama seperti Inception dan VGG, terutama saat diuji pada dataset lapangan yang mencerminkan kondisi nyata [31]. Penelitian ini menegaskan pentingnya pemilihan arsitektur yang sesuai dengan karakteristik data.

Potato Leaf Disease Detection Based on a Lightweight Deep Learning Model (2024) secara spesifik menangani klasifikasi tujuh kelas penyakit daun kentang: Bacteria, Fungi, Healthy, Nematode, Pest, Phytophthora, dan Virus. Model RegNetY_400MF yang digunakan berhasil mencapai akurasi 89,87% dengan dukungan augmentasi data berupa rotasi dan penyesuaian kecerahan [32]. Studi ini

relevan karena menggunakan jumlah kelas dan jenis penyakit yang sama dengan penelitian ini.

Vision Transformer (ViT) sebagai pendekatan baru turut dikaji dalam *An Efficient Approach for Potato Leaf Disease Classification Using Cascaded CNN-Transformers* (2024). Penelitian ini menggabungkan ResNet50V2 dan ViT, menghasilkan akurasi sebesar 97% dengan waktu inferensi kurang dari satu detik [33]. Hasil ini menunjukkan bahwa ViT tidak hanya akurat tetapi juga efisien untuk klasifikasi citra penyakit tanaman.

DSCSkipNet: An Accuracy-Complexity Trade-off for Effective Potato Disease Identification in Uncontrolled Environments (2024) mengkaji efektivitas klasifikasi penyakit kentang di lingkungan pertanian terbuka. Model yang dikembangkan mencapai akurasi 83,23% meskipun menghadapi tantangan seperti ketidakseimbangan kelas dan kemiripan visual antar penyakit [34]. Ini menguatkan pentingnya pengujian pada dataset dari lingkungan nyata seperti yang digunakan dalam penelitian ini.

Dari berbagai penelitian tersebut, dapat disimpulkan bahwa keberhasilan model klasifikasi sangat dipengaruhi oleh kualitas data, strategi preprocessing, dan arsitektur model yang digunakan. Teknik seperti augmentasi data, transfer learning, dan fine-tuning terbukti dapat meningkatkan performa model secara signifikan. Oleh karena itu, penelitian ini memilih Vision Transformer (ViT) sebagai pendekatan untuk mengklasifikasikan tujuh kelas penyakit daun kentang, dengan fokus pada akurasi dan kemampuan adaptasi terhadap citra dari lingkungan tidak terkontrol.