BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi saat ini telah mengalami kemajuan pesat dalam kurun waktu yang relatif singkat. Salah satu indikator utama dari kemajuan ini adalah pengembangan kecerdasan buatan (*Artificial Intelligence*/AI) [1]. *Artificial Intelligence* (AI) atau kecerdasan buatan merupakan cabang ilmu komputer yang berfokus pada pengembangan sistem cerdas yang mampu meniru kemampuan berpikir dan bertindak seperti manusia dalam menyelesaikan berbagai tugas kompleks. Sistem AI dirancang untuk dapat menganalisis data, belajar dari pengalaman, menarik kesimpulan logis, serta melakukan koreksi mandiri terhadap kesalahan yang terjadi, sehingga menghasilkan keputusan yang lebih akurat di masa mendatang. Proses ini mencakup tiga komponen utama, yaitu learning (pembelajaran dari data historis), reasoning (penalaran terhadap informasi yang tersedia), dan *self-correction* (kemampuan untuk memperbaiki kesalahan secara otomatis). Kemampuan tersebut menjadikan AI sebagai teknologi yang sangat potensial untuk diimplementasikan di berbagai bidang, termasuk kesehatan, industri, pendidikan, dan pertanian [2].

Salah satu teknologi yang banyak mendapat perhatian dalam kecerdasan buatan adalah analisis citra [3]. Analisis citra digital adalah suatu teknik dalam bidang komputer yang bertujuan untuk memproses, memanipulasi, dan menganalisis citra digital guna memperoleh informasi yang berguna dari citra tersebut [4] . Salah satu contoh teknik analisis citra adalah klasifikasi gambar. Klasifikasi gambar adalah proses dalam bidang analsis citra yang bertujuan untuk mengidentifikasi dan memberikan label kategori pada suatu gambar berdasarkan fitur-fitur visual yang terkandung di dalamnya. Dalam proses ini, sistem komputer dilatih untuk mengenali pola atau karakteristik tertentu dari gambar (seperti bentuk, warna, tekstur, atau struktur objek) agar dapat menentukan kelas atau kategori yang sesuai [5].

Dalam melakukan klasifikasi gambar, teknik yang umum dilakukan adalah mengimplementasikan algoritma *deep learning*. *Deep Learning* adalah cabang dari *machine learning* (pembelajaran mesin) yang menggunakan jaringan saraf tiruan (*artificial neural networks*) dengan banyak lapisan (disebut *deep neural networks*) untuk mempelajari dan mengekstrak pola atau representasi kompleks dari data. Berbeda dengan *machine learning* tradisional yang membutuhkan fitur-fitur khusus yang dirancang secara manual (*feature engineering*), *deep learning* mampu belajar langsung dari data mentah (seperti gambar, teks, atau suara) dengan cara otomatis, melalui proses pembelajaran berlapis-lapis yang menyerupai cara kerja otak manusia [6].

Vision Transformer (ViT) merupakan salah satu algoritma *deep learning* yang efektif untuk menyelesaikan permasalahan berbasis analisis citra, termasuk dalam pengelolaan dan pemantauan tanaman. *Vision Transformer* merupakan arsitektur pemrosesan citra yang didasarkan pada prinsip dasar dari arsitektur Transformer [7]. *Vision Transformer* (ViT) adalah arsitektur *deep learning* yang dirancang khusus untuk tugas-tugas pengolahan citra, dan merupakan adaptasi dari arsitektur Transformer yang awalnya dikembangkan untuk pemrosesan bahasa alami (*Natural Language Processing*/NLP). Transformer pertama kali diperkenalkan oleh Vaswani, dkk. [8] pada tahun 2017 melalui paper berjudul "Attention is All You Need". Model ini merevolusi NLP dengan mekanisme *self-attention*, yang mampu memahami hubungan antar elemen dalam sebuah sekuens tanpa menggunakan struktur rekursif seperti LSTM (*Long Short-Term Memory*) atau RNN (*Recurrent Neural Network*). Transformer kemudian digunakan dalam model-model besar seperti BERT, GPT, dan T5. Kesuksesan Transformer di NLP mendorong peneliti untuk mengeksplorasi kemampuannya di bidang visi komputer. Pada tahun 2020, sebuah tim dari Google Research memperkenalkan *Vision Transformer* (ViT) dalam paper berjudul "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" [9] .

Model *Transformer* telah mulai diterapkan dalam bidang visi komputer, menggantikan arsitektur *Long Short-Term Memory* (LSTM) yang sebelumnya

digunakan. Perbedaan utama antara *Vision Transformer* dan LSTM terletak pada pendekatan pemrosesan data citra. LSTM dirancang untuk menangani data berurutan dengan kemampuan bawaan dalam memahami hubungan temporal, sedangkan Vision Transformer dirancang untuk mengidentifikasi hubungan spasial dalam data citra dan dioptimalkan untuk tugas pengolahan gambar [10]. Arsitektur *Vision Transformer* terdiri dari beberapa komponen utama, yaitu proses pembentukan *patch embeddings*, mekanisme *multi-head attention*, dan lapisan *multi-layer perceptron* [7]. Kemampuannya dalam mempelajari pola-pola kompleks pada data citra memberikan keunggulan dalam hal akurasi dan kecepatan identifikasi dibandingkan dengan pendekatan manual atau metode konvensional lainnya.

Terdapat beberapa penelitian yang membahas tentang klasifikasi gambar menggunakan algoritma *vision transformer*. Seperti penelitian yang dilakukan oleh Arya Pangestu [11] di mana melakukan pembandingan antara model *Vision Transformer* (ViT) dan *Convolutional Neural Network* (CNN) sebagai pembandingnya dan mendapatkan hasil ViT lebih unggul dalam hal generalisasi dan CNN lebih unggul dalam waktu pelatihan.

Selain itu, penelitian yang dilakukan oleh Ganjar Gingin Tahyudin, [12] yang melakukan klasifikasi gender melakukan penggunaan algoritma *Vision Transformer* di mana didapatkan akurasi validasi dan pengujian di atas 90%, dengan akurasi tertinggi 0.9843 pada gambar 224 x 224 piksel dan 28 *patch* serta untuk evaluasi *cross-dataset* menunjukkan konfigurasi optimal dengan gambar 224 x 224 piksel dan 14 *patch*, menghasilkan akurasi 0.8174 dan skor F1 sebesar 0.8189.

Penelitian lain yang menguji Algoritma Vision Transformer yaitu seperti yang ditulis oleh Rayhan Uthama, [13] di mana dilakukan klasifikasi mengenali perbedaan variasi pada ikan koi. Didapatkan hasil bahwa model *Vision Transformer* (ViT) untuk klasifikasi 15 variasi ikan koi mencapai akurasi 89% setelah diuji dengan 300. Hasil ini mengungguli metode CNN yang mencapai 84% pada penelitian sebelumnya yang dijadikan acuan.

Dari penelitian-penelitian yang sudah ada mengenai implementasi vision transformer untuk klasifikasi gambar, penulis belum menemukan implementasi arsitektur *vision transformer* untuk klasifikasi terhadap tanaman singkong. Maka dari itu, penulis merasa perlu untuk mengimplementasikan *vision transformer* terhadap salah satu komoditas pertanian yang memiliki peranan krusial dalam bidang pangan yaitu tanaman singkong. Implementasi *Vision Transformer* untuk klasifikasi penyakit memungkinkan deteksi dini terhadap penyakit atau gangguan pada tanaman singkong. Penulis berencana untuk menguji sejauh mana performa algoritma vision transformer dalam melakukan proses klasifikasi gambar pada penyakit tanaman singkong.

Pada penelitian ini penulis akan menggunakan algoritma *deep learning* dengan arsitektur *Vision Transformer* (ViT) untuk membuat model yang nantinya akan digunakan untuk melakukan klasifikasi pada data gambar tanaman daun singkong yang sudah terkena penyakit ataupun masih sehat untuk kemudian mengklasifikasikannya. Penyakit yang akan diklasifikasikan yaitu *Cassava Bacterial Blight, Cassava Brown Streak Disease, Cassava Green Mottle, Cassava Mosaic Disease*, dan daun yang masih sehat.

1.1. Rumusan masalah

1. Bagaimana mengimplementasikan model klasifikasi citra daun singkong menggunakan arsitektur *Vision Transformer* (ViT)?
2. Bagaimana pengaruh variasi *hyperparameter* seperti jumlah *epoch*, ukuran *batch*, jenis optimizer, dan rasio pembagian data terhadap performa model Vision Transformer dalam klasifikasi penyakit daun singkong?

1.2. Pertanyaan Penelitian

1. Bagaimana mengimplementasikan model dengan arsitektur *Vision Transformer* yang mampu mengklasifikasikan jenis penyakit pada tanaman daun singkong?

2. Bagaimana hasil performa model yang dirancang dalam mengklasifikasikan jenis penyakit pada tanaman daun singkong?
3. Faktor-faktor apa saja yang mempengaruhi performa model *Vision Transformer* dalam mengklasifikasikan jenis penyakit pada tanaman daun singkong?

1.3. Tujuan Penelitian

1. Untuk mengimplementasikan model dengan arsitektur *Vision Transformer* yang mampu mengklasifikasikan jenis penyakit pada tanaman daun singkong.
2. Untuk mengetahui hasil performa model yang diimplementasikan dalam mengklasifikasikan jenis penyakit pada tanaman daun singkong.
3. Untuk mengetahui faktor-faktor apa saja yang mempengaruhi performa Model *Vision Transformer* dalam mengklasifikasikan jenis penyakit pada tanaman daun singkong.

1.4. Batasan Masalah

1. Variasi Parameter yang diujikan ke masing-masing model berupa *epoch, batch size, optimizer* dan variasi rasio *splitting* data
2. Hasil penelitian tidak sampai membuat *prototype* untuk *website* dan *mobile.*
3. *Dataset* yang digunakan berupa *dataset* sekunder yang terdiri dari 5 kelas yaitu *Cassava Bacterial Blight, Cassava Brown Streak Disease, Cassava Green Mottle, Cassava Mosaic Disease*, dan *Healthy.*

1.5. Manfaat Penelitian

1. Penelitian ini diharapkan berkontribusi terhadap pengembangan ilmu pengetahuan, khususnya dalam bidang *Computer Visi*on dan *Deep Learning*, dengan mengimplementasikan arsitektur *Vision Transformer* (ViT) untuk klasifikasi citra daun singkong

2. Penelitian ini diharapkan dapat menjadi landasan dan referensi bagi peneliti lain yang tertarik untuk mengembangkan atau membandingkan metode *deep learning* lainnya untuk identifikasi penyakit pada tanaman singkong atau komoditas pertanian lainnya.
3. Penelitian ini menunjukkan performa model *Vision Transformer* dalam tugas klasifikasi citra, serta mengkaji pengaruh variasi *hyperparameter* (*epoch, batch size, optimizer*, dan *data split*) terhadap performa model. Temuan ini diharapkan dapat menjadi acuan dalam perancangan model ViT yang lebih efisien dan optimal di masa mendatang.