

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang digunakan pada penelitian dengan judul Optimasi Arsitektur VGG16 dengan *Bayesian Optimization* pada Klasifikasi Penyakit Daun Tomat.

1.1. Latar Belakang

Tomat memiliki nama ilmiah *Lycopersicon esculentum* dan tergolong dalam famili *Solanaceae* (Mauseth, 2016). Adanya revisi yang dilakukan terhadap klasifikasi filo genetik *Solanaceae*, membuat genus *Lycopersicon* diintegrasikan kembali ke dalam genus *Solanum* dengan nomenklatur yang baru (Bai & Lindhout, 2007). Pada akhirnya, nama ilmiah tomat diubah menjadi *Solanum lycopersicum* dan merupakan satu-satunya spesies dalam genus *Solanum* yang dibudidayakan (Peralta dkk., 2005). Budidaya tomat merupakan salah satu industri pertanian utama dan ditanam secara luas di seluruh dunia (Foolad, 2007). Tomat merupakan sayuran kedua yang paling banyak dikonsumsi di dunia setelah kentang (Adhikari dkk., 2017). Mengonsumsi tomat maupun produk olahannya yang kaya akan antioksidan seperti karotenoid (terutama likopen), asam askorbat, dan senyawa fenolik, secara teratur telah terbukti dapat mengurangi risiko berbagai jenis kanker dan penyakit kardiovaskular (Borguini & Torres, 2009). Menurut data dari Badan Pusat Statistik, produksi tomat pada tahun 2021 mengalami peningkatan sekitar 2,71% terhadap produksi pada tahun 2020 dan terus meningkat hingga 4,876% pada tahun 2022 (BPS, 2023). Melihat peningkatan produksi tomat yang terus meningkat dari tahun ke tahun, diperlukan sebuah tindakan yang dapat mempertahankan bahkan meningkatkan produktivitas pertanian yang berkelanjutan ini. Seiring berkembangnya pertanian ekologis modern, peningkatan mutu dan kualitas dari suatu tanaman budidaya menjadi lebih penting (D. Jiang dkk., 2020), disamping kuantitas produk yang harus tetap ditingkatkan.

Kehadiran Organisme Pengganggu Tumbuhan (OPT) pada hakikatnya dapat menurunkan kualitas maupun kuantitas produksi dari suatu tanaman budidaya (Setiawati dkk., 2001). Penyakit yang menyerang tanaman budidaya kian hari semakin beranekaragam dan semakin kompleks. Sumber penyakit yang biasa menyerang tanaman tomat berasal dari

bakteri, jamur, *oomycete*, nematoda, parasit, virus, dan lain sebagainya (Suresh, 2017). Penyakit-penyakit ini dapat menyerang bagian daun, batang, buah, maupun akarnya. Salah satu cara yang dapat dilakukan untuk mengetahui adanya penyakit/kelainan pada tanaman tomat adalah melalui bagian daunnya. Beberapa penyakit daun tomat yang umum dijumpai adalah *Bacterial Spot*, *Early Blight*, *Late Blight*, *Leaf Mold*, *Septoria Leaf Spot*, *Two-spotted Spider Mites*, *Target Spot*, *Tomato Mosaic Virus*, *Tomato Yellow Leaf Curl Virus*, dan masih banyak lagi (Suresh, 2017). Pengaruh penyakit-penyakit tersebut terhadap pertumbuhan tanaman tomat sangatlah besar (Chakravarthy & Raman, 2020). Semakin beranekaragam jenis penyakitnya, maka semakin rumit pula proses pencegahan maupun penanganannya. Sebagai contoh, tindakan pencegahan terhadap penyakit *Bacterial Spot* adalah hanya dengan menaburkan benih yang telah diuji dan disertifikasi bebas dari bakteri ini. Penyemprotan dengan *copper fungicide spray* atau *copper-resistant strains* bisa dilakukan untuk memberikan perlindungan yang moderat/sedang. Terkait tindakan penanganan apabila telah terinfeksi penyakit ini, petani dapat menghindari melakukan *overhead irrigation* karena bakteri ini dapat menyebar dengan cepat melalui pergerakan air dari satu daun ke daun tanaman lainnya (HGIC, 2021; Suresh, 2017). Di sisi lain, tindakan pencegahan terhadap penyakit *Tomato Mosaic Virus* adalah dengan menghindari penanaman di lahan yang sebelumnya telah terinfeksi, mencuci tangan sampai bersih setelah merokok (karena virus ini bermula dari *Tobacco Mosaic Virus*), serta tetap menjaga kebersihan berupa sanitasi peralatan perkebunan dan rutin menjalankan inspeksi tanaman. Terkait tindakan penanganan apabila telah terinfeksi penyakit ini adalah dengan menggunakan *dry milk* tanpa lemak yang mengandung 3,5% protein serta mengaplikasikan disinfektan kimia selama perawatan tanaman, pemangkasan, dan pencabutan daun pada lahan budidaya yang steril (HGIC, 2021; Suresh, 2017).

Perbedaan jenis organisme yang menyerang tanaman membuat para petani budidaya harus melakukan tindakan yang berbeda pula. Dibutuhkan adanya bantuan para ahli untuk mengamati suatu area tanam tertentu dengan tujuan agar perlakuan terhadap tanaman yang terinfeksi penyakit dapat dilakukan dengan tepat. Meski demikian, pada praktiknya para ahli patologi pun tetap bergantung dengan metode pengamatan melalui laboratorium untuk memastikan diagnosis penyakit/kelainan pada tanaman tertentu (Suresh, 2017). Proses pengamatan penyakit daun tomat secara langsung dengan mata telanjang sangat rentan terhadap faktor lingkungan dan intervensi kesalahan manusianya (*human error*), terlebih juga bergantung pada wawasan maupun pengalaman dari para ahli. Oleh karenanya,

pelaksanaan penelitian untuk merancang sebuah metode pencegahan dan diagnosis penyakit tanaman yang lebih efektif dan efisien menjadi sangat penting (D. Jiang dkk., 2020).

Artificial Intelligence (AI) berkembang pesat dan telah diterapkan dalam berbagai bidang seperti automasi pekerjaan manusia, pengenalan suara atau gambar, diagnosis ilmu medis dan penelitian riset ilmiah. *AI deep learning* merupakan konsep yang memungkinkan komputer untuk mempelajari persoalan rumit dengan cara menyusunnya dari bentuk yang lebih sederhana. Sebuah sistem yang dibangun dengan *hard-code knowledge* memiliki keterbatasan dalam mendapatkan pengetahuan yang lebih luas. *Machine Learning* (ML) diperkenalkan untuk mengatasi masalah ini, sehingga mesin dapat mempelajari dan memperoleh berbagai informasi dengan cara mengekstraksi *pattern* dari suatu data mentah. Mayoritas algoritma *machine learning* dikategorikan dalam *supervised* dan *unsupervised learning*. Terdapat berbagai macam *task* yang dapat diselesaikan dengan *machine learning*, seperti klasifikasi, regresi, transkripsi, mesin penerjemah, deteksi anomali, dan lain sebagainya. Permasalahan yang harus diselesaikan oleh komputer untuk dapat menentukan suatu *input* termasuk dalam *output/class* tertentu dikenal sebagai *classification task* (Goodfellow dkk., 2016). Oleh karena penelitian terkait penyakit daun tomat ini ditujukan untuk membangun sebuah model yang dapat mengklasifikasikan suatu *input* ke dalam jenis penyakit tertentu dengan tepat, maka *classification task* merupakan jenis metode *machine learning* yang paling sesuai.

Implementasi *machine learning* untuk mengklasifikasi penyakit daun tomat telah banyak dilakukan oleh para peneliti di berbagai tempat di penjuru dunia. Pada tahun 2020, D. Jiang dkk menggunakan model ResNet-50 untuk mengklasifikasi penyakit daun tomat ke dalam 3 penyakit populer yakni *spot blight*, *late blight*, dan *yellow leaf curl disease*. Pelatihan model ResNet-50 dengan modifikasi penggunaan fungsi aktivasi Leaky-ReLU dan ukuran *kernel* 11x11 pada lapisan konvolusi pertama juga dilakukan sebagai pembanding. Penggunaan Leaky-ReLU dimaksudkan untuk mengurangi dampak inaktivasi neuron oleh ReLU dan meningkatkan kinerja *network* sampai batas tertentu dengan memperbaiki kemampuan menangkap fitur secara detail. Selain itu, perubahan ukuran *kernel* menjadi 11x11 pada lapisan konvolusi pertama dimaksudkan untuk memperluas bidang reseptif ekstraksi fitur. *Dataset* yang digunakan adalah *AI Challenger* dan diambil 3000 data gambar. Implementasi *random* augmentasi seperti *gaussian noise* dan transformasi *affine* pada data latih sehingga berjumlah 6794 gambar bertujuan untuk menghindari masalah *overfitting*.

Hasil yang diperoleh yaitu adanya peningkatan skor akurasi pelatihan dari 97,7% menjadi 98,3% dan peningkatan skor akurasi uji dari 95,7% menjadi 98,0%.

J. Waleed dkk. (2021) mengusulkan sebuah model *Convolutional Neural Network* (CNN) yang efektif dan akurat untuk mendeteksi 10 penyakit daun tomat yang terdiri dari 9 jenis penyakit dan 1 kondisi daun yang sehat. Beberapa *hyperparameter* yang diatur nilainya adalah jumlah lapisan konvolusi, jumlah *filter*, jumlah lapisan *max-pooling*, jumlah *fully-connected layer*, ukuran *dropout*, *optimizer*, *learning rate*, dan sebagainya. Menggunakan *dataset Tomato Leaf Disease Detection* dari *Kaggle*, pembagian terhadap 11000 data gambar dilakukan sehingga menjadi 9900 data latih dan 1100 data uji. Hasil akurasi uji yang diperoleh adalah 98,0%. Meski demikian, J. Waleed dkk. (2021) kurang memberikan informasi yang lebih detail terkait arsitektur model yang diusulkannya, sehingga mempersulit peneliti lain dalam mengimplementasikan kembali arsitektur tersebut. S. Omar dan R. Jain (2022) juga mengusulkan sebuah model CNN untuk mengklasifikasi 10 penyakit daun tomat namun dengan sumber *dataset* dan pembagian rasio antara data latih dan data uji yang berbeda. *Dataset* dari *PlantVillage* dipilih dan dibagi dengan perbandingan 60:40 antara data latih dan data uji. Ukuran *input* gambar untuk arsitektur model CNN yang diusulkan adalah 64x64 piksel. Meskipun telah melakukan augmentasi seperti *rescale*, *shear*, *zoom*, dan *horizontal flip* pada data latih, serta augmentasi *rescale* pada data uji, hasil akurasi uji yang didapatkan hanya sekitar 92,0%.

Penggunaan model CNN juga pernah dilakukan oleh D. Malunao dkk. (2022) dengan arsitektur YOLOv3 untuk mengklasifikasi 3 penyakit umum pada daun tomat, yakni *healthy*, *early blight*, dan *septoria leaf spot*. D. Malunao dkk. (2022) menggunakan *dataset* yang sama dengan J. Waleed dkk. (2021) namun hanya mengambil 300 data gambar saja, yang kemudian dibagi menjadi 240 data latih dan 60 data uji. Hasil percobaan yang diperoleh berupa *mean Average Precision* (mAP) sebesar 98,28% dengan akurasi hasil pelatihan model yang berkisar antara 75% - 99%. Sementara itu, M. Mamun dkk. (2020) pernah mengusulkan sebuah model CNN yang diberi nama TLNet (*Tomato Leaf Net*). Meskipun menggunakan sumber *dataset* yang sama dengan S. Omar dan R. Jain (2022), M. Mamun dkk. (2020) hanya melakukan klasifikasi penyakit daun tomat ke dalam 8 kelas penyakit saja, yang terdiri dari 1 kondisi sehat dan 7 penyakit. Cara pembagian *dataset*-nya cukup unik, dimulai dengan mengambil 50 data di setiap kelas untuk data uji (total 400 data), kemudian mengambil 80% sisanya untuk data latih (total 9123 data) dan 20% untuk data validasi (total 2280 data).

Implementasi penghilangan *outliers*, *resize images*, normalisasi, augmentasi, dan *class label encoding*, mendapat hasil akurasi uji sebesar 98,77%.

Terkait upaya pencarian suatu desain yang paling optimal, manusia membutuhkan waktu dan sumber daya yang tidak sedikit untuk mengevaluasi bentuk-bentuk desain yang kurang optimal. Pentingnya penyesuaian *hyperparameter* serta arsitektur *network* dengan cermat saat perancangan sebuah *deep neural network* sangat berpengaruh terhadap performa model yang dibangun. Meski demikian, pengaturan yang efektif hanya bisa didapatkan melalui *trial and error*, seperti melatih beberapa *network* dengan pengaturan yang berbeda kemudian dievaluasi kinerjanya terhadap data validasi (Garnett, 2023). *Bayesian optimization* merupakan suatu metode optimisasi yang menggunakan pendekatan probabilistik berdasarkan teorema *Bayes* untuk menentukan konfigurasi *hyperparameter* paling optimal dari suatu fungsi objektif yang membutuhkan waktu lama untuk dievaluasi (Frazier, 2018). *Bayesian optimization* akan memodelkan kinerja generalisasi algoritma pembelajaran sebagai sampel dari *Gaussian Process* (GP)(Larochelle dkk., 2012).

Arsitektur VGG16 (*Visual Geometry Group 16*) pernah digunakan sebelumnya oleh Iffaty dkk. (2023) untuk mengklasifikasi 3 penyakit daun tomat. Tanpa adanya modifikasi berupa penggunaan *callback function*, hasil akurasi validasi yang diperoleh sebesar 96,27%. Sementara itu, implementasi *callback function* bernama *FastAI* berhasil meningkatkan skor akurasi validasi hingga mencapai 99,03%. Hasil penelitian tersebut membuktikan bahwa potensi arsitektur VGG16 sangatlah tinggi dalam menangani tugas klasifikasi suatu citra. Terlebih apabila dilakukan suatu modifikasi maupun optimasi terhadap arsitekturnya. Arsitektur VGG16 juga cukup terkenal di kalangan peneliti dikarenakan struktur jaringan yang sederhana serta relatif mudah dipahami dan dimodifikasi (Z. Jiang, 2019). Oleh karenanya, arsitektur VGG16 digunakan pada penelitian ini untuk mengklasifikasikan 10 penyakit daun tomat dengan metode *transfer learning*. Modifikasi arsitektur tersebut dilakukan, dilanjutkan dengan mengoptimasi beberapa *hyperparameter* yang terdapat dalam struktur jaringan VGG16 termodifikasi menggunakan *bayesian optimization* agar mendapat hasil yang lebih baik. Kombinasi nilai-nilai terbaik untuk setiap *hyperparameter* dicari secara otomatis dengan tujuan agar proses pencarian tidak perlu dilakukan secara *trial and error*. Pada akhirnya, hasil pelatihan arsitektur VGG16 termodifikasi yang telah dioptimalkan dibandingkan dengan hasil pelatihan arsitektur VGG16 orisinal.

1.2. Rumusan Masalah

Berdasarkan permasalahan yang telah disampaikan pada latar belakang, rumusan masalah pada penelitian ini adalah bagaimana mengoptimasi arsitektur VGG16 dengan *bayesian optimization* untuk mengklasifikasikan penyakit daun tomat agar menghasilkan skor akurasi yang tinggi.

1.3. Tujuan dan Manfaat

Tujuan dari penelitian ini adalah sebagai berikut.

1. Memodifikasi arsitektur VGG16 yang dioptimasi dengan *bayesian optimization* untuk mengklasifikasikan penyakit daun tomat pada *dataset Tomato Leaf Disease Detection*.
2. Mencari dan menentukan kombinasi nilai-nilai *hyperparameter* untuk mendapatkan model klasifikasi penyakit daun tomat yang terbaik menggunakan pencarian *bayesian optimization*.
3. Membandingkan performa arsitektur VGG16 termodifikasi yang paling optimal dengan arsitektur VGG16 orisinal dalam mengklasifikasikan penyakit daun tomat berdasarkan skor *test accuracy* dan *test loss*.

Manfaat dari penelitian ini adalah bahwa model hasil modifikasi arsitektur VGG16 yang telah dioptimasi, dapat digunakan untuk pengembangan aplikasi kedepannya, yang mampu mengklasifikasi penyakit daun tomat dengan tepat. Dalam lingkup masyarakat, para petani atau penanam tanaman tomat dapat menggunakan aplikasi yang telah dikembangkan untuk mendeteksi dan mengklasifikasi penyakit daun tomat, sehingga mereka dapat melakukan tindakan pencegahan maupun penanganan dini yang sesuai dengan penyakit yang menyerang tanaman tomat tersebut.

1.4. Ruang Lingkup

Ruang lingkup penelitian ini dibatasi pada beberapa aspek agar pembahasan lebih terarah dan tidak melebihi topik yang diteliti, diantaranya adalah sebagai berikut :

1. *Dataset* yang digunakan merupakan *dataset* dari *Kaggle* bernama “*Tomato Leaf Disease Detection*” dengan total 11000 data gambar berwarna (RGB).

2. *Dataset* memiliki 10 kelas penyakit yaitu : *bacterial spot, early blight, healthy, late blight, leaf mold, septoria leaf spot, two-spotted spider mites, target spot, tomato mosaic virus*, dan *tomato yellow leaf curl virus*.
3. Arsitektur yang dikembangkan dalam penelitian ini adalah dari arsitektur VGG16.
4. Optimasi arsitektur VGG16 termodifikasi adalah dengan *bayesian optimization*.
5. *Hyperparameter* yang dioptimasi adalah ukuran *input* gambar, jumlah *filter* tiap lapisan konvolusi, ukuran setiap *dropout rate* setelah lapisan konvolusi, dan ukuran setiap *dense unit* berdasarkan rentang/pilihan nilai tertentu.
6. Perbandingan performa model melalui skor *test accuracy* dan *test loss* terhadap modifikasi arsitektur VGG16 paling optimal hasil *bayesian optimization* dengan arsitektur VGG16 orisinal.

1.5. Sistematika Penulisan

Untuk memberikan suatu gambaran yang urut, runtut, dan jelas mengenai pembahasan penyusunan laporan penelitian dengan judul “Optimasi Arsitektur VGG16 dengan *Bayesian Optimization* pada Klasifikasi Penyakit Daun Tomat”, maka dibuatlah sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Bab ini membahas terkait latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan pada penelitian yang berjudul “Optimasi Arsitektur VGG16 dengan *Bayesian Optimization* pada Klasifikasi Penyakit Daun Tomat”.

BAB II TINJAUAN PUSTAKA

Bab ini menjabarkan studi pustaka serta teori-teori yang berhubungan dengan topik atau masalah pada penelitian ini. Pokok bahasan yang dibahas meliputi : *State of The Art*, Penyakit Daun Tomat, Klasifikasi Citra, *Deep Learning*, Optimasi pada *Deep Models*, *Convolutional Neural Network*, *Adam Optimizer*, *Bayesian Optimization*, *Transfer Learning*, *Activation Function*, Arsitektur VGG16, *Loss Function*, *Metric Function*, *Confusion Matrix*, serta *Tools and Library*.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tentang metodologi yang digunakan dalam penelitian “Optimasi Arsitektur VGG16 dengan *Bayesian Optimization* pada Klasifikasi Penyakit Daun Tomat”. Proses yang dilakukan dalam menyelesaikan topik permasalahan ini meliputi : garis besar penyelesaian masalah, pengumpulan data, pengecekan dan perubahan data, pra-pemrosesan dan pembagian data, optimasi arsitektur VGG16 termodifikasi menggunakan *bayesian optimization*, pelatihan model, pengujian model, serta evaluasi model.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang lingkungan dan perangkat yang digunakan dalam penelitian ini, skenario penelitian model klasifikasi penyakit daun tomat, serta menyajikan hasil dan analisis dari penelitian “Optimasi Arsitektur VGG16 dengan *Bayesian Optimization* pada Klasifikasi Penyakit Daun Tomat”.

BAB V PENUTUP

Bab ini berisi kesimpulan berdasarkan seluruh pembahasan yang telah dijabarkan di bab-bab sebelumnya, serta adanya beberapa saran untuk pengembangan penelitian kedepannya.