

BAB I PENDAHULUAN

1.1 Latar Belakang

Di Indonesia, jamur merupakan salah satu bahan pangan yang banyak ditemui dan dikonsumsi karena memiliki nilai ekonomi yang cukup tinggi. Namun, kemiripan bentuk dan warna antara jamur yang dapat dikonsumsi (*edible*) dan yang beracun (*poisonous*) sering menyebabkan kesalahan dalam proses identifikasi, sehingga berisiko menimbulkan keracunan [1]. Dalam beberapa tahun terakhir, kasus keracunan akibat konsumsi jamur liar menjadi perhatian serius karena tingginya angka kejadian serta dampak kesehatannya, yang umumnya disebabkan oleh kurangnya pengetahuan masyarakat dalam membedakan jenis jamur yang aman dan beracun [2]. Kondisi ini menunjukkan perlunya sistem identifikasi jamur yang akurat dan mudah diakses untuk mencegah risiko kesehatan, di mana dengan pendekatan berbasis citra digital proses identifikasi dapat dilakukan secara otomatis dan lebih konsisten dibandingkan penilaian manual [3]. Oleh karena itu, pengembangan metode klasifikasi jamur berbasis kecerdasan buatan menjadi topik penting dalam mendukung kesehatan dan keselamatan masyarakat.

Seiring dengan kemajuan teknologi kecerdasan buatan, metode *Convolutional Neural Network (CNN)* telah banyak digunakan dalam klasifikasi citra, termasuk untuk identifikasi jamur [4]. Penerapan *CNN* berbasis *mobile* terbukti dapat meningkatkan akurasi dalam proses klasifikasi jamur [5], sedangkan penggunaan model *InceptionV3* juga menunjukkan peningkatan performa yang signifikan dalam mengenali jamur yang dapat dikonsumsi [6]. Keunggulan utama *CNN* terletak pada kemampuannya mengekstraksi fitur visual yang kompleks secara otomatis [7]. Namun, penerapan *CNN* umumnya membutuhkan sumber daya komputasi yang besar, terutama pada tahap pelatihan dan *inference* [8]. Oleh karena itu, dibutuhkan model dengan arsitektur yang lebih ringan namun tetap mampu mempertahankan tingkat akurasi yang tinggi.

Salah satu solusi untuk mengatasi keterbatasan tersebut adalah dengan menggunakan arsitektur MobileNetV3, yang dirancang khusus untuk efisiensi tinggi pada perangkat *mobile* [9]. Model MobileNetV3 yang dioptimalkan terbukti mampu mengklasifikasikan citra jamur dengan latar belakang kompleks secara efisien tanpa penurunan akurasi [10]. Model ini memanfaatkan blok *depthwise separable convolution* dan mekanisme *squeeze-and-excitation* untuk mempercepat proses komputasi [11]. Penggunaan model ringan seperti MobileNetV3 juga mendukung implementasi sistem klasifikasi pada perangkat terbatas, seperti *smartphone* [12].

Meskipun MobileNetV3 sudah efisien, arsitektur ini masih memiliki redundansi pada *channel* dan fitur, sehingga ukuran model dan kebutuhan komputasinya masih dapat dioptimalkan agar *inference* menjadi lebih cepat dan hemat memori [13]. Tingkat beban komputasi, latensi *inference*, dan konsumsi daya perangkat *mobile* meningkat sebagai akibat dari redundansi *channel* dan fitur. Kondisi ini dapat menghambat sistem dan menurunkan efisiensi operasional secara keseluruhan ketika digunakan secara real-time pada *smartphone* atau perangkat *edge* dengan keterbatasan memori dan baterai. Oleh karena itu, bukan hanya peningkatan tambahan, optimalisasi melalui teknik pemangkasan parameter menjadi sangat penting. Beberapa teknik digunakan untuk mengurangi beban komputasi, seperti *Channel Pruning*, *Structured Pruning*, dan *Layer-Adaptive Magnitude-based Pruning (LAMP)* [14], tetapi metode pemangkasan tersebut bersifat statis dan tidak mempertimbangkan kontribusi tiap fitur secara adaptif. Hal ini membuat redundansi *channel* dan fitur masih tersisa sehingga efisiensi model belum optimal [15]. Dibutuhkan pendekatan yang lebih selektif dan adaptif dalam menentukan parameter yang benar-benar penting bagi proses klasifikasi karena keterbatasan metode *pruning* konvensional ini menunjukkan perbedaan dalam penelitian sebelumnya. Oleh karena itu, diperlukan pendekatan yang lebih selektif seperti *Attention-Guided Pruning*, dengan memanfaatkan mekanisme perhatian (*attention*) untuk menentukan tingkat kepentingan setiap fitur, sehingga parameter yang kurang berkontribusi dapat dipangkas secara lebih selektif.

Teknik *Attention-Guided Pruning* dikembangkan untuk meningkatkan efisiensi model *deep learning* dengan memangkas neuron yang kurang berpengaruh berdasarkan peta perhatian dari lapisan model [16]. Metode ini mampu mengurangi kompleksitas model tanpa menyebabkan penurunan akurasi yang signifikan [17]. Pendekatan ini memungkinkan model menjadi lebih ringan dan cepat dalam proses *inference*, terutama pada perangkat dengan keterbatasan memori [18]. Ketika diterapkan pada arsitektur MobileNetV3, metode ini berpotensi menghasilkan model klasifikasi jamur yang lebih efisien sekaligus tetap akurat [19]. Meskipun *Attention-Guided Pruning* dapat mengurangi jumlah parameter dan operasi komputasi, model umumnya masih menggunakan representasi numerik *floating-point* 32-bit yang membutuhkan memori dan energi cukup besar. Teknik optimasi lanjutan diperlukan untuk semakin menekan kebutuhan sumber daya tanpa mengorbankan akurasi model karena penggunaan presisi 32-bit pada perangkat mobile dapat memperlambat proses komputasi dan meningkatkan konsumsi memori. Oleh karena itu, diperlukan teknik optimasi tambahan berupa *Quantization*.

Quantization bertujuan menurunkan presisi bobot dan aktivasi, misalnya dari *floating-point* 32-bit menjadi 16-bit atau 8-bit, sehingga ukuran model menjadi lebih kecil dan proses *inference* lebih cepat serta hemat energi [20]. Teknik ini sangat relevan untuk implementasi model *deep learning* pada perangkat *mobile*. Penggunaan *Post-Training Quantization* (PTQ) memungkinkan model yang telah dilatih dikonversi ke presisi lebih rendah tanpa proses pelatihan ulang yang kompleks [21]. Beberapa skema seperti *PTQ INT8*, *PTQ INT16*, dan *PTQ FP16* mampu mengurangi kebutuhan memori dan latensi *inference* dengan penurunan akurasi yang minimal [22]. Oleh karena itu, penelitian ini bertujuan menerapkan *Attention-Guided Pruning* dan *Quantization* pada arsitektur MobileNetV3 untuk meningkatkan efisiensi komputasi dan akurasi klasifikasi jamur *edible* dan *poisonous*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja model *baseline* MobileNetV3 dalam melakukan klasifikasi jamur *edible* dan *poisonous* ditinjau dari akurasi, ukuran model, dan waktu *inference*?
2. Sejauh mana penerapan *Attention-Guided Pruning* dan *Quantization* pada arsitektur MobileNetV3 dapat meningkatkan efisiensi komputasi dibandingkan dengan model *baseline*?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini ditetapkan untuk memperjelas ruang lingkup agar fokus pada tujuan penelitian. Adapun batasannya sebagai berikut:

1. Penelitian ini berfokus pada penerapan *Attention-Guided Pruning* dan *Quantization* pada arsitektur MobileNetV3 untuk meningkatkan efisiensi komputasi serta mempertahankan akurasi dalam klasifikasi citra jamur *edible* dan *poisonous*.
2. Data yang digunakan dari sumber publik dan hanya mencakup dua kelas utama jamur, yaitu jamur *edible* dan *poisonous*. Dataset ini terdiri dari 25 spesies jamur *edible* dan 22 spesies jamur *poisonous*. Namun penelitian ini tidak melakukan klasifikasi berdasarkan spesies, melainkan hanya membedakan jamur ke dalam dua kelas utama tersebut.
3. Faktor lingkungan eksternal pada gambar, seperti pencahayaan ekstrem, bayangan tajam, atau latar belakang kompleks, tidak dianggap sebagai faktor yang mempengaruhi hasil klasifikasi dalam penelitian ini.
4. Untuk mengevaluasi performa model digunakan metrik evaluasi klasifikasi standar seperti akurasi, *precision*, *recall*, dan *confusion matrix*. Selain itu, efisiensi model juga dievaluasi melalui ukuran model dan waktu *inference* untuk menilai kinerja komputasi dan kesiapan implementasi pada perangkat terbatas.

5. Proses pelatihan dan pengujian model dilakukan pada lingkungan komputasi lokal, dengan asumsi bahwa konfigurasi perangkat keras dan perangkat lunak yang digunakan tidak memberikan pengaruh signifikan terhadap hasil penelitian.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini sebagai berikut:

1. Menganalisis kinerja model *baseline* MobileNetV3 dalam klasifikasi jamur *edible* dan *poisonous* berdasarkan akurasi, ukuran model, dan waktu *inference*.
2. Menganalisis pengaruh penerapan *Attention-Guided Pruning* dan *Quantization* pada arsitektur MobileNetV3 terhadap peningkatan efisiensi komputasi klasifikasi jamur *edible* dan *poisonous* dibandingkan dengan model *baseline*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini dapat dijelaskan sebagai berikut:

1. Secara Teoritis

Dari segi teoritis, penelitian ini membantu dapat menambah pemahaman tentang cara kecerdasan buatan beroperasi, khususnya dalam pengolahan gambar dengan *deep learning*. Penelitian ini menunjukkan bagaimana model dapat dibuat lebih ringan dan cepat tanpa mengorbankan akurasi melalui kombinasi *Attention-Guided Pruning* dan *Quantization* pada arsitektur MobileNetV3. Selain itu, penggunaan teknik ini menunjukkan bagaimana *attention mechanism* dapat membantu model fokus pada bagian penting dari gambar selama proses klasifikasi jamur *edible* dan *poisonous*. Diharapkan hasil penelitian ini akan berfungsi sebagai referensi untuk pengembangan model klasifikasi yang efektif dan akurat, terutama untuk digunakan di perangkat dengan kemampuan komputasi terbatas.

2. Secara Praktis

Dari segi praktis, penelitian ini diharapkan dapat memberikan solusi yang efektif dan akurat untuk mengklasifikasikan jamur *edible* dan *poisonous* dengan menggunakan *Attention-Guided Pruning* dan *Quantization* pada arsitektur *MobileNetV3*. Model yang dihasilkan menjadi lebih cepat dan ringan tanpa mengurangi akurasi, sehingga dapat digunakan pada perangkat seperti *smartphone*. Hal ini memudahkan masyarakat mengenali jenis jamur hanya melalui gambar dengan cepat dan hasil yang akurat. Sistem ini diharapkan dapat meningkatkan keamanan masyarakat dalam konsumsi jamur dan mengurangi risiko kesalahan dalam identifikasi jamur beracun.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN, Bab ini mencakup latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan. Pada bagian ini berisi alasan dilakukannya penelitian, penjelasan rumusan permasalahan yang dihadapi, serta tujuan akhir yang ingin dicapai melalui penelitian terkait.

BAB II TINJAUAN PUSTAKA, Bab ini berisi teori-teori pendukung yang digunakan dalam penelitian, seperti konsep dasar klasifikasi citra, prinsip kerja *Convolutional Neural Network (CNN)*, *MobileNetV3*, teknik *Attention Mechanism*, *Pruning*, dan *Quantization*. Selain itu, bagian ini juga menyajikan penelitian terdahulu yang relevan untuk menjadi acuan dalam pengembangan penelitian ini.

BAB III METODE PENELITIAN, Bab ini membahas metode yang digunakan dalam penelitian, meliputi tahapan pengumpulan dan pengolahan *dataset* gambar jamur, *preprocessing* data, perancangan model *MobileNetV3* dengan menerapkan *Attention-Guided Pruning*, pelatihan model, *Quantization*, dan evaluasi model. Pada bagian ini juga menjelaskan alat, bahan, serta perangkat keras dan perangkat lunak yang digunakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN, Bab ini menyajikan hasil implementasi model *Attention-Guided Pruning* dan *Quantization* pada *MobileNetV3* serta

analisis hasil yang diperoleh dari model. Analisis ini dilakukan untuk membandingkan hasil klasifikasi sebelum dan sesudah teknik *pruning* diterapkan guna melihat peningkatan efisiensi dan akurasi model.

BAB V PENUTUP, Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan penelitian selanjutnya untuk melakukan pengembangan lebih jauh pada model *deep learning* untuk klasifikasi jamur dan objek citra lainnya.

