

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*) mengalami kemajuan yang sangat signifikan dalam beberapa tahun terakhir, khususnya pada bidang *generative model* dalam pengolahan citra digital. Model generatif modern, seperti 4o Image Generation, mampu menghasilkan citra dengan tampilan fotorealistis melalui masukan berupa *prompt* teks [2, 3, 4]. Kemampuan ini menunjukkan perubahan signifikan dalam proses pembuatan konten visual, di mana citra tidak lagi harus diperoleh melalui pengambilan gambar secara langsung, melainkan dapat dihasilkan secara sintesis oleh model berbasis AI.

Pemanfaatan teknologi *image generator* berbasis AI saat ini telah meluas ke berbagai sektor, seperti industri kreatif, media sosial, pengembangan karakter digital, hingga produksi konten visual interaktif. Citra yang dihasilkan mampu menampilkan detail visual yang kompleks, seperti tekstur kulit, struktur wajah, serta pencahayaan yang realistis. Seiring dengan peningkatan kualitas model generatif, perbedaan antara citra wajah asli dan citra sintesis menjadi semakin sulit dikenali melalui pengamatan visual biasa [5]. Kondisi ini menimbulkan tantangan dalam menentukan autentisitas konten visual digital, serta berpotensi disalahgunakan untuk berbagai tindakan seperti pemalsuan identitas, penipuan digital, dan penyebaran informasi yang menyesatkan. Dengan demikian, diperlukan metode yang secara otomatis mampu membedakan citra wajah asli dan citra wajah *AI-generated*.

Pada tugas klasifikasi citra, *Convolutional Neural Network* (CNN) merupakan salah satu arsitektur yang umum diterapkan. Arsitektur ini dapat mempelajari representasi visual secara bertingkat melalui proses konvolusi, sehingga mampu mengenali pola tekstur dan struktur lokal pada citra [6]. Dalam konteks deteksi citra *AI-generated*, CNN telah terbukti mampu mengidentifikasi karakteristik distribusi visual tertentu yang sulit dikenali secara manual [7]. Akan tetapi, arsitektur CNN konvensional tetap memiliki sejumlah kendala, terutama dalam hal efisiensi parameter dan proses penskalaan arsitektur, di mana peningkatan performa sering kali diikuti dengan meningkatnya kompleksitas komputasi serta risiko *overfitting* [1, 8].

Untuk mengatasi keterbatasan tersebut, berbagai pengembangan arsitektur CNN telah diusulkan, seperti ResNet dan Xception. ResNet memperkenalkan konsep *residual learning* untuk mengurangi degradasi performa ketika jaringan dibuat semakin dalam [9], sedangkan Xception mengoptimalkan proses ekstraksi fitur melalui penggunaan *depthwise separable convolution* [10]. Meskipun demikian, peningkatan performa pada arsitektur tersebut umumnya diikuti dengan peningkatan jumlah parameter dan kompleksitas model yang cukup signifikan.

Sebagai solusi atas permasalahan tersebut, EfficientNet diperkenalkan dengan metode *compound scaling*, yaitu metode penskalaan jaringan yang dilakukan secara seimbang terhadap lebar, kedalaman, dan resolusi [1]. Melalui strategi tersebut, EfficientNet dilaporkan mampu menghasilkan kinerja klasifikasi citra yang lebih baik daripada beberapa arsitektur CNN terdahulu, seperti ResNet dan Xception, dengan penggunaan parameter dan beban komputasi yang lebih hemat [1]. Karakteristik tersebut menjadi dasar pemilihan EfficientNet dalam penelitian ini karena arsitektur ini menyediakan beberapa varian dengan peningkatan kapasitas secara bertahap, sehingga sesuai untuk digunakan dalam analisis perbandingan performa tanpa mengabaikan efisiensi parameter dan komputasi.

EfficientNet terdiri atas beberapa varian, mulai dari EfficientNet-B0 sebagai varian dasar hingga EfficientNet-B7 dengan kapasitas representasi yang lebih besar [1]. Dalam penelitian ini, analisis difokuskan pada EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, dan EfficientNet-B3 untuk merepresentasikan peningkatan kapasitas varian awal hingga menengah secara bertahap. Pada arsitektur asli EfficientNet, jumlah parameter total pada EfficientNet-B0 hingga EfficientNet-B3 meningkat dari sekitar 5,3 juta menjadi 12 juta, sedangkan EfficientNet-B4, EfficientNet-B5, EfficientNet-B6, dan EfficientNet-B7 masing-masing memiliki sekitar 19 juta, 30 juta, 43 juta, dan 66 juta parameter [1]. Pemilihan EfficientNet-B0 hingga EfficientNet-B3 mempertimbangkan dataset penelitian yang berjumlah 2.000 citra serta peningkatan kebutuhan komputasi pada varian yang lebih besar. Selain itu, model dengan kapasitas parameter yang semakin besar cenderung memiliki risiko *overfitting* yang lebih tinggi ketika dilatih pada jumlah sampel yang terbatas [8].

Varian EfficientNet-B0 hingga EfficientNet-B3 dipilih secara berurutan untuk membandingkan performa pada tingkat kapasitas yang berdekatan dalam kondisi eksperimen yang sama. Penggunaan varian dengan jarak skala yang lebih besar memang memberikan kontras kapasitas yang lebih lebar sehingga

perbedaan antartitik dapat terlihat lebih jelas, tetapi tidak memberikan informasi mengenai pola performa pada tingkat kapasitas di antaranya. Oleh karena itu, pemilihan varian berurutan ditujukan untuk mempertahankan keterincian analisis pada setiap peningkatan kapasitas model, bukan untuk memperbesar selisih performa. Sebagai konteks perbandingan antarvarian, penelitian asli EfficientNet melaporkan bahwa EfficientNet-B0 memperoleh akurasi 77,1%, EfficientNet-B1 sebesar 79,1%, EfficientNet-B2 sebesar 80,1%, dan EfficientNet-B3 sebesar 81,6% pada dataset ImageNet [1]. Keempat varian tersebut juga memiliki relevansi dengan tugas deteksi citra sintetis dan *deepfake* pada penelitian terdahulu [11, 12, 13].

Namun, pola peningkatan performa seiring bertambahnya skala arsitektur tidak selalu berlaku secara universal pada setiap dataset maupun tugas klasifikasi. Hal tersebut terlihat pada penelitian *Detection and Classification of Brain Tumor Using Hybrid Deep Learning Models*, yang menemukan bahwa EfficientNet-B0, B1, B2, dan B3 masing-masing memperoleh akurasi 97,43%, 98,66%, 99,06%, dan 98,08% pada klasifikasi citra MRI otak setelah augmentasi data [14], di mana varian dengan skala lebih besar tidak selalu menghasilkan performa tertinggi. Temuan tersebut menegaskan bahwa performa antarvarian dapat berbeda bergantung pada karakteristik dataset dan tugas yang digunakan. Dengan demikian, analisis terhadap masing-masing varian dalam penelitian ini menjadi penting, bukan hanya untuk mengetahui apakah EfficientNet secara umum dapat diterapkan pada tugas deteksi citra wajah *AI-generated*, tetapi juga untuk melihat bagaimana performa setiap varian pada konteks dataset dan tugas yang digunakan secara lebih terstruktur dan terkontrol.

Kebutuhan analisis antarvarian tersebut menjadi semakin relevan mengingat masih terbatasnya penelitian yang secara spesifik membandingkan varian EfficientNet dalam konteks deteksi wajah *AI-generated*. Sebagian besar penelitian deteksi citra sintetis masih berfokus pada citra hasil GAN atau *deepfake* berbasis video, sedangkan perbandingan EfficientNet-B0 hingga EfficientNet-B3 dalam mendeteksi wajah *AI-generated* berbasis *prompt* pada citra statis masih terbatas. Selain itu, peningkatan kapasitas model tidak selalu menghasilkan performa yang lebih tinggi pada setiap dataset [14]. Tanpa pengujian antarvarian pada dataset dan konfigurasi eksperimen yang sama, belum diperoleh dasar empiris untuk menentukan varian yang paling sesuai pada konteks penelitian ini. Oleh karena itu, analisis performa antarvarian diperlukan untuk memperoleh hasil perbandingan yang terukur serta mendukung pemilihan varian EfficientNet secara objektif.

Sejalan dengan hal tersebut, penelitian ini difokuskan pada analisis performa

varian EfficientNet dalam mendeteksi citra wajah *AI-generated*. Proses deteksi tersebut diterapkan melalui klasifikasi biner pada tingkat citra, dengan citra wajah asli sebagai kelas *real* dan citra wajah *AI-generated* sebagai kelas *fake*. Varian yang dianalisis meliputi EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, dan EfficientNet-B3. Analisis dilakukan pada citra statis yang telah diketahui mengandung wajah manusia sebagai objek utama. Citra kelas *fake* merupakan citra wajah *AI-generated* berbasis *prompt*. Penelitian ini mengeksplorasi kemampuan masing-masing varian EfficientNet dalam mengolah representasi fitur visual pada citra wajah sintesis. Dalam penelitian ini, data dibatasi pada citra wajah dengan karakteristik Asia dalam rentang usia 18 sampai dengan 60 tahun untuk menjaga konsistensi dan mengurangi bias distribusi data. Hasil evaluasi digunakan untuk mengetahui performa masing-masing varian EfficientNet dalam membedakan citra wajah asli dan citra wajah *AI-generated* berdasarkan *confusion matrix* serta metrik turunannya, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, serta sebagai referensi dalam pengembangan metode identifikasi konten visual berbasis kecerdasan buatan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini disusun sebagai berikut:

1. Bagaimana membangun model berbasis EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, dan EfficientNet-B3 yang mampu mendeteksi wajah *AI-generated*?
2. Bagaimana performa masing-masing model EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, dan EfficientNet-B3 dalam mendeteksi wajah *AI-generated* berdasarkan hasil evaluasi *confusion matrix*?

1.3 Batasan Permasalahan

Agar penelitian yang dilakukan memiliki fokus yang jelas dan terarah, maka batasan masalah dalam penelitian ini ditetapkan sebagai berikut:

1. Varian model yang dianalisis dalam penelitian ini dibatasi pada EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, dan EfficientNet-B3.

2. Data penelitian difokuskan pada citra statis berupa foto wajah asli dan wajah hasil generasi AI (*AI-generated faces*), tanpa mencakup citra tanpa wajah, data berbentuk video, maupun manipulasi wajah nyata atau *deepfake*.
3. Citra wajah yang dianalisis dibatasi pada wajah dengan karakteristik Asia dengan rentang usia 18 sampai dengan 60 tahun.
4. Citra wajah *AI-generated* yang digunakan merupakan citra yang dihasilkan melalui fitur *image generator* pada ChatGPT menggunakan masukan berupa *prompt* teks.
5. Penelitian ini tidak membahas implementasi sistem dalam lingkungan produksi atau aplikasi nyata, melainkan difokuskan pada analisis performa model dalam lingkup penelitian akademik.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Membangun model berbasis EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, dan EfficientNet-B3 yang mampu mendeteksi citra wajah *AI-generated*.
2. Menganalisis performa masing-masing model EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, dan EfficientNet-B3 dalam mendeteksi citra wajah *AI-generated* berdasarkan *confusion matrix* dan metrik turunannya seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Berkontribusi pada pengembangan kajian pengolahan citra digital, khususnya terkait deteksi citra wajah *AI-generated* menggunakan pendekatan EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, dan EfficientNet-B3.
2. Menjadi referensi akademik mengenai analisis performa EfficientNet-B0, EfficientNet-B1, EfficientNet-B2, dan EfficientNet-B3 dalam mendeteksi citra wajah *AI-generated* berdasarkan metrik evaluasi klasifikasi.

1.6 Sistematika Penulisan

Bagian ini berisi uraian singkat mengenai struktur isi laporan penelitian, mulai dari Bab I hingga Bab V. Sistematika penulisan skripsi ini disusun sebagai berikut:

- **Bab I PENDAHULUAN**

Bab ini mencakup latar belakang masalah, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

- **Bab II LANDASAN TEORI**

Bab ini membahas teori yang menjadi dasar penelitian, meliputi konsep wajah *AI-generated*, model generatif 4o Image Generation, arsitektur *Convolutional Neural Network* (CNN), EfficientNet sebagai model yang digunakan, pembagian data (*data splitting*), serta metrik evaluasi berupa *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*.

- **Bab III METODOLOGI PENELITIAN**

Bab ini menguraikan metode penelitian yang digunakan, termasuk desain penelitian, variabel penelitian, proses pengumpulan dataset, tahapan pra-pemrosesan data, implementasi model EfficientNet, skenario pembagian data, dan prosedur evaluasi performa model.

- **Bab IV HASIL DAN PEMBAHASAN**

Bab ini memaparkan hasil implementasi dan analisis performa varian EfficientNet dalam mendeteksi citra wajah *AI-generated*. Pembahasan mencakup spesifikasi sistem, proses pelatihan, hasil pengujian berdasarkan skenario pembagian data, analisis metrik evaluasi, dan interpretasi terhadap hasil eksperimen.

- **Bab V KESIMPULAN DAN SARAN**

Bab ini memuat simpulan berdasarkan hasil penelitian serta saran yang dapat digunakan sebagai arahan untuk penelitian berikutnya.