

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi kecerdasan artifisial (*Artificial Intelligence/AI*) telah menghasilkan berbagai inovasi, salah satunya dalam bidang pengolahan gambar. Model generatif seperti *Deepfake*, *Generative Adversarial Network* (GAN), dan alat berbasis *diffusion* telah memungkinkan penciptaan gambar yang tampak nyata, padahal tidak pernah benar-benar ada. Kemampuan ini, meskipun bermanfaat di beberapa bidang, juga menimbulkan ancaman serius terhadap keaslian informasi visual. *AI Generatif* adalah kemampuan untuk menghasilkan data baru dan orisinal. *AI Generatif* merupakan puncak inovasi dalam dunia kecerdasan artifisial sebagai kekuatan kreatif dalam menciptakan dunia baru diluar data pelatihan yang telah ada[1][2].

Salah satu dampak negatif yang mencolok adalah meningkatnya kasus penipuan yang melibatkan gambar wajah hasil AI. Gambar-gambar ini sering digunakan untuk menipu publik, menyebarkan disinformasi, atau menciptakan identitas palsu yang sulit dibedakan dari gambar asli. Terdapat 100 orang dari 20 provinsi menjadi korban dengan total kerugian mencapai Rp 65 juta yang disebabkan oleh penipuan gambar palsu dengan menggunakan wajah Presiden Prabowo dan Wakil Presiden Gibran Rakabuming Raka serta Menteri Keuangan Sri Mulyani. Tingginya akurasi visual gambar palsu membuat masyarakat umum kesulitan untuk mendeteksi keasliannya secara kasat mata[3][4].

Teknologi AI dalam pembuatan gambar telah disalah gunakan untuk menghasilkan konten pornografi manipulatif, seperti mengganti wajah seseorang ke dalam gambar tubuh telanjang. Kasus Taylor Swift yang menjadi korban dampak negatif AI merupakan contoh nyata dari bahaya yang terjadi dalam teknologi ini. Konten semacam ini tidak hanya melanggar etika dan privasi, tetapi juga dapat merusak reputasi dan psikologis korban. Oleh karena itu, diperlukan pendekatan sistematis untuk mengidentifikasi dan mengklasifikasi gambar wajah buatan AI[3][5].

Penelitian ini bertujuan untuk menganalisis performa pada arsitektur ResNet50 dalam klasifikasi untuk membedakan antara citra wajah asli dan citra wajah yang dihasilkan oleh *AI Generatif*. ResNet50 merupakan salah satu arsitektur model CNN yang dikembangkan untuk mengatasi *vanishing gradient* pada jaringan saraf konvolusional yang mendalam. *Vanishing gradient* adalah masalah dalam pelatihan *deep neural network* di mana gradien menjadi kecil merambat mundur melalui lapisan-lapisan jaringan. Penelitian ini akan mengukur efektivitas ResNet50 menggunakan metrik evaluasi standar yang berfungsi sebagai alat objektif untuk menguji performa model terhadap data uji. Parameter metrik evaluasi yang digunakan sebagai indikator yaitu *accuracy*, *precision*, *recall*, *F1-score* sebagai tolak ukur kuantitatif yang dapat berkontribusi pada sistem pengembangan deteksi dan klasifikasi gambar palsu yang dihasilkan oleh AI[6][7].

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana hasil metrik performa model ResNet50, yang meliputi *accuracy*, *precision*, *recall*, dan *F1-score*, dalam klasifikasi untuk deteksi wajah *AI Generatif*?
2. Bagaimana pengujian model ResNet50 untuk membedakan antara citra wajah manusia asli dan citra wajah yang dihasilkan oleh *AI*?

1.3 Batasan Masalah

Untuk memfokuskan penelitian, maka batasan masalah ditentukan sebagai berikut:

1. Penelitian hanya difokuskan pada gambar palsu yang dihasilkan oleh model *AI-generated image*, seperti *Deepfake* dan *Generative Adversarial Network (GAN)*, tanpa mencakup video atau konten bergerak.
2. Metode yang digunakan untuk deteksi dan klasifikasi adalah algoritma *Convolutional Neural Network (CNN)* tanpa perbandingan langsung dengan metode lain seperti SVM, Random Forest, atau arsitektur *deep*

learning lainnya.

3. Dataset yang digunakan merupakan dataset gambar yang bersifat publik dan open-source, tidak mencakup data dari sumber pribadi atau data dengan pembatasan akses.
4. Evaluasi kinerja model hanya difokuskan pada metrik *accuracy*, *precision*, *recall*, dan *F1-score*, sedangkan aspek keamanan sistem dan waktu komputasi tidak dijadikan fokus utama.
5. Penelitian dilakukan secara terbatas dalam lingkungan perangkat lunak (*software*) tanpa menguji implementasi sistem dalam bentuk aplikasi atau perangkat keras (*hardware*).

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengembangkan model CNN dengan arsitektur ResNet50 untuk mendeteksi dan mengklasifikasikan antara wajah asli dan buatan AI.
2. Menganalisis dan mengevaluasi performa model ResNet50 yang telah dilatih menggunakan *confusion matrix* beserta metrik turunannya, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Manfaat Teoritis

Menambah kajian ilmiah dalam bidang pengolahan citra digital dan kecerdasan artifisial, khususnya penerapan algoritma *Convolutional Neural Network* (CNN) dalam mendeteksi dan mengklasifikasikan gambar wajah hasil *AI-generated*. Hasil penelitian ini dapat menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan pendeteksian konten visual manipulatif.

2. Manfaat Praktis

Memberikan solusi teknologi yang dapat dimanfaatkan oleh institusi, organisasi media, platform digital, dan masyarakat dalam mengidentifikasi gambar wajah secara otomatis. Aplikasi dari penelitian ini berpotensi

membantu mencegah penyebaran informasi palsu dan penyalahgunaan gambar yang merugikan secara sosial dan hukum.

1.6 Sistematika Penulisan

Adapun sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN: berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA: berisi kajian literatur terdahulu dan dasar teori yang mendukung penelitian.

BAB III METODOLOGI PENELITIAN: berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat penelitian, dan sistematika penulisan.

BAB IV HASIL DAN PEMBAHASAN: berisi hasil implementasi sistem, pengujian yang dilakukan, serta analisis terhadap hasil yang diperoleh berdasarkan tujuan penelitian dan referensi yang relevan.

BAB V PENUTUP: berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan atau penelitian lanjutan.