BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi kecerdasan buatan (*Artificial Intelligence*, *AI*) khususnya dalam bidang pembelajaran mendalam (*Deep Learning*), telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk dalam manipulasi citra digital. Salah satu perkembangan yang menonjol adalah teknologi *deepfake*, yang memungkinkan pembuatan citra wajah manusia yang dimanipulasi dengan sangat realistis menggunakan model berbasis *Generative Adversarial Networks* (GANs) dan *Convolutional Neural Networks* (CNNs). Teknologi ini telah digunakan untuk berbagai tujuan, mulai dari hiburan, efek visual dalam industri film, hingga pembuatan konten edukatif. Namun, penyalahgunaan *deepfake* dalam berbagai disinformasi menimbulkan kekhawatiran yang semakin besar terhadap dampaknya bagi integritas informasi

Deteksi *deepfake* menjadi tantangan utama dalam keamanan siber dan forensik digital, mengingat metode manipulasi wajah semakin canggih dan sulit dibedakan dari citra asli oleh manusia. Oleh karena itu, pengembangan sistem deteksi berbasis kecerdasan buatan menjadi kebutuhan yang mendesak untuk mengidentifikasi citra *deepfake* dengan akurasi yang cuku tinggi dan efisiensi yang baik. Berbagai pendekatan telah dikembangkan, salah satunya adalah pemanfaatan

Convolutional Neural Networks (CNN) sebagai metode utama dalam analisis pola dan fitur citra digital.

Dengan arsitekturnya yang dirancang khusus untuk menganalisis data visual, *Convolutional Neural Networks* (CNN) adalah salah satu pendekatan utama dalam mendeteksi dan mengklasifikasikan citra. Banyak metode yang dapat memecahkan permasalahan ini, tetapi ada satu yaitu model *Convolutional Neural Network* atau CNN yang berulang kali digunakan dan cukup terkenal. CNN memiliki tingkat akurasi yang cukup tinggi dan telah terbukti efektif dalam analisis citra, tetapi efektivitasnya dalam mendeteksi manipulasi AI dalam konteks keamanan siber masih membutuhkan penelitian lebih lanjut.

Dengan meningkatnya ancaman yang ditimbulkan teknologi *deepfake* dalam berbagai bidang dan aspek, penelitian berharap dapat berkontribusi dalam segi informasi dalam mendukung pengembangan sistem deteksi yang lebih akurat dan efisien. Penelitian berharap memberikan dasar bagi studi lebih lanjut dalam eksplorasi model CNN yang lebih optimal dalam mendeteksi manipulasi wajah berbasis kecerdasan buatan.

1.2 Rumusan Masalah

Dengan mempertimbangkan masalah-masalah yang telah diidentifikasi, perumusan masalah berikut dapat dibuat:

1. Bagaimana performa MesoNet dalam mendeteksi *facial anomalies* pada citra deepfake?

2. Bagaimana performa ResNet dalam mendeteksi *facial anomalies* pada citra deepfake?
3. Arsitektur Convolutional Neural Networks mana yang paling efektif dalam mendeteksi *facial anomalies* manipulasi citra, MesoNet atau ResNet?

1.3 Tujuan Penelitian

Studi ini bertujuan untuk:

1. Mengevaluasi kemampuan MesoNet dalam mendeteksi *facial anomalies* dalam citra *deepfake.*
2. Mengevaluasi kemampuan ResNet dalam mendeteksi *facial anomalies* dalam citra *deepfake.*
3. Membandingkan kemampuan ResNet dan MesoNet dalam mendeteksi *facial anomalies* dalam citra *deepfake*, menggunakan metrik evaluasi yang paling tepat untuk menentukan model CNN mana yang paling efisien dalam mendeteksi *facial anomalies* dalam citra *deepfake*.

1.4 Batasan Penelitian

1. Dataset publik yang tersedia di Kaggle berisi sejumlah citra *deepfake* dengan *facial anomalies* yang telah diproses sebelumnya, dan citra asli. Ada juga dataset tambahan yang tersedia, yang terdiri dari citra asli yang dikumpulkan sendiri dan citra *deepfake* yang dikumpulkan sendiri dan citra *deepfake* yang dimanipulasi oleh *Grok AI* menggunakan citra asli yang dikumpul sendiri.

2. Dataset yang digunakan hanya akan terdiri dari citra yang berasal dari dataset publik yang terdiri dari citra asli dan citra yang telah dimanipulasi oleh kecerdasan buatan. Dataset lain yang akan digunakan adalah dataset yang dikumpul sendiri, dan dimanipulasi
3. Model yang dibandingkan dilatih dan diuji dengan menggunakan *hyperparameter default* yang disesuaikan secara minimal untuk memastikan keadilan perbandingan, dan menggunakan arsitektur standar MesoNet dan ResNet.
4. Eksperimen ini dilakukan dengan perangkat lunak dan perangkat keras yang terbatas, seperti *central processing unit* (CPU) kelas menengah dan framework pembelajaran mendalam *Torchvision*. *Jupyter Notebook* adalah platform utama untuk menerapkan model dan menganalisis hasil dari data.
5. Eksperimen komparatif antara model MesoNet dan model ResNet akan dilakukan di perangkat laptop yang memiliki *Intel CPU I7 Gen 8th, Random Access Memory* (RAM) 16GB, dan *Graphic Processing Unit* (GPU) *Nvidia GTX 1060 6GB*.
6. Demi menghasilkan metrik performa yang andil terhadap kedua model akan diukur berdasarkan metrik yang dipilih yaitu, akurasi, presisi, recall, F1-score, dan waktu pemrosesan. Keputusan metrik ini diambil untuk memastikan tidak ada hasil berpihak ke satu sisi pada kedua model dan membuatnya cukup sederhana sehingga mudah untuk dibandingkan.

1.5 Manfaat Penelitian

Diharapkan bahwa penelitian ini akan bermanfaat dalam:

1. Memberikan informasi yang dapat diandalkan untuk menemukan bukti digital yang telah dimanipulasi oleh AI.
2. Memberikan informasi lebih dalam bagaimana performa kedua model arsitektur CNN yang terpilih, dalam melakukan tugas secara *real-time* dan skenario spesifik di sebuah perangkat/*hardware* kelas menengah.
3. Meningkatkan pemahaman tentang kekuatan dan keterbatasan arsitektur CNN, seperti model ResNet dan model MesoNet dalam tugas mendeteksi visual.