

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Dalam era digital yang berkembang pesat, aplikasi kecerdasan buatan telah menjadi elemen kunci dalam berbagai aspek kehidupan manusia. Salah satu subdisiplin Artificial Intelligence yang muncul dengan sangat signifikan adalah Deep Learning (Pembelajaran Mendalam), yang telah mengubah cara kita memproses data kompleks dan mengambil keputusan. Metode ini memungkinkan sistem komputer untuk memahami serta menggunakan informasi yang terkandung dalam data yang umumnya besar dan kompleks (M. Buhmester and F. Yang, "Deep Learning and Its Impact on Data Processing and Decision Making in the Digital Age," *Journal of Artificial Intelligence Research*, 2021.) [1], sesuatu yang biasanya sulit diolah dengan metode konvensional.

Pada perkembangan teknologi saat ini, Artificial Intelligence (AI) berkembang dengan sangat pesat. AI banyak diterapkan dalam teknologi pintar untuk menghasilkan suatu sistem otomatisasi. [2] Selain itu, AI juga dapat diterapkan dalam teknologi yang melakukan pengolahan citra dengan tingkat yang sangat memuaskan. Akan tetapi, tidak semua teknologi yang menggunakan AI berdampak positif. AI juga dapat menimbulkan dampak negatif. Kecerdasan buatan generatif (Generative AI atau GenAI) adalah bentuk AI yang dapat secara instan menghasilkan konten baru, seperti teks, gambar, audio, dan video [3]. Dengan model tersebut, Generative AI telah mengubah cara mencari dan mengakses informasi secara daring. Hanya dengan mengungkapkan pertanyaan ataupun prompt, penggunaanya dapat langsung memperoleh jawaban atau respon yang telah terstruktur dan lebih efektif. Sehingga tidak perlu lagi memilah dan merangkul informasi secara mandiri. Hal tersebut memberikan pengalaman pencarian yang lebih dinamis dan interaktif bagi para penggunaanya [3]. Salah satu teknologi AI yang dapat menimbulkan dampak negatif adalah deepfake. Deepfake merupakan istilah yang digunakan untuk memanipulasi citra maupun suara seseorang dari gambar atau video. Banyak dampak negatif yang dapat ditimbulkan dari

penggunaan deepfake ini, seperti menyebarkan informasi palsu, penyalahgunaan identitas, dan pornografi [4]. Banyaknya dampak negatif yang disebabkan oleh deepfake tersebut tentunya membuat kekhawatiran di masyarakat. Terlebih, deepfake sangat berbahaya apabila diterapkan pada gambar manusia. Teknologi deepfake pada saat ini sudah dapat membuat gambar manusia yang sulit dibedakan dengan aslinya. Oleh karena itu, diperlukan sebuah model yang dapat mengidentifikasi gambar asli dan gambar buatan deepfake.

Penelitian yang dilakukan oleh Chatterjee et al. (2022), yang menggunakan kombinasi data asli dan sintetis untuk meningkatkan akurasi klasifikasi citra. Dalam studi ini, mereka menggunakan data gambar botol plastik sintetis yang dihasilkan oleh *Generative Adversarial Networks* (GAN) untuk melatih model klasifikasi berbasis *transfer learning*. Hasilnya menunjukkan peningkatan akurasi klasifikasi hingga 99,06% dibandingkan dengan penggunaan data asli saja, membuktikan bahwa data sintetis dapat berperan penting dalam memperluas variasi data pelatihan dan meningkatkan performa model (Chatterjee et al., 2022) [5]. Temuan ini memperkuat relevansi penelitian tentang klasifikasi gambar *FAKE* dan *REAL* menggunakan model berbasis *EfficientNet*, yang juga berfokus pada pengelolaan data sintetis untuk meningkatkan akurasi model klasifikasi.

Tujuan penelitian ini adalah untuk merancang, membangun, dan mengevaluasi model klasifikasi citra *Generative Artificial Intelligence* menggunakan arsitektur *EfficientNet*. Model ini bertujuan untuk membedakan gambar *REAL* (gambar asli) dan *FAKE* (gambar hasil generatif) secara akurat dengan memanfaatkan transfer learning dari model pra-latih *EfficientNet-B0* yang telah teruji pada dataset *ImageNet*. Selain itu, penelitian ini juga bertujuan untuk mengukur kinerja model dengan berbagai metrik evaluasi, seperti akurasi, presisi, *recall*, dan *F1-score*, serta menganalisis hasil klasifikasi menggunakan metode visualisasi *Grad-CAM* untuk memahami area fokus model dalam melakukan klasifikasi. Dengan demikian, diharapkan penelitian ini dapat menghasilkan model yang efisien, akurat, dan mudah diterapkan untuk deteksi gambar sintetis dalam berbagai aplikasi, termasuk keamanan digital dan validasi konten media.

## 1.2 Rumusan Masalah

1. Bagaimana performa algoritma EfficientNet dalam mengklasifikasi citra hasil Generative Artificial Intelligence (FAKE) dan citra asli (REAL)?
2. Seberapa besar tingkat akurasi model klasifikasi gambar FAKE dan REAL yang diperoleh dari training model menggunakan teknik augmentasi, oversampling, dan evaluasi berbasis Grad-CAM?
3. Sejauh mana metode visualisasi Grad-CAM mampu menjelaskan keputusan klasifikasi pada model EfficientNet terhadap citra FAKE dan REAL?
4. Bagaimana sebelum dan sesudah oversampling?

## 1.3 Batasan Masalah

Agar penelitian lebih terfokus dan terarah, maka dilakukan pembatasan masalah sebagai berikut:

1. Penelitian hanya difokuskan pada klasifikasi citra hasil dari Generative Artificial Intelligence (Gambar FAKE) dan citra asli (Gambar REAL).
2. Model yang digunakan dalam penelitian ini terbatas pada arsitektur EfficientNet-B0 sampai EfficientNet-B2 karena keterbatasan sumber daya, maka penelitian ini tidak membandingkan performanya dengan arsitektur lain seperti ResNet, VGG, atau DenseNet.
3. Teknik visualisasi interpretabilitas model dibatasi hanya pada metode Grad-CAM untuk membantu menjelaskan hasil klasifikasi.
4. Dataset yang digunakan adalah dataset khusus yang berisi gambar hasil generatif dan gambar nyata, yang diperoleh secara manual dan dimasukkan ke Google Colab oleh peneliti, tanpa proses pengumpulan data secara otomatis.
5. Eksperimen dilakukan dengan asumsi lingkungan training yang stabil (hardware dan software), serta tidak mempertimbangkan faktor seperti waktu komputasi, konsumsi daya, dan efisiensi resource sistem secara mendetail.

#### **1.4 Tujuan Penelitian**

Tujuan penelitian ini adalah sebagai berikut :

1. Menganalisis performa model EfficientNet dalam mengklasifikasikan citra yang dihasilkan oleh Generative Artificial Intelligence (gambar FAKE) dan citra asli (gambar REAL).
2. Mengukur akurasi, presisi, recall, dan F1-score sebagai parameter evaluasi utama untuk mengidentifikasi efektivitas model dalam melakukan klasifikasi citra.
3. Memvalidasi penerapan teknik Grad-CAM untuk memberikan interpretabilitas terhadap keputusan yang diambil oleh model dalam klasifikasi gambar.
4. Membandingkan hasil klasifikasi menggunakan model EfficientNet dengan standar evaluasi yang ada, guna mengetahui kelebihan dan kekurangan dari model tersebut dalam konteks klasifikasi citra generatif.
5. Memberikan rekomendasi untuk pengembangan lebih lanjut dalam aplikasi klasifikasi citra generatif menggunakan algoritma berbasis deep learning.

#### **1.5 Manfaat Penelitian**

Manfaat dalam penelitian ini diharapkan untuk mencapai hasil yang baik dan benar, antara lain :

1. Penelitian ini memberikan kontribusi dalam meningkatkan akurasi sistem deteksi citra generatif dengan menggunakan model EfficientNet, yang dikenal dengan kemampuannya dalam mengklasifikasikan gambar dengan efisiensi tinggi. Dengan menggunakan pendekatan ini, penelitian ini memungkinkan sistem untuk lebih efektif membedakan antara citra asli (REAL) dan citra palsu (FAKE) yang dihasilkan oleh algoritma generative artificial intelligence (AI). Penerapan model ini dapat meningkatkan performa deteksi dalam berbagai aplikasi praktis, seperti pemfilteran konten digital, identifikasi gambar palsu, dan pengawasan keamanan digital.
2. Penelitian ini memberikan wawasan baru dalam mendeteksi citra palsu yang

dibuat oleh algoritma generative seperti GAN (Generative Adversarial Networks) atau model AI lainnya. Dengan kemajuan teknologi AI yang semakin canggih dalam menciptakan citra yang sangat mirip dengan gambar asli, deteksi citra palsu menjadi tantangan besar dalam keamanan digital, terutama dalam bidang forensik digital, media sosial, dan pengenalan wajah. Penelitian ini berpotensi memperkuat sistem keamanan dan mencegah penyalahgunaan teknologi AI dalam penyebaran informasi palsu.

3. Salah satu manfaat signifikan dari penelitian ini adalah penerapan Grad-CAM (Gradient-weighted Class Activation Mapping), yang memungkinkan visualisasi area gambar yang paling berkontribusi terhadap keputusan model. Ini memberikan interpretabilitas yang lebih baik terhadap keputusan yang diambil oleh model EfficientNet dalam mengklasifikasikan citra. Dengan pemahaman yang lebih baik tentang bagaimana model bekerja, para peneliti dan pengembang dapat meningkatkan akurasi serta mengurangi bias dalam model AI, sehingga menjadikan sistem ini lebih transparan dan dapat diandalkan dalam pengambilan keputusan otomatis.

## **1.6 Sistematika Penulisan**

### **1. BAB 1 PENDAHULUAN**

Pada bab ini menjelaskan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan. Bab ini memberikan gambaran umum mengenai alasan dan urgensi penelitian dilakukan, serta ruang lingkup yang dibatasi agar penelitian tetap fokus dan terarah.

### **2. BAB 2 TINJAUAN PUSTAKA**

Pada bab ini meneliti kajian literatur yang relevan dengan topik penelitian, meliputi teori-teori dasar terkait klasifikasi citra, deep learning, transfer learning, arsitektur EfficientNet, serta metode interpretabilitas Grad-CAM. Bab ini juga menguraikan hasil-hasil penelitian terdahulu yang menjadi

landasan atau pembandingan dalam studi ini.

### **3. BAB 3 METODE PENELITIAN**

Pada bab ini menjelaskan langkah-langkah penelitian yang dilakukan, mulai dari objek penelitian, tahapan pengumpulan data, preprocessing dan augmentasi citra, proses oversampling, pelatihan model, evaluasi kinerja model, hingga visualisasi Grad-CAM. Selain itu, bab ini juga mencantumkan alat dan bahan yang digunakan dalam pelaksanaan penelitian.

### **4. BAB 4 HASIL DAN PEMBAHASAN**

Dalam bab ini, menyajikan hasil dari proses pelatihan dan evaluasi model klasifikasi citra menggunakan EfficientNet. Bab ini mencakup hasil preprocessing data, visualisasi augmentasi, performa akurasi model, evaluasi menggunakan confusion matrix dan classification report, serta interpretasi hasil klasifikasi menggunakan visualisasi Grad-CAM.

### **5. BAB 5 PENUTUP**

Pada bab ini, berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan lebih lanjut. Kesimpulan merangkum temuan utama penelitian, sedangkan saran ditujukan untuk perbaikan dan perluasan cakupan penelitian pada masa yang akan datang.