

Penerapan Convolutional Neural Network untuk Prototipe Sistem Deteksi Foto Sintetis Generative Artificial Intelligence dan Foto Asli

Yuliana Permata Sari ^{1*}, Agung Riyadi ^{2**},

* Teknik Informatika, Politeknik Negeri Batam

** Teknik Informatika, Politeknik Negeri Batam

permatasari27y@gmail.com ¹, agung@polibatam.ac.id ²,

Article Info

Article history:

Received ...

Revised ...

Accepted ...

Keyword:

CNN, generative artificial intelligence, image classification, deep learning

ABSTRACT

The advancement of Generative Artificial Intelligence (GAI) has enabled the creation of synthetic images that closely resemble real photographs, raising challenges in verifying digital content authenticity. This study aims to develop a classification model based on Convolutional Neural Network (CNN) to distinguish between synthetic and real images accurately, as well as to implement it into a prototype detection system. The dataset consists of AI-generated and real images processed through preprocessing, training, and testing stages. Model performance is evaluated using accuracy, precision, recall, and F1-score. The results indicate that CNN effectively extracts visual features and achieves good classification performance. The model is then integrated into a web-based prototype to support real-time detection. The implementation results show that the system performs well; however, it still has limitations in handling unseen data outside the training distribution.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Perkembangan teknologi Artificial Intelligence (AI) telah membawa perubahan signifikan dalam berbagai bidang, khususnya pada pengolahan citra digital. Salah satu cabang yang berkembang pesat adalah Generative Artificial Intelligence (GAI), yaitu teknologi yang mampu menghasilkan citra sintetis dengan tingkat kemiripan yang sangat tinggi terhadap foto asli. Kemajuan ini didukung oleh model generatif seperti Generative Adversarial Network (GAN) dan diffusion model yang mampu menciptakan konten visual realistis dalam berbagai domain, seperti desain, media, dan industri kreatif [1].

Di sisi lain, perkembangan tersebut juga menimbulkan permasalahan baru terkait keaslian informasi visual. Citra hasil generative AI semakin sulit dibedakan dari citra asli yang dihasilkan oleh kamera, sehingga berpotensi disalahgunakan untuk manipulasi informasi, penyebaran disinformasi, serta pelanggaran hak cipta [2]. Oleh sebab itu, penting dilakukan penelitian yang mampu membedakan gambar hasil generatif dan gambar asli.

Berbagai penelitian sebelumnya telah mengkaji deteksi citra sintetis menggunakan pendekatan deep learning. Model berbasis Convolutional Neural Network (CNN) terbukti efektif dalam mengekstraksi fitur spasial dari citra dan mencapai performa tinggi dalam berbagai tugas klasifikasi, termasuk deteksi deepfake dan citra sintetis. Penelitian oleh [3] mengembangkan pendekatan self-supervised learning menggunakan 3D CNN untuk klasifikasi penyakit Alzheimer pada dataset citra otak sintetis berskala besar. Hasil penelitian menunjukkan bahwa model CNN mampu mempelajari representasi fitur yang baik dari citra sintetis dan mempertahankan performa yang sebanding dengan data nyata. Selain itu, penelitian [4] mengusulkan pendekatan berbasis multi-model ensemble untuk mendeteksi face deepfake hasil GAN dengan mengintegrasikan beberapa arsitektur seperti ResNet50V2, DenseNet201, dan Gram-Net melalui teknik transfer learning. Pendekatan ini terbukti mampu meningkatkan akurasi deteksi hingga di atas 97% serta mengurangi ketergantungan pada metode forensik manual. Penelitian [5] juga mengembangkan metode forensik berbasis deep learning yang menggabungkan fitur spasial dan temporal menggunakan arsitektur hybrid seperti CNN-LSTM

dan CNN-GRU untuk meningkatkan akurasi deteksi deepfake. Pendekatan ini menunjukkan performa tinggi dengan akurasi mendekati 99% serta mampu menangani berbagai variasi manipulasi citra dan video. Namun demikian, sebagian besar penelitian masih berfokus pada domain tertentu, seperti wajah manusia atau citra medis, sehingga belum sepenuhnya mencakup variasi citra umum yang lebih luas.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pembangunan model klasifikasi berbasis CNN untuk membedakan foto sintetis hasil generative AI dan foto asli secara lebih umum. Selain itu, penelitian ini tidak hanya berfokus pada pengembangan model, tetapi juga pada implementasi model ke dalam sebuah prototipe sistem berbasis web yang memungkinkan proses deteksi dilakukan secara langsung oleh pengguna. Pendekatan ini diharapkan dapat meningkatkan nilai praktis dari penelitian serta mendukung penerapan teknologi dalam kehidupan nyata.

Kontribusi utama dari penelitian ini adalah pengembangan model CNN untuk klasifikasi citra sintetis dan citra asli, serta implementasi model dalam bentuk prototipe sistem deteksi berbasis web. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam bidang computer vision serta menjadi dasar pengembangan sistem verifikasi keaslian citra digital yang andal.

II. LANDASAN TEORI

A. Tinjauan Pustaka

Penelitian mengenai deteksi citra hasil Generative Artificial Intelligence telah berkembang pesat seiring meningkatnya kualitas citra sintetis. Berbagai pendekatan berbasis deep learning, khususnya Convolutional Neural Network (CNN), telah banyak digunakan untuk membedakan citra asli dan citra sintetis. Penelitian oleh [6] melakukan analisis komparatif terhadap berbagai arsitektur CNN seperti VGG16, ResNet50, VGG19, dan MobileNetV2 dalam mendeteksi citra deepfake dari berbagai model generatif. Hasil penelitian menunjukkan bahwa model CNN memiliki kemampuan yang baik dalam mengenali pola manipulasi, namun performanya dapat menurun ketika dihadapkan pada jenis data generatif yang berbeda, sehingga menunjukkan adanya tantangan dalam generalisasi model. Selanjutnya, penelitian oleh [7] membandingkan performa CNN dengan pendekatan berbasis Vision Transformer dalam klasifikasi citra sintetis dan citra asli. Hasil penelitian menunjukkan bahwa transformer memiliki kemampuan lebih baik dalam menangkap hubungan global antar fitur citra, sementara CNN cenderung fokus pada fitur local. Selain itu, studi oleh [8] mengkaji berbagai metode deteksi citra hasil AI dan deepfake secara komprehensif, serta menekankan bahwa tantangan utama terletak pada kemampuan model dalam menghadapi variasi data dan teknik generatif yang terus berkembang. Penelitian lain oleh [9] mengusulkan pendekatan CNN ringan (lightweight CNN) yang mampu mencapai performa tinggi dengan kompleksitas yang lebih rendah, serta menunjukkan

perbedaan pola aktivasi antara citra asli dan sintetis. Selain itu, [10] mengembangkan model klasifikasi multi-kelas berbasis CNN untuk membedakan citra AI-generated, deepfake, dan citra asli, yang menunjukkan efektivitas CNN dalam tugas klasifikasi citra modern. Berikut detail penelitian terkait :

Tabel 1 Penelitian Terdahulu

No	Judul	Peneliti	Algoritma
1	Comparative Analysis of Deepfake Detection Models on Diverse GAN-Generated Images	Medha Wyawahare, Vrinda Parkhi, Siddharth Bhorge, Mayank Jha, Milind Rane, Narendra Muhal	CNN (VGG16, ResNet50, VGG19, MobileNet V2)
2	Synthetic vs Real Image Detection using Vision Transformers and CNN-based Architectures	Noor ul Huda, Ayesha Fayyaz, Uswa Asad, Yasir Saleem Afridi	CNN dan Vision Transformer (ViT)
3	Unmasking AI-created visual content: a review of generated images and	Yupeng Zhang, Zongwei Pang, Shanyuan	Deep Learning (CNN, Frequency-based,

	deepfake detection technologies	Huang, Chengyou Wang, Xiao Zhou	Multimodal Approaches)
4	Detection of AI- Generated Synthetic Images with a Lightweight CNN	Adrian Lokner La devi'c, Tin Kramberger, Renata Kramberger, DinoVlahek	Lightweight CNN
5	Classification of AI-Generated, Deepfake and Real Images Using CNN	Shashank Mani Tripathi, Devansh Yadav, Priyanshu Srivastava, Prachi Verma (Assistant Professor)	CNN (EfficientNe t / Optimized CNN)

B. Dasar Teori

1) Artificial Intelligence (AI)

Artificial Intelligence (AI) merupakan cabang ilmu komputer yang berfokus pada pengembangan sistem yang mampu meniru kemampuan kognitif manusia, seperti belajar, berpikir, bernalar, serta mengambil keputusan secara mandiri. Menurut [11], kecerdasan buatan merupakan hasil transformasi kecerdasan otak manusia ke dalam perangkat komputer melalui proses matematis dan pemrograman yang

memungkinkan mesin melakukan pekerjaan layaknya manusia.

2) Deep Learning

Deep learning merupakan cabang dari machine learning yang berfokus pada penggunaan jaringan saraf tiruan dengan banyak lapisan untuk mempelajari representasi data secara hierarkis dan kompleks. Menurut [12], deep learning memiliki kemampuan luar biasa dalam memproses dan menganalisis data dalam jumlah besar serta mendeteksi pola-pola yang sulit dikenali oleh manusia.

3) Generative Artificial Intelligence

Generative Artificial Intelligence merupakan cabang dari artificial intelligence yang berfokus pada kemampuan sistem komputer untuk menciptakan data atau konten baru yang menyerupai data asli. Menurut [13], penerapan Generative Adversarial Networks (GANs) dalam pembuatan karya seni digital membuktikan bahwa algoritma ini dapat menciptakan gambar abstrak baru yang dihasilkan dari proses pelatihan terhadap kumpulan citra asli seperti lukisan alam dan pemandangan.

4. Klasifikasi

klasifikasi merupakan teknik penting dalam machine learning yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola dan karakteristik yang terdapat di dalamnya. Dalam konteks pengolahan citra digital, klasifikasi dilakukan dengan mengenali fitur visual seperti warna, tekstur, bentuk, dan pola piksel guna menentukan label atau kategori suatu gambar [14].

5. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) merupakan salah satu arsitektur deep learning yang dirancang secara khusus untuk memproses data visual seperti foto dan video. CNN terinspirasi dari cara kerja sistem penglihatan biologis manusia yang mampu mengenali pola visual secara bertahap, mulai dari bentuk sederhana hingga struktur kompleks. Menurut [15], Convolutional Neural Network (CNN) merupakan arsitektur deep learning yang tersusun atas beberapa lapisan yang bekerja secara terstruktur untuk mengekstraksi dan memproses informasi dari data citra, di antaranya meliputi convolutional layer, pooling layer, dan fully connected layer. Setiap lapisan dalam CNN berperan dalam membangun representasi fitur secara hierarkis, di mana fitur yang lebih kompleks dipelajari pada lapisan yang lebih dalam sehingga meningkatkan kemampuan model dalam mengenali pola visual pada citra digital. Secara matematis, proses ini dapat dinyatakan melalui persamaan berikut :

Operasi Convolution :

$$S(i, j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X(i+m, j+n) K(m, n)$$

Keterangan:

- $S(i,j)$: nilai output pada posisi (i,j)
- $X(i,j)$: nilai piksel masukan
- $K(m,n)$: kernel/filter berukuran $M \times N$

Max Pooling :

$$P(i,j) = \max(X(i : i + h, j : j + w))$$

Keterangan:

- $P(i,j)$: nilai hasil pooling
- X : feature map
- h, w : ukuran jendela pooling (misalnya 2×2)

Fully Connected Layer :

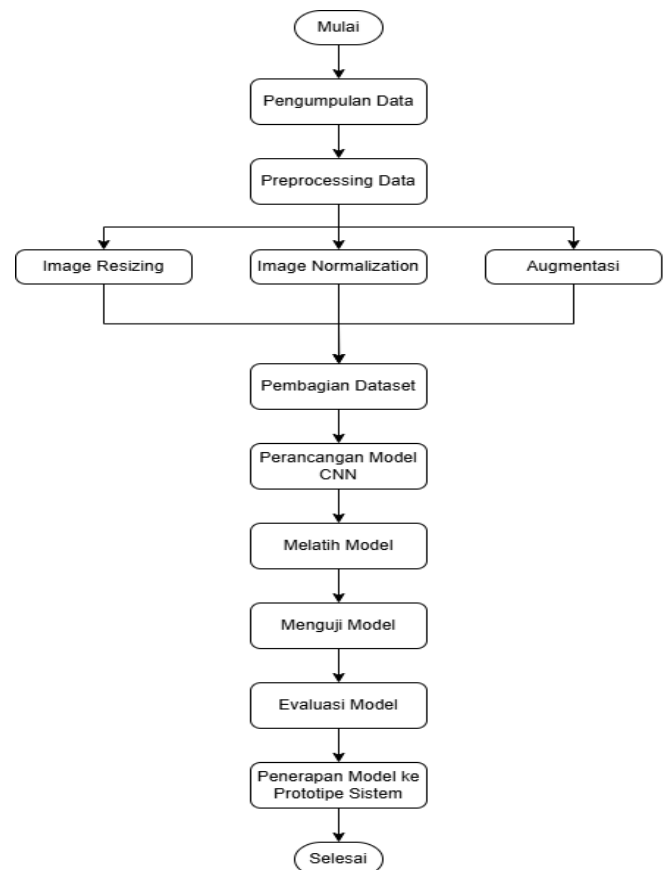
$$y = f\left(\sum_{i=1}^n (w_i x_i) + b\right)$$

Keterangan:

- x_i : input dari layer sebelumnya
- w_i : bobot
- b : bias
- f : fungsi aktivasi (ReLU, Softmax, dll.)

III. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan deep learning menggunakan metode Convolutional Neural Network (CNN) untuk mengklasifikasikan citra menjadi dua kategori, yaitu foto hasil Generative Artificial Intelligence dan foto asli. Tahapan penelitian meliputi pengumpulan data, preprocessing, perancangan model, pelatihan, pengujian, evaluasi, serta implementasi model ke dalam prototipe sistem. Berikut detail alur tahapan penelitian :



Gambar 1 Alur Tahapan Penelitian

A. Pengumpulan Data

Data pada penelitian ini diperoleh dari dataset publik di Kaggle. Dataset yang digunakan berjumlah 100.000 gambar, terdiri dari gabungan foto asli dan foto sintetis yang dihasilkan oleh model generative AI.

B. Preprocessing Data

Tahap preprocessing bertujuan untuk mempersiapkan data sebelum digunakan dalam proses pelatihan model. Adapun tahapan yang dilakukan meliputi:

1) Resizing

Semua citra diubah ke ukuran yang seragam (misalnya 150×150 piksel) agar sesuai dengan input model CNN.

2) Normalisasi

Nilai piksel dinormalisasi ke rentang $[0,1]$ untuk mempercepat proses pelatihan dan meningkatkan stabilitas model.

3) Augmentasi Data

Dilakukan augmentasi seperti rotasi, flipping, dan zoom untuk meningkatkan variasi data dan mengurangi overfitting.

C. Pembagian Dataset

Dataset dibagi menjadi data latih (training set) dan data uji (testing set) dengan rasio tertentu, misalnya 80:20.

D. Perancangan Model CNN

Model yang digunakan dalam penelitian ini adalah Convolutional Neural Network (CNN) yang dirancang untuk mengekstraksi fitur visual dari citra. Parameter yang digunakan meliputi jumlah epoch, batch size, dan learning rate yang disesuaikan untuk memperoleh performa optimal.

E. Pelatihan Model

Model CNN dilatih menggunakan data latih untuk mempelajari pola yang membedakan citra sintetis dan citra asli. Proses pelatihan dilakukan menggunakan library TensorFlow/Keras pada Jupyter Notebook.

F. Pengujian Model

Setelah pelatihan selesai, model diuji menggunakan data uji yang belum pernah dilihat sebelumnya. Tujuan pengujian ini adalah untuk mengetahui kemampuan generalisasi model terhadap data baru.

G. Evaluasi Model

Evaluasi model dilakukan untuk menilai performa model CNN dalam melakukan klasifikasi foto. Beberapa metrik yang digunakan antara lain sebagai berikut :

1) Confusion Matrix

Confusion Matrix merupakan metode evaluasi yang umum digunakan untuk menilai performa model klasifikasi seperti Convolutional Neural Network (CNN), karena mampu memberikan representasi rinci terhadap hasil prediksi model pada setiap kelas. Matriks ini terdiri dari empat komponen utama, yaitu true positive (TP), true negative (TN), false positive (FP), dan false negative (FN), yang digunakan sebagai dasar dalam menghitung berbagai metrik evaluasi seperti accuracy, precision, dan recall [11]. Untuk memberikan gambaran yang lebih jelas mengenai struktur dan interpretasinya, berikut disajikan contoh bentuk Confusion Matrix yang digunakan dalam penelitian ini.

Tabel 2 Confusion Matrix

	True Class		
		Positive	Negative
		TP (True Positive)	FP (False Positive)
Predicted Class	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

2) Akurasi (accuracy)

Akurasi (accuracy) digunakan untuk mengukur sejauh mana model Convolutional Neural Network (CNN) mampu melakukan prediksi yang benar terhadap keseluruhan data uji. Akurasi menunjukkan proporsi jumlah prediksi yang benar dibandingkan dengan total keseluruhan data, baik yang

termasuk dalam kelas positif maupun negatif. Nilai akurasi yang tinggi menandakan bahwa model memiliki kemampuan klasifikasi yang baik secara umum. Dalam konteks evaluasi model klasifikasi, akurasi didefinisikan sebagai perbandingan antara jumlah prediksi benar (true positive dan true negative) terhadap seluruh jumlah data [12]. Rumus akurasi dapat dituliskan sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3) Precision

Precision digunakan untuk mengukur tingkat ketepatan model Convolutional Neural Network (CNN) dalam memberikan prediksi terhadap kelas positif. Precision menilai seberapa banyak data atau citra yang diprediksi sebagai positif benar-benar termasuk dalam kelas positif. Dengan demikian, nilai precision yang tinggi menunjukkan bahwa model memiliki tingkat kesalahan yang rendah dalam mengklasifikasikan citra negatif sebagai positif (false positive). Dalam evaluasi model klasifikasi, precision merupakan salah satu metrik performa yang umum digunakan bersama dengan accuracy dan recall untuk menilai kualitas prediksi model [13]. Rumus precision dapat dituliskan sebagai berikut.

$$Precision = \frac{TP}{TP + FP}$$

4) Recall

Recall digunakan untuk mengukur kemampuan model Convolutional Neural Network (CNN) dalam menemukan seluruh data positif yang ada di dalam dataset. Recall menunjukkan seberapa banyak citra positif yang berhasil dideteksi dengan benar dari semua citra yang seharusnya termasuk dalam kelas positif. Nilai recall yang tinggi menandakan bahwa model mampu mengenali sebagian besar data positif dengan baik. Rumus recall dapat dituliskan sebagai berikut.

$$Recall = \frac{TP}{TP + FN}$$

5) F1-score

F1-score digunakan untuk mengukur keseimbangan antara precision dan recall dalam model Convolutional Neural Network (CNN). Nilai F1-score menjadi indikator penting ketika terdapat ketidakseimbangan data antara kelas positif dan negatif, karena metrik ini memberikan rata-rata harmonis dari kedua nilai tersebut [16]. Semakin tinggi nilai F1-score, semakin baik keseimbangan model dalam mendeteksi data positif secara akurat dan menyeluruh. Rumus F1-score dapat dituliskan sebagai berikut.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

H. Implementasi Prototipe Sistem

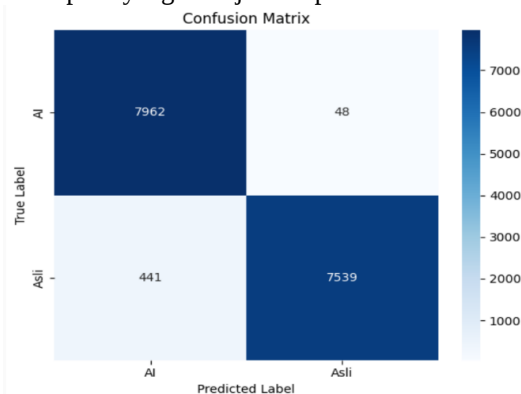
Model CNN yang telah dilatih kemudian diimplementasikan ke dalam sebuah prototipe sistem berbasis web. Prototipe ini dirancang untuk memungkinkan pengguna mengunggah gambar dan memperoleh hasil deteksi keaslian citra secara langsung berdasarkan model yang telah dikembangkan.

IV. HASIL DAN PEMBAHASAN

Penelitian ini mengevaluasi kinerja model Convolutional Neural Network (CNN) dalam mengklasifikasikan citra menjadi dua kategori, yaitu foto hasil Generative Artificial Intelligence (AI) dan foto asli. Evaluasi dilakukan menggunakan confusion matrix, metrik evaluasi, serta analisis proses pelatihan model.

A. Hasil Confusion Matrix

Berdasarkan hasil pengujian model, diperoleh confusion matrix seperti yang ditunjukkan pada Gambar 2.



Gambar 2 Confusion Matrix - CNN

Hasil tersebut menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar, yang ditandai dengan dominannya jumlah prediksi benar dibandingkan kesalahan prediksi.

Kesalahan klasifikasi masih ditemukan, terutama pada citra asli yang diklasifikasikan sebagai citra AI (false positive). Hal ini menunjukkan bahwa model memiliki kecenderungan lebih sensitif dalam mendeteksi pola yang menyerupai citra sintesis. Sensitivitas ini mengindikasikan bahwa model mampu menangkap fitur-fitur tertentu yang dianggap sebagai karakteristik citra AI, namun pada beberapa kasus fitur tersebut juga muncul pada citra asli sehingga menyebabkan kesalahan prediksi.

B. Metrik Evaluasi

Selain menggunakan confusion matrix, performa model juga dievaluasi menggunakan beberapa metrik evaluasi untuk memperoleh gambaran yang lebih komprehensif terhadap kinerja model. Metrik yang digunakan meliputi accuracy, precision, recall, dan F1-score. Hasil evaluasi performa model ditunjukkan pada Tabel 3.

Tabel 3 Hasil Metrik Evaluasi

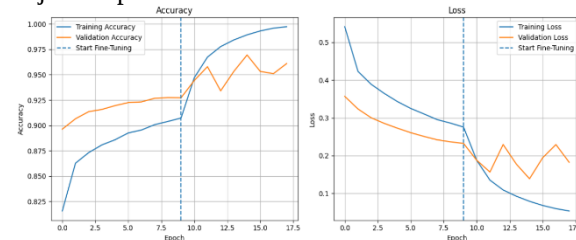
Metrik	Nilai
Accuracy	0.97
Precision	0.97
Recall	0.97
F1-Score	0.97

Berdasarkan Tabel 3, model menunjukkan performa yang sangat baik pada seluruh metrik evaluasi. Nilai accuracy yang tinggi menunjukkan bahwa model mampu melakukan klasifikasi citra secara tepat, sementara nilai precision dan recall yang seimbang mengindikasikan bahwa model tidak hanya akurat, tetapi juga konsisten dalam mendeteksi kedua kelas.

Selain itu, nilai F1-score yang tinggi menunjukkan bahwa model mampu menjaga keseimbangan antara kesalahan false positive dan false negative, sehingga kinerja model dapat dikatakan stabil dalam berbagai kondisi data.

C. Analisis Proses Pelatihan Model

Proses pelatihan model dianalisis berdasarkan grafik akurasi dan loss yang dihasilkan selama proses training dan validasi. Grafik tersebut memberikan gambaran mengenai kemampuan model dalam mempelajari pola data serta tingkat konvergensi selama pelatihan. Hasil proses pelatihan model ditunjukkan pada Gambar 3.



Gambar 3 Grafik Akurasi dan Loss Model CNN

Berdasarkan grafik akurasi, terlihat bahwa nilai akurasi training mengalami peningkatan secara bertahap dari awal pelatihan hingga mencapai nilai yang tinggi pada epoch akhir. Hal ini menunjukkan bahwa model mampu mempelajari pola data secara efektif. Sementara itu, akurasi validasi juga menunjukkan tren peningkatan yang relatif stabil, meskipun terdapat fluktuasi pada beberapa epoch. Fluktuasi tersebut merupakan hal yang wajar dan menunjukkan adanya variasi pada data validasi.

Pada grafik loss, terlihat bahwa nilai loss training mengalami penurunan secara konsisten seiring bertambahnya epoch, yang menandakan bahwa model semakin baik dalam meminimalkan kesalahan prediksi. Nilai loss validasi juga cenderung menurun, meskipun terdapat peningkatan sementara setelah tahap fine-tuning. Hal ini menunjukkan bahwa model sedang menyesuaikan parameter untuk menangkap fitur yang lebih kompleks.

Selain itu, ditandai dengan adanya garis vertikal pada grafik, proses fine-tuning memberikan dampak positif terhadap performa model. Setelah tahap tersebut, terlihat peningkatan akurasi yang lebih signifikan serta penurunan loss yang lebih tajam. Hal ini mengindikasikan bahwa proses fine-tuning mampu meningkatkan kemampuan model dalam mengenali pola yang lebih kompleks pada citra.

Perbedaan antara kurva training dan validasi relatif kecil, sehingga menunjukkan bahwa model tidak mengalami overfitting yang signifikan. Dengan demikian, model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya, meskipun masih terdapat sedikit fluktuasi pada data validasi.

D. Implementasi Prototipe

Model Convolutional Neural Network (CNN) yang dikembangkan dalam penelitian ini menunjukkan kemampuan yang baik dalam mendeteksi dan membedakan citra sintetis dan citra asli. Pada tahap implementasi prototipe, model menunjukkan konsistensi dalam menghasilkan prediksi, di mana citra yang sama cenderung menghasilkan hasil klasifikasi dan nilai confidence yang relatif stabil pada pengujian berulang. Hal ini menunjukkan bahwa model memiliki stabilitas yang baik dalam proses deteksi.

Untuk memberikan gambaran implementasi sistem, tampilan antarmuka prototipe ditunjukkan pada Gambar 4, sedangkan contoh hasil deteksi citra ditampilkan pada Gambar 5.



Gambar 4 Tampilan Prototipe Deteksi Foto



Gambar 5 Tampilan Prototipe Hasil Deteksi Foto

Konsistensi tersebut juga terlihat pada pengujian terhadap citra baru di luar dataset pelatihan, di mana hasil deteksi tidak menunjukkan perubahan yang signifikan ketika dilakukan pengujian berulang. Hal ini mengindikasikan bahwa model

bekerja secara stabil dan tidak dipengaruhi secara dominan oleh faktor acak dalam proses prediksi.

Namun demikian, pada beberapa kondisi tertentu masih ditemukan kesalahan klasifikasi, terutama pada citra dengan karakteristik visual yang mirip antara citra sintetis dan citra asli. Hal ini menunjukkan bahwa meskipun model memiliki stabilitas yang baik, kemampuan generalisasi terhadap variasi data yang kompleks masih memiliki keterbatasan.

Sebagai upaya untuk meningkatkan performa model, prototipe sistem juga dilengkapi dengan fitur umpan balik (feedback) berupa pilihan “sesuai” dan “tidak sesuai” terhadap hasil deteksi. Fitur ini memungkinkan pengguna untuk memberikan koreksi terhadap hasil prediksi yang dihasilkan oleh sistem. Data umpan balik tersebut selanjutnya dapat dimanfaatkan sebagai tambahan dataset untuk proses pelatihan ulang (retraining) model di masa mendatang.

Dengan demikian, hasil implementasi prototipe tidak hanya menunjukkan bahwa model stabil dalam memberikan prediksi, tetapi juga menyediakan mekanisme peningkatan berkelanjutan melalui interaksi pengguna, sehingga diharapkan mampu meningkatkan kemampuan generalisasi model pada penggunaan nyata.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model Convolutional Neural Network (CNN) mampu digunakan untuk membedakan citra sintetis hasil Generative Artificial Intelligence dan citra asli dengan performa yang baik. Hal ini ditunjukkan melalui hasil evaluasi menggunakan confusion matrix serta metrik accuracy, precision, recall, dan F1-score yang menunjukkan kinerja model yang tinggi dan seimbang. Selain itu, hasil analisis proses pelatihan menunjukkan bahwa model mampu belajar secara optimal, ditandai dengan peningkatan akurasi dan penurunan loss selama proses pelatihan tanpa mengalami overfitting yang signifikan. Model juga berhasil diimplementasikan ke dalam prototipe sistem berbasis web yang dapat digunakan untuk melakukan deteksi keaslian citra secara langsung.

Pada tahap implementasi, model menunjukkan konsistensi dalam memberikan hasil deteksi, baik pada citra yang sama maupun pada pengujian berulang. Namun demikian, masih ditemukan kesalahan klasifikasi pada beberapa citra dengan karakteristik visual yang mirip, yang menunjukkan bahwa kemampuan generalisasi model terhadap data baru masih memiliki keterbatasan. Dengan demikian, penelitian ini tidak hanya berhasil mengembangkan model klasifikasi citra, tetapi juga mengimplementasikannya dalam bentuk sistem deteksi yang memiliki nilai praktis dalam penggunaan nyata.

Berdasarkan temuan tersebut, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih beragam guna meningkatkan kemampuan generalisasi model dan mempertimbangkan penerapan arsitektur yang lebih kompleks atau pendekatan lain seperti transfer learning,

ensemble model, maupun Vision Transformer. Selain itu, optimalisasi pada tahap preprocessing dan augmentasi data juga perlu dilakukan agar model mampu mengenali pola yang lebih variatif. Serta pemanfaatan data umpan balik pengguna melalui fitur “sesuai” dan “tidak sesuai” juga dapat digunakan sebagai bagian dari proses pelatihan ulang (retraining) untuk meningkatkan performa model secara berkelanjutan.

DAFTAR PUSTAKA

- [1] P. Liyanage, N. Ranasinghe, and L. Kruglova, “Advancements in Generative AI and its Applications,” in *2025 5th International Conference on Advanced Research in Computing (ICARC)*, 2025, pp. 1–6. doi: 10.1109/ICARC64760.2025.10963239.
- [2] B. V. Moyo, T. Tuyikeze, F. Matsebula, and I. C. Obagbuwa, “An AI-driven conceptual framework for detecting fake news and deepfake content: a systematic review,” *Front. Artif. Intell.*, vol. 9, no. March, p. 1737790, 2026, doi: 10.3389/frai.2026.1737790.
- [3] C. Zheng, “Self-Supervised Pretext Tasks for Alzheimer’s Disease Classification using 3D Convolutional Neural Networks on Large-Scale Synthetic Neuroimaging Dataset,” pp. 1–9, 2024.
- [4] S. Ali Raza, U. Habib, M. Usman, A. Ashraf Cheema, and M. Sajid Khan, “MMGANGuard: A Robust Approach for Detecting Fake Images Generated by GANs Using Multi-Model Techniques,” *IEEE Access*, vol. 12, no. March, pp. 104153–104164, 2024, doi: 10.1109/ACCESS.2024.3393842.
- [5] S. Sohail, S. M. Sajjad, A. Zafar, Z. Iqbal, Z. Muhammad, and M. Kazim, “Deepfake Image Forensics for Privacy Protection and Authenticity Using Deep Learning,” *Inf.*, vol. 16, no. 4, pp. 1–30, 2025, doi: 10.3390/info16040270.
- [6] M. Wyawahare, V. Parkhi, S. Bhorge, M. Jha, M. Rane, and N. Muhal, “Comparative Analysis of Deepfake Detection Models on Diverse GAN-Generated Images,” *Proc. - Int. Conf. Comput. Intell. Networks*, pp. 9–18, 2024, doi: 10.1109/CINE63708.2024.10881519.
- [7] N. Huda, A. Fayyaz, U. Asad, and Y. S. Afridi, “Synthetic vs Real Image Detection using Vision Transformers and CNN-based Architectures,” 2025.
- [8] Y. Zhang, Z. Pang, S. Huang, C. Wang, and X. Zhou, “Unmasking AI-created visual content: a review of generated images and deepfake detection technologies,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 37, no. 6, 2025, doi: 10.1007/s44443-025-00154-8.
- [9] A. Lokner Lađević, T. Kramberger, R. Kramberger, and D. Vlahek, “Detection of AI-Generated Synthetic Images with a Lightweight CNN,” *AI*, vol. 5, no. 3, pp. 1575–1593, 2024, doi: 10.3390/ai5030076.
- [10] S. M. Tripathi, “Classification of AI-Generated , Deepfake and Real Images Using CNN,” vol. 12, no. 6, pp. 555–560, 2025.
- [11] R. F. Ramadhan *et al.*, *Kecerdasan buatan digital*, no. August 2025. 2023.
- [12] N. Puspita Sari, “Analisis Performa Algoritma CNN dalam Klasifikasi Citra Medis Berbasis Deep Learning Analysis Of CNN Algorithm In Deep Learning-Based Medical Image Classification,” *J. Komput.*, vol. 2, no. 2, pp. 87–92, 2024.
- [13] E. S. Purba and Hendry, “Implementation of Generative Adversarial Networks for Penerapan Generative Adversarial Networks Dalam Pembuatan,” *J. Tek. Inform.*, vol. 3, no. 3, pp. 707–715, 2022, [Online]. Available: <https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/262/125>
- [14] N. M. Noori and O. S. Qasim, “Enhancing image classification with hybrid deep learning and optimized feature selection via an improved metaheuristic algorithm,” *OPSEARCH*, 2026, doi: 10.1007/s12597-025-01084-9.
- [15] B. R. Lamichhane, G. Srijuntongsiri, and T. Horanont, “CNN based 2D object detection techniques: a review,” *Front. Comput. Sci.*, vol. 7, 2025, doi: 10.3389/fcomp.2025.1437664.
- [16] F. Z. Ünal, M. S. Güzel, M. Ünal, F. Ekinici, T. Aşuroğlu, and K. Açıcı, “Hybrid deep learning approach for multi-label classification problem: genre prediction,” *Neural Comput. Appl.*, vol. 38, no. 5, 2026, doi: 10.1007/s00521-026-11852-3.