

Analisis Perbandingan Kinerja Arsitektur CNN VGG19, ResNet50, EfficientNetB0, dan MobileNetV2 untuk Deteksi Wajah Asli dan Wajah Buatan AI

Performance Comparison of CNN Architectures VGG19, ResNet50, EfficientNetB0, and MobileNetV2 for Real and AI-Generated Face Detection

Erika Yanti^{a,1}, Muhammad Faisal^{a,2,*}, Titin Wahyuni^{a,3}, Abd Rakhim Nanda^{b,4}, Nurnawaty^{b,5}, Rizki Yusliana Bakti^{a,6}

^a Informatika, Universitas Muhammadiyah Makassar, Makassar, Indonesia

^b Teknik Pengairan, Universitas Muhammadiyah Makassar, Makassar, Indonesia

¹105841104322@student.unismuh.ac.id; ²muhfaisal@unismuh.ac.id; ³titinwahyuni@unismuh.ac.id;

⁴abd.rakhimnanda@unismuh.ac.id; ⁶nurnawaty@unismuh.ac.id; ⁵rizkiyusliana@unismuh.ac.id

*corresponding author

Informasi Artikel	ABSTRAK
<p>Diserahkan : 29 April 2026 Diterima : 14 Mei 2026 Direvisi : 18 Mei 2026 Diterbitkan : 22 Mei 2026</p> <p>Kata Kunci: deepfake Convolutional Neural Network transfer learning deteksi wajah citra sintetis</p>	<p>Perkembangan Generative Artificial Intelligence (GenAI) memungkinkan pembuatan citra wajah sintetis yang sangat menyerupai wajah asli, sehingga menimbulkan tantangan terhadap keaslian informasi digital, privasi, dan keamanan identitas. Penelitian ini mengevaluasi kinerja empat arsitektur Convolutional Neural Network (CNN), yaitu VGG19, ResNet50, EfficientNetB0, dan MobileNetV2, dalam klasifikasi wajah asli dan wajah hasil generasi AI. Dataset yang digunakan adalah HFD-8000 yang terdiri atas 8.000 citra wajah dengan skenario klasifikasi biner. Tahapan penelitian meliputi prapemrosesan data, pembagian dataset, augmentasi, penanganan ketidakseimbangan kelas, serta pelatihan model menggunakan transfer learning. Evaluasi dilakukan menggunakan accuracy, precision, recall, F1-score, ROC-AUC, dan confusion matrix. Hasil penelitian menunjukkan bahwa ResNet50 dan VGG19 memperoleh performa terbaik dengan akurasi 99,50% dan macro F1-score 99,22%. EfficientNetB0 mencapai akurasi 97,83% dan F1-score 96,61%, sedangkan MobileNetV2 memperoleh akurasi 92,58% dan F1-score 86,40%. Secara keseluruhan, ResNet50 menjadi model paling optimal karena menunjukkan keseimbangan antara akurasi, stabilitas, efisiensi, dan keandalan dalam klasifikasi wajah asli dan sintetis.</p> <p>ABSTRACT</p> <p><i>The development of Generative Artificial Intelligence (GenAI) allows the creation of synthetic facial images that closely resemble real faces, thus posing challenges to the authenticity of digital information, privacy, and identity security. This study evaluates the performance of four Convolutional Neural Network (CNN) architectures, namely VGG19, ResNet50, EfficientNetB0, and MobileNetV2, in the classification of native faces and faces generated by AI. The dataset used is HFD-8000 which consists of 8,000 facial images with binary classification scenarios. The research stages include data preprocessing, dataset sharing, augmentation, handling class imbalances, and model training using transfer learning. Evaluation was carried out using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix. The results showed that ResNet50 and VGG19 obtained the best performance with an accuracy of 99.50% and a macro F1-score of 99.22%. EfficientNetB0 achieved an accuracy of 97.83% and an F1-score of 96.61%, while MobileNetV2 achieved an accuracy of 92.58% and an F1-score of 86.40%. Overall, the ResNet50 is the most optimal model because it shows the balance between accuracy, stability, efficiency, and reliability in the classification of real and synthetic faces.</i></p>
<p>Keywords: deepfake Convolutional Neural Network transfer learning face detection synthetic image</p> <p>This is an open access article under the CC-BY-SA license.</p> 	

I. Pendahuluan

Kemajuan teknologi *Generative Artificial Intelligence* (GenAI) telah mempercepat perkembangan dalam penciptaan media sintetis, khususnya citra wajah buatan yang memiliki tingkat kemiripan sangat tinggi dengan wajah asli. Kemajuan ini membuka peluang pemanfaatan pada bidang kreatif, hiburan, dan komunikasi digital. Namun, di sisi lain, teknologi tersebut juga menimbulkan ancaman terhadap keaslian media karena dapat

dimanfaatkan untuk disinformasi, penipuan identitas, pelanggaran privasi, dan manipulasi opini publik. Oleh sebab itu, deteksi wajah sintetis menjadi isu penting dalam forensik digital dan keamanan informasi [1], [2].

Manipulasi wajah berbasis AI menjadi perhatian khusus karena wajah merupakan identitas visual utama manusia. Seiring meningkatnya kualitas hasil sintesis, perbedaan visual antara wajah asli dan wajah buatan AI menjadi semakin halus, sehingga tidak selalu dapat dikenali secara langsung oleh manusia. Kondisi ini menunjukkan perlunya pendekatan komputasional yang lebih objektif, konsisten, dan terukur untuk mengidentifikasi citra wajah palsu secara lebih andal [3], [4].

Salah satu metode yang banyak digunakan untuk mengatasi masalah ini adalah *Convolutional Neural Network* (CNN). Pendekatan ini dianggap efektif karena mampu mengekstraksi berbagai fitur visual penting, seperti tekstur, pola spasial, serta artefak yang muncul pada citra hasil manipulasi. Sejumlah studi sebelumnya mengungkapkan bahwa arsitektur berbasis CNN tetap menunjukkan performa yang kompetitif dalam deteksi deepfake. Namun demikian, performa yang dihasilkan dapat bervariasi tergantung pada jenis arsitektur yang digunakan, karakteristik dataset, serta strategi evaluasi yang diterapkan [5], [6].

Meskipun banyak penelitian melaporkan performa tinggi, hasil deteksi *deepfake* belum selalu konsisten pada berbagai kondisi. Beberapa studi menunjukkan bahwa performa model dapat menurun ketika data mengalami kompresi, gangguan visual, atau perubahan distribusi data. Selain itu, efisiensi komputasi juga menjadi faktor penting, karena model dengan akurasi tinggi belum tentu praktis diterapkan pada perangkat atau lingkungan dengan sumber daya terbatas. Temuan ini mengindikasikan bahwa penilaian kinerja model tidak dapat hanya bergantung pada akurasi, melainkan juga harus mempertimbangkan kemampuan generalisasi, keseimbangan klasifikasi, serta efisiensi dalam implementasi [7], [8].

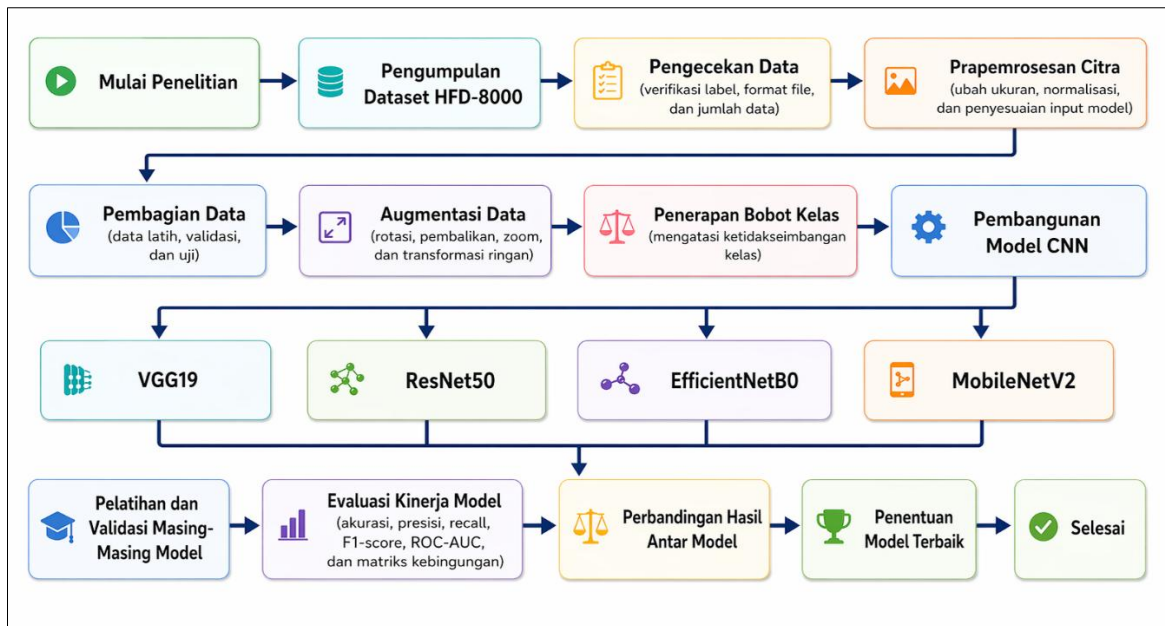
Penelitian terdahulu menunjukkan bahwa perbedaan arsitektur CNN dapat menghasilkan performa yang berbeda pada tugas deteksi wajah palsu [9]. Model berbasis ResNet dilaporkan memiliki kemampuan yang baik dalam klasifikasi citra wajah hasil manipulasi [10]. Arsitektur ringan seperti EfficientNet dan MobileNet juga dinilai menarik karena menawarkan efisiensi komputasi yang lebih tinggi [11]. Namun, sebagian besar penelitian terdahulu masih menitikberatkan pada satu arsitektur model tertentu atau menggunakan pengaturan eksperimen yang berbeda-beda, sehingga hasilnya sulit dibandingkan secara langsung [12]. Hal ini menunjukkan adanya kesenjangan penelitian, yaitu masih terbatasnya penelitian yang membandingkan beberapa arsitektur CNN pada dataset dan skema evaluasi yang sama. Akibatnya, belum diketahui secara jelas arsitektur yang memberikan keseimbangan terbaik antara akurasi, kemampuan generalisasi, dan efisiensi komputasi dalam deteksi wajah sintetis. Oleh karena itu, masih diperlukan kajian komparatif yang mengevaluasi beberapa arsitektur CNN dalam skema eksperimen yang seragam agar karakteristik performa masing-masing model dapat dianalisis secara lebih objektif.

Berdasarkan kondisi tersebut, penelitian ini berfokus pada klasifikasi citra wajah statis ke dalam dua kelas, yaitu *real* dan *fake*, menggunakan dataset HFD-8000 [13]. Penelitian ini membandingkan empat arsitektur CNN, yaitu VGG19, ResNet50, EfficientNetB0, dan MobileNetV2, dengan konfigurasi pelatihan dan evaluasi yang seragam. Keempat arsitektur tersebut dipilih karena mewakili karakteristik CNN yang berbeda, yaitu VGG19 sebagai baseline dengan struktur sederhana, ResNet50 dengan residual connection, EfficientNetB0 yang dirancang untuk mencapai keseimbangan antara akurasi dan efisiensi, serta MobileNetV2 sebagai model ringan untuk perangkat dengan sumber daya terbatas. Perbandingan dilakukan menggunakan *accuracy*, *precision*, *recall*, *F1-score*, *ROC-AUC*, *confusion matrix*, dan efisiensi model. Dengan pendekatan ini, penelitian tidak hanya menilai model berdasarkan akurasi, tetapi juga berdasarkan keseimbangan performa klasifikasi dan efisiensi implementasinya [14].

Dengan demikian, kebaruan penelitian ini terletak pada analisis komparatif empat arsitektur CNN pada dataset HFD-8000 menggunakan skema eksperimen yang seragam serta evaluasi yang mempertimbangkan akurasi, keseimbangan klasifikasi, dan efisiensi model secara bersamaan. Penelitian ini bertujuan untuk menentukan arsitektur yang paling efektif dalam mendeteksi wajah asli dan wajah buatan AI, sekaligus memberikan dasar yang lebih jelas bagi pengembangan sistem deteksi *deepfake* yang andal, efisien, dan aplikatif.

II. Metode

Penelitian ini menerapkan pendekatan eksperimen kuantitatif untuk mengevaluasi dan membandingkan performa empat arsitektur Convolutional Neural Network (CNN), yaitu VGG19, ResNet50, EfficientNetB0, dan MobileNetV2, dalam mendeteksi wajah asli dan wajah buatan AI. Pendekatan komparatif dipilih karena setiap arsitektur memiliki karakteristik representasi fitur, kompleksitas komputasi, dan kemampuan generalisasi yang berbeda pada tugas deteksi *deepfake* [15], [16]. Tahapan penelitian mencakup pengumpulan dataset, pengecekan data, pra-pemrosesan citra, pembagian data, augmentasi data, penerapan class weight, pembangunan model, pelatihan model, evaluasi kinerja, serta perbandingan hasil untuk menentukan model terbaik. Alur penelitian digambarkan pada Gambar 1.



Gambar 1. Alur Penelitian

A. Dataset dan Pra-pemrosesan

Dataset yang digunakan adalah HFD-8000, yang terdiri atas 8.000 citra wajah untuk klasifikasi biner dengan dua kelas, yaitu *real* dan *fake* [13]. Dataset ini dipilih karena relevan dengan tujuan penelitian yang ingin dicapai, yaitu membedakan citra wajah asli dan citra wajah buatan AI dalam skenario klasifikasi biner. Seluruh citra diperiksa, diberi label, dan disusun ke dalam dua kelas sebelum diproses lebih lanjut.

Tahapan pra-pemrosesan mencakup pemeriksaan file, penyesuaian ukuran citra menjadi 224×224 piksel, normalisasi nilai piksel, serta penerapan fungsi *preprocessing* yang disesuaikan dengan model yang digunakan. Selanjutnya, dataset dibagi menjadi data latih, validasi, dan uji dengan rasio 70:15:15 menggunakan metode *stratified split* untuk memastikan proporsi kelas tetap sama pada setiap subset. Untuk memperkaya variasi data latih sekaligus meminimalkan risiko overfitting, dilakukan teknik augmentasi seperti *horizontal flip*, rotasi ringan, zoom dalam skala kecil, serta penyesuaian tingkat kecerahan [17]. Karena distribusi kelas pada dataset tidak seimbang, penelitian ini juga menggunakan *class weight* agar model tetap memperhatikan kelas minoritas selama pelatihan. Penerapan strategi ini penting untuk mencegah model menunjukkan bias terhadap kelas mayoritas, yang dapat menurunkan kemampuan dalam mendeteksi kelas minoritas secara optimal [18].

B. Pengembangan Model

Dalam penelitian ini, empat arsitektur *Convolutional Neural Network* (CNN) digunakan, yaitu VGG19, ResNet50, EfficientNetB0, dan MobileNetV2, untuk membandingkan performa deteksi wajah asli dan wajah buatan AI. Keempat model dipilih karena mewakili karakteristik arsitektur yang berbeda, yaitu model CNN klasik yang dalam, model berbasis residual learning, model efisien, dan model ringan untuk komputasi terbatas.

Seluruh model dikembangkan dengan pendekatan *transfer learning* menggunakan bobot awal yang berasal dari *ImageNet*. Pendekatan ini dipilih karena mampu memanfaatkan fitur visual umum yang telah dipelajari dari dataset berskala besar, sehingga proses pelatihan menjadi lebih efisien dan stabil pada tugas klasifikasi baru. Penggunaan *transfer learning* pada model VGG19 dan ResNet50 juga telah dilaporkan efektif untuk deteksi citra deepfake [19].

Pada setiap model, lapisan klasifikasi akhir disesuaikan untuk tugas klasifikasi biner dengan satu *neuron* keluaran dan fungsi aktivasi *sigmoid*. VGG19 digunakan sebagai representasi arsitektur CNN klasik yang dalam dan memiliki struktur konvolusi bertingkat yang sederhana. ResNet50 dipilih karena memiliki mekanisme residual connection yang membantu proses pelatihan jaringan secara lebih stabil. EfficientNetB0 digunakan karena menawarkan keseimbangan antara akurasi dan efisiensi, dan arsitektur EfficientNet juga telah menunjukkan performa yang baik pada tugas deteksi deepfake [20].

Sementara itu, MobileNetV2 digunakan sebagai representasi model ringan yang dirancang untuk kebutuhan komputasi terbatas dan implementasi yang lebih efisien. Karakteristik ini menjadikan MobileNet relevan untuk dibandingkan ketika aspek efisiensi model juga dipertimbangkan selain performa klasifikasi [21].

Dengan pemilihan empat arsitektur tersebut, penelitian ini diharapkan dapat memberikan perbandingan yang lebih representatif terhadap pengaruh perbedaan desain model terhadap performa deteksi wajah asli dan wajah buatan AI.

C. Prosedur Pelatihan dan Evaluasi

Untuk menjaga keadilan dalam perbandingan, seluruh model dilatih menggunakan konfigurasi pelatihan yang seragam. Lapisan klasifikasi akhir pada masing-masing model dimodifikasi agar sesuai dengan tugas klasifikasi biner, yaitu membedakan citra wajah ke dalam dua kelas, *real* dan *fake*. Pelatihan dilakukan dalam dua tahap, yaitu pelatihan awal (phase 1) dan fine-tuning (phase 2). Pada phase 1, model dilatih menggunakan *optimizer* Adam, *loss function* binary crossentropy, *learning rate* 0,0001, *batch size* 32, dan 20 epoch. Selanjutnya, pada phase 2 dilakukan *fine-tuning* dengan *learning rate* 0,00001 selama 5 epoch. Dengan demikian, seluruh model dilatih selama total 25 epoch menggunakan konfigurasi hyperparameter yang sama agar hasil perbandingan lebih objektif dan dapat direplikasi [22].

Tabel 1. Konfigurasi Pelatihan Seluruh Model

Parameter	Nilai
Ukuran input citra	224 × 224 piksel
Jumlah kelas	2 kelas (<i>real</i> dan <i>fake</i>)
<i>Optimizer</i>	Adam
<i>Loss function</i>	Binary crossentropy
<i>Batch size</i>	32
<i>Learning rate phase 1</i>	0,0001
<i>Learning rate phase 2 (fine-tuning)</i>	0,00001
<i>Epoch phase 1</i>	20
<i>Epoch phase 2</i>	5
<i>Total epoch</i>	25
Strategi pembagian data	70% train, 15% test, 15% validation
Penanganan ketidakseimbangan kelas	Class weight
Augmentasi data	Horizontal flip, rotasi ringan, zoom, penyesuaian kecerahan

Seluruh konfigurasi hyperparameter pada Tabel 1 diterapkan secara identik pada VGG19, ResNet50, EfficientNetB0, dan MobileNetV2 untuk memastikan perbandingan dilakukan secara adil dan objektif.

Kinerja model dinilai menggunakan sejumlah metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, *F1-score*, *ROC-AUC*, dan *confusion matrix*. *Accuracy* berfungsi untuk menghitung persentase prediksi yang tepat secara keseluruhan, tetapi pada dataset yang tidak seimbang, metrik ini kurang mampu menggambarkan performa model secara komprehensif. Oleh sebab itu, *precision*, *recall*, dan *F1-score* digunakan untuk mengevaluasi keseimbangan performa model pada masing-masing kelas. Di sisi lain, *ROC-AUC* dimanfaatkan untuk mengukur kemampuan model dalam membedakan antarkelas pada berbagai nilai *threshold*. Selain itu, *confusion matrix* digunakan untuk memberikan informasi yang lebih detail mengenai distribusi prediksi yang benar maupun salah pada setiap kelas [23].

Berdasarkan *confusion matrix*, nilai *accuracy* dihitung menggunakan Persamaan (1):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Nilai *precision* dihitung menggunakan Persamaan (2):

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Nilai *recall* dihitung menggunakan Persamaan (3):

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Selanjutnya, *F1-score* dihitung menggunakan Persamaan (4):

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

dengan *TP*= true positive, *TN*= true negative, *FP*= false positive, dan *FN*= false negative.

Selain itu, *ROC-AUC* digunakan untuk mengevaluasi kemampuan model dalam membedakan dua kelas secara keseluruhan. Nilai *AUC* yang semakin mendekati 1 menunjukkan kemampuan diskriminatif model yang semakin baik [23].

Tabel 2. Metrik evaluasi yang digunakan

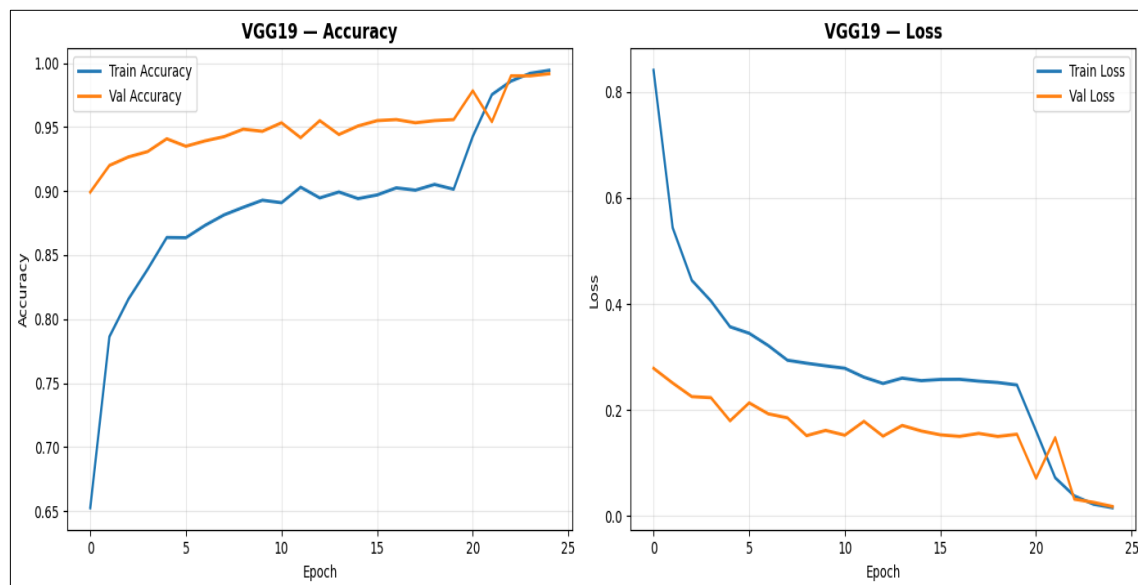
Metrik	Fungsi utama
<i>Accuracy</i>	Menilai proporsi prediksi benar secara keseluruhan
<i>Precision</i>	Menilai ketepatan prediksi pada kelas positif
<i>Recall</i>	Menilai kemampuan menangkap seluruh data positif
<i>F1-score</i>	Menilai keseimbangan <i>precision</i> dan <i>recall</i>
<i>ROC-AUC</i>	Menilai kemampuan diskriminatif model
<i>Confusion matrix</i>	Menunjukkan pola prediksi benar dan salah per kelas

Penentuan model terbaik dilakukan berdasarkan evaluasi secara menyeluruh dengan mempertimbangkan akurasi, keseimbangan antara *precision* dan *recall*, nilai F1-score, ROC-AUC, pola kesalahan yang terlihat pada *confusion matrix*, serta efisiensi model. Dengan demikian, model yang terpilih tidak hanya unggul secara kuantitatif, tetapi juga lebih andal dan sesuai untuk penerapan di dunia nyata.

III. Hasil dan Pembahasan

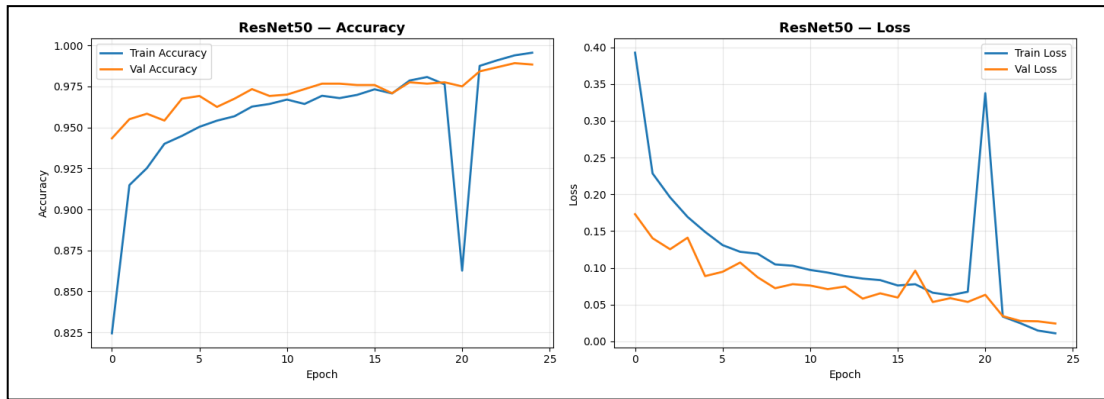
Hasil pengujian menunjukkan bahwa keempat model *Convolutional Neural Network* (CNN), yaitu VGG19, ResNet50, EfficientNetB0, dan MobileNetV2, memiliki performa yang berbeda dalam mendeteksi wajah asli dan wajah buatan AI. Evaluasi dilakukan melalui kurva pelatihan dan validasi, *confusion matrix*, metrik kuantitatif, serta analisis efisiensi model.

Berdasarkan kurva pelatihan, VGG19 dan ResNet50 menunjukkan proses pembelajaran yang paling stabil. Pada kedua model tersebut, nilai training accuracy dan validation accuracy meningkat secara konsisten hingga mendekati 1,00, sedangkan training loss dan validation loss menurun sampai nilai yang sangat kecil. EfficientNetB0 juga menunjukkan tren pelatihan yang baik, meskipun masih mengalami fluktuasi yang sedikit lebih besar dibandingkan dengan VGG19 dan ResNet50. Sebaliknya, MobileNetV2 memperlihatkan pola pelatihan yang paling fluktuatif, terutama pada bagian akhir pelatihan ketika *validation accuracy* menurun dan *validation loss* meningkat. Kondisi ini menunjukkan bahwa kemampuan generalisasi MobileNetV2 lebih rendah dibandingkan dengan model lainnya.



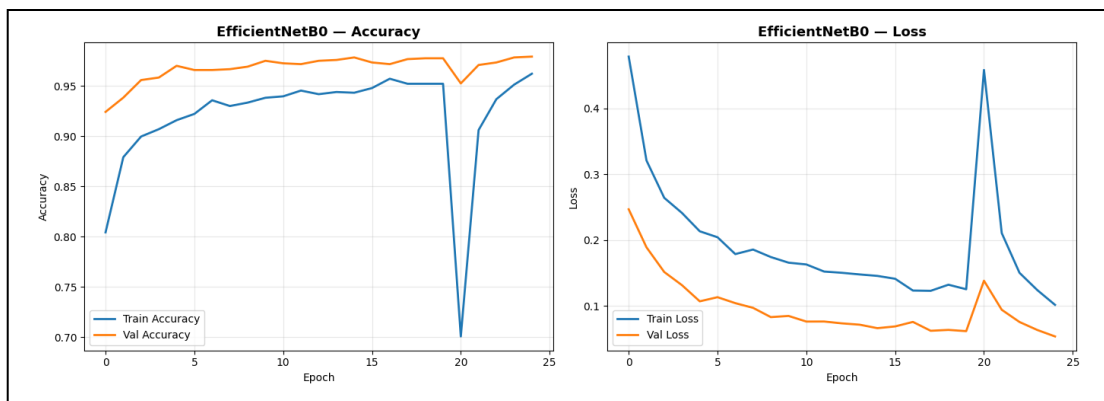
Gambar 2. Kurva *accuracy* dan loss pelatihan dan validasi model VGG19.

Pada VGG19, peningkatan *accuracy* dan penurunan *loss* berlangsung stabil hingga akhir epoch. Pola ini menunjukkan bahwa model mampu mempelajari fitur pembeda antara citra wajah asli dan citra wajah buatan AI secara efektif.



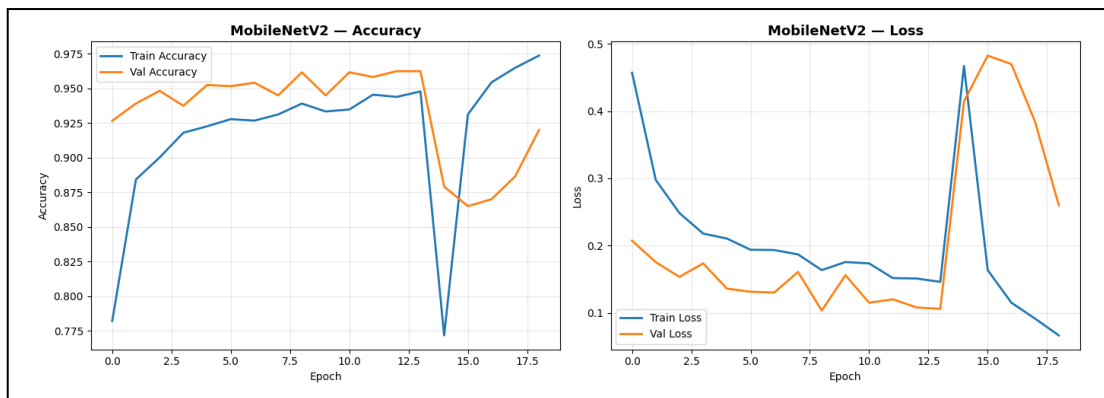
Gambar 3. Kurva *accuracy* dan loss pelatihan dan validasi model ResNet50.

ResNet50 menunjukkan performa pelatihan yang sangat baik. Meskipun terdapat fluktuasi sesaat pada sekitar epoch ke-20, model kembali stabil pada epoch berikutnya. Secara umum, pola ini menunjukkan bahwa ResNet50 memiliki konvergensi yang cepat dan kestabilan pelatihan yang tinggi.



Gambar 4. Kurva *accuracy* dan loss pelatihan dan validasi model EfficientNetB0.

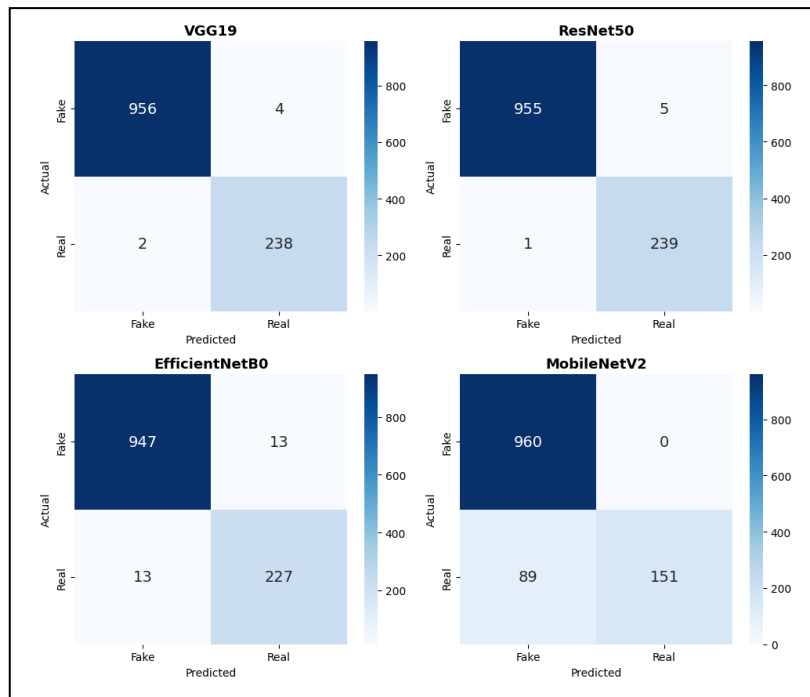
EfficientNetB0 juga mampu mencapai validation accuracy yang tinggi, namun masih memperlihatkan fluktuasi lebih besar dibandingkan dengan dua model terbaik. Hal ini menunjukkan bahwa model masih mampu mempelajari pola data dengan baik, tetapi kestabilannya sedikit berada di bawah VGG19 dan ResNet50.



Gambar 5. Kurva *accuracy* dan loss pelatihan dan validasi model MobileNetV2.

MobileNetV2 menunjukkan pola pelatihan yang paling tidak stabil. Meskipun training accuracy meningkat, validation accuracy menurun pada bagian akhir pelatihan. Hasil ini mengindikasikan bahwa model ringan tersebut memiliki keterbatasan dalam mempertahankan kemampuan generalisasi.

Selain melalui kurva pelatihan, performa model dianalisis menggunakan confusion matrix untuk melihat distribusi prediksi benar dan salah pada masing-masing kelas.



Gambar 6. Confusion matrix hasil klasifikasi seluruh model.

Berdasarkan *confusion matrix*, VGG19 berhasil mengklasifikasikan 956 citra fake dan 238 citra real dengan benar, dengan total kesalahan yang sangat kecil. ResNet50 menunjukkan hasil yang sedikit lebih baik, yaitu 955 citra fake dan 239 citra real terklasifikasi benar, dengan hanya 6 kesalahan secara keseluruhan. EfficientNetB0 juga memberikan performa yang baik, tetapi jumlah kesalahannya lebih besar dibandingkan dengan VGG19 dan ResNet50. Sementara itu, MobileNetV2 berhasil mengenali seluruh citra fake dengan benar, tetapi masih banyak salah mengklasifikasikan citra real sebagai fake, sehingga menunjukkan kecenderungan bias terhadap kelas mayoritas.

Untuk memperoleh gambaran performa yang lebih lengkap, evaluasi kuantitatif dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan ROC-AUC.

Tabel 3. Hasil evaluasi kinerja seluruh model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
VGG19	0.9950	0.9835	0.9917	0.9876	0.9998
ResNet50	0.9950	0.9795	0.9958	0.9876	0.9999
EfficientNetB0	0.9783	0.9458	0.9458	0.9458	0.9971
MobileNetV2	0.9258	1.0000	0.6292	0.7724	0.9946

Berdasarkan Tabel 3, VGG19 dan ResNet50 memperoleh nilai *accuracy* tertinggi yang sama, yaitu 0,9950. Namun, ResNet50 sedikit unggul pada *recall* dan ROC-AUC. EfficientNetB0 menempati posisi berikutnya dengan performa yang masih tinggi, sedangkan MobileNetV2 memperoleh *accuracy* terendah. Walaupun MobileNetV2 memiliki *precision* sempurna, nilai *recall* yang rendah menunjukkan bahwa model tersebut gagal mengenali sebagian besar citra pada kelas *real*.

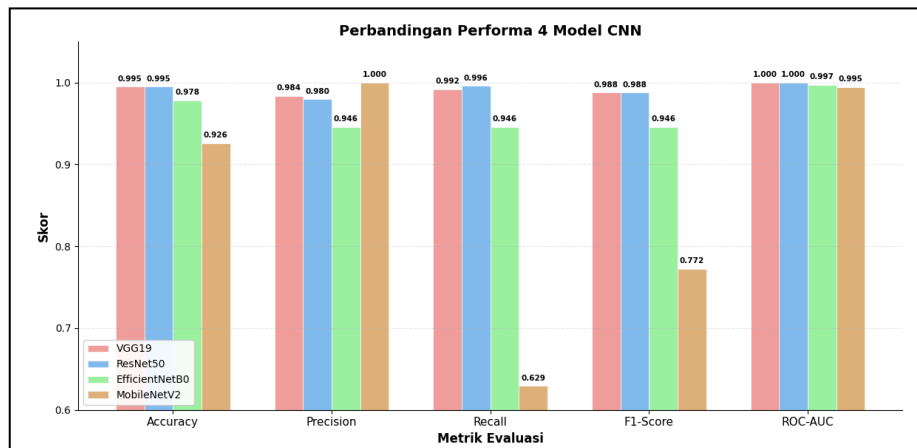
Selain performa klasifikasi, penelitian ini juga mengevaluasi efisiensi model berdasarkan waktu pelatihan dan jumlah parameter.

Tabel 4. Perbandingan efisiensi model

Model	Training Time (s)	Total Parameters	Trainable Parameters
VGG19	11005.0	20.155.969	9.570.817
ResNet50	3681.7	24.112.513	16.619.521
EfficientNetB0	1794.5	4.377.764	3.000.257
MobileNetV2	1045.2	2.586.177	2.009.537

Berdasarkan Tabel 4, VGG19 merupakan model dengan waktu pelatihan paling lama. ResNet50 memiliki jumlah parameter terbesar, tetapi waktu pelatihannya jauh lebih singkat dibandingkan dengan VGG19.

EfficientNetB0 menunjukkan kompromi yang baik antara performa dan efisiensi, sedangkan MobileNetV2 merupakan model paling ringan dengan waktu pelatihan tercepat.



Gambar 7. Perbandingan performa empat model CNN berdasarkan metrik evaluasi.

Grafik perbandingan pada Gambar 7. menunjukkan bahwa ResNet50 dan VGG19 menempati posisi terbaik pada hampir seluruh metrik utama. EfficientNetB0 tetap kompetitif, sedangkan MobileNetV2 tertinggal terutama pada *recall* dan *F1-score*.

Hasil penelitian menunjukkan bahwa ResNet50 dan VGG19 merupakan dua model dengan performa terbaik untuk deteksi wajah asli dan wajah buatan AI pada dataset HFD-8000. Kedua model menghasilkan *accuracy* yang sama tinggi, tetapi ResNet50 sedikit lebih unggul pada *recall* dan ROC-AUC. Hal ini menunjukkan bahwa ResNet50 memiliki kemampuan yang lebih baik dalam mengenali seluruh data pada kelas target serta memiliki kemampuan diskriminatif yang paling tinggi dibandingkan dengan model lainnya.

Keunggulan ResNet50 dapat dijelaskan oleh penggunaan residual connection yang memungkinkan aliran informasi dan gradien tetap terjaga pada jaringan yang dalam. Mekanisme ini membantu model mempelajari fitur kompleks tanpa mengalami degradasi performa, sehingga lebih efektif dalam menangkap artefak halus pada citra wajah sintesis. VGG19 juga menunjukkan performa tinggi karena memiliki struktur yang sederhana dan konsisten, sehingga mampu mengekstraksi fitur visual secara stabil, meskipun membutuhkan biaya komputasi yang lebih besar. EfficientNetB0 menerapkan pendekatan compound scaling yang menyeimbangkan kedalaman, lebar, dan resolusi input, sehingga menghasilkan efisiensi yang baik dengan performa yang tetap kompetitif. Sementara itu, MobileNetV2 menggunakan depthwise separable convolution untuk mengurangi jumlah parameter dan mempercepat pelatihan, namun penyederhanaan ini dapat membatasi kapasitas model dalam menangkap pola visual yang sangat halus, sehingga performanya lebih rendah dibandingkan model lainnya.

VGG19 menunjukkan performa yang hampir setara dengan ResNet50. Kurva pelatihannya juga merupakan yang paling stabil, sehingga menunjukkan bahwa arsitektur CNN klasik yang dalam masih sangat efektif untuk tugas klasifikasi ini. Namun, dari sisi efisiensi komputasi, VGG19 membutuhkan waktu pelatihan paling lama sehingga kurang unggul untuk implementasi yang menuntut sumber daya lebih hemat.

EfficientNetB0 menempati posisi menengah dengan performa yang tetap tinggi dan biaya komputasi yang jauh lebih rendah dibandingkan VGG19 dan ResNet50. Hal ini menunjukkan bahwa EfficientNetB0 dapat menjadi alternatif yang baik apabila efisiensi komputasi menjadi salah satu pertimbangan utama dalam implementasi sistem.

MobileNetV2 merupakan model paling efisien dari sisi jumlah parameter dan waktu pelatihan. Akan tetapi, efisiensi tersebut disertai penurunan performa klasifikasi yang cukup besar, terutama pada nilai *recall* dan *F1-score*. Temuan ini menunjukkan adanya *trade-off* yang jelas antara efisiensi komputasi dan kualitas klasifikasi. Dalam konteks dataset yang tidak seimbang, kemampuan model dalam mengenali kelas minoritas menjadi sangat penting, sehingga model yang terlalu ringan dapat mengalami penurunan kemampuan generalisasi.

Jika dibandingkan secara menyeluruh, ResNet50 memberikan keseimbangan terbaik antara performa klasifikasi, kestabilan pelatihan, dan efisiensi komputasi. VGG19 sangat kuat dari sisi performa, tetapi kurang efisien. EfficientNetB0 memberikan kompromi yang baik, sedangkan MobileNetV2 lebih sesuai ditempatkan sebagai model ringan dengan kompromi performa. Oleh karena itu, ResNet50 dapat direkomendasikan sebagai model terbaik dalam penelitian ini karena memberikan kombinasi paling optimal antara akurasi tinggi, distribusi kesalahan yang kecil, dan efisiensi komputasi yang masih baik.

IV. Kesimpulan dan saran

Berdasarkan hasil penelitian, dapat disimpulkan bahwa perbandingan empat arsitektur Convolutional Neural Network (CNN), yaitu VGG19, ResNet50, EfficientNetB0, dan MobileNetV2, menunjukkan bahwa perbedaan desain arsitektur berpengaruh terhadap performa deteksi wajah asli dan wajah buatan AI. ResNet50 dan VGG19 memberikan hasil klasifikasi terbaik, dengan nilai akurasi tertinggi sebesar 99,50%. Namun, ResNet50 ditetapkan sebagai model terbaik karena memiliki keseimbangan performa yang paling optimal, ditinjau dari recall, F1-score, ROC-AUC, confusion matrix, serta efisiensi waktu pelatihan yang lebih baik dibandingkan dengan VGG19. Temuan ini menunjukkan bahwa arsitektur berbasis residual learning mampu memberikan kombinasi yang kuat antara akurasi tinggi, kestabilan pelatihan, dan kemampuan generalisasi pada dataset HFD-8000. Hasil penelitian ini juga menegaskan bahwa evaluasi model deteksi wajah buatan AI tidak cukup hanya didasarkan pada akurasi, tetapi perlu mempertimbangkan keseimbangan klasifikasi dan efisiensi implementasi. Adapun saran untuk penelitian selanjutnya adalah memperluas pengujian pada dataset yang lebih beragam dan lebih menantang, termasuk dataset lintas sumber agar kemampuan generalisasi model dapat dievaluasi secara lebih mendalam. Selain itu, penelitian berikutnya dapat mengeksplorasi teknik augmentasi data yang lebih beragam untuk meningkatkan ketangguhan model terhadap variasi kualitas citra, kompresi, dan gangguan visual. Penelitian selanjutnya juga dapat mengembangkan pendekatan deteksi pada data video atau multimodal, serta mengeksplorasi arsitektur yang lebih baru, seperti model berbasis transformer atau pendekatan hibrida, untuk meningkatkan performa deteksi sekaligus mempertahankan efisiensi komputasi. Penggunaan teknik interpretabilitas model juga dapat dipertimbangkan agar proses pengambilan keputusan model menjadi lebih transparan dan mudah dianalisis.

Daftar Pustaka

- [1] S. M. Qureshi, A. Saeed, S. H. Almotiri, F. Ahmad, and M. A. A. Ghamdi, "Deepfake forensics: a survey of digital forensic methods for multimodal deepfake identification on social media," *PeerJ Comput. Sci.*, vol. 10, pp. 1–40, 2024, doi: 10.7717/PEERJ-CS.2037.
- [2] L. Y. Gong and X. J. Li, "A Contemporary Survey on Deepfake Detection: Datasets, Algorithms, and Challenges," 2024, doi: 10.3390/electronics13030585.
- [3] G. Gupta, K. Raja, M. Gupta, T. Jan, S. T. Whiteside, and M. Prasad, "A Comprehensive Review of DeepFake Detection Using Advanced Machine Learning and Fusion Methods," 2024. doi: 10.3390/electronics13010095.
- [4] A. Raza, K. Munir, and M. Almutairi, "A Novel Deep Learning Approach for Deepfake Image Detection," *Appl. Sci.*, vol. 12, no. 19, 2022, doi: 10.3390/app12199820.
- [5] V. L. L. Thing, "Deepfake Detection with Deep Learning: Convolutional Neural Networks versus Transformers," in *Proceedings of the 2023 IEEE International Conference on Cyber Security and Resilience, CSR 2023*, 2023, pp. 246–253. doi: 10.1109/CSR57506.2023.10225004.
- [6] A. Angeline and H. Kusniyati, "Komparasi Performa VGG19, ResNet50, DenseNet121 dan MobileNetV2 Dalam Mendeteksi Gambar Deepfake," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 9, no. 2, p. 397, 2024, doi: 10.24114/cess.v9i2.58671.
- [7] Y. Lu and T. Ebrahimi, "Assessment framework for deepfake detection in real - world situations," *EURASIP J. Image Video Process.*, vol. 8, 2024, doi: 10.1186/s13640-024-00621-8.
- [8] I. Balafrej and M. Dahmane, "Enhancing practicality and efficiency of deepfake detection," *Sci. Rep.*, vol. 14, no. 1, pp. 1–11, 2024, doi: 10.1038/s41598-024-82223-y.
- [9] Y. Ralhen and S. Sharma, "Enhanced deepfake detection using an ensemble of convolutional neural networks," *IAES Int. J. Artif. Intell.*, vol. 14, no. 5, pp. 4043–4049, 2025, doi: 10.11591/ijai.v14.i5.pp4043-4049.
- [10] L. Kroiß and J. Reschke, "Deepfake Detection of Face Images based on a Convolutional Neural Network," 2025, [Online]. Available: <http://arxiv.org/abs/2503.11389>
- [11] E. Şafak and N. Barişçi, "Detection of fake face images using lightweight convolutional neural networks with stacking ensemble learning method," *PeerJ Comput. Sci.*, vol. 10, pp. 1–20, 2024, doi: 10.7717/PEERJ-CS.2103.
- [12] S. AlMuhaideb, H. Alshaya, L. Almutairi, D. Alomran, and S. T. Alhamed, "LightFakeDetect: A Lightweight Model for Deepfake Detection in Videos That Focuses on Facial Regions," *Mathematics*, vol. 13, no. 19, pp. 1–22, 2025, doi: 10.3390/math13193088.
- [13] K. R. Ahmed, A. K. M. Masum, and M. Z. Uddin, "The HFD-8000," Mendeley Data. [Online]. Available: <https://data.mendeley.com/datasets/r88c3hr2x6/1>
- [14] S. M. Yasir and H. Kim, "Lightweight Deepfake Detection Based on Multi-Feature Fusion," *Appl. Sci.*, vol. 15, no. 4, 2025, doi: 10.3390/app15041954.
- [15] A. Al-Adwan, H. Alazzam, N. Al-Anbaki, and E. Alduweib, "Detection of Deepfake Media Using a Hybrid CNN–RNN Model and Particle Swarm Optimization (PSO) Algorithm," *Computers*, vol. 13, no. 4, pp. 1–16, 2024, doi: 10.3390/computers13040099.
- [16] S. Lei, J. Song, F. Feng, Z. Yan, and A. Wang, "Deepfake Face Detection and Adversarial Attack

- Defense Method Based on Multi-Feature Decision Fusion,” *Appl. Sci.*, vol. 15, no. 12, 2025, doi: 10.3390/app15126588.
- [17] T. Raharjo *et al.*, “Analisis Forensik Deepfake Berbasis Convolutional Neural Network (Cnn) Untuk Deteksi Inkonsistensi Tekstur Dan Pola Pada Citra Wajah,” *JATI (Jurnal Mhs. Tek. Inform.,* vol. 9, no. 2, pp. 2731–2738, 2025, doi: 10.36040/jati.v9i2.13058.
- [18] A. Bhattacharjee *et al.*, “CAE-Net: Generalized deepfake image detection using convolution and attention mechanisms with spatial and frequency domain features,” *J. Vis. Commun. Image Represent.*, vol. 115, pp. 1–35, 2026, doi: 10.1016/j.jvcir.2025.104679.
- [19] T. N. Rohmah, P. Dewi, Kurniawati, H. Didin, M. A. Yunifa, and C.-C. Santiago, “Deepfake Image Detection using Deep Learning,” *Int. J. Sci. Res. Eng. Manag.*, vol. 09, no. 05, pp. 1–9, 2025, doi: 10.55041/ijrsrem46579.
- [20] E. M. Dharma, M. S. I. Ni, and P. R. A. A. I, “Deepfake Image Detection Using Convolutional Neural Network,” in *Lecture Notes in Networks and Systems*, 2025, pp. 51–62. doi: 10.1007/978-981-96-1981-8_5.
- [21] M. Amen and M. L. Ranam, “Lightweight Deepfake Detection on Mobile Devices Using Attention-Enhanced MobileNet and Frequency Domain Analysis,” *J. Technol. Informatics Eng.*, vol. 4, no. 1, pp. 95–114, 2025, doi: 10.51903/jtie.v4i1.275.
- [22] A. Karathanasis, J. Violos, and I. Kompatsiaris, “A Comparative Analysis of Compression and Transfer Learning Techniques in DeepFake Detection Models,” *Mathematics*, vol. 13, no. 5, Mar. 2025, doi: 10.3390/math13050887.
- [23] M. Abbasi, P. Váz, J. Silva, and P. Martins, “Comprehensive Evaluation of Deepfake Detection Models: Accuracy, Generalization, and Resilience to Adversarial Attacks,” *Appl. Sci.*, vol. 15, no. 3, pp. 1–16, 2025, doi: 10.3390/app15031225.