Jurnal Rekayasa Sistem Informasi dan Teknologi
Volume 2, No 3 – Februari 2025
e-ISSN : 3025-888X

SISTEM DETEKSI GAMBAR *DEEPFAKE* MENGGUNAKAN CNN DENSENET-121 DENGAN *WATERMARKING LEAST SIGNIFICANT BIT* (LSB)

Fatwa Akbar Jiwani[1*], Bagus Satrio Waluyo Poetro[2]
Universitas Islam Sultan Agung[1,2]
fatwaakbar@std.unissula.ac.id* bagusswp@unissula.ac.id[2]

Received: 06-02-2025	Revised: 12-02-2025	Approved: 26-02-2025

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model deteksi deepfake menggunakan arsitektur Convolutional Neural Network (CNN) DenseNet-121 serta mengevaluasi efektivitas teknik watermarking Least Significant Bit (LSB) dalam meningkatkan keamanan citra digital. Dataset yang digunakan terdiri dari citra asli dan citra deepfake yang diperoleh dari sumber terbuka Kaggle, yang dibagi menjadi subset pelatihan, validasi, dan pengujian dengan total 22.382 gambar. Proses preprocessing melibatkan resizing ke ukuran 128x128 piksel, konversi ke grayscale, normalisasi, serta augmentasi data untuk meningkatkan generalisasi model. Model DenseNet-121 dikompilasi menggunakan optimizer Adam dan loss function categorical crossentropy, dengan evaluasi menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil pelatihan menunjukkan bahwa model mampu mendeteksi deepfake dengan akurasi tinggi. Selain itu, evaluasi watermarking menggunakan PSNR menunjukkan bahwa penyisipan watermark dengan metode LSB tidak mengurangi kualitas visual citra secara signifikan. Penelitian ini memberikan kontribusi dalam meningkatkan deteksi deepfake dan keamanan digital melalui kombinasi metode CNN dan watermarking.

Kata Kunci: Deepfake, CNN, DenseNet-121, Watermarking, LSB, Deteksi Citra Digital

PENDAHULUAN

Pemalsuan citra digital, termasuk teknologi deepfake, telah berkembang pesat seiring kemajuan kecerdasan buatan (AI) dan pembelajaran mendalam. Deepfake memungkinkan manipulasi citra dan video dengan tingkat ketelitian yang sangat tinggi, sehingga sering kali sulit dideteksi secara visual oleh manusia. Deepfake dapat mengubah wajah, ekspresi, atau keseluruhan adegan dalam citra dan video, yang berdampak pada privasi, keamanan informasi, dan kepercayaan publik terhadap media digital (Almars, 2021)(Karnouskos, 2020), Seiring dengan semakin canggihnya teknik manipulasi ini, muncul kebutuhan mendesak untuk mengembangkan metode pendeteksian yang lebih andal, akurat, dan efisien dalam membedakan citra asli dari citra hasil manipulasi. Menurut penelitian yang dilakukan oleh Tolosana (Tolosana et al., 2020) dan Karnouskos (Karnouskos, 2020), sudah banyak metode yang dikembangkan untuk mendeteksi deepfake, baik yang berbasis analisis piksel maupun teknik pembelajaran mesin. Meskipun demikian, tantangan utama dalam pendeteksian deepfake adalah kemampuan teknologi ini yang terus meningkat dalam meniru pola wajah, pencahayaan, dan tekstur secara realistis (Masood et al., 2023). Penelitian lanjutan juga mengeksplorasi metode hybrid yang menggabungkan teknik deep learning dengan analisis statistik untuk meningkatkan akurasi deteksi serta mengurangi laju kesalahan identifikasi.

Convolutional Neural Network (CNN) merupakan salah satu metode deeplearning yang cocok dan efektif untuk mengenali pola dalam citra. CNN bekerja dengan mengekstraksi fitur spasial dari citra, sehingga mampu membedakan citra asli dari citra yang telah dimanipulasi. Penelitian yang dilakukan oleh (License & Rizvee, 2023) menggunakan delapan arsitektur CNN untuk mendeteksi gambar deepfake dari dataset

berukuran besar, dengan hasil yang terbukti andal dan akurat. Dari penelitian yang dilakukan oleh (Patel et al., 2023) menunjukkan bahwa CNN sangat efektif dalam mendeteksi gambar deepfake. sedangkan studi yang dilakukan oleh Alzahrani, Ahmed (Alzahrani, 2024) melaporkan bahwa model DenseNet121 mencapai akurasi hingga 92.32% dalam mengidentifikasi deepfake. Yang mana akan dilakukan evaluasi dengan menggunakan metrik seperti akurasi, presisi, recall, F1-score, dan area under the ROC curve (AUC), sehingga memberikan wawasan mendalam mengenai kelebihan dan kekurangan tiap pendekatan (Khan & Dang-Nguyen, 2023). Dengan demikian, pemilihan arsitektur CNN yang tepat sangat menentukan keandalan sistem deteksi deepfake.

Namun, salah satu tantangan utama dalam sistem ini adalah bagaimana mengamankan hasil deteksi model agar informasi mengenai keaslian suatu citra dapat dipertahankan dan diverifikasi oleh pihak lain. Meskipun CNN telah terbukti efektif dalam mendeteksi deepfake, hasil deteksi ini masih dapat dimanipulasi atau dipalsukan kembali tanpa mekanisme perlindungan tambahan. Selain itu, metode konvensional dalam pendeteksian deepfake umumnya hanya menghasilkan output dalam bentuk nilai prediksi atau label oleh karna itu di samping pendeteksian deepfake, *watermarking* steganografi adalah metode yang digunakan untuk melindungi keaslian citra digital dengan menambahkan tanda air yang dapat diverifikasi. Teknik ini menggunakan algoritma steganografi untuk menyisipkan tanda air atau hash ke dalam citra digital, yang kemudian dapat diverifikasi untuk memastikan integritas citra (Sanivarapu et al., 2022)(Swaminathan et al., 2006). Menurut penelitian (Memon & Wong, 1998), *watermarking* memberikan perlindungan tambahan pada citra, yang membantu mengidentifikasi setiap perubahan atau manipulasi yang terjadi pada citra sejak pertama kali dibuat. sedangkan studi oleh Boato et al. (Boato et al., 2009) mengevaluasi kekuatan dan ketahanan *watermarking* kriptografi terhadap berbagai bentuk manipulasi.

Dengan kemajuan teknologi dan kemudahan akses ke media digital, isu manipulasi serta pemalsuan gambar digital semakin mengemuka (A C & M T, 2023). Metode *watermarking* berfungsi melindungi keaslian gambar dengan menyisipkan informasi yang dapat mengidentifikasi sumber atau pemilik gambar, sehingga memungkinkan deteksi terhadap perubahan atau penggunaan yang tidak sah (Begum & Uddin, 2020). Bose dan Sabramanyam (Bose et al., 2014),(Subramanyam et al., 2012) telah mengeksplorasi dampak kompresi terhadap efektivitas *watermarking*. salah satu metode watermarking adalah menggunakan metode LSB (*Least Significant Bit*), metode LSB mampu menghasilkan kualitas citra yang baik dengan rata-rata PSNR mencapai 65 dB, serta memungkinkan penyisipan watermark secara tak terlihat tanpa mengubah citra asli secara signifikan (Chopra, 2012). Studi lebih lanjut oleh Parthasarathy. (Parthasarathy & Nagar, 2009),(Wan et al., n.d.) menyoroti potensi peningkatan ketahanan teknik LSB terhadap gangguan eksternal seperti kompresi dan manipulasi, sehingga dapat meningkatkan keandalannya untuk aplikasi keamanan data digital. Perkembangan teknologi ini mengindikasikan bahwa solusi integratif antara pendeteksian deepfake dan *watermarking* digital merupakan arah masa depan yang menjanjikan, terutama dalam menghadapi tantangan manipulasi media yang semakin kompleks.

Selain itu, penelitian ini juga mengintegrasikan digital watermarking sebagai mekanisme validasi hasil deteksi, sehingga informasi keaslian citra dapat disisipkan langsung ke dalam gambar. Evaluasi efektivitas pendekatan ini dilakukan dengan mengukur akurasi deteksi deepfake menggunakan CNN serta kualitas watermark yang

disisipkan menggunakan metrik seperti PSNR (Setiawati et al., 2023). Dengan pendekatan ini, diharapkan sistem yang dikembangkan dapat meningkatkan keamanan dan kepercayaan terhadap informasi digital, sekaligus menyediakan mekanisme validasi otomatis yang memastikan hasil deteksi deepfake tidak dapat dimanipulasi dengan mudah

METODE PENELITIAN

Dataset yang digunakan dalam penelitian ini terdiri dari dua jenis citra, yaitu citra asli (*original images*) yang tidak mengalami modifikasi atau manipulasi, serta citra *deepfake* (*deepfake images*) yang telah dimanipulasi menggunakan algoritma deepfake dengan tanda-tanda modifikasi yang halus dan sulit dikenali. Dataset ini diperoleh dari sumber terbuka (*opensource*) Kaggle, yang mencakup face forensic images dan deepfake images. Seluruh dataset dibagi menjadi tiga subset, yaitu: Data pelatihan (*training* data), Data validasi (*validation* data) dan, Data pengujian (*testing* data). Pembagian ini dilakukan untuk memastikan proses pengembangan model yang terstruktur dan komprehensif. Dataset disimpan dalam struktur folder terpisah dan diunggah ke Google Drive guna memudahkan integrasi dengan Google Colab dalam proses pelatihan serta evaluasi model. Pembagian dataset dilakukan dengan proporsi yang seimbang untuk memastikan akurasi dan generalisasi model.

Tabel 1.
Pembagian Dataset

No.	Dataset	Deepake Image	Real Image
1.	Dataset Pelatihan	8000	8000
2.	Dataset validasi	2000	2000
3.	Dataset Pengujian	1191	1191
	Total	11191	11191

Preprocessing Dataset

Tahapan ini mencakup beberapa langkah penting dalam pengolahan citra sebelum digunakan untuk pelatihan model:

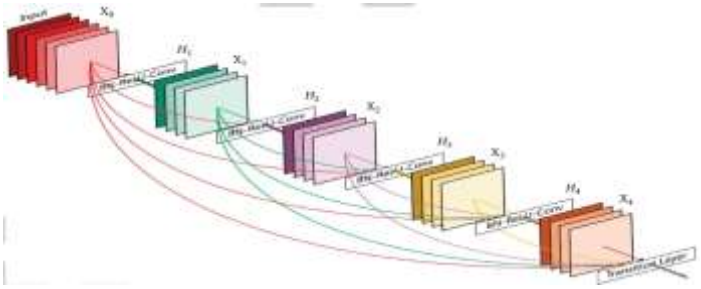
1) *Resize* : Menyesuaikan ukuran gambar menjadi 128x128 piksel agar lebih optimal dalam proses pelatihan dan mengurangi beban komputasi.
2) Konversi ke *Grayscale* : Setiap gambar dikonversi menjadi skala abu-abu (grayscale) untuk mengurangi dimensi fitur dan meningkatkan efisiensi pemrosesan.
3) Normalisasi : Nilai piksel dipetakan ke dalam rentang 0-1 dengan membagi setiap piksel dengan 255 (rescale=1./255) guna mempercepat konvergensi model.
4) Augmentasi Data (Data Augmentation) : Dilakukan hanya pada data pelatihan untuk menambah variasi data secara virtual melalui transformasi berikut:
    - Rotasi hingga 20°
    - Perubahan ukuran lebar & tinggi ±20%
    - Shear transformation hingga 20%
    - Zoom in/out hingga 20%
    - Flipping horizontal
    - Variasi kecerahan (0.8 – 1.2)
    - Perubahan intensitas channel hingga 10 unit

Gambar yang telah diproses disimpan dalam bentuk array NumPy, kemudian diberikan dimensi tambahan untuk menyesuaikan format input model (1 channel untuk grayscale). Dataset dibagi menjadi data pelatihan, validasi, dan pengujian/testing, dengan proses normalisasi yang sama untuk memastikan distribusi data tetap konsisten.

Gambar 1. *Flowchart* Preprocessing *data*

Pelatihan Model

1) Model CNN DenseNet-121
   Pada penelitain ini, model CNN Dengan arsitektur DenseNet-121 digunakan untuk mengklasifikasikan apakah gambar asli atau gambar telah dimanipulasi.Dense Convolutional Network (DenseNet) adalah arsitektur jaringan saraf konvolusional yang menghubungkan setiap lapisan atau blok secara langsung dengan semua lapisan atau blok sebelumnya melalui koneksi umpan maju. Berbeda dengan jaringan konvolusional tradisional yang memiliki koneksi antara lapisan yang berurutan, DenseNet memiliki koneksi langsung antara setiap lapisan, memungkinkan informasi dari lapisan sebelumnya digunakan sebagai input untuk lapisan berikutnya (License & Rizvee, 2023). DenseNet menggunakan tiga blok utama yang masing-masing terdiri dari lapisan-lapisan dengan batch normalization, fungsi aktivasi ReLU, dan konvolusi dengan filter 3×3. Lapisan yang menghubungkan dua blok disebut lapisan transisi, yang berfungsi untuk mengubah ukuran fitur melalui proses konvolusi dan pooling.



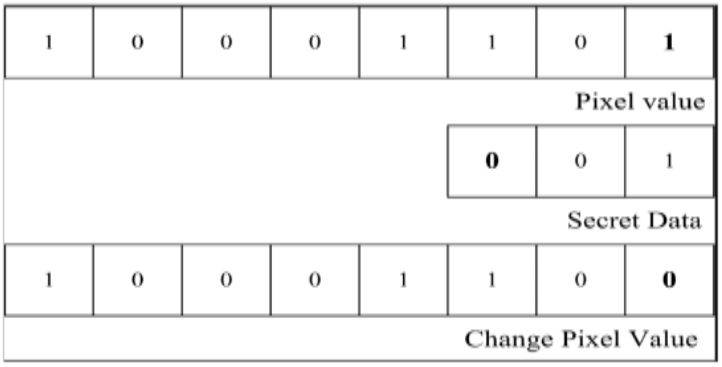
Gambar 2. Gambar Arsitektur DenseNet121 (License & Rizvee, 2023)

Pada tahap trainning, model ditrainning menggunakan dataset yang telah melewati tahap preprocessing untuk memastikan kualitas data yang optimal dalam proses pembelajaran. Model dikompilasi dengan menggunakan optimizer Adam, serta loss function categorical crossentropy untuk menangani klasifikasi multi-kelas, dengan metrik evaluasi berupa akurasi. Proses training dilakukan dengan menentukan jumlah epoch dan batch size yang sesuai agar model dapat belajar secara optimal tanpa mengalami overfitting. Setelah tahap pelatihan selesai, model dievaluasi menggunakan data validasi untuk mengukur performanya dan memastikan bahwa model bisa optimal dalam mengeneralisasi data yang belum pernah dilihat sebelumnya.

2) *Watermarking Least Significant Bit* (LSB)

Teknik *Least Significant Bit* (LSB) adalah salah satu teknik *watermarking* pada domain spasial yang paling sederhana dan populer. Teknik ini bekerja dengan memanfaatkan bit paling kecil/tidak signifikan (*Least Significant Bit*) dari piksel citra digital. Dalam metode ini, watermark disisipkan dengan mengganti LSB piksel citra host dengan data watermark (Faheem et al., 2022).

Dalam konteks representasi biner dari sebuah data, bit yang besar/paling signifikan (Most Significant Bit atau MSB) adalah bit yang terletak paling kiri dalam suatu angka biner dan memiliki nilai paling besar. Sebaliknya, bit yang paling tidak signifikan (*Least Significant Bit* atau LSB) adalah bit yang terletak paling kanan dan memiliki kontribusi nilai terkecil terhadap nilai keseluruhan data. Sebagai contoh, jika sebuah byte (8 bit) memiliki nilai biner 10110011, bit pertama (1) adalah MSB dan bit terakhir (1) adalah LSB.



Gambar 3. contoh 1 bit LSB

Evaluasi

A. Evaluasi Model CNN

Mengukur kinerja model dalam mendeteksi citra deepfake pada dataset yang belum pernah dilihat sebelumnya (*validation dan testing set*). Validation set digunakan selama pelatihan untuk memantau performa model. Testing set digunakan untuk mengevaluasi model setelah pelatihan selesai dengan prosedur model dilatih menggunakan dataset *training*. Dataset *validation* digunakan untuk memilih hyperparameter terbaik, seperti *learning rate* dan jumlah epoch. Dataset *testing* digunakan untuk mengevaluasi performa akhir model.

a) Metrik Evaluasi

Model dievaluasi menggunakan beberapa matrik evaluasi yakni *Accuracy* (Akurasi), *precision* (presisi), *recall* dan *F1 score* dengan rumus seperti berikut :

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

Akurasi digunakan untuk menunjukkan persentase prediksi yang benar dari seluruh prediksi

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

Presisi digunakan untuk mengukur seberapa baik model menghindari kesalahan dalam mendeteksi citra asli sebagai deepfake. Semakin tinggi nilai presisi, semakin sedikit jumlah citra asli yang secara keliru diklasifikasikan sebagai deepfake, sehingga mengurangi false positives.

Hal ini menunjukkan bahwa model tersebut secara konsisten menghasilkan deteksi yang benar, yang merupakan hal penting dalam menjaga keutuhan dan kepercayaan dalam sistem verifikasi gambar digital.

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

Recall digunakan untuk mengukur kemampuan model untuk mendeteksi semua citra deepfake dengan benar. Semakin tinggi nilai recall, semakin sedikit citra deepfake yang tidak terdeteksi (false negatives), sehingga memastikan bahwa hampir semua citra yang dimanipulasi berhasil diidentifikasi oleh sistem.

$$F1\ SCORE = 2 * \frac{Precision * Recall}{Precision + Recall}$$

F1-Score berfungsi untuk menyeimbangkan Precision dan Recall dalam evaluasi model klasifikasi, terutama saat terdapat ketidakseimbangan kelas. Metrik ini membantu mengukur seberapa baik model dalam mendeteksi deepfake tanpa terlalu banyak kesalahan False Positives atau False Negatives.

B. Evaluasi Watermark

Mengukur kualitas visual gambar setelah watermark disisipkan serta ketahanan watermark terhadap manipulasi. Gambar asli tanpa watermark dibandingkan dengan gambar yang telah diberi watermark. Untuk mengevaluasi watermark, seberapa baik watermark disisipkan kedalam gambar dan tidak mengurangi kualitas dari gambar maka PSNR digunakan untuk mengukur kualitas gambar setelah watermark disisipkan. Nilai PSNR yang tinggi menunjukkan bahwa watermark tidak memengaruhi kualitas visual gambar. Rumus PSNR :

$$PSNR = 10 * \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}}\right)$$

HASIL PENELITIAN DAN PEMBAHASAN

Hasil Preprocessing Dataset

Pada tahap ini dataset dari data data yang sudah dikumpuilkan akan melewati beberapa langkah penting sebelum data digunakan untuk melatih model yakni resize,konversi ke grayscale,augmentasi dan normalisasi.

Gambar 4. Hasil preprocessing

Gambar diatas menunjukkan tahapan pemrosesan citra yang dilakukan sebelum digunakan untuk melatih model. Gambar pertama adalah gambar asli dalam format berwarna (RGB). Ini adalah citra yang diambil langsung dari dataset sebelum

mengalami perubahan atau pemrosesan. Gambar kedua adalah hasil augmentasi data, di mana berbagai transformasi seperti rotasi, pergeseran, shear, zoom, dan flipping diterapkan untuk meningkatkan variasi data. Augmentasi ini bertujuan untuk membuat model lebih robust terhadap variasi input, sehingga tidak hanya mengandalkan pola statis dalam mendeteksi deepfake. Gambar ketiga menunjukkan hasil konversi ke grayscale, di mana citra berwarna diubah menjadi hitam-putih. Konversi ini menghilangkan informasi warna dan hanya mempertahankan intensitas cahaya dari setiap piksel. Dengan cara ini, model lebih fokus pada pola tekstur dan bentuk tanpa terganggu oleh variasi warna. Gambar keempat adalah hasil normalisasi dari gambar grayscale. Proses normalisasi dilakukan dengan membagi nilai piksel dengan 255, sehingga semua nilai berada dalam rentang [0,1]. Ini membantu model dalam konvergensi lebih cepat saat pelatihan karena menghindari skala nilai yang terlalu besar atau kecil. Urutan pemrosesan ini memastikan bahwa gambar yang dimasukkan ke dalam model sudah dalam format yang optimal, dengan menghilangkan informasi yang tidak relevan dan meningkatkan variasi data agar model dapat mendeteksi deepfake dengan lebih akurat.

Hasil Pelatihan Model

A. Model Deepfake

Pada proses ini model dilatih menggunakan dataset yang sudah melewati tahap preprocessing dimana dari hasil training model kita uji coboa untuk melakukan prediksi dan didapatkan hasil seperti berikut

Gambar 5 Gambar hasil prediksi

Gambar yang ditampilkan merupakan hasil prediksi model deteksi deepfake terhadap lima gambar wajah dengan menggunakan teknik pemrosesan gambar, di mana setiap gambar memiliki label prediksi berupa "Fake" atau "Real" di bagian atasnya yang menunjukkan apakah gambar tersebut diklasifikasikan sebagai deepfake atau asli oleh model. Warna hijau kebiruan dengan pola khas menunjukkan bahwa gambar telah melalui teknik pemrosesan

B. Watermarking LSB

Hasil dari proses *watermarking* LSB menunjukkan bahwa watermark dapat disisipkan dan diekstraksi dengan tingkat keberhasilan tinggi tanpa mengubah kualitas visual gambar secara signifikan. Gambar di bawah ini menampilkan perbandingan antara gambar asli dan gambar yang sudah disisipi watermark

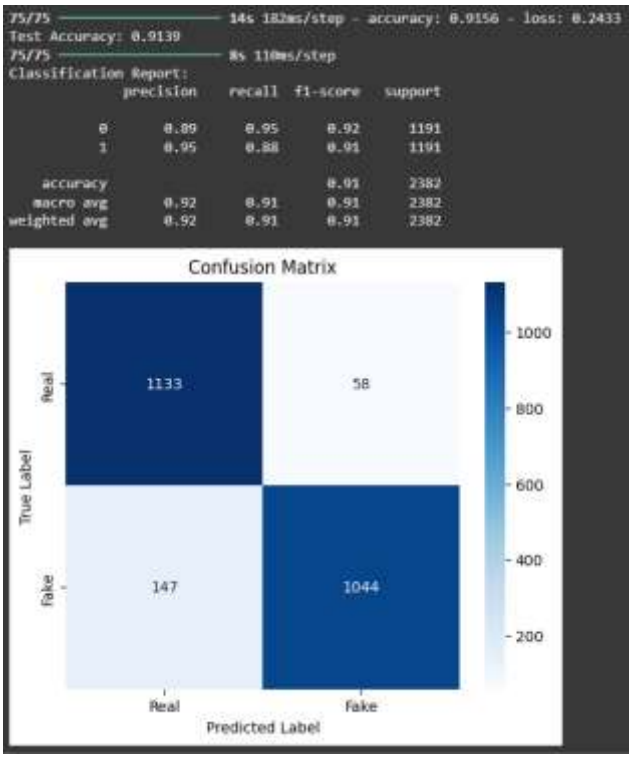
Gambar 6 . Gambar asli dan *watermark*

Gambar yang ditampilkan menunjukkan hasil dari proses *watermarking* digital pada sebuah citra. Pada bagian atas, terdapat tiga gambar yang merepresentasikan tahapan dalam proses ini. Gambar pertama adalah gambar asli sebelum diberikan watermark, yang berfungsi sebagai referensi awal. Gambar kedua adalah gambar yang telah disisipkan watermark tersembunyi. Secara visual, gambar ini tampak hampir identik dengan gambar asli, menandakan bahwa watermark disisipkan dengan metode yang tidak mengganggu tampilan utama. Gambar ketiga menunjukkan hasil ekstraksi watermark dari gambar yang telah dimodifikasi. Teks "Watermark Tersembunyi" mengindikasikan bahwa watermark berhasil disisipkan dan dapat diambil kembali dari gambar watermarked.

Hasil Evaluasi

A. Evaluasi Model CNN

Model CNN DenseNet121 diterapkan untuk mendeteksi apakah sebuah gambar merupakan deepfake atau asli. Model ini diuji menggunakan dataset yang terdiri dari 2.382 gambar, dengan komposisi 50% gambar asli dan 50% gambar deepfake.



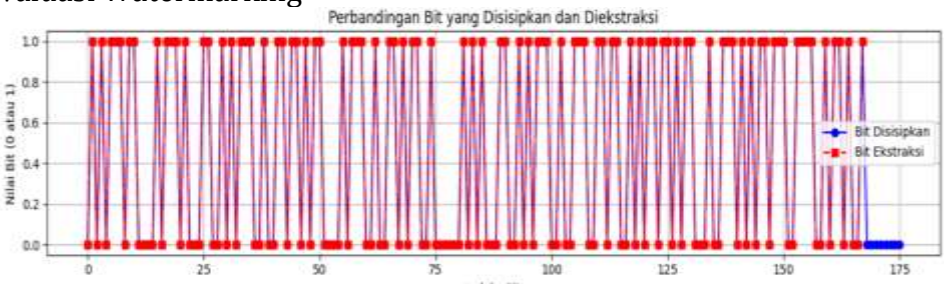
Gambar 7. Hasil *accuracy,precission recall* dan *f1-score*

Gambar 7 menunjukkan ringkasan metrik evaluasi model. Hasil pelatihan model menunjukkan akurasi sebesar 91.56% dengan loss 0.2433 pada dataset uji. Dari hasil ini, dapat disimpulkan bahwa model mampu membedakan deepfake dengan baik.

Tabel 2.
Tabel evaluasi

Matrix	Real	Fake
Akurasi	$\frac{1044+1133}{1044+1133+58+147} = \frac{2177}{2382} = 0.91$	
Presisi	$\frac{1133}{1133+58} = \frac{1133}{1191} = 0.89$	$\frac{1044}{1044+147} = \frac{1044}{1191} = 0.95$
Recall	$\frac{1133}{1133+147} = \frac{1133}{1280} = 0.95$	$\frac{1044}{1044+58} = \frac{1044}{1102} = 0.88$
F1-Score	$2 * \frac{0.89 * 0.95}{0.89+0.95} =$ $2 * \frac{0.8455}{1.84} = 0.92$	$2 * \frac{0.95 * 0.88}{0.95+0.88} =$ $2 * \frac{0.836}{1.83} = 0.91$

B. Evaluasi Watermarking



Gambar 8 . Gambar asli dan *watermark*

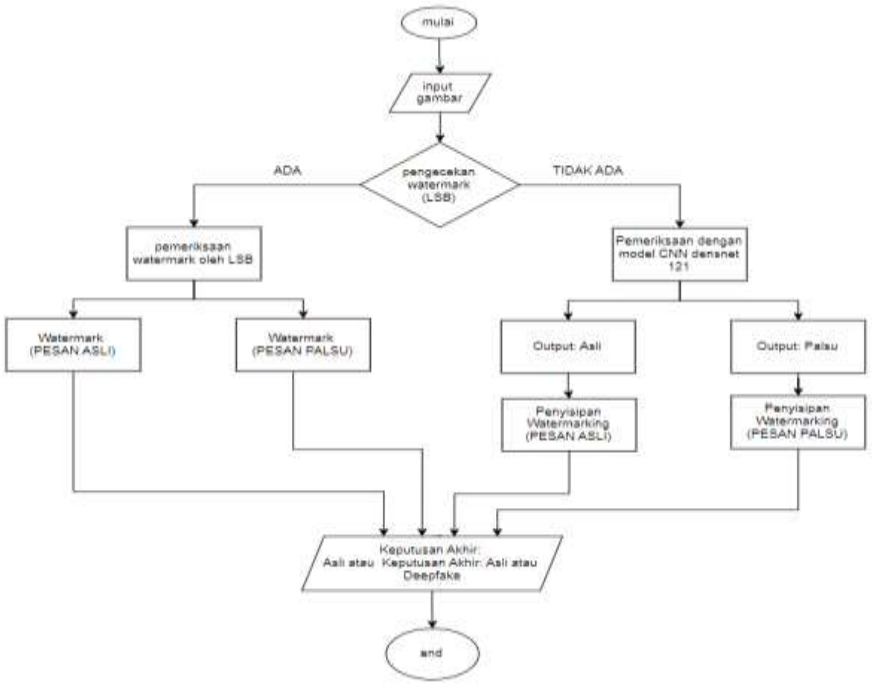
Di bagian bawah, terdapat grafik yang membandingkan bit-bit watermark yang disisipkan dengan bit-bit yang berhasil diekstraksi. Grafik ini memiliki sumbu X yang menunjukkan indeks bit dalam watermark dan sumbu Y yang merepresentasikan nilai biner dari masing-masing bit (0 atau 1). Dua jenis data ditampilkan dalam grafik, yaitu bit yang disisipkan (ditandai dengan garis biru dan titik bulat) serta bit yang diekstraksi (ditandai dengan garis merah dan kotak merah). Dari pola grafik, terlihat bahwa sebagian besar bit yang diekstraksi sesuai dengan bit yang disisipkan, menunjukkan bahwa proses watermarking berhasil.

Secara keseluruhan, hasil ini menunjukkan bahwa metode watermarking yang digunakan cukup efektif dalam menyisipkan informasi tanpa mengubah tampilan gambar secara signifikan. Selain itu, watermark dapat diekstraksi kembali dengan tingkat keberhasilan yang tinggi. Sedikit perbedaan dalam hasil

ekstraksi bisa menjadi indikasi adanya faktor eksternal yang mempengaruhi proses pengambilan watermark, tetapi secara umum, metode ini bekerja dengan baik dalam menjaga keutuhan informasi yang disisipkan.

Implementasi Model Deteksi Deepfake dan *Watermarking*

Pada tahap ini, model yang telah dilatih akan dikombinasikan dengan teknik *watermarking Least Significant Bit* (LSB). Metode ini digunakan untuk menyisipkan informasi ke dalam citra dengan cara menggantikan bit paling tidak signifikan, sehingga tidak mengubah struktur utama gambar secara signifikan. Integrasi model deep learning dengan *watermarking* LSB bertujuan untuk meningkatkan keamanan dan keandalan sistem dalam mendeteksi citra deepfake

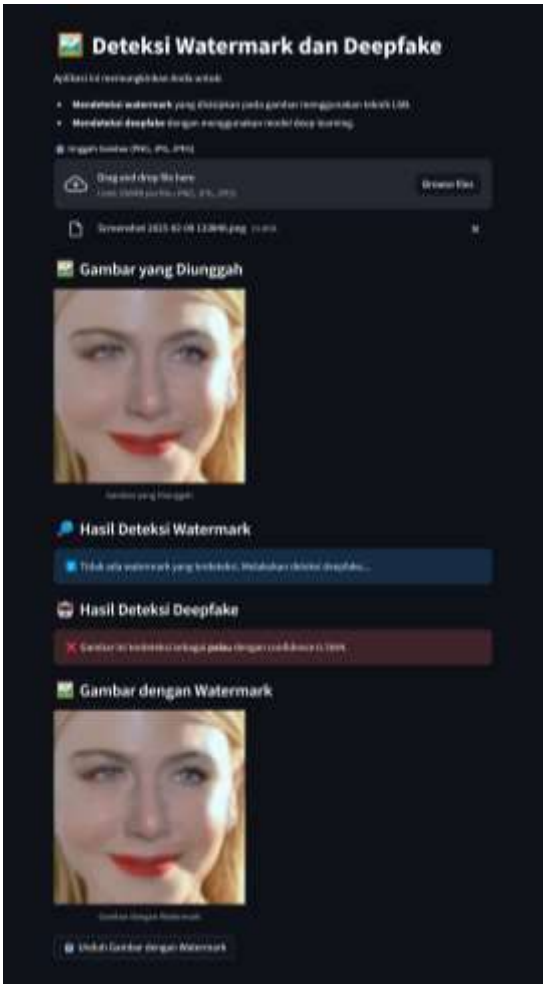


Gambar 9. Flowchart Alur Sistem

Dari alur system sesuai flowchart tersebut maka akan di implementasikan dalam bentuk tampilan website yang nantinya user bisa mengecek gambar apakah gambar yang diperiksa termasuk kedalam deepfake atau asli,berikut adalah tampilan halaman utama dari system tersebut

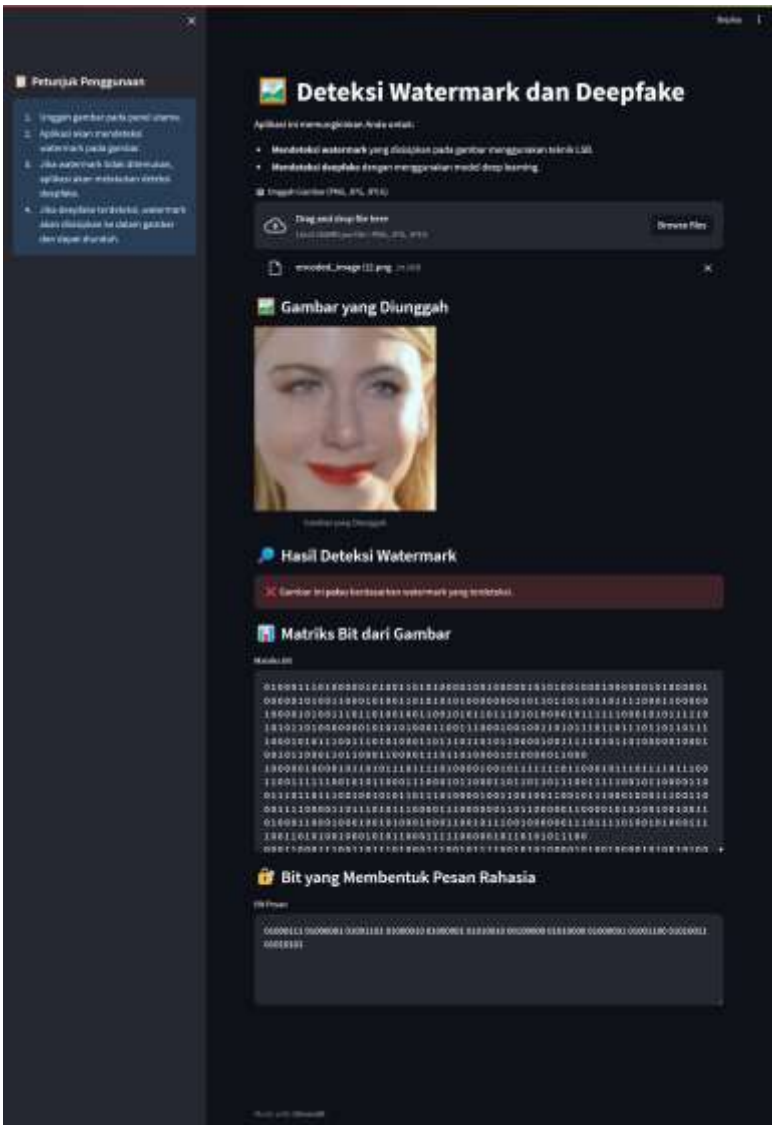
Gambar 10. Tampilan website

Dari tampilan tersebut user hanya perlu memasukan gambar dalam format PNG,JPG,JPEG dll, yang nantinya sistem akan memeriksa watermark didalam gambar jika gambar yang dimasukkan tidak memiliki watermark maka akan dilanjutkan pada proses deteksi gambar oleh model CNN densenet-121.



Gambar 11 . Hasil deteksi dan pengecekan watermark

Ketika hasil deteksi keluar maka user akan diberi pilihan untuk mendownload gambar yang sudah diberikan watermark oleh system untuk validasi bahwa gambar yang dimasukkan sudah dicek keasliannya dan nantinya Ketika user memasukkan lagi gambar yang sudah didownload maka system hanya perlu memberikan hasil berdasarkan watermark yang sudah disisipi sebelumnya sehingga disini fungsi watermark akan mempercepat proses deteksi selain menjadi validasi bahwa gambar asli atau palsu.



Gambar 12. Hasil deteksi gambar

Dari gambar diatas didapatkan bit yang membentuk pesan rahasia yang disisipkan oleh sistem sebelumnya yang mana jika di konversi maka akan membentuk pesan “GAMBAR ASLI” atau “GAMBAR PALSU”

KESIMPULAN

Berdasarkan implementasi aplikasi ini, dapat disimpulkan bahwa integrasi antara model deepfake detection berbasis CNN DenseNet121 dan teknik *watermarking* LSB berhasil membentuk sistem yang efektif untuk memverifikasi keaslian gambar digital. Sistem yang dikembangkan mampu menampilkan gambar asli, menyisipkan watermark secara imperceptible ke dalam gambar, serta mengekstraksi kembali watermark tersebut dengan akurasi tinggi, sehingga memberikan jaminan integritas dan perlindungan terhadap manipulasi konten. Selain itu, pada kondisi ketika watermark tidak terdeteksi, model deepfake memberikan prediksi yang andal, sehingga memungkinkan sistem untuk secara otomatis mengklasifikasikan gambar sebagai asli atau palsu dan melakukan penyisipan watermark yang sesuai. Keseluruhan, pendekatan integratif ini menunjukkan potensi besar dalam meningkatkan keamanan dan kepercayaan terhadap data digital, serta membuka jalan bagi pengembangan sistem

verifikasi otomatis yang dapat mengatasi tantangan manipulasi media yang semakin kompleks

DAFTAR PUSTAKA

A C, P., & M T, S. (2023). An Overview on Research Trends, Challenges, Applications and Future Direction in Digital Image Watermarking. *International Research Journal on Advanced Science Hub*, *5*(01), 8–14. https://doi.org/10.47392/irjash.2023.002

Almars, A. M. (2021). Deepfakes detection techniques using deep learning: a survey. *Journal of Computer and Communications*, *9*(05), 20–35.

Alzahrani, A. (2024). Digital Image Forensics: An Improved DenseNet Architecture for Forged Image Detection. *Engineering, Technology and Applied Science Research*, *14*(2), 13671–13680. https://doi.org/10.48084/etasr.7029

Begum, M., & Uddin, M. S. (2020). Digital image watermarking techniques: A review. *Information (Switzerland)*, *11*(2). https://doi.org/10.3390/info11020110

Boato, G., Conotter, V., De Natale, F. G. B., & Fontanari, C. (2009). Watermarking robustness evaluation based on perceptual quality via genetic algorithms. *IEEE Transactions on Information Forensics and Security*, *4*(2), 207–216. https://doi.org/10.1109/TIFS.2009.2020362

Bose, S., Madhulika, Acharjee, S., Chowdhury, S. R., Chakraborty, S., & Dey, N. (2014). Effect of watermarking in vector quantization based image compression. *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies, ICCICCT 2014*, *December*, 503–508. https://doi.org/10.1109/ICCICCT.2014.6993014

Chopra, D. (2012). Lsb Based Digital Image Watermarking For Gray Scale Image. *IOSR Journal of Computer Engineering*, *6*(1), 36–41. https://doi.org/10.9790/0661-0613641

Faheem, Z. Bin, Ali, M., Raza, M. A., Arslan, F., Ali, J., Masud, M., & Shorfuzzaman, M. (2022). Image Watermarking Scheme Using LSB and Image Gradient. *Applied Sciences (Switzerland)*, *12*(9), 1–12. https://doi.org/10.3390/app12094202

Karnouskos, S. (2020). Artificial Intelligence in Digital Media: The Era of Deepfakes. *IEEE Transactions on Technology and Society*, *1*(3), 138–147. https://doi.org/10.1109/tts.2020.3001312

Khan, S. A., & Dang-Nguyen, D.-T. (2023). *Deepfake Detection: A Comparative Analysis*. *1*(1), 1–28.

License, C. A., & Rizvee, M. M. (2023). *Retracted : Comparative Analysis of Deepfake Image Detection*. *2021*. https://doi.org/10.1155/2021/3111676

Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection: state-of-the-art, open challenges, countermeasures, and way forward. *Applied Intelligence*, *53*(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

Memon, N., & Wong, P. W. (1998). Protecting Digital Media Content. *Communications of the ACM*, *41*(7), 35–43. https://doi.org/10.1145/278476.278485

Parthasarathy, C., & Nagar, J. (2009). Increased Robustness of Lsb Audio Steganography By Reduced Distortion Lsb. *Audio*.

Patel, Y., Tanwar, S., Bhattacharya, P., Gupta, R., Alsuwian, T., Davidson, I. E., & Mazibuko, T. F. (2023). An Improved Dense CNN Architecture for Deepfake Image Detection. *IEEE Access*, *11*(March), 22081–22095. https://doi.org/10.1109/ACCESS.2023.3251417

Sanivarapu, P. V., Rajesh, K. N. V. P. S., & Hosny, K. M. (2022). *applied sciences Digital*

Watermarking System for Copyright Protection and Authentication of Images Using Cryptographic Techniques.

Setiawati, I., Hermanto, M. T., & Ujianto, E. (2023). Digital Watermarking Implementation Of Digital Watermarking On Images Using The Least Significant Bit Method. *International Journal of Engineering Technology and Natural Sciences*, *5*(1), 10–18. https://doi.org/10.46923/ijets.v5i1.191

Subramanyam, A. V., Emmanuel, S., & Kankanhalli, M. S. (2012). Robust watermarking of compressed and encrypted JPEG2000 images. *IEEE Transactions on Multimedia*, *14*(3 PART 2), 703–716. https://doi.org/10.1109/TMM.2011.2181342

Swaminathan, A., Mao, Y., & Wu, M. (2006). Robust and secure image hashing. *IEEE Transactions on Information Forensics and Security*, *1*(2), 215–230. https://doi.org/10.1109/TIFS.2006.873601

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A Survey of face manipulation and fake detection. *Information Fusion*, *64*, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014

Wan, W., Wang, J., Zhang, Y., Li, J., Yu, H., & Sun, J. (n.d.). *A Comprehensive Survey on Robust Image Watermarking A R T I C L E I N F O*. 1–26.