Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 464-473

KLASIFIKASI *PAYLOAD MALWARE* TERSEMBUNYI PADA CITRA DIGITAL (STEGANALISIS) MENGGUNAKAN ALGORITMA *CONVOLUTIONAL NEURAL NETWORK* (CNN)

Muh. Dirham Rahim[1*], Muhyiddin A.M Hayat[2], Rizki Yusliana Bakti[3], Darniati[4]
[1,2,3,4]Universitas Muhammadiyah Makassar, Indonesia
[1]105841105622@student.unismuh.ac.id , [2]muhyiddin@unismuh.ac.id,
[3]rizkiyusliana@unismuh.ac.id , [4]darniati@unismuh.ac.id
*Coresponding Author

Info Artikel

Pengajuan : 05-08-2026
Peninjauan : 15-08-2026
Revisi : 30-08-2026
Diterima : 07-09-2026
Diterbitkan : 20-09-2026

Kata Kunci:
Steganalisis, Steganografi JPEG, Discrete Cosine Transform (DCT), Convolutional Neural Network 1D, Multi-class Payload Classification.

Abstrak

Evolusi ancaman keamanan siber saat ini semakin mengarah pada teknik penetrasi senyap, salah satunya pemanfaatan steganografi menggunakan aplikasi Steghide pada citra JPEG. Sistem steganalisis konvensional saat ini masih bertumpu pada klasifikasi biner, sehingga menimbulkan titik buta forensik yang gagal membedakan tingkat keparahan anomali, seperti antara sisipan teks biasa dan skrip malware destruktif. Penelitian ini mengusulkan model 1-Dimensional Convolutional Neural Network (CNN 1D) untuk mengklasifikasikan tingkat ancaman steganografi ke dalam tiga kelas: Normal, Anomali Ringan (teks 20 KB sebagai representasi eksfiltrasi data minor), dan Anomali Berat (malware berekstensi .bat 40 KB sebagai representasi ancaman skrip destruktif). Ambang batas payload spesifik ini dipilih secara empiris karena mampu merepresentasikan margin perubahan densitas koefisien DCT yang signifikan untuk menguji sensitivitas model. Untuk mengatasi overfitting dan kebutaan fitur spasial, ekstraksi fitur dilakukan pada domain frekuensi menggunakan koefisien Discrete Cosine Transform (DCT) yang menghasilkan 2.466 fitur statistik. Pengujian dilakukan menggunakan skenario Paired Dataset dan Grouped Split pada 1.500 citra beresolusi 1024x1024 piksel untuk mencegah bias visual. Hasil penelitian menunjukkan bahwa model CNN 1D yang diusulkan sukses mencatatkan tingkat akurasi pengujian sebesar 83,11% dengan rata-rata F1-Score 82,98%. Dari aspek keamanan siber preventif, model menunjukkan keandalan yang tinggi. Meskipun terdapat tumpang tindih (overlap) prediksi antara kelas Malware dan kelas Teks yang menyebabkan nilai Recall kelas Anomali Berat berada di angka 76,00%, sistem mencatatkan ketiadaan False Negative (0 kejadian) pada persilangan menuju kelas Normal. Hal ini mengindikasikan bahwa pada pengujian ini, tidak ada satupun sampel citra bermuatan skrip destruktif yang lolos dan terklasifikasi sebagai citra aman.

How to Cite : Rahim, M. D., Hayat, M. A. M., Bakti, R. Y., & Darniati, D. (2026). Klasifikasi *payload malware* tersembunyi pada citra digital (steganalisis) menggunakan algoritma *convolutional neural network* (CNN). *Journal of Computer Science and Information Technology (JCSIT), 3*(4), 464–474.

DOI : 10.70248/jcsit.v3i4.4400

PENDAHULUAN

Perkembangan pesat teknologi informasi dan komunikasi di era digital telah memicu transformasi lanskap keamanan siber (cyber security) secara signifikan. Ancaman siber modern tidak lagi didominasi oleh metode serangan konvensional yang agresif dan destruktif secara langsung, melainkan telah berevolusi menuju teknik infiltrasi yang sangat senyap dan persisten. Salah satu metode penyamaran data tingkat tinggi yang paling menantang bagi analis forensik digital saat ini adalah steganografi.

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 464-473

Pada dasarnya, steganografi memiliki sifat penggunaan ganda (dual-use nature); di satu sisi bermanfaat untuk perlindungan hak cipta (watermarking) dan komunikasi militer rahasia, namun di sisi lain semakin masif dieksploitasi oleh aktor kriminal siber untuk mendistribusikan data ilegal dan payload serangan(Bhardwaj & Sharma, 2016). Fenomena integrasi antara steganografi dan penyebaran perangkat perusak ini melahirkan sebuah tren ancaman baru yang dikenal dengan istilah stegware, di mana malware diselundupkan di dalam media pembawa untuk menghindari deteksi perimeter keamanan (Jelodar et al., 2026).

Dalam praktiknya, para penyerang sering kali memanfaatkan format citra JPEG sebagai media pembawa (cover image) karena popularitasnya yang masif dalam lalu lintas data internet, dipadukan dengan algoritma penyisipan canggih seperti Steghide(Liu et al., 2023). Berbeda dengan metode steganografi tradisional yang menyisipkan data secara langsung pada domain spasial, Steghide beroperasi dengan memanipulasi bit-bit pada koefisien Discrete Cosine Transform (DCT) di domain frekuensi. Teknik ini dinilai sangat adaptif karena mampu mengeksploitasi karakteristik kompresi lossy dan membaurkan jejak distorsi biner ke dalam deretan noise bawaan gambar (G.D.Sanjaya et al., 2018). Dampaknya, citra yang telah terinfeksi tidak mengalami degradasi visual secara kasat mata, memungkinkannya dengan mudah mem-bypass sistem pertahanan konvensional, seperti Intrusion Detection System (IDS), firewall, dan perangkat lunak antivirus standar yang hanya bergantung pada pencocokan tanda tangan (signature-based detection).

Ketika berhadapan dengan tahap investigasi dan mitigasi, sistem steganalisis forensik digital saat ini sering kali mengalami kebuntuan operasional akibat penggunaan arsitektur klasifikasi biner(Hemamalini, 2025). Pendekatan tradisional ini terbatas pada identifikasi dua kondisi absolut saja, yakni kondisi bersih (normal) atau kondisi terinfeksi (tersisipi data). Keterbatasan ini memunculkan titik buta (blind spot) yang sangat fatal karena sistem kehilangan kapabilitas untuk membedakan dan mengukur secara instan derajat keparahan anomali struktural pada citra yang terdeteksi. Akibatnya, analis keamanan kesulitan menentukan apakah anomali tersebut sekadar dipicu oleh sisipan dokumen teks pasif yang tidak mengancam stabilitas sistem, atau merupakan enkapsulasi skrip malware destruktif berskala besar yang siap tereksekusi(Shehab & Alhaddad, 2025). Ketidakmampuan membedakan skala ancaman ini pada akhirnya memicu lonjakan angka alarm palsu (false alarm), menghamburkan beban dan sumber daya komputasi, serta mendegradasi waktu respons tim penanganan insiden siber secara drastis dalam menetralisir ancaman nyata.

Berbagai penelitian state-of-the-art telah berupaya mengatasi tantangan deteksi ini dengan beragam inovasi. Sebagai contoh, penelitian oleh Veerakumar et al. memperkenalkan pendekatan fusi fitur domain ganda spasial dan frekuensi untuk steganalysis(Veerakumar et al., 2026). sementara Hong et al, mengoptimalkan efisiensi model deteksi Deep Learning melalui teknik pemangkasan parameter secara selektif(Hong et al., 2023). Penelitian lain oleh De La Croix Ntivuguruzwa & Ahmad juga menerapkan filter Spatial Rich Model (SRM) yang diintegrasikan ke dalam arsitektur CNN(De La Croix Ntivuguruzwa & Ahmad, 2023). Meskipun berbagai peningkatan arsitektur telah diusulkan, mayoritas fokus penelitian tersebut masih terperangkap dalam ranah deteksi biner dan mengabaikan urgensi identifikasi level ancaman. Pendekatan unimodal dan biner tersebut dinilai tidak lagi memadai untuk menjawab kebutuhan praktis tim analis forensik yang tidak hanya bertugas mendeteksi keberadaan muatan, tetapi juga dituntut untuk langsung mengukur tingkat keparahan

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 464-473

ancaman siber tersebut secara tepat sasaran.

Untuk memecahkan kesenjangan metodologis tersebut, transisi menuju sistem deteksi berbasis klasifikasi multi-kelas (multi-class classification) merupakan langkah solutif yang mutlak. Penelitian ini mengusulkan penerapan kecerdasan buatan berbasis Computer Vision dengan memanfaatkan ketangguhan algoritma 1-Dimensional Convolutional Neural Network (CNN 1D)(Purwono et al., 2022). Mengingat jejak steganografi pada format JPEG bersembunyi di domain frekuensi, ekstraksi fitur secara spasial rentan mengalami "kebutaan fitur" akibat kompleksitas noise gambar. Oleh karena itu, langkah solutif yang mutlak diambil adalah menggeser fokus analisis ke domain frekuensi. Ekstraksi fitur dilakukan dengan mengubah matriks koefisien Discrete Cosine Transform (DCT) dari citra JPEG menjadi representasi fitur statistik tingkat tinggi. Kombinasi antara fitur Histogram Terkalibrasi (yang mengukur pergeseran probabilitas nilai koefisien individual) dan matriks probabilitas transisi Markov (yang mengukur dependensi antar-koefisien yang bertetangga) menghasilkan sebuah vektor 1-Dimensi yang padat berisi 2.466 elemen statistik. Dimensi spesifik ini diusulkan karena terbukti secara teoretis mampu menangkap rekam jejak distorsi sekecil apapun yang ditinggalkan oleh Steghide. Vektor fitur 1-Dimensi tersebut kemudian dipelajari menggunakan klasifikasi multi-kelas dengan arsitektur 1-Dimensional Convolutional Neural Network (CNN 1D). Pemilihan arsitektur CNN 1D didasarkan pada superioritasnya dalam mengekstrak fitur hierarkis dari data sekuensial secara lokal, dibandingkan metode machine learning konvensional (seperti Support Vector Machine atau Random Forest) yang cenderung kehilangan relasi spasial antar-fitur bertetangga. Selain itu, operasi matriks pada CNN 1D jauh lebih ringan secara komputasi dibandingkan CNN 2D, menjadikannya sangat ideal untuk analisis forensik secara real-time tanpa mengorbankan performa(Zhao et al., 2024).

Lebih jauh, model CNN 1D yang dirancang diinstruksikan untuk mampu menangkap kontras kerusakan mikro yang diakibatkan oleh penyisipan dokumen teks berukuran 20 KB (Anomali Ringan) dibandingkan dengan distorsi masif akibat injeksi skrip simulasi malware berukuran 40 KB (Anomali Berat). Namun demikian, perancangan model Deep Learning dalam bidang klasifikasi visual ini kerap dihadapkan pada masalah fundamental, yakni tingginya risiko overfitting yang dipicu oleh keterbatasan suplai data latih berkualitas serta karakteristik jaringan yang rakus data (data-hungry)(Rustad et al., 2022). Karena teknik rekayasa augmentasi spasial konvensional (seperti rotasi atau cropping) dilarang keras dalam forensik citra akibat kemampuannya yang dapat merusak integritas koefisien kompresi secara permanen, penelitian ini merumuskan penanganan khusus melalui pendekatan Paired Dataset(De La Croix Ntivuguruzwa & Ahmad, 2024). Pendekatan ini diformulasikan untuk memastikan jaringan memusatkan generalisasinya secara eksklusif pada kerusakan biner yang diinjeksi, bukan pada visual gambar.

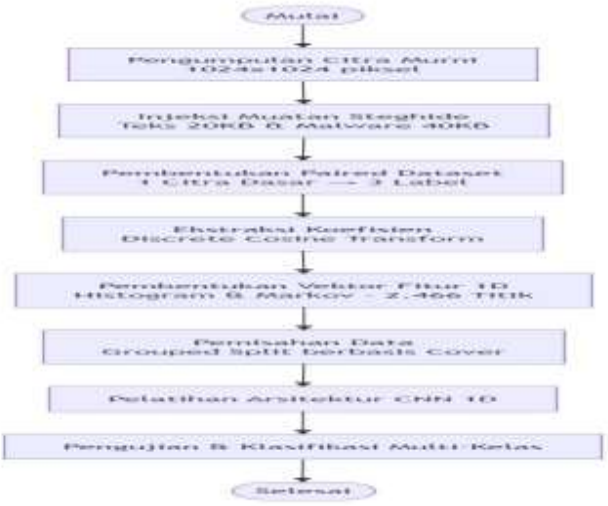
Guna menjamin reliabilitas, memertahankan integritas distribusi noise sensor kamera, serta memastikan orisinalitas riset, penelitian ini menolak penggunaan data sekunder dengan mengkurasi dan membangun dataset primer secara mandiri dalam resolusi tinggi 1024x1024 piksel. Tujuan akhir dari penelitian ini adalah memaksimalisasi tingkat akurasi klasifikasi jaringan ke dalam tiga tingkat signifikansi kategorikal yang terukur. Tiga kelas tersebut meliputi citra normal tanpa modifikasi, citra bersit dengan anomali tidak berbahaya berupa teks biasa, serta citra yang memuat skrip malware berbahaya. Melalui inovasi yang menggabungkan ketangguhan ekstraksi fitur domain frekuensi dan pemodelan klasifikasi multi-kelas, penelitian ini diharapkan

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 464-473

mampu memberikan referensi baru dalam metodologi investigasi insiden siber digital modern dan menghasilkan instrumen forensik yang sangat presisi secara real-time(Agarwal et al., 2023).

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan merancang dan menguji model klasifikasi Deep Learning berbasis arsitektur CNN 1D. Alur pelaksanaan penelitian dimulai dari pengumpulan 500 citra digital murni berformat JPEG secara mandiri yang mencakup beragam variasi visual (seperti lanskap, potret, dan objek bertekstur tinggi) untuk merepresentasikan kondisi noise kompleks di dunia nyata. Citra tersebut diseragamkan (resizing) menjadi resolusi tinggi 1024x1024 piksel guna memastikan ruang matriks koefisien DCT yang cukup luas untuk menampung payload berkapasitas besar secara lossless. Secara keseluruhan, alur perancangan sistem dan metodologi penelitian ini divisualisasikan pada Gambar 1.



Gambar 1. Flowchart Sistem Penelitian

Berdasarkan Gambar 1, sistem menerapkan skenario Paired Dataset, yaitu setiap citra murni diduplikasi menjadi tiga versi: Normal (Kosong), Anomali Ringan yang disisipi teks plaintext 20 KB, dan Anomali Berat yang diinjeksi skrip simulasi malware (.bat) 40 KB menggunakan Steghide. Pendekatan ini memastikan model berfokus pada anomali koefisien DCT, bukan variasi konten visual. Seluruh hasil injeksi divalidasi untuk menjamin integritas data, kemudian 1.500 citra diekstraksi ke domain frekuensi. Dekomposisi matematis dilakukan untuk membentuk dua representasi fitur: pertama, Histogram Terkalibrasi dari 63 koefisien Alternating Current (AC) DCT dengan rentang ambang batas $T = 8$; dan kedua, matriks probabilitas transisi Markov orde pertama yang dihitung dari selisih nilai koefisien bertetangga dengan ambang batas $T_m = 4$. Penggabungan kedua rumusan parameter statistik ini secara presisi menghasilkan matriks vektor 1-Dimensi yang padat berisi 2.466 titik fitur per citra. Dataset selanjutnya dibagi menggunakan Grouped Split menjadi data latih (70%), validasi (15%), dan uji (15%) untuk mencegah data leakage, sedangkan kinerja model dievaluasi menggunakan Confusion Matrix dengan metrik Accuracy, Precision, Recall, dan F1-Score.

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 464-473

HASIL PENELITIAN DAN PEMBAHASAN

Arsitektur dan Pelatihan Model CNN 1D

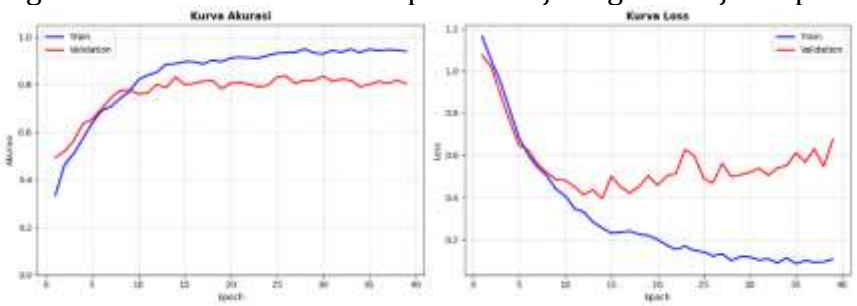
Rincian hierarki lapisan dan jumlah parameter komputasi dari pemodelan 1-Dimensional Convolutional Neural Network (CNN 1D) yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Arsitektur Model CNN 1D

Layer (type)	Output Shape	Param #
conv1d_6 (Conv1D)	(None, 2466, 32)	128
conv1d_7 (Conv1D)	(None, 2466, 32)	3,104
max_pooling1d_4 (*MaxPooling*1D)	(None, 1233, 32)	0
conv1d_8 (Conv1D)	(None, 1233, 64)	6,208
max_pooling1d_5 (*MaxPooling*1D)	(None, 616, 64)	0
flatten_2 (Flatten)	(None, 39424)	0
dense_4 (Dense)	(None, 128)	5,046,400
*Dropout*_2 (*Dropout*)	(None, 128)	0
dense_5 (Dense)	(None, 3)	387

Berdasarkan Tabel 1, vektor fitur berdimensi 2.466 diumpankan ke dalam model CNN 1D. Arsitektur jaringan diawali dengan blok Convolutional Layer (Conv1D) dengan 32 dan 64 filter yang beroperasi menggunakan fungsi aktivasi Rectified Linear Unit (ReLU) untuk mengeliminasi distorsi negatif. Reduksi dimensi spasial dilakukan oleh lapisan MaxPooling1D, menyusutkan fitur hingga menjadi 616 elemen. Matriks ini diratakan (Flatten) menjadi 39.424 elemen menuju Fully Connected Layer (Dense) berukuran 128 neuron, dilengkapi lapisan Dropout untuk regularisasi, sebelum berlabuh pada lapisan penutup 3 neuron beraktivasi Softmax. Secara keseluruhan, arsitektur ini memproses 5.056.227 parameter yang dapat dilatih dengan ukuran memori yang sangat efisien yakni sekitar 19,29 MB.

Pelatihan model dijalankan menggunakan pengoptimal Adam dengan learning rate 0,0003 dan fungsi loss sparse_categorical_crossentropy. Menerapkan mekanisme Early Stopping, model secara dinamis menghentikan pelatihan pada iterasi ke-39. Visualisasi pergerakan metrik selama fase pelatihan jaringan disajikan pada Gambar 2.



Gambar 2. Kurva Akurasi dan Loss pada Pelatihan CNN 1D

Berdasarkan Gambar 2, mekanisme Early Stopping secara dinamis menghentikan pelatihan pada epoch ke-39. Analisis lebih lanjut pada kurva Loss validasi memperlihatkan adanya tren peningkatan nilai loss secara perlahan setelah

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 464-473

epoch ke-15, meskipun kurva akurasi validasi tetap stabil di atas 80%. Fenomena loss-accuracy decoupling ini merupakan indikasi awal terjadinya overfitting, di mana model mulai memberikan penalti kesalahan (cross-entropy loss) yang lebih besar terhadap beberapa sampel misklasifikasi akibat kerumitan kamuflase algoritma Steghide. Menyikapi tren tersebut, penggunaan Early Stopping dengan parameter restore_best_weights terbukti sangat krusial, karena sistem berhasil mengamankan dan merestorasi bobot (weights) paling optimal yang dicapai pada epoch ke-26, tepat sebelum model kehilangan kemampuan generalisasinya secara signifikan.

Pengujian dan Evaluasi Kinerja

Model optimal dievaluasi pada data uji (225 citra) yang diisolasi ketat. Sistem mencatatkan akurasi pengujian (Test Accuracy) sebesar 83,11% dengan Test Loss 0,3614. Rincian kalkulasi parameter matriks kebingungan disajikan pada Tabel 2, sedangkan hasil performa klasifikasi untuk masing-masing kelas target ditunjukkan pada Tabel 3.

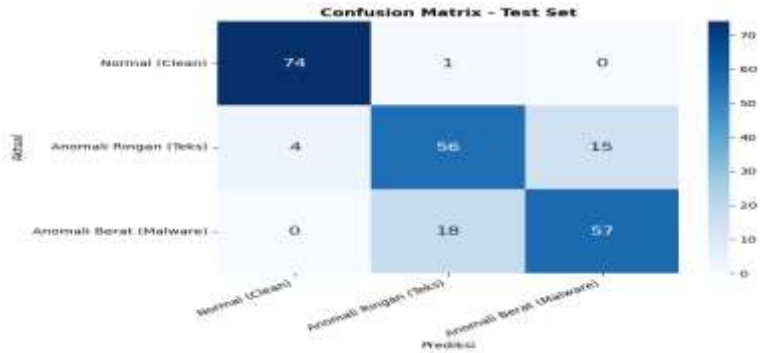
Tabel 2. Kalkulasi Parameter Prediksi pada Data Uji

Kelas	TP	TN	FP	FN	Acc
Normal	74	146	4	1	0.9778
Anomali Ringan	56	131	19	19	0.8311
Anomali Berat	57	135	15	18	0.8533

Tabel 3. Metrik Evaluasi Klasifikasi Multi-Kelas

	Precision	Recall	F1-Score	Support
Normal	0.9487	0.9867	0.9673	75
Anomali Ringan	0.7467	0.7467	0.7467	75
Anomali Berat	0.7917	0.7600	0.7755	75
Accuracy			0.8311	225
Macro avg	0.8290	0.8311	0.8298	225
Weighted avg	0.8290	0.8311	0.8298	225

Pemetaan distribusi tebakan model secara spesifik terhadap label aktual divisualisasikan melalui Confusion Matrix pada Gambar 3.



Gambar 3. Confusion Matrix pada Pengujian Data Uji

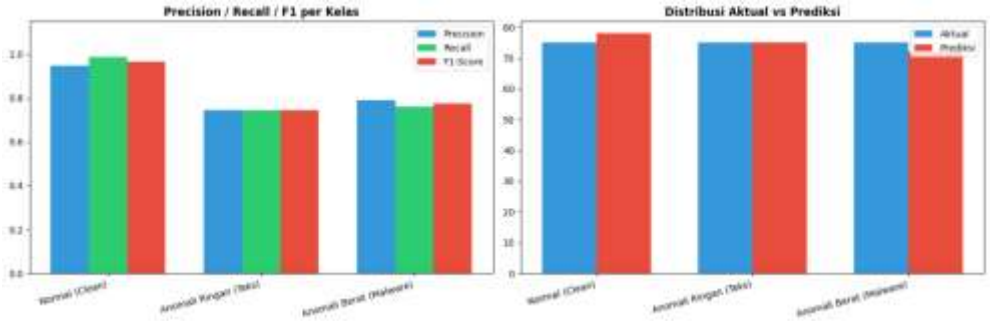
Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 464-473

Berdasarkan Tabel 2, Tabel 3, dan Gambar 3, analisis metrik kinerja disajikan sebagai berikut:

1. Kelas Normal: Membuktikan sensitivitas tertinggi dengan Recall 98,67% dan F1-Score 96,73%. Dari 75 citra, 74 ditebak secara presisi, menekan angka False Positive.
2. Anomali Ringan (Teks): Mencapai F1-Score 74,67%. Dari 75 citra, 56 berhasil diprediksi tepat. Terdapat sedikit misclassification, di mana 15 citra diprediksi sebagai anomali berat.
3. Anomali Berat (Malware): Menghasilkan tingkat presisi 79,17% dan F1-Score 77,55%. Sebanyak 57 citra malware diidentifikasi dengan benar.

Diskusi dan Analisis Temuan

Pemanfaatan ekstraksi fitur dari domain frekuensi yang dikombinasikan dengan CNN 1D terbukti sangat optimal membedakan tingkat ancaman steganografi. Visualisasi komparasi metrik dan keseimbangan distribusi prediksi model disajikan pada Gambar 4.



Gambar 4. Visualisasi Metrik Performa (Kiri) dan Distribusi Prediksi (Kanan)

Keberhasilan akurasi 83,11% divalidasi oleh grafik sisi kanan pada Gambar 4, di mana distribusi prediksi sangat seimbang dan proporsional (78 Normal, 75 Teks, 72 Malware) berbanding dengan distribusi aktual (75 per kelas), yang membuktikan absennya bias mayoritas (class imbalance bias) pada model.

Dari perspektif forensik keamanan siber, temuan paling krusial terletak pada angka False Negative untuk deteksi malware. Model sama sekali tidak pernah meloloskan citra bermuatan malware berkapasitas 40 KB menjadi gambar "Normal" (FN = 0). Hal ini secara empiris menegaskan bahwa arsitektur yang dirancang memiliki keandalan preventif yang sangat tangguh untuk menangkal infiltrasi skrip destruktif ke dalam jaringan. Adapun tumpang tindih prediksi antara kelas Anomali Ringan dan Anomali Berat merupakan implikasi teknis dari algoritma Steghide yang memanfaatkan metode Graph-Theoretic Matching. Algoritma ini secara cerdas menukar (swap) nilai koefisien DCT pada blok frekuensi tinggi yang berdekatan untuk menyamarkan sisipan. Pada citra dengan tekstur visual asli yang sangat kompleks, sebaran noise DCT akibat payload 40 KB dapat terkompresi sedemikian rupa hingga pola fluktuasi probabilitas Markov-nya mengalami tumpang tindih (overlapping) dengan pola kerusakan 20 KB pada citra bertekstur halus. Fluktuasi margin fitur statistik yang sangat tipis inilah yang menyebabkan filter konvolusi sesekali mengalami keraguan dalam memisahkan batas keparahan anomali. Namun, stabilitas F1-Score rata-rata di angka 82,98% membuktikan keberhasilan sistem dalam mendobrak batas deteksi biner dengan mengukur skala keparahan ancaman payload.

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 464-473

KESIMPULAN

Penerapan 1-Dimensional Convolutional Neural Network (CNN 1D) pada ekstraksi koefisien Discrete Cosine Transform (DCT) terbukti tangguh dalam mendeteksi dan membedakan tingkat bahaya steganografi pada citra JPEG. Penggunaan strategi Paired Dataset dengan resolusi tinggi berhasil memandu model untuk mengabaikan bias visual dan fokus secara eksklusif pada anomali struktural koefisien frekuensi akibat injeksi aplikasi Steghide, menanggulangi kendala overfitting yang sering dijumpai pada model Deep Learning konvensional.

Hasil evaluasi mengonfirmasi bahwa arsitektur usulan sukses mendobrak batas deteksi biner melalui klasifikasi multi-kelas dengan keandalan identifikasi malware yang sangat presisi (0 kejadian False Negative menuju kelas Normal). Dari segi implikasi praktis, efisiensi beban komputasi CNN 1D membuka jalan bagi instrumen forensik digital ini untuk diimplementasikan sebagai plug-in pemindaian lalu lintas data secara real-time terintegrasi pada perangkat Intrusion Detection System (IDS). Sistem dapat secara proaktif memblokir serangan eksfiltrasi data maupun menetralisir injeksi backdoor sebelum payload dieksekusi di komputer pengguna akhir.

Meskipun model usulan menunjukkan performa yang menjanjikan, penelitian ini masih memiliki sejumlah keterbatasan teknis. Evaluasi sistem saat ini difokuskan secara eksklusif pada format pembawa JPEG dan satu jenis algoritma penyisipan steganografi, yaitu Steghide, dengan batasan muatan tetap (20 KB dan 40 KB). Sebagai rekomendasi untuk riset mendatang, pemodelan ini perlu diekspansi dan dilatih menggunakan dataset hibrida yang mengintegrasikan berbagai variasi algoritma steganografi termutakhir lainnya (seperti JPHide, F5, atau OutGuess) dengan rentang kapasitas payload yang lebih dinamis. Hal ini penting guna mewujudkan instrumen deteksi Zero-Day Steganography yang tahan banting untuk kebutuhan forensik digital di dunia nyata.

Pernyataan mengenai Penggunaan Kecerdasan Buatan (AI).

AI digunakan dalam penyusunan naskah ini. Penulis menggunakan perangkat kecerdasan buatan semata-mata untuk membantu perbaikan bahasa, pemeriksaan tata bahasa, serta penyuntingan demi kejelasan dan keterbacaan. Penulis tetap bertanggung jawab penuh atas akurasi, orisinalitas, integritas, dan konten akademis naskah tersebut. Seluruh konten yang dibantu oleh AI telah ditinjau dan diverifikasi oleh penulis sebelum naskah diserahkan.

Konflik kepentingan

Penulis menyatakan tidak ada konflik kepentingan terkait dengan penelitian, kepenulisan, dan/atau publikasi artikel ini.

Kontribusi penulis

Konseptualisasi, Muh. Dirham Rahim, Muhyiddin A.M Hayat, dan Rizki Yusliana Bakti; pengumpulan data, Muh. Dirham Rahim; validasi data, Muhyiddin A.M Hayat, Rizki Yusliana Bakti, dan Darniati; analisis formal, Muh. Dirham Rahim; penulisan—penyusunan draf awal, Muh. Dirham Rahim; penulisan—peninjauan dan penyuntingan, Muhyiddin A.M Hayat, Rizki Yusliana Bakti, dan Darniati; visualisasi, Muh. Dirham Rahim. Seluruh penulis telah membaca dan menyetujui versi akhir naskah ini.

Ucapan terima kasih

Penelitian ini didukung oleh Program Studi Teknik Informatika, Universitas Muhammadiyah Makassar. Penulis mengucapkan terima kasih yang sebesar-besarnya atas dukungan, fasilitas laboratorium, serta bimbingan yang telah diberikan selama proses penelitian hingga penyelesaian penyusunan naskah ini.

DAFTAR PUSTAKA

Agarwal, S., Kim, H., & Jung, K. (2023). High-Pass-Kernel-Driven Content-Adaptive Image Steganalysis Using Deep Learning. *Mathematics*. https://doi.org/10.3390/math11204322

Bhardwaj, R., & Sharma, V. (2016). Image Steganography Based on Complemented Message and Inverted Bit LSB Substitution. *Procedia Computer Science*, *93*(September), 832–838. https://doi.org/10.1016/j.procs.2016.07.245

De La Croix Ntivuguruzwa, J., & Ahmad, T. (2023). A convolutional neural network to detect possible hidden data in spatial domain images. *Cybersecurity*, *6*, 1–16. https://doi.org/10.1186/s42400-023-00156-x

De La Croix Ntivuguruzwa, J., & Ahmad, T. (2024). FuzConvSteganalysis: Steganalysis via fuzzy logic and convolutional neural network. *SoftwareX*, *26*, 101713. https://doi.org/10.1016/j.softx.2024.101713

G.D.Sanjaya, R.Hadi, N.Luh, & G.Pivin. (2018). Kompresi Citra Digital Menggunakan Metode Discrete Cosine Transform. *Prosiding Sintak*, 38–44.

Hemamalini. (2025). A Novel Hybrid framework of NAdamBound optimized Dilated Depthwise Separable CNN for deep learning based image steganalysis in digital forensics. *Journal of Information Systems Engineering and Management*. https://doi.org/10.52783/jisem.v10i42s.7892

Hong, E., Lim, K., Oh, T.-W., & Jang, H. (2023). Lightweight image steganalysis with block-wise pruning. *Scientific Reports*, *13*(1), 16148. https://doi.org/10.1038/s41598-023-43386-2

Jelodar, H., Bai, S., Hamedi, P., Mohammadian, H., Razavi-Far, R., & Ghorbani, A. (2026). Large Language Model (LLM) for Software Security: Code Analysis, Malware Analysis, Reverse Engineering. *Journal of Information Security and Applications*, *98*(February), 104390. https://doi.org/10.1016/j.jisa.2026.104390

Liu, Q., Yang, Z., & Wu, H.-Y. (2023). JPEG steganalysis based on steganographic feature enhancement and graph attention learning. *Journal of Electronic Imaging*, *32*, 33032. https://doi.org/10.1117/1.jei.32.3.033032

Purwono, Ma'arif, A., Rahmaniar, W., Fathurrahman, H. I. K., Frisky, A. Z. K., & Haq, Q. M. U. (2022). Understanding of Convolutional Neural Network (CNN): A Review. *International Journal of Robotics and Control Systems*, *2*(4), 739–748. https://doi.org/10.31763/ijrcs.v2i4.888

Rustad, S., Setiadi, D. R. I. M., Syukur, A., & Andono, P. N. (2022). Inverted LSB image steganography using adaptive pattern to improve imperceptibility. *Journal of King Saud University - Computer and Information Sciences*, *34*(6), 3559–3568. https://doi.org/10.1016/j.jksuci.2020.12.017

Shehab, D., & Alhaddad, M. (2025). Image steganalysis using LSTM fused convolutional neural networks for secure telemedicine. *Frontiers in Medicine*, *12*. https://doi.org/10.3389/fmed.2025.1619706

Veerakumar, J., Prasad, H., & Muthulingam, A. (2026). Cross-Algorithm Steganalysis via Dual-Domain Feature Fusion: A Hybrid Deep Learning Approach for Payload

Journal : Journal of Computer Science and Information Technology (JCSIT)
Publisher : Yayasan Nuraini Ibrahim Mandiri
ISSN : 3031-8467
Volume / Issue / Year/ Page : Vol. 3 No. 4 (2026): September/ Hal 464-473

Detection. *International Journal of Bioinformatics and Intelligent Computing*. https://doi.org/10.61797/ijbic.v5i1.656

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. In *Artificial Intelligence Review* (Vol. 57, Number 4). Springer Netherlands. https://doi.org/10.1007/s10462-024-10721-6