FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 16, June, 2026, pp 551 – 556
DOI: https://doi.org/10.33003/fjs-2026-1016-5739

Deep Learning for Image Steganalysis: A Structured Comparative Review of Dataset, Domain, and Robustness Coverage

***[1]Shehu Asma'u, [2]Kirubakaran Prema, [3]Muhammad-Bello L. Bilkisu and [4]Ibrahim Nurudeen**

[1]Department of Computer Science, Nile University of Nigeria, FCT Abuja, Nigeria.
[2]Department of Information Technology, Nile University of Nigeria, FCT Abuja, Nigeria.
[3]Department of Software Engineering, Nile University of Nigeria, FCT Abuja, Nigeria.
[4]Department of Cyber Security, Nile University of Nigeria, FCT Abuja, Nigeria.

*Corresponding authors' email: asmausidi4@gmail.com

Received: 19 Aug. 2026
Accepted: 30 Aug. 2026
Published: 07 Sep. 2026

ABSTRACT
Deep learning has substantially advanced image steganalysis, yet inconsistent evaluations across datasets, embedding domains, payload rates, and threat models obscure the field's readiness for real-world forensic deployment. This structured comparative review analyses 26 deep-learning steganalysis studies (2015–2025) across six standardized dimensions, synthesizing findings into three critical structural gaps. First, adversarial robustness is severely under-addressed: only 19% (5 of 26) of studies evaluate robustness, with just a single study testing pixel-level, norm-bounded perturbations rather than feature-space attacks. Second, single-domain architectural specialization predominates, as no reviewed study evaluates a single unified model across both spatial and frequency embedding domains; limiting multi-domain efforts to separate models or content heterogeneity. Third, cover-source mismatch remains widespread due to dataset homogeneity, with most studies relying solely on single-source benchmarks like BOSSBase 1.01. Ultimately, while individual studies partially address single limitations, none solve all three simultaneously; closing these gaps requires a unified architecture combining multi-source training, dual-domain coverage, and pixel-level adversarial testing, representing an essential open challenge for practical deployment.

Keywords: Adversarial Robustness, Deep Learning, Image Steganalysis

INTRODUCTION

Image steganalysis, the task of detecting whether a digital image carries a covertly embedded message has progressed substantially since the introduction of deep convolutional architectures, which now consistently outperform earlier statistical and handcrafted-feature approaches across most published benchmarks (Boroumand et al., 2019; Fridrich & Kodovsky, 2012). This progress has, however, occurred across a fragmented set of evaluation conditions: studies differ in which benchmark datasets they use, whether they evaluate spatial-domain or frequency-domain embedding, what payload rates they test, and whether they consider an adversary capable of perturbing the input image to evade detection. This fragmentation makes it difficult to assess, from the published literature taken as a whole, how close the field is to a steganalysis system that would perform reliably under the heterogeneous, adversarial conditions of real-world forensic deployment, as opposed to performing well only under the specific, often narrower conditions any single study happens to test.

This paper presents a structured comparative review of 26 deep-learning-based image steganalysis studies, coding each study consistently against six comparison dimensions chosen to characterise exactly this fragmentation: benchmark dataset(s) used, embedding domain coverage, payload rate range tested, model architecture family, whether and how adversarial robustness is evaluated, and the study's own reported or inferred cross-source generalisation capability. Our aim is not to rank the reviewed studies against one another. The heterogeneity in their evaluation conditions makes direct ranking unreliable, a point we return to in section 3.2, but to characterise, across the reviewed set as a whole, which combinations of dataset diversity, domain coverage, and adversarial evaluation are common, which are rare, and which do not appear to have been jointly addressed within any single reviewed study.

MATERIALS AND METHODS

Scope and Positioning of This Review

This review covers 26 deep-learning-based image steganalysis studies published between 2015 and 2025, assembled through purposive sampling of the steganalysis literature rather than a documented multi-database search protocol. Explicitly, this does not report a documented multi-database search string, a stated date range applied uniformly at the point of search, or a count of records identified, screened, and excluded at each stage. Readers expecting a PRISMA-compliant search and screening log should treat this review as a structured comparative synthesis of a purposively assembled set of studies rather than as evidence of an exhaustive or reproducible literature search.

What this review does provide, and what is considered is the methodological contribution. It is a consistent six-dimension coding scheme applied uniformly across all 26 studies, extracted from each study's own reported methodology and results rather than from secondary summaries, and a synthesis organised around that coding scheme rather than around a chronological or architecture-family narrative.

Comparison Dimensions

Each of the 26 reviewed studies is coded against six dimensions, summarised in Table 1.

Table 1: The Six Comparison Dimensions Applied consistently across all 26 Reviewed Studies

Dimension	Definition	Possible Values
Dataset	Benchmark image collection(s) used for training and evaluation	Named benchmark (e.g. BOSSBase 1.01) or custom collection
Embedding domain	Whether embedding is evaluated in the spatial (pixel) or frequency (DCT/JPEG) domain	Spatial; Frequency; Spatial + Frequency
Payload rate range	Range of embedding rates (bpp or bpnzAC) evaluated	Single rate or range, as reported
Model architecture family	The principal architectural paradigm used	CNN; CNN + attention; GAN-based; Transformer/hybrid; Classical ML
Adversarial robustness	Whether, and how, the study evaluates robustness to adversarial perturbation	None; Feature-space; Pixel-level; Custom adversarial-noise defence
Generalisation capability	The study's own reported or inferable capability to generalise across sources, as assessed by this review	Low; Moderate; High (see section 2.3)

Generalization Capability Coding

The generalization capability dimension is the most interpretive of the six and warrants explicit definition. A study as High only if it evaluates its model on two or more structurally distinct image sources (differing in at least acquisition device diversity, compression format, or embedding domain) within a single trained model and training run. A study as Moderate if it evaluates on a single dataset but with some attention to diversity within that dataset (for instance, multiple payload rates, multiple embedding algorithms, or an explicit discussion of generalization as a limitation), and Low if the study evaluates a single architecture on a single dataset, single embedding algorithm, and single payload condition without discussion of cross-source generalization. This coding is necessarily a judgment applied by the present authors based on each study's own reported scope, not a measured quantity, and we flag it accordingly wherever it is used in section 4 and section 5.

RESULTS AND DISCUSSION

Overview of the Reviewed Studies

Table 2 presents the full coded comparison across all 26 reviewed studies, spanning publications from 2015 to 2025. The table is organised in the order the studies are discussed in section 4, grouped loosely by their primary contribution (architecture, adversarial robustness, or evaluation methodology).

Table 2: Full Coded Comparison across the 26 Reviewed Studies (2015–2025)

Study	Dataset(s)	Domain	Architecture Family	Adversarial Eval.
Yu et al. (2024)	BOSSBase 1.01, ALASKA II	Spatial	CNN + Feature Selection	None
Bravo-Ortiz et al., (2024)	BOSSBase 1.01	Spatial	CNN + Vision Transformer	None
Fu et al. (2022)	BOSSBase 1.01	Spatial	CNN + SE attention	None
Agarwal & Jung (2024)	BOSSBase 1.01, BOWS2	Spatial	Entropy-driven CNN + SRM	None
Dwaik & Belkhouche (2024)	BOSSBase 1.01, ALASKA II	Spatial	CNN + SRM + Gabor	None
Liu et al., (2023)	BOSSBase 1.01	Spatial	CNN + ECA + Transfer Learning	None
Ke & Sheng (2019)	BOSS, RAISE	Spatial	Co-occurrence + RealAdaBoost	None
Lu et al. (2019)	BOSSBase 1.01	Spatial + JPEG (two models)	CNN + TLU preprocessing	None
Mohamed et al., (2020)	Custom (colour)	Frequency	Statistical + CNN (survey)	None
Li & Liu (2015)	Custom (SS stego)	Spatial	Unsupervised IGLS	None
Shankar & Azhakath (2019)	Custom (JPEG)	Spatial + Frequency	Hybrid SVM-PSO	None
De La Croix et al., (2024)	BOSSBase	Spatial	CNN + genetic-algorithm pooling	None
Mo et al., (2023)	BOSSBase, BOWS2, ALASKA II	Spatial (J-UNIWARD)	Wavelet + CNN clustering	None
Akram et al., (2024)	Custom (colour)	Spatial	Curvelet + SVM	None
Peng et al., (2024)	FFHQ, LSUN	Spatial	RS-GAN (disentanglement)	FGSM, PGD
Hu & Wang (2025)	Custom (adv. stego)	Spatial	Directional adversarial noise	Custom adv. Noise
Al-Obaidi et al., (2024)	CNN benchmarks	Spatial	RL + GAN framework	GAN-based
Hu & Wang (2023)	BOSSBase, BOWS2	Spatial	TStegNet (two-stream CNN)	ADVEMB, MAE
Li et al., (2024)	BOSSBase 1.01	Spatial	FPFNet (pyramid fusion)	None
Huang et al., (2024)	Heterogeneous mix	Mixed (routed)	FACSNet (content routing)	None
Alrusaini (2025)	BOSSBase 1.01	Spatial	5-architecture robustness study	None
Li & Dong (2024)	Custom	Spatial	CNN + channel attention	None
Yang et al., (2024)	Standard benchmarks	Spatial	Hybrid CNN + attention fusion	None
Bohang et al., (2025)	BOSSBase, BOWS2	Spatial (S-UNIWARD)	CNN + DRL active learning	None
Zhengliang et al., (2025)	Custom (stego)	Spatial	SG-ResNet (dual-stream Gabor)	None
Jawad et al., (2025)	CelebA	Spatial (LSB)	EfficientNet + adv. Training	FGSM (feature-space)

Why Direct Numerical Ranking Is Unreliable

The reviewed studies differ across dataset, embedding algorithm, payload rate, image resolution, and train/test split methodology; a reported accuracy of, for instance, 96% in one study and 89% in another is not evidence that the first method is superior, since the two figures may reflect a substantially easier evaluation condition (a single dataset, a single high payload rate, a favourable split) rather than a better detector. Therefore, the comparative claims in this review are restricted to the presence or absence of specific evaluation conditions,

which datasets, which domains, whether adversarial testing was performed rather than to accuracy figures themselves, a position consistent with general best practice in benchmarking surveys where evaluation conditions are not uniformly controlled across studies (Selvaraj et al., 2021).

Dimension 1: Dataset Coverage

Table 3: Dataset Coverage Pattern across the 26 Reviewed Studies

Dataset Pattern	Count	Share of 26
Single named benchmark only (BOSSBase 1.01 alone)	11	42%
Multiple named benchmarks (BOSSBase + ALASKA II)	8	31%
Custom or non-standard dataset	7	27%

Eight of 26 studies (31%) evaluate on more than one named benchmark dataset, which is a meaningfully larger share than an impression of universal single-dataset evaluation might suggest. However, evaluating on multiple datasets is not the same claim as training a single model on a deliberately balanced, combined mixture of heterogeneous sources and testing its cross-source generalisation gap directly: of the eight multi-dataset studies, six are coded in their own reported methodology (via the generalisation-capability judgement defined in section 2.3) as achieving only Moderate generalisation, typically because the datasets are evaluated as separate conditions rather than combined into a single joint training distribution (Muhammed et al., 2026). Only one study, Ke & Sheng (2019), is coded as achieving good cross-source generalisation, having evaluated a single classifier across the BOSS and RAISE datasets.

Dimension 2: Embedding Domain Coverage

Of the 26 reviewed studies, 22 (85%) evaluate exclusively spatial-domain embedding. One study (Mohamed et al., 2020) evaluates frequency-domain (transform-domain) techniques exclusively, as a survey rather than a novel architecture. Three studies touch on multiple domains in a narrower sense than a single shared model evaluated across both: Lu et al., (2019) evaluates two separate, domain-specific architectures (Yedroudj-Net for spatial, Dense-Net for JPEG) rather than one architecture handling both domains; Shankar & Azhakath (2019) evaluates various spatial and transform domain steganographic techniques using a single classifier, though the reported methodology does not unambiguously establish whether this constitutes a single model tested on genuinely separate spatial and frequency-domain datasets or a single domain's images analysed with features drawn from both spatial and transform representations; and Huang et al., (2024) addresses heterogeneity in image content and source type via a routing mechanism, which is a different and narrower notion of mixed than spatial-versus-frequency embedding-domain coverage, since FACSNet's routing operates on image complexity categories within what remains, as far as the reported methodology indicates, a single embedding domain. No reviewed study trains and evaluates one shared architecture on two structurally distinct cover sources differing in embedding domain (one spatial, one frequency-domain JPEG) within a single training run. This is the most consistently under-addressed of the six comparison dimensions in the reviewed set.

Dimension 3: Adversarial Robustness Evaluation

Only 5 of the 26 reviewed studies (19%) report any form of adversarial robustness evaluation. Table 4 details these five studies and the nature of their adversarial evaluation.

Table 4: The Adversarial Robustness Evaluation Reviews

Study	Adversarial Approach	Threat Model
Peng et al., (2024)	RS-GAN: disentangle and reconstruct via GAN	FGSM, PGD (pixel-level, on non-standard datasets)
Hu & Wang (2025)	Directional adversarial noise as defensive training signal	Custom adversarial-noise generation, not a standard bounded attack
Al-Obaidi et al., (2024)	RL-guided GAN framework	GAN-based; not a norm-bounded perturbation attack
Hu & Wang (2023)	TStegNet: gradient and image subnets targeting adversarial artefacts	ADVEMB, MAE (pixel-level, spatial-domain only)
Jawad et al., (2025)	EfficientNet + adversarial training	FGSM applied to internal EfficientNet feature representations, not input pixels

Of these five, three (Peng et al., 2024; Hu & Wang, 2023; Hu & Wang, 2025) evaluate pixel-level or near-pixel-level threats, though Peng et al., (2024) does so on non-standard face and scene datasets (FFHQ, LSUN) rather than standard steganalysis benchmarks, and Hu & Wang (2025) uses a custom adversarial-noise generation procedure rather than a standard norm-bounded attack such as FGSM or PGD, making direct comparison to norm-bounded threat models difficult. Jawad et al., (2025) is explicit that its adversarial evaluation targets internal EfficientNet feature representations rather than the input image directly, a methodological choice with a clear practical consequence: a detector evaluated only against feature-space perturbations may appear robust under that evaluation while remaining untested against an adversary who can only manipulate the pixels of a transmitted image, which is the realistic threat model for steganalysis deployment. Al-Obaidi et al., (2024) uses GANs as part of the detection framework's own architecture rather than as an evaluation-time adversary, and so, it is better characterised as adversarially-inspired architecture design than adversarial robustness evaluation in the threat-model sense used by the other four studies.

Taken together, no reviewed study evaluates adversarial robustness via a standard norm-bounded, pixel-level, multi-step attack (the PGD threat model) on a standard steganalysis benchmark dataset (BOSSBase, ALASKA II, or BOWS2) using both spatial and frequency domain embedding. This combination, pixel-level, multi-step adversarial evaluation on standard benchmarks spanning both embedding domains does not appear to have been jointly addressed by any architecture in the reviewed literature, despite each individual element (pixel-level attacks, standard benchmarks, or single-domain robustness testing) appearing separately across different studies.

Dimension 4: Architecture Families

Table 5: Architecture Family Distribution across the 26 Reviewed Studies

Architecture Family	Count	Example Studies
CNN (with or without attention)	14	Fu et al., (2022); Liu et al., (2023); Li et al., (2024); Li & Dong (2024); Yang et al., (2024)
GAN-based or RL+GAN	3	Peng et al., (2024); Al-Obaidi et al. (2024); Hu & Wang (2025)
CNN + Transformer hybrid	1	Bravo-Ortiz et al., (2024)
Classical ML (SVM, AdaBoost, co-occurrence features)	3	Ke & Sheng (2019); Shankar & Azhakath (2019); Akram et al., (2024)
Unsupervised / statistical	1	Li & Liu (2015)
Survey (no single proposed architecture)	1	Mohamed et al., (2020)
Reinforcement learning for data/training strategy	1	Bohang et al., (2025)
Specialised dual-stream / routing architectures	2	Huang et al., (2024); Zhengliang et al., (2025)

CNN-based architectures, with or without an attention mechanism layered on top, account for 14 of 26 studies (54%), confirming CNNs as the dominant architectural paradigm in the reviewed period. Squeeze-and-Excitation or comparable channel-attention mechanisms appear in several of these (Fu et al., 2022; Liu et al., 2023; Li & Dong, 2024), consistent with channel attention's broad adoption in the wider image classification literature (Hu et al., 2018) and its specific relevance to steganalysis, where the embedding signal is concentrated in particular feature channels that benefit from adaptive reweighting.

Dimension 5 and 6: Payload Range and Generalisation Capability

Payload rate ranges varied widely across the reviewed studies, from as low as 0.05 bpp (Liu et al., 2023) to as high as 0.5 bpp (Jawad et al., 2025; Bohang et al., 2025), with most studies testing a range rather than a single fixed rate. Several studies explicitly identify low-payload detection as a residual difficulty even within their own results (Liu et al., 2023; Agarwal & Jung, 2024), consistent with the broader pattern documented elsewhere for low-payload steganalysis specifically (Selvaraj et al., 2021).

Applying the generalisation-capability coding defined in §2.3 across all 26 studies, we find 13 studies (50%) coded as Low, 11 studies (42%) coded as Moderate, and 2 studies (8%) coded above Moderate in some form. One of these two, Ke & Sheng (2019), meets our good threshold by evaluating a single classifier across the BOSS and RAISE datasets. The second, Huang et al. (2024), was coded "High-heterogeneous-aware" in the original source table this review draws on, reflecting that FACSNet is architecturally designed to address cover-source heterogeneity through content-based routing; applying our stricter section 2.3 definition, which requires demonstrated evaluation across two or more structurally distinct sources rather than an architectural design intended to handle heterogeneity, we do not consider Huang et al., (2024) to meet the High threshold as we have defined it, since its routing mechanism requires test-time knowledge of image type and the reported evaluation does not establish cross-source testing in the same sense as Ke & Sheng (2019). This recoding was flagged because it illustrates a genuine ambiguity in the generalisation-capability dimension: a study can be designed with heterogeneity in mind without thereby demonstrating measured cross-source generalisation, and the two are easy to conflate (Muhammed et al., 2026). No reviewed study, under our stricter reading, meets the High threshold defined in section 2.3, which requires evaluation of a single trained model across two or more structurally distinct sources differing in at least acquisition device diversity, compression format, or embedding domain.

Synthesis: Three Structural Gaps

The six-dimension comparison in section 3 converges on three structural gaps in the reviewed literature, each addressed partially by a small number of individual studies but, within the 26 studies reviewed, never addressed jointly within a single architecture.

Gap 1: Cover-Source Mismatch from Homogeneous Training

The majority of reviewed studies (18 of 26, combining the 11 single-named-benchmark and 7 custom-dataset studies in Table 3) train and evaluate on a single image source. Even among the 8 multi-dataset studies, most evaluate datasets as separate conditions rather than combining them into a single heterogeneous training distribution and measuring the resulting cross-source accuracy gap directly. Only Ke & Sheng (2019) is coded as demonstrating good cross-source generalisation under section 2.3 criteria. This pattern is consistent with cover-source mismatch being recognised as a named, persistent challenge in the broader steganalysis literature (as discussed narratively, though not always tested directly, in several reviewed studies) without yet being addressed through systematic heterogeneous-training evaluation in most of the reviewed primary studies themselves.

Gap 2: Single-Domain Architectural Specialisation

As detailed in section 3.4, no reviewed study trains and evaluates a single shared architecture across both spatial-domain and frequency-domain (JPEG) embedding. The closest approaches, Lu et al., (2019), use two domain-specific architectures rather than one shared model, and Shankar & Azhakath (2019), whose domain coverage could not be unambiguously established from the reported methodology, both fall short of single-model dual-domain evaluation. This gap may partly reflect a reasonable engineering response to genuine differences between spatial and frequency-domain images (resolution, colour depth, compression artefacts, and the differing physical meaning of a spatial bpp payload versus a frequency-domain bpnzAC payload), which would need to be resolved deliberately by any architecture attempting single-model coverage of both domains. Whether this complexity is better addressed by domain-specific architectures, as in Lu et al., (2019), or by a single architecture designed to generalise across both, is not resolved by the present review and would benefit from direct experimental comparison in future work.

Gap 3: Absence of Pixel-Level Adversarial Evaluation

As detailed in section 3.5, adversarial robustness evaluation is reported in only 5 of 26 studies, and among those five,

evaluation against a standard norm-bounded, multi-step, pixel-level attack (PGD) on a standard benchmark dataset is reported in none. This is, in our assessment, the most consistently under-addressed of the three gaps, both in terms of the share of studies addressing it at all (19%) and the rigour of the threat model among those that do. Jawad et al., (2025) is noted as a study whose adversarial evaluation, while a meaningful contribution in other respects, evaluates a methodologically different and arguably less realistic threat (feature-space perturbation) than the pixel-level threat model a real-world adversary would actually be positioned to mount.

No Reviewed Study Addresses All Three Gaps Jointly

No study among the 26 reviewed addresses cover-source mismatch (via demonstrated multi-source heterogeneous training), single-domain specialisation (via single-model dual-domain evaluation), and pixel-level adversarial robustness (via norm-bounded, multi-step attack evaluation) simultaneously. The studies that come closest to any individual gap, Ke & Sheng (2019) for cross-source generalisation, Lu et al., (2019) for domain coverage (albeit via two separate architectures), and Hu & Wang (2023) or Peng et al., (2024) for pixel-level or near-pixel-level adversarial evaluation do not overlap with one another, and none addresses more than one of the three gaps within the same study. We are not aware of a published architecture, among the studies reviewed here, that combines multi-source heterogeneous training, single-model dual-domain coverage, and pixel-level adversarial robustness evaluation within one training run, and we present this combination as a concretely specified, currently open direction for future architecture design in image steganalysis.

Limitations

First, and most importantly, this review's set of 26 studies was not assembled through a documented, reproducible multi-database search protocol; it was compiled through purposive sampling of the deep learning steganalysis literature, and a consistent coding scheme is applied and extended to it rather than independently verifying its completeness against the wider steganalysis literature. A reader requiring PRISMA-level reproducibility guarantees should treat the dimension-level statistics reported in section 3 and section 4 (for instance, "5 of 26 studies") as descriptive of this specific, non-exhaustively-assembled set, not as an unbiased estimate of the true proportion across all published steganalysis literature in the relevant period.

Second, the generalisation capability dimension (section 2.3) required interpretive judgement by the present authors, and we found at least one instance (Huang et al., 2024, section 3.7) where the coding diverged from the original source table's coding for the same study, indicating the dimension is more subjective than the other five and should be weighted accordingly when drawing conclusions.

Third, several reviewed studies described methodologies (e.g., Shankar & Azhakath, 2019, on domain coverage) were not sufficiently detailed in the available secondary description to code with full confidence; we have flagged these ambiguities explicitly in section 3 rather than resolving them by assumption, but this means a small number of dimensions coding in this review carry genuine residual uncertainty.

Fourth, this review covers only image steganalysis; cover-source mismatch, domain specialisation, and adversarial robustness are active concerns in audio, video, and text steganalysis as well, and there is no claim that the gaps identified here generalise to those other media types.

CONCLUSION

This structured comparative review coded 26 deep-learning-based image steganalysis studies (2015–2025) against six consistent dimensions: dataset coverage, embedding domain, payload range, architecture family, adversarial robustness evaluation, and generalisation capability. The review identifies three structural gaps in the reviewed literature, widespread reliance on homogeneous single-source training data, near-universal confinement to a single embedding domain within any one architecture, and a strong concentration of adversarial robustness evaluation (where present at all, in only 19% of studies) on feature-space rather than pixel-level threat models. No study among the 26 reviewed addresses all three gaps jointly within a single architecture and training run.

This review does not follow a formal PRISMA search protocol and should be read as a structured comparative synthesis of a purposively assembled set of studies rather than an exhaustive systematic review in the strictest methodological sense. Within that scope, the consistency of the six-dimension coding scheme across all 26 studies supports the conclusion that a steganalysis architecture combining multi-source heterogeneous training, single-model dual-domain coverage, and pixel-level adversarial robustness evaluation remains, at the time of this review, a combination addressed by very few if any studies in the reviewed deep learning steganalysis literature, representing a genuine and specific direction for continued research.

REFERENCES

Agarwal, S., & Jung, K.-H. (2024). Digital image steganalysis using entropy driven deep neural network. Journal of Information Security and Applications, 84, 103799.

Akram, A., Khan, I., Rashid, J., Saddique, M., Idrees, M., Ghadi, Y. Y., & Algarni, A. (2024). Enhanced steganalysis for color images using curvelet features and support vector machine. Computers, Materials & Continua, 78(1).

Al-Obaidi, S. A. R., Lighvan, M. Z., & Asadpour, M. (2024). Enhanced image steganalysis through reinforcement learning and generative adversarial networks. Intelligent Decision Technologies, 18(2), 1077–1100.

Alrusaini, O. A. (2025). Deep learning for steganalysis: Evaluating model robustness against image transformations. Frontiers in Artificial Intelligence, 8, 1532895.

Bohang, L., Ningxin, L., Osama, A., Fahad, A., Amr, T., Zaffar, A. S., Rohallah, A., Pawel, P., & Pol, L. Y. (2025). Image steganalysis using active learning and hyperparameter optimization. Scientific Reports, 15, 7340.

Boroumand, M., Chen, M., & Fridrich, J. (2019). Deep residual network for steganalysis of digital images. IEEE Transactions on Information Forensics and Security, 14(5), 1181–1193.

Bravo-Ortiz, M. A., Mercado-Ruiz, E., Villa-Pulgarin, J. P., Hormaza-Cardona, C. A., Quiñones-Arredondo, S., Arteaga-Arteaga, H. B., & Tabares-Soto, R. (2024). CvT-StegoNet: A convolutional vision transformer architecture for spatial image steganalysis. Journal of Information Security and Applications, 81, 103695.

De La Croix, N. J., Putra, M. A. R., & Ahmad, T. (2024). Toward the confidential data location in spatial domain

images via a genetic-based pooling in a convolutional neural network. In 2024 16th International Conference on Computer and Automation Engineering (ICCAE), 283–288.

Dwaik, A., & Belkhouche, Y. (2024). Enhancing the performance of convolutional neural network image-based steganalysis in spatial domain using spatial rich model and 2D Gabor filters. Journal of Information Security and Applications, 85, 103864.

Fridrich, J., & Kodovsky, J. (2012). Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security, 7(3), 868–882.

Fu, T., Chen, L., Fu, Z., Yu, K., & Wang, Y. (2022). CCNet: CNN model with channel attention and convolutional pooling mechanism for spatial image steganalysis. Journal of Visual Communication and Image Representation, 88, 103633.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7132–7141.

Hu, M., & Wang, H. (2023). Image steganalysis against adversarial steganography by combining confidence and pixel artifacts. IEEE Signal Processing Letters, 30, 987–991.

Hu, M., & Wang, H. (2025). Directional adversarial noise-based universal steganalysis method for detecting adversarial steganography. IEEE Signal Processing Letters.

Huang, S., Zhang, M., Kong, Y., Ke, Y., & Di, F. (2024). FACSNet: Forensics aided content selection network for heterogeneous image steganalysis. Scientific Reports, 14, 26258.

Jawad, T. A., Mohasefi, J. B., & Abdelghany, M. S. R. (2025). Adversarial-robust steganalysis system leveraging adversarial training and EfficientNet. TELKOMNIKA (Telecommunication Computing Electronics and Control), 23(2), 393–401.

Ke, Q., & Sheng, L. (2019). Content adaptive image steganalysis in spatial domain using selected co-occurrence features. In 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 28–33.

Li, H., & Dong, S. (2024). Image steganalysis algorithm based on deep learning and attention mechanism for computer communication. Journal of Electronic Imaging, 33(1), 013015.

Li, J., Wang, X., Song, Y., & Wang, P. (2024). FPFNet: Image steganalysis model based on adaptive residual extraction and feature pyramid fusion. Multimedia Tools and Applications, 83, 48539–48561.

Li, M., & Liu, Q. (2015). Steganalysis of SS steganography: Hidden data identification and extraction. Circuits, Systems, and Signal Processing, 34(10), 3305–3324.

Liu, S., Zhang, C., Wang, L., Yang, P., Hua, S., & Zhang, T. (2023). Image steganalysis of low embedding rate based on the attention mechanism and transfer learning. Electronics, 12(4), 969.

Lu, Y. Y., Yang, Z. L. O., Zheng, L., & Zhang, Y. (2019). Importance of truncation activation in pre-processing for spatial and JPEG image steganalysis. In 2019 IEEE International Conference on Image Processing (ICIP), 689–693.

Mo, C., Liu, F., Zhu, M., Yan, G., Qi, B., & Yang, C. (2023). Image steganalysis based on deep content features clustering. Computers, Materials & Continua, 76(3).

Mohamed, N., Rabie, T., & Kamel, I. (2020). A review of color image steganalysis in the transform domain. In 2020 14th International Conference on Innovations in Information Technology (IIT), 45–50.

Muhammed, F. O., Suleiman, M. A., Abdullahi, S. E., & Ogar, A. O. (2026). An explainable hybrid CNN-LSTM-Random Forest framework for early epidemic outbreak detection from multimodal data, validated by real corpora and controlled simulation. *FUDMA Journal of Sciences*, *10*(10), 26–36.

Peng, Y., Yu, Q., Fu, G., Zhang, W., & Duan, C. (2024). Improving the robustness of steganalysis in the adversarial environment with generative adversarial network. Journal of Information Security and Applications, 82, 103743.

Selvaraj, A., Ezhilarasan, A., Wellington, S. L. J., & Sam, A. R. (2021). Digital image steganalysis: A survey on paradigm shift from machine learning to deep learning based techniques. IET Image Processing, 15(2), 504–522.

Shankar, D. D., & Azhakath, A. S. (2019). Steganalysis of minor embedded JPEG image in transform and spatial domain system using SVM-PSO. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 46–49.

Yang, S., Jia, X., Zou, F., Zhang, Y., & Yuan, C. (2024). A novel hybrid network model for image steganalysis. Journal of Visual Communication and Image Representation, 103, 104251.

Yu, X., Ma, Y., Zhang, Y., Li, X., & Zhao, Y. (2024). Fast dominant feature selection with compensation for efficient image steganalysis. Signal Processing, 220, 109475.

Zhengliang, L., Chenyi, W., Zhu, X., Jianhua, W., & Guiqin, D. (2025). SG-ResNet: Spatially adaptive Gabor residual networks with density-peak guidance for joint image steganalysis and payload location. Mathematics, 13, 1460.