Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

VIDEO STEGANOGRAPHY TECHNIQUES: A COMPARATIVE STUDY OF SPATIAL, TRANSFORM-DOMAIN, CODEC-AWARE AND DEEP NEURAL LEARNING APPROACHES

Mueen ud Din[*1], Muhammad Mohsin[2], Muhammad Usman Ghani[3]

[*1,2]*Riphah International University, Faisalabad Campus, Faisalabad*
[3]*University of Management Sciences (UMT), Lahore*

[*1]mueen80@gmail.com, [2]mohsinscholar98@gmail.com, [3]mug0001985@gmail.com

[*1]https://orcid.org/0009-0009-0122-1087

DOI: https://doi.org/10.5281/zenodo.21974638

Keywords
video steganography; information hiding; motion vectors; deep learning; secure communication

Article History
Received: 28 January 2026
Accepted: 13 March 2026
Published: 27 March 2026

Copyright @Author
Corresponding Author: *
Mueen ud Din

Abstract

Video steganography conceals secret information within digital video while attempting to preserve visual quality, extraction reliability and resistance to statistical detection or signal-processing attacks. The field has progressed from direct least-significant-bit substitution to transform-domain embedding, motion-vector and codec-element manipulation, region-of-interest selection, optimization-based schemes, robust social-media transmission and deep neural video hiding. This paper presents a structured comparative review of representative approaches published from foundational work through 2025. The methods are analyzed using a common framework comprising embedding domain, payload capacity, imperceptibility, robustness, security against steganalysis, computational complexity, codec dependence and practical deployment. Particular attention is given to the randomized frame selection method which combines a 64-bit secret-key-derived frame index, payload encryption and LSB embedding. The method reports very high visual fidelity in its experiments, including a maximum PSNR of 74.15 dB and an MSE of 0.0002, while improving location secrecy over fixed-frame LSB schemes. However, its robustness against transcoding, frame deletion, resizing and modern learning-based steganalysis remains insufficiently established. The comparative synthesis shows that spatial methods offer high capacity and low cost but weak robustness; transform and codec-aware methods improve resilience at the expense of complexity and codec dependence; and neural methods achieve flexible high-capacity hiding but require training resources and more rigorous security evaluation. The paper concludes that future systems should combine key-controlled temporal selection, content-adaptive embedding, codec-aware robustness, error correction and steganalysis-aware learning under standardized evaluation protocols.

1. Introduction

Steganography aims to conceal the existence of a communication by embedding a secret payload inside an apparently ordinary carrier. This objective differs from cryptography, which transforms the message content but usually leaves the existence of protected communication visible. In practice, the two mechanisms are

complementary: encryption protects semantic content, whereas steganography attempts to reduce suspicion by hiding the communication channel itself (Petitcolas et al., 1999; Provos and Honeyman, 2003).

Digital video is an attractive carrier because it contains many frames, multiple color components, spatial redundancy within each frame, temporal redundancy across frames and a rich collection of coding elements in compressed bitstreams. These characteristics create substantially more embedding opportunities than a single still image. At the same time, they produce additional constraints. An effective method must preserve frame quality, temporal consistency, bitrate behavior, codec compliance, payload recoverability and resistance to video steganalysis (Sadek et al., 2015; Liu et al., 2019; Kunhoth et al., 2023).

The evolution of video steganography can be understood as a sequence of responses to the capacity-security-robustness trade-off. Early spatial methods modified pixel values directly and were easy to implement, but were vulnerable to compression and statistical analysis. Transform-domain approaches embedded data in frequency coefficients to improve resistance to signal processing. Compressed-domain methods targeted motion vectors, prediction modes, transform coefficients and other syntax elements so that hiding could be integrated with video coding. Content-adaptive methods selected edges, moving objects, faces or skin regions to exploit perceptual masking. More recent neural methods learn distributed representations and optimize imperceptibility, recovery, capacity and adversarial security jointly (Aly, 2011; Sadek et al., 2017; Mou et al., 2023).

A recurring limitation of conventional schemes is predictable embedding location. Din et al. (2025) addressed this issue by deriving a randomized frame position from a 64-bit secret key, encrypting the payload and inserting it through LSB replacement in the selected frame. This design retains the simplicity of spatial embedding while adding key-dependent temporal uncertainty. Because this paper is particularly relevant to secure frame selection, it is examined as a principal case study in the comparative analysis.

The contribution of this article is fourfold. First, it organizes the principal approaches used in video steganography into a unified taxonomy. Second, it compares representative methods using common performance dimensions rather than relying on incomparable headline PSNR values alone. Third, it positions randomized frame selection relative to fixed spatial, transform, codec-aware, robust and neural alternatives. Fourth, it proposes a standardized evaluation framework and identifies research directions for secure, practical and reproducible video steganography.

2. Materials and Methods

2.1 Review design and source selection

This study uses a structured narrative review and comparative synthesis rather than a statistical meta-analysis. A meta-analysis is not appropriate because the published methods use different cover videos, codecs, resolutions, payload definitions, attacks and quality metrics. Representative peer-reviewed studies were identified from IEEE Xplore, SpringerLink, ScienceDirect, the Computer Vision Foundation repository and journal publisher platforms. Foundational studies and influential reviews were included to establish the historical taxonomy, while recent studies were included to represent HEVC, social-media robustness, invertible networks, implicit neural representations and randomized frame selection.

Studies were considered relevant when they proposed, evaluated or systematically analyzed a method that hides data in raw or compressed video or when an image-steganography architecture directly influenced modern video-hiding networks. Watermarking-only studies were excluded unless their architecture or robustness mechanism was directly relevant to steganographic design. The resulting evidence base covers spatial, transform, motion-vector, prediction-mode, region-of-interest, optimization, encryption-assisted, robust-transcoding, deep-learning, generative, invertible and implicit-representation approaches.

2.2 Comparative dimensions

Each approach was evaluated using seven dimensions: (1) embedding capacity, (2) imperceptibility, (3) robustness to processing and transmission, (4) security against steganalysis, (5) computational complexity, (6) codec and platform compatibility and (7) extraction reliability. These dimensions are interdependent. Increasing payload frequently increases distortion or detection risk; improving robustness can reduce capacity; and adding encryption protects the message content without necessarily preventing detection of the steganographic channel.

Common distortion measures include mean squared error (MSE) and peak signal-to-noise ratio (PSNR). For an original frame I and a stego frame K of size M x N, MSE = (1/MN) sum[(I(i,j) - K(i,j))^2] and PSNR = 10 log10(MAX^2/MSE), where MAX is the maximum possible pixel value. Structural similarity (SSIM) provides a perceptually oriented image-quality measure. Payload may be expressed in total bits, bits per frame, bits per pixel or bits per second. Extraction reliability is commonly reported as bit error rate (BER), where lower values indicate more accurate payload recovery.

2.3 Interpretation rules

Quantitative values are reported only when they were explicitly available in the cited study and directly relevant to the analysis. Qualitative ratings in the comparison tables represent typical behavior of an approach family rather than universal guarantees. Terms such as high capacity or strong robustness should therefore be interpreted relative to the embedding domain, attack model and payload type used by each study.

3. Taxonomy of Video Steganography Approaches

Table 1. Main families of video steganography approaches.

Approach family	Typical embedding location	Main strengths	Main weaknesses
Spatial-domain	Pixel LSBs or selected color components	High payload; simple and fast; easy blind extraction	Weak against compression, filtering, frame manipulation and statistical analysis
Transform-domain	DCT, DWT, SVD or hybrid coefficients	Improved perceptual masking and signal-processing robustness	Lower payload; transform overhead; coefficient selection is sensitive
Motion-vector / codec-aware	Motion vectors, prediction indices, modes or transform syntax	Codec integration; low visible distortion; possible real-time operation	Codec dependence; syntax changes may be detected by specialized steganalysis
ROI / content-adaptive	Edges, skin, faces, moving objects or textured regions	Better masking and selective distortion control	Requires reliable detection/tracking; content-dependent capacity
Randomized / key-controlled	Secret-key-selected frames, blocks or coefficients	Reduces predictable embedding location; supports access control	Randomization alone does not ensure robustness or undetect-ability

Approach family	Typical embedding location	Main strengths	Main weaknesses
Optimization / hybrid	Optimized pixels, frames, regions or coefficients	Balances multiple objectives; can combine encryption and error correction	Higher computation; parameter and fitness-function dependence
Robust transmission	Stable spatial/transform regions or side-channel-assisted locations	Designed for transcoding, down sampling or social-media pipelines	Often lower payload and more synchronization overhead
Deep neural / generative	Learned spatial-temporal residuals, latent variables or network parameters	High flexibility, distributed hiding, end-to-end optimization	Training cost, generalization risk, difficult security interpretation

3.1 Spatial-domain and LSB methods

Spatial-domain methods directly alter pixel intensities, commonly by replacing one or more least significant bits. Their principal advantage is payload capacity: every eligible pixel can carry one or more bits without requiring a transform or codec modification. Their low computational cost also makes them suitable for prototypes, low-power systems and applications in which the carrier is transmitted losslessly.

The central weakness is fragility. Lossy compression, resizing, filtering, color conversion, frame dropping and re-encoding can modify the embedded bits. Sequential use of frames or pixels also creates predictable embedding paths. Statistical steganalysis may detect abnormal bit-plane distributions, especially at high payloads. Hash-based or adaptive LSB schemes attempt to improve location selection, but the underlying modified-bit pattern remains vulnerable when the video passes through a lossy platform (Sadek et al., 2015; Liu et al., 2019).

3.2 Transform-domain methods

Transform-domain methods first represent a frame in a frequency or matrix-decomposition domain. DCT-based schemes alter selected frequency coefficients, while DWT-based schemes distribute information across wavelet sub-bands. SVD-based schemes modify singular values that are relatively stable under some image-processing operations. Hybrid methods combine these transforms with error-correcting codes, encryption or adaptive region selection.

Mstafa and Elleithy (2017) combined DCT-domain embedding with error-correcting codes to improve secure communication and recovery. Pilania et al. (2022) used ROI selection with SVD in the wavelet domain, illustrating how transform stability and perceptual regions can be combined. Transform methods usually improve robustness and imperceptibility over direct high-payload LSB substitution, but their effectiveness depends on coefficient choice, quantization, reconstruction accuracy and the attack model.

3.3 Motion-vector and compressed-domain methods

Compressed video contains coding decisions that can serve as carriers, including motion vectors, motion-vector differences, prediction indices, partition modes and quantized transform coefficients. Embedding in these elements may avoid full pixel-domain manipulation and can be integrated into the encoder. Aly (2011) selected motion vectors according to their associated prediction error, establishing an influential direction for motion-vector data hiding.

Modern codec-aware methods increasingly preserve properties exploited by steganalysis. Yang et al. (2023) designed adaptive HEVC embedding with distortion compensation, while

Li et al. (2025) combined the motion-vector prediction index and motion-vector difference to improve security under HEVC. Mohamed et al. (2025) embedded during motion estimation while preserving local optimality and consistency, demonstrating that real-time operation and anti-steganalysis design can be considered together. The disadvantage is dependence on the codec, encoder configuration and specialized bitstream knowledge.

3.4 Region-of-interest and content-adaptive methods

Content-adaptive approaches select visually complex or semantically meaningful regions instead of distributing payload uniformly. Edges, textures, moving objects, faces and skin regions can mask small modifications more effectively than smooth backgrounds. Mstafa and Elleithy (2016) combined Kanade-Lucas-Tomasi tracking with error-correcting codes, whereas Sadek et al. (2017) used adaptive skin-tone detection as the embedding region.

These methods can improve imperceptibility and focus the payload in areas likely to survive motion or object tracking. However, capacity varies with content and failures in detection or tracking can cause synchronization errors. There is also a security concern: repeatedly embedding in the same semantic class of regions can create a selection-channel signature for an informed steganalyzer.

3.5 Optimization-based and hybrid methods

Optimization techniques use a search procedure to select embedding positions, bit patterns or parameter values that balance visual distortion and payload. Dasgupta et al. (2013) improved a 3-3-2 LSB video scheme using a genetic algorithm and an anti-steganalysis objective. Hybrid systems may also combine encryption, transforms, object tracking and error-correcting codes.

The advantage of optimization is explicit multi-objective control. The disadvantage is computation, dependence on the selected fitness function and limited comparability when studies optimize different metrics. An optimization procedure that maximizes PSNR, for example, does not automatically maximize resistance to a trained steganalyzer.

3.6 Robust social-media and resampling-resistant methods

Practical video transmission frequently includes transcoding, bitrate reduction, scaling, frame-rate conversion or thumbnail generation. Robust methods therefore model the processing channel rather than assuming an unchanged stego bitstream. Fan et al. (2022) used principal component analysis to select robust embedding regions and transmitted compressed side information to synchronize extraction after social-media transcoding. Wang (2024) focused on downsampling-resilient hiding in thumbnail videos.

Such approaches are more deployment-oriented than methods evaluated only on unprocessed stego videos. Their cost is usually lower payload, additional synchronization information and dependence on the specific platform transformation. A method robust to one service or encoder profile may fail when the service changes its processing pipeline.

3.7 Deep-learning, invertible and implicit-representation methods

Deep learning changed the design objective from manually selecting embedding locations to learning a distributed encoding and decoding transformation. Baluja (2017) demonstrated full-image hiding with neural preparation, hiding and reveal networks. Hayes and Danezis (2017) introduced adversarial interaction among an encoder, decoder and steganalyzer. HiDDeN learned robust message hiding under simulated distortions (Zhu et al., 2018), while SteganoGAN emphasized high-capacity adversarial embedding (Zhang et al., 2019). StegaStamp trained an end-to-end system for recovery after physical imaging distortions (Tancik et al., 2020).

Invertible networks provide a reversible mapping that reduces information loss during hiding and recovery. HiNet used wavelet-domain invertible blocks for image hiding (Jing et al., 2021) and RIIS added robustness to noise and lossy processing (Xu et al., 2022). These image-oriented

developments directly informed modern video hiding.

LF-VSN used an invertible network to hide multiple secret videos in one cover video and introduced key-controllable recovery and scalable capacity (Mou et al., 2023). StegaNeRV explored embedding within implicit neural video representations, shifting the carrier from a conventional video file toward a learned representation (Biswal et al., 2024). Neural systems provide flexible payload types and can optimize temporal losses, but they require training data, hardware, careful distortion modeling and rigorous testing against adaptive steganalysis.

4. Randomized Frame Selection as a Case Study

Din et al. (2025) proposed a randomized frame selection based video steganography method for secure embedding of secret data. The method uses a 64-bit secret key and combines XOR, complement and logarithmic operations to derive a frame index within the available frame range. The secret plaintext is encrypted before being converted to bits and embedded into pixel LSBs of the selected frame. At extraction, the receiver uses the same key to reproduce the frame index, recover the embedded bits and decrypt the message.

The principal contribution is not a new pixel-modification primitive; it is the addition of key-dependent temporal uncertainty to a spatial LSB method. A fixed or sequential frame strategy exposes a predictable search space, whereas a key-derived frame index makes the embedding location unknown to an observer who lacks the secret key. Encryption provides a separate protection layer by preventing immediate interpretation of an extracted bitstream.

The reported experiments show very small pixel-domain distortion, including a maximum PSNR of 74.15 dB and an MSE of 0.0002. These values indicate excellent imperceptibility for the tested samples. However, visual quality should not be equated with complete steganographic security. The method remains based on LSB modification and the published evaluation does not establish resilience against H.264/H.265 transcoding, frame deletion, frame reordering, frame-rate conversion, resizing, cropping, noise or learning-based steganalysis.

The method therefore occupies an important middle position. It is more secure than a deterministic LSB baseline because it conceals the frame location and encrypts the payload, but it is less robust than transform-domain or channel-aware methods under lossy processing. It is also much lighter than neural methods because it requires no training or GPU inference. Its most promising role is as a temporal keying component within a hybrid architecture rather than as a complete solution to all robustness and detectability problems.

5. Results and Comparative Analysis

The comparative result is a qualitative synthesis because the reviewed studies do not share a common dataset, payload definition, codec, resolution or attack protocol. Tables 2 and 3 summarize representative methods and the typical trade-offs of each approach family.

Table 2. Representative studies and their principal contributions.

Study	Domain / carrier	Principal contribution	Primary evaluation focus
Aly (2011)	Compressed video motion vectors	Selects motion vectors using associated prediction error	Visual quality and motion-vector suitability
Dasgupta et al. (2013)	Spatial LSB plus GA	Optimizes a 3-3-2 LSB scheme using a genetic algorithm	Imperceptibility and anti-steganalysis objective

Study	Domain / carrier	Principal contribution	Primary evaluation focus
Mstafa and Elleithy (2016)	Tracked ROI plus ECC	KLT feature tracking with error correction	Capacity, tracking and extraction reliability
Mstafa and Elleithy (2017)	DCT plus ECC	Transform-domain embedding with error-correcting codes	Robustness and secure communication
Sadek et al. (2017)	Adaptive skin-tone ROI	Blind adaptive embedding in detected skin regions	Imperceptibility and ROI robustness
Fan et al. (2022)	PCA-selected spatial regions	Social-media transcoding robustness with side information	BER after platform processing
Pilania et al. (2022)	ROI, wavelet and SVD	Combines semantic selection with stable transform components	Capacity, imperceptibility and robustness
Yang et al. (2023)	HEVC coding domain	Adaptive embedding with distortion compensation	Codec-aware distortion and security
Mou et al. (2023)	Invertible neural video hiding	Multiple-video hiding, scalable capacity, key-controllable recovery	Capacity, flexibility and recovery quality
Biswal et al. (2024)	Implicit neural video representation	Hides distinct information in INR-generated video frames	Representation quality and extraction
Wang (2024)	Thumbnail / downsampled video	Adaptive downsampling-resilient hiding	Recovery after thumbnail generation
Li et al. (2025)	HEVC MVP index and MVD	Preserves security-related motion-vector properties	Resistance to HEVC motion-vector steganalysis
Mohamed et al. (2025)	H.264 motion estimation	Real-time embedding preserving consistency and local optimality	Runtime and anti-steganalysis behavior
Din et al. (2025)	Key-selected frame plus LSB	Randomized frame selection, encryption and lightweight embedding	Imperceptibility and location secrecy

Table 3. Typical qualitative trade-offs among approach families.

Family	Capacity	Imperceptibility	Processing robustness	Steganalysis resistance	Complexity
Basic LSB	High	Medium-high at low payload	Low	Low	Low
Randomized/keyed LSB	High	High in tested lossless settings	Low	Medium for location secrecy; LSB traces remain	Low-medium

Family	Capacity	Imperceptibility	Processing robustness	Steganalysis resistance	Complexity
DCT/DWT/SVD	Medium	High	Medium-high	Medium	Medium-high
Motion-vector / codec-aware	Low-medium	High	Medium	Medium-high if coding statistics are preserved	High
ROI / content-adaptive	Content-dependent	High	Medium	Medium	Medium
Optimization / hybrid	Medium-high	High	Method-dependent	Method-dependent	High
Social-media robust	Low-medium	High	High for modeled channels	Medium-high	High
Deep neural / invertible	High to very high	High to very high	Medium-high when trained with distortions	Potentially high but attack-dependent	Very high
Implicit / generative	High and flexible	Very high	Emerging evidence	Not yet fully established	Very high

5.1 Capacity

Spatial LSB schemes usually provide the highest straightforward bit payload because many pixels can be used. Neural methods can exceed conventional payload definitions by hiding full images or multiple videos, as demonstrated by LF-VSN. However, full-video hiding and binary message hiding are not directly comparable. Future papers should report both normalized binary capacity and the size and quality of reconstructed multimedia payloads.

5.2 Imperceptibility

Most reviewed methods report high PSNR or SSIM under their selected operating points. Very high PSNR values, such as those reported by Din et al. (2025), indicate low average pixel distortion but do not prove resistance to steganalysis. Neural methods may optimize perceptual losses that distribute residuals more naturally, but their security must be measured using independent detectors rather than visual metrics alone.

5.3 Robustness and practical transmission

Transform-domain, social-media-aware and distortion-trained neural approaches are generally stronger under compression and resampling than unprotected LSB embedding. Fan et al. (2022) and Wang (2024) directly model realistic delivery transformations. Codec-aware schemes can survive within a controlled encoding pipeline, although re-encoding by another codec may destroy syntax-level payloads. Randomized frame selection does not itself protect against loss of the selected frame.

5.4 Security against steganalysis

Security depends on whether embedding preserves the statistical distribution of the carrier. Motion-vector methods are increasingly evaluated against detectors that exploit local optimality, consistency, skipped blocks and coding decisions. Recent HEVC and H.264 schemes explicitly preserve these properties (Li et al., 2025; Mohamed et al., 2025). Neural methods can include an adversarial steganalyzer during training, but they remain vulnerable to detector mismatch, cover-source mismatch and adaptive

attacks. Key secrecy and encryption should not be used as substitutes for steganalysis evaluation.

5.5 Computational and deployment cost

Basic LSB and randomized-frame methods are lightweight and can be implemented without training. Transform and codec-aware methods require more signal-processing or encoder integration. Deep neural systems offer the greatest flexibility but introduce model storage, GPU or accelerated inference, training-data requirements and reproducibility challenges. The best method therefore depends on whether the application prioritizes lightweight lossless exchange, robust platform transmission, real-time encoding or high-capacity multimedia concealment.

6. Discussion

The literature demonstrates that video steganography is not a single optimization problem. A method that is optimal for lossless local transfer may be unsuitable for social media. A method that is robust to transcoding may sacrifice payload. A codec-integrated method may be secure within H.264 but incompatible with HEVC or VVC. Similarly, high visual quality does not ensure low detectability and encryption does not prevent a steganalyzer from identifying that hidden communication exists.

Randomized frame selection is a valuable security mechanism because predictable carrier selection is a genuine weakness. The Din et al. (2025) method addresses this issue clearly and with low computational cost. Its security contribution should nevertheless be separated into three components: message confidentiality from encryption, location uncertainty from key-derived frame selection and perceptual invisibility from low-distortion LSB modification. An ablation study would show how much each component contributes.

A direct comparison with modern alternatives requires a common testbed. The same cover sequences should be encoded with fixed resolutions, frame rates and payload rates. Each method should then be tested under lossless transfer, H.264 and H.265 compression, bitrate reduction, scaling, frame dropping, frame-rate conversion, cropping, color conversion and social-media transcoding. Security should be measured using classical features and independently trained deep steganalyzers. Without this protocol, comparisons based only on the highest reported PSNR are scientifically weak.

Deep video hiding introduces new possibilities and new risks. Invertible networks reduce information loss and enable large payloads, while implicit representations can store information in network parameters or generation processes. However, high model capacity may memorize training distributions and reported robustness can depend heavily on simulated distortions. Neural systems should therefore be evaluated across unseen datasets, codecs and steganalyzers, with open code and exact preprocessing settings.

7. Research Gaps and Future Directions

7.1 Standardized datasets and payload units

Studies should use public video sets with diverse motion, texture, scene cuts, resolutions and codecs. Payload should be reported in bits per pixel, bits per frame and bits per second, together with the number and type of modified frames.

7.2 Temporal-security analysis

Frame-level PSNR does not reveal temporal artifacts. Future work should measure optical-flow consistency, temporal SSIM, flicker, inter-frame residual changes and detector features derived from adjacent frames.

7.3 Robust randomized selection

Randomized frame selection should be extended from a single key-selected frame to a keyed sequence with redundancy, interleaving and error-correcting codes so that frame loss or reordering does not destroy the payload.

7.4 Steganalysis-aware hybrid design

A promising architecture would combine key-controlled frame and region selection, transform or codec-aware embedding, payload encryption, error correction and an adversarial detector trained on both spatial and temporal features.

7.5 Realistic platform channels
Training and testing should model current social-media transformations, including variable bitrate, resolution ladders, thumbnail extraction, format conversion and repeated upload-download cycles.

7.6 Explainability and security proofs
Neural approaches require clearer analysis of where information is stored, how keys control access, how security changes under model leakage and whether the system remains secure against adaptive white-box steganalysis.

7.7 Efficiency and edge deployment
Lightweight neural networks, quantized models and encoder-integrated approaches are needed for mobile, IoT, telemedicine and real-time communication environments.

7.8 Reproducibility
Authors should release code, trained weights, payload-generation scripts, cover-video lists, codec commands, random seeds and steganalysis models. This is essential for credible comparison.

8. Recommended Evaluation Protocol
Table 4. Minimum evaluation protocol for future comparative studies.

Dimension	Required tests	Recommended reporting
Visual quality	MSE, PSNR, SSIM and a perceptual metric	Mean, standard deviation and per-sequence range
Payload	Multiple normalized payload levels	Bits/pixel, bits/frame, bits/second, payload type
Recovery	Clean and attacked channels	BER, exact-message recovery rate, confidence intervals
Temporal quality	Adjacent-frame and motion-consistency analysis	Temporal SSIM, flicker or flow-residual measure
Robustness	H.264/H.265, resizing, crop, noise, frame drop, frame-rate conversion	Performance curves against attack severity
Security	Classical and deep steganalysis; unseen detectors	Detection accuracy, AUC, false-positive and false-negative rates
Key security	Wrong-key extraction and key sensitivity	Key space, collision analysis, extraction failure rate
Efficiency	Encoding and decoding on defined hardware	Runtime, memory, model size, bitrate overhead
Reproducibility	Public code and exact codec commands	Dataset split, seeds, parameters, software versions

9. Conclusion
This paper compared the principal approaches developed for video steganography, including spatial LSB, transform-domain, motion-vector and codec-aware, ROI-based, optimization-driven, robust-transmission, randomized-frame and deep neural methods. No approach dominates every criterion. Spatial methods provide simplicity and capacity but are fragile; transform and codec-aware methods improve robustness and statistical control but increase complexity; content-adaptive methods improve perceptual masking but have variable capacity; and neural methods offer flexible high-capacity hiding while introducing substantial training, generalization and security-evaluation requirements.

The randomized frame selection method of Din et al. (2025) is a useful advancement over predictable LSB embedding. Its key-derived frame index and encrypted payload strengthen location secrecy and confidentiality and the reported PSNR and MSE indicate very low distortion in the evaluated videos. Nevertheless, its dependence on LSB modification and its limited attack evaluation mean that robustness and resistance to modern steganalysis remain open questions.

The most credible future direction is a hybrid, reproducible framework that unifies key-controlled temporal selection, content-adaptive or codec-aware embedding, error correction, realistic channel modeling and steganalysis-aware optimization. Progress should be judged using standardized datasets, normalized payloads, temporal-quality measures, attack-based recovery, independent steganalysis and transparent reporting rather than isolated visual-quality values.

Acknowledgements
Removed from the blinded manuscript. See the separate title-page file.

Funding
Removed from the blinded manuscript. See the separate title-page file.

Conflict of Interest
The authors declare no conflict of interest.

Author Contributions
Removed from the blinded manuscript to protect author identity. The final accepted version will include a CRediT author-contribution statement.

Data Availability Statement
No new data were created or analyzed. The comparative findings are based on the published literature cited below.

Acknowledgements
The authors acknowledge no external assistance.

Funding Statement
This research received no external funding.

Declaration of Interest
The authors declare that they have no known financial or personal conflicts of interest that could have influenced the work reported in this manuscript.

Author Contributions
All authors have equal contributions.

Data Availability
No new dataset was generated for this comparative review. All findings are derived from the published studies cited in the manuscript.

References

Aly, H. A. (2011). Data hiding in motion vectors of compressed video based on their associated prediction error. IEEE Transactions on Information Forensics and Security, 6(1), 14-18. https://doi.org/10.1109/TIFS.2010.2090520

Baluja, S. (2017). Hiding images in plain sight: Deep steganography. In Advances in Neural Information Processing Systems 30 (pp. 2066-2076).

Biswal, M., Shao, T., Rose, K., Yin, P. and McCarthy, S. (2024). StegaNeRV: Video steganography using implicit neural representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 888-898). https://doi.org/10.1109/CVPRW63382.2024.00094

Dasgupta, K., Mondal, J. K. and Dutta, P. (2013). Optimized video steganography using genetic algorithm (GA). Procedia Technology, 10, 131-137. https://doi.org/10.1016/j.protcy.2013.12.345

Din, M. ud, Shoukat, I. A., Arif, E., Amjad, M., Mohsin, M., Mumtaz, M. and Rauf, M. A. (2025). Randomized frame selection based video steganography method for secure embedding of secret data. Journal of Computing & Biomedical Informatics, 8(2). https://doi.org/10.56979/802/2025

IEER Institute for Excellence in Education & Research

Fan, P., Zhang, H. and Zhao, X. (2022). Robust video steganography for social media sharing based on principal component analysis. EURASIP Journal on Information Security, 2022, Article 4. https://doi.org/10.1186/s13635-022-00130-z

Hayes, J. and Danezis, G. (2017). Generating steganographic images via adversarial training. In Advances in Neural Information Processing Systems 30.

Jing, J., Deng, X., Xu, M., Wang, J. and Guan, Z. (2021). HiNet: Deep image hiding by invertible network. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4733-4742). https://doi.org/10.1109/ICCV48922.2021.00469

Kunhoth, J., Subramanian, N., Al-Maadeed, S. and Bouridane, A. (2023). Video steganography: Recent advances and challenges. Multimedia Tools and Applications, 82, 41943-41985. https://doi.org/10.1007/s11042-023-14844-w

Li, J., Zhang, M., Niu, K., Zhang, Y., Ke, Y. and Yang, X. (2025). High-security HEVC video steganography method using the motion vector prediction index and motion vector difference. Tsinghua Science and Technology, 30(2), 813-829. https://doi.org/10.26599/TST.2024.9010016

Liu, Y., Liu, S., Wang, Y., Zhao, H. and Liu, S. (2019). Video steganography: A review. Neurocomputing, 335, 238-250. https://doi.org/10.1016/j.neucom.2018.09.091

Mohamed, H., Elliethy, A., Abdelaziz, A. and Aly, H. (2025). Real-time motion estimation based video steganography with preserved consistency and local optimality. Multimedia Tools and Applications, 84, 5001-5024. https://doi.org/10.1007/s11042-024-18651-9

Mou, C., Xu, Y., Song, J., Zhao, C., Ghanem, B. and Zhang, J. (2023). Large-capacity and flexible video steganography via invertible neural network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 22606-22615).

Mstafa, R. J. and Elleithy, K. M. (2016). A video steganography algorithm based on Kanade-Lucas-Tomasi tracking algorithm and error correcting codes. Multimedia Tools and Applications, 75, 10311-10333. https://doi.org/10.1007/s11042-015-3060-0

Mstafa, R. J. and Elleithy, K. M. (2017). An ECC/DCT-based robust video steganography algorithm for secure data communication. Journal of Cyber Security and Mobility, 5(3), 167-194. https://doi.org/10.13052/2245-1439.531

Mstafa, R. J., Elleithy, K. M. and Abdelfattah, E. (2017). Video steganography techniques: Taxonomy, challenges and future directions. In 2017 IEEE Long Island Systems, Applications and Technology Conference (LISAT) (pp. 1-6). https://doi.org/10.1109/LISAT.2017.8001965

Petitcolas, F. A. P. anderson, R. J. and Kuhn, M. G. (1999). Information hiding - A survey. Proceedings of the IEEE, 87(7), 1062-1078. https://doi.org/10.1109/5.771065

Pilania, U., Tanwar, R. and Gupta, P. (2022). An ROI-based robust video steganography technique using SVD in wavelet domain. Open Computer Science, 12(1), 1-16. https://doi.org/10.1515/comp-2020-0229

Provos, N. and Honeyman, P. (2003). Hide and seek: An introduction to steganography. IEEE Security & Privacy, 1(3), 32-44. https://doi.org/10.1109/MSECP.2003.1203220

Sadek, M. M., Khalifa, A. S. and Mostafa, M. G. M. (2015). Video steganography: A comprehensive review. Multimedia Tools and Applications, 74, 7063-7094. https://doi.org/10.1007/s11042-014-1952-z

Sadek, M. M., Khalifa, A. S. and Mostafa, M. G. M. (2017). Robust video steganography algorithm using adaptive skin-tone detection. Multimedia Tools and Applications, 76, 3065-3085. https://doi.org/10.1007/s11042-015-3170-8

Tancik, M., Mildenhall, B. and Ng, R. (2020). StegaStamp: Invisible hyperlinks in physical photographs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2117-2126).

Wang, Y. (2024). Hiding data within thumbnail videos: An adaptive downsampling-resilient video steganography method. IEEE Access, 12, 52963-52977. https://doi.org/10.1109/ACCESS.2024.3386798

Xu, Y., Mou, C., Hu, Y., Xie, J. and Zhang, J. (2022). Robust invertible image steganography. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7875-7884). https://doi.org/10.1109/CVPR52688.2022.00772

Yang, L., Xu, D., Wang, R. and He, S. (2023). Adaptive HEVC video steganography based on distortion compensation optimization. Journal of Information Security and Applications, 73, 103442. https://doi.org/10.1016/j.jisa.2023.103442

Zhang, K. A., Cuesta-Infante, A., Xu, L. and Veeramachaneni, K. (2019). SteganoGAN: High capacity image steganography with GANs. arXiv:1901.03892.

Zhu, J., Kaplan, R., Johnson, J. and Fei-Fei, L. (2018). HiDDeN: Hiding data with deep networks. In Proceedings of the European Conference on Computer Vision (pp. 657-672).

IEER
Institute of Excellence in Education & Research