2026 Volume: 02 Issue: 11 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-16

# A REVIEW ON IMAGE TAMPERING DETECTION: FROM TRADITIONAL TO ADVANCED COMPUTATIONAL METHODS

***Shravani S. Bansod, Srushti P. Ghom, Prof. Sumit B. Dhande**

P.G Department of Computer Science and Technology, HVPM Amravati, Maharashtra, India.

**Article Received: 27 September 2026**
**Article Revised: 17 October 2026**
**Published on: 06 November 2026**

***Corresponding Author: Shravani S. Bansod**
P.G Department of Computer Science and Technology, HVPM Amravati, Maharashtra, India.

DOI: https://doi-doi.org/101555/ijrpa.6888

## ABSTRACT

The rapid development of accessible editing software and modern generative AI tools has made creating visually imperceptible digital manipulations effortless, posing severe threats to digital media authenticity, journalism, and forensics. Image tampering detection encompassing splicing, copy-move duplication, and localized retouching relies on identifying underlying statistical, physical, or semantic inconsistencies. Historically, the field leaned heavily on handcrafted traditional mechanisms, such as keypoint matching, frequency-domain transformations, compression artifact analysis, and sensor pattern noise. Over the past decade, the domain has pivoted toward advanced data-driven paradigms, progressing from standard convolutional networks to dual-stream architectures, attention mechanisms, vision transformers, and multi-modal foundation models.This paper provides a comprehensive review tracking the technical transition from traditional image forensics to advanced deep learning paradigms. We examine the core operational principles of both families of methods, classifying traditional approaches by their underlying physical and statistical assumptions, and categorizing advanced architectures by feature extraction, boundary refinement, and context-modelling strategies. Furthermore, we survey widely utilized benchmark datasets, standard evaluation metrics (such as pixel-level F1-score and Intersection over Union), and cross-dataset performance drops. Finally, we highlight pressing open problems, including detector vulnerability to social media re-compression, counter-forensics, and the detection of hybrid tampering involving generative inpainting.

**KEYWORDS:** Image Tampering Detection, Digital Image Forensics, Splicing, Copy-Move, Handcrafted Features, Deep Learning, Vision Transformers, Artifact Analysis.

# 1. INTRODUCTION

## 1.1 Background and Problem Significance

The rapid growth of digital imaging technologies and social media platforms has led to an unprecedented increase in the creation, sharing, and consumption of visual content. Millions of images are circulated daily across public channels for personal expression, social communication, journalism, and legal documentation . However, the concurrent evolution of accessible, advanced photo-editing software (such as Photoshop and GIMP) and modern automated manipulation tools has made photo tampering remarkably effortless and visually imperceptible.[1], [2].

When manipulated visual content is deployed maliciously, it undermines the credibility of digital evidence in judicial processes, compromises law enforcement investigations, and accelerates the viral spread of online misinformation and fake news [3]. Consequently, digital image forensics has emerged as a critical discipline in cyber security and multimedia integrity, aiming to design reliable methods that can verify image authenticity, uncover tampering, and preserve the digital chain of custody against anti-forensic countermeasures [4], [5], [6], [7].

## 1.2 Taxonomy of Image Tampering

Image manipulations range from subtle post-processing modifications to complex multi-source composition. Passive image forensics broadly classifies visual tampering into three foundational categories:

- **Image Splicing:** This technique involves taking one or more regions from donor images and combining them into a separate target image to construct a misleading composite. Because donor regions originate from different devices or environments, splicing frequently introduces anomalies in camera sensor patterns, compression artifacts, and boundary edges.
- **Copy-Move Forgery:** An intra-image manipulation where a specific region is duplicated and pasted elsewhere within the very same image. This is typically performed to either conceal an unwanted object or clone an existing entity to alter semantic meaning. Detecting copy-move manipulations is particularly challenging because the cloned segment shares identical lighting, noise profiles, and camera sensor characteristics with the authentic image background, often compounded by geometric adjustments like scaling, rotation, or JPEG compression.
- **Removal and Imprinting:** This involves erasing an object entirely from an image and infilling the missing area using neighboring textures or learning-based generative imprinting algorithms, effectively removing visual evidence without obvious geometric discontinuities.

While whole-image binary classification determines whether an image has been manipulated, practical forensic investigation requires fine-grained localization to pinpoint the precise pixel-level boundaries of the altered regions [8], [9].

### 1.3 Evolution of Detection Paradigms and Paper Objectives

To expose image tampering, classical forensic approaches historically relied on handcrafted, physical, and statistical features. These techniques extracted block-level frequency transformations (such as the Discrete Cosine Transform or Discrete Wavelet Transform), matched keypoint descriptors (e.g., SIFT, SURF), analyzed Photo-Response Non-Uniformity (PRNU) sensor fingerprints, or evaluated double-compression inconsistencies via Error Level Analysis (ELA) and JPEG ghost traces [10]. Although theoretically sound and interpretable, these handcrafted algorithms often require manual parameter tuning, incur high computational latency on large resolutions, and degrade severely when tampered media undergoes routine post-processing operations such as lossy re-compression, spatial resizing, or Gaussian blurring .

To overcome these constraints, recent research has shifted decisively toward advanced data-driven deep learning architectures [9], [11], [12], [13]. Modern frameworks leverage convolutional neural networks (CNNs), transfer learning models, two-stream architectures, attention modules, and graph neural networks (GNNs) to automatically extract rich, hierarchical forensic representations directly from raw pixels and noise residuals without explicit feature engineering [8], [9], [14], [15]. These data-driven systems have achieved superior classification accuracy and greater adaptability to complex, multi-type forgeries [9].

The primary objective of this review paper is to trace this technological transition in digital image forensics. We comprehensively examine:

1. The mathematical principles and inherent limitations of classical, handcrafted tampering detection techniques [10].
2. The evolution of advanced deep learning and hybrid architectures designed for automated detection and pixel-level localization [8], [11].
3. The benchmark datasets, standard performance evaluation metrics, and comparative trade-offs between traditional and advanced paradigms [9].
4. The remaining open challenges confronting the field, including cross-dataset generalization, counter-forensics, and social media compression degradation [7], [16], [17].

## 2. Traditional Image Tampering Detection Techniques

Before the widespread adoption of data-driven neural networks, digital image forensics relied on handcrafted mathematical representations and intrinsic hardware artifacts.These traditional techniques operated under the assumption that any manipulation—such as duplicating an intra-image element or splicing an external object—disturbs the underlying statistical consistency, physical sensor footprints, or compression history of the authentic image .

### 2.1 Block- and Keypoint-Based Approaches

Classical forgery localization, particularly for copy-move manipulations, primarily developed along two algorithmic directions: block-based partitioning and sparse keypoint matching .

- **Block-Based Methods:** The input image is divided into overlapping square pixel blocks. Mathematical transforms, such as the Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT), are applied to each block to extract compact, low-dimensional feature vectors. Similar blocks are subsequently identified by lexicographical sorting or Euclidean distance matching. Because identical features appearing in non-contiguous regions signify duplication, these methods effectively spot cloned regions. However, standard block matching assumes identical orientation and scale, degrading sharply when copied fragments undergo spatial transformations such as rotation or resizing .
- **Keypoint-Based Descriptors:** To provide resilience against affine distortions, keypoint extraction algorithms such as Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) were introduced. These techniques detect interest points across multiple image scales and compute local gradient descriptors that remain invariant to illumination shifts, scaling, and rotation. Cloned regions are located by computing feature similarities between pairs of keypoints and clustering corresponding geometric transformation vectors. Nevertheless, keypoint approaches fail when tampered segments are small or fall within smooth, textureless regions where stable keypoints cannot be extracted.

### 2.2 Physical and Hardware Sensor Fingerprints (PRNU)

Every digital image acquisition pipeline leaves hardware-specific artifacts embedded in the recorded pixel values. The most prominent physical identifier is **Photo-Response Non-Uniformity (PRNU)**, an intrinsic, deterministic noise pattern caused by microscopic silicon irregularities in the camera's charge-coupled device (CCD) or complementary metal-oxide-semiconductor (CMOS) sensor.

PRNU functions as a unique camera fingerprint. In classical forensic analysis, the sensor noise residual is estimated by subtracting a denoised version of the image from the original query image using high-pass or wavelet-based denoising filters. In an authentic photograph,

the PRNU pattern remains uniform and mutually correlated across all regions. When a splicing attack occurs, the spliced fragment originates from a foreign camera sensor with an alien PRNU profile. By computing the cross-correlation between local image patches and the estimated camera reference fingerprint, investigators can localize foreign objects that lack sensor noise correlation.Despite its theoretical rigor, PRNU estimation requires either prior access to multiple authentic reference images from the host device or flat, well-exposed query scenes, making it impractical for blind investigations of single, compressed web images .

**2.3 Compression Artifacts and Inconsistencies (JPEG Ghosts & ELA)**

Because the vast majority of digital photographs are stored and shared in the lossy JPEG format, compression-specific statistical anomalies serve as essential forensic cues [10].

- JPEG Ghosts and Double Compression: When a region from one JPEG image is spliced into another, the composite is re-saved, subjecting the foreign patch to double quantization. If the secondary saving process uses a compression quality factor different from the original, distinct periodic errors emerge in the Discrete Cosine Transform (DCT) coefficient distributions [10]. JPEG ghost analysis re-compresses the tampered image across a sequence of simulated quality factors; the difference between the re-compressed versions and the suspicious image highlights the forged region—termed a "ghost"—which reaches minimum displacement when the re-compression quality matches that of the host cover image .
- Error Level Analysis (ELA): ELA operates on a similar principle by re-saving the entire image at a fixed, predetermined compression rate (e.g., 95%) and calculating the absolute difference map between the original and the re-compressed versions [2]. Because previously edited or inserted areas compress at different error rates compared to surrounding unaltered regions, the resulting error map displays visible brightness discrepancies around forged boundaries [2].

**2.4 Fundamental Limitations of Handcrafted Methods**

While traditional handcrafted methods offer high interpretability and do not require massive amounts of labeled training data, they exhibit severe technical bottlenecks that limit their deployment in modern forensic scenarios:

**1. Vulnerability to Post-Processing and Noise:** Handcrafted features rely on fragile pixel-level statistical assumptions [18]. When tampered images undergo aggressive post-processing such as Gaussian blurring, noise contamination, or median filtering—the subtle high-frequency traces, compression ghosts, and PRNU patterns are often erased or severely corrupted [18] .

**2. Degradation Under Social Media Transcoding:** Uploading images to social media platforms subjects them to automated pipelines involving severe lossy re-compression, metadata stripping, and spatial downscaling. Traditional descriptors like ELA and block-based DCT coefficients degrade significantly under repeated quantization, yielding high false-positive rates.

**3. High Computational Complexity:** Dense block-matching operations across high-resolution imagery require pairwise comparisons that scale quadratically with image dimensions, creating substantial memory and execution bottlenecks that preclude real-time analysis.

**4. Lack of End-to-End Automation:** Traditional algorithms are siloed; a PRNU estimator cannot detect copy-move duplication within the same sensor profile, while a keypoint matcher cannot identify smooth spliced backgrounds. This lack of generalization and unified analysis motivated the recent transition toward adaptive, multi-scale deep learning models [9], [12].

## 3. Advanced Deep Learning Methods

To overcome the fragility, manual intervention, and high false-alarm rates characteristic of handcrafted forensic techniques, recent research has shifted decisively toward deep learning frameworks [9], [11], [12], [13], . Rather than relying on manually designed rules, advanced data-driven models automatically learn hierarchical, discriminative feature representations directly from input pixel arrays, residual noise layers, and frequency maps [8], [14], . This shift has transformed passive forensic analysis from isolated binary categorization into unified, automated frameworks capable of high-precision pixel-level forgery localization [8], [9], .

### 3.1 Convolutional Neural Networks (CNNs) and Semantic Segmentation Models

Convolutional neural networks serve as the foundational backbone for modern image tampering detection due to their inherent ability to capture spatial linkages, local texture statistics, and high-resolution feature hierarchies [9], [11], [15], . Standard architectures—such as VGG16, MobileNet, and ResNet-50—have been extensively repurposed and fine-tuned for forensic classification tasks [9], [15], .

- Feature Learning and Transfer Learning: Training deep networks from scratch requires vast volumes of densely annotated forensic data [9], . To resolve data scarcity, transfer learning models pre-trained on large-scale natural image repositories are adapted for tampering identification [2], [9], [15]. By replacing generic classification heads with tailored

binary classifiers and fine-tuning convolutional filters, deep models achieve rapid convergence and elevated detection accuracy with fewer training parameters [9], [15], .

● Pixel-Level Localization Pipelines: Moving beyond whole-image labels, semantic segmentation architectures—such as Fully Convolutional Networks (FCNs), U-Net, and Mask R-CNN—are employed to generate binary masks that highlight tampered boundaries with pixel-level precision [8], [9], . Through encoder-decoder pathways and skip connections, these networks preserve fine-grained spatial resolutions, allowing forensic investigators to precisely isolate forged regions and quantify the exact area compromised by manipulation [8], [9], .

● Constrained Forensic Convolutions: Standard computer vision CNNs naturally extract high-level semantic objects (e.g., distinguishing a person from a vehicle) rather than imperceptible manipulation traces [9], . To suppress dominant semantic scene content and force the network to focus on low-level statistical discrepancies, constrained convolutional layers or high-pass residual filtering are positioned at the network's front end, amplifying tampering artifacts prior to deep feature extraction .

### 3.2 Dual-Stream and Multi-Branch Architectures

Because digital tampering modifies both semantic structures (such as object edges) and latent statistical properties (such as compression history and sensor noise), single-stream networks often struggle to resolve conflicting forensic cues [8], . To address this challenge, multi-branch and dual-stream architectures have been widely adopted [8], .

● RGB and Noise Stream Fusion: Dual-channel models systematically decouple the detection task across two parallel streams: an RGB stream that analyzes spatial anomalies, lighting inconsistencies, and boundary artifacts, alongside a specialized noise stream [8] . The noise stream extracts residual patterns using pre-filtering tools such as Spatial Rich Models (SRM) or median filter residuals, thereby highlighting high-frequency inconsistencies introduced during splicing or cloning [8], [18] .

● Multimodal and Cross-Modal Attention: Intermediate representations from both streams are fused through multi-resolution hybrid modules or cross-attention mechanisms [8] . By adaptively weighting the most reliable stream across different image patches, multi-branch models maintain high detection precision even when one modality is compromised by post-processing or aggressive JPEG compression [8], [10].

### 3.3 Graph Neural Networks (GNNs) and Relational Modeling

While convolutional networks excel at extracting local patch characteristics, they are inherently limited by local receptive fields, making it difficult to capture long-range

contextual relationships across distant parts of an image [9], . This presents a major obstacle in copy-move forgery, where an identical object is cloned into a non-adjacent background region [9], .

To bridge this gap, Graph Neural Networks (GNNs) have been integrated into image forensics to model long-range dependencies and structural relationships [9], .

- Graph-Based Structural Formulation: The input image is segmented into distinct superpixels, patches, or visual regions that serve as graph nodes, while spatial proximity and feature similarities (such as color, texture, and compression traits) are encoded as graph edges [9].
- Contextual Aggregation: Through iterative message passing and node feature aggregation, each region's representation is refined based on the properties of its surrounding neighbors [9]. Under this formulation, GNNs readily identify structural anomalies—such as an isolated patch exhibiting an unusually high similarity to a distant region—enabling the network to expose copy-move duplications that evade standard CNN-based local descriptors [9]. Combining CNN feature extractors with relational GNN classifiers has established state-of-the-art benchmarks on standard datasets [9].

### 3.4 Hybrid Frameworks and Advanced Attention Mechanisms

Recent forensic systems increasingly combine multiple paradigms to construct comprehensive, end-to-end tampering pipelines [8], [9], .

- Attention and Transformer Modules: Dual self-attention networks and hierarchical vision transformers model long-range interactions across multiple scales, focusing specifically on boundary transitions between authentic backgrounds and spliced components [8], .
- Compression Discrepancy Analysis: Advanced hybrid networks leverage re-compression differences, analyzing discrepancies between query images and their artificially re-compressed counterparts [9], [10]. Feeding these differential error maps into deep networks allows models to identify double-compression anomalies without manual threshold calibration [9], [10].
- Integration of Denoising and DSNI Architectures: To handle noisy transmission environments, specialized networks such as DSNI-Net combine median denoising layers with deep feature extraction pipelines, maintaining reliable detection performance even when tampered imagery is corrupted by dense salt-and-pepper or Gaussian noise [18].

## 4. Datasets, Performance Metrics, and Comparative Analysis

Evaluating image tampering detection and localization frameworks requires standardized benchmark corpora and rigorous statistical metrics [9], . Because manipulated images exhibit subtle artifacts that vary across manipulation types and compression qualities, comprehensive cross-model benchmarking provides empirical validation for classical and advanced paradigms [9] .

### 4.1 Benchmark Datasets

A variety of public benchmark datasets have been developed to evaluate copy-move, splicing, and multi-type tampering detection algorithms under diverse imaging conditions [9], :

● CASIA Tampering Evaluation Database (CASIA v1.0 and v2.0): CASIA v1.0 contains 1,754 images categorized into 800 authentic images, 480 copy-move tampered images, and 474 spliced composite images, incorporating post-processing variations such as rotation, scaling, and JPEG re-compression [9], . CASIA v2.0 significantly scales the benchmark to 12,614 images, comprising 7,491 authentic and 5,123 tampered images with unconstrained scene complexities and refined edge feathering .

● Columbia Image Splicing Dataset: Focuses exclusively on uncompressed splicing anomalies across homogeneous indoor and outdoor camera settings, providing a controlled baseline for camera-inconsistency evaluation .

● COVERAGE (Copy-Move Forgery Database with Similar but Genuine Objects): Designed specifically for copy-move evaluations, featuring tampered images accompanied by authentic images that contain naturally similar, non-duplicated objects to challenge keypoint and block-matching algorithms [9].

● CoMoFoD: Tailored for copy-move forgery detection under aggressive post-processing transformations, including multi-scale rotation, color space distortions, blurring, and heavy JPEG quantization [2], [9].

● NIST and MICC Benchmarks (NIST16, MICC-F2000): Developed for advanced forensic evaluations, offering diverse spatial resolutions, realistic tampering scenarios, and corresponding ground-truth binary masks for pixel-level boundary localization [9].

### 4.2 Standard Performance Evaluation Metrics

To assess classification reliability and pixel-level segmentation accuracy, image forensic systems are evaluated using standard confusion-matrix measures [9].

- **Accuracy (): Represents the overall proportion of correctly identified authentic and tampered images over the entire test set [9]:** $Acc$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} [14], [19][13]$$

where, , , and denote true positives, true negatives, false positives, and false negatives, respectively [7], [9]. $TPTNFPFN$

- **Precision (): Quantifies the proportion of images flagged as tampered that are genuinely forged, measuring the framework's ability to suppress false-positive alarms [9] :** $Prec$

$$Precision = \frac{TP}{TP + FP} [14], [19][13]$$

- **Recall ( / True Positive Rate): Reflects the capability of the detector to identify all forged instances without missing manipulated content [9] :** $Rec$

$$Recall = \frac{TP}{TP + FN} [14], [19][13]$$

- **F1-Score (): Formulated as the harmonic mean of precision and recall, providing a balanced metric when authentic and tampered sample classes are imbalanced [9] :** $F_1$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} [14], [19][13]$$

Intersection over Union (IoU) and Localization Masks: Evaluates the spatial overlap between predicted binary forgery masks and ground-truth tampered regions, measuring boundary segmentation accuracy [8], [9].

### 4.3 Direct Comparative Synthesis & Analysis

Empirical evaluations demonstrate that while handcrafted methods establish stable mathematical baselines, deep learning and hybrid architectures achieve superior detection and localization performance [9].

**Table: Comparative Synthesis of Image Forgery Detection and Localization Paradigms.**

| Methodology & Architecture | Model Backbone / Feature Mechanism | Primary Target Dataset | Detection Accuracy (%) | F1-Score (%) | Invariance & Robustness | Key Operational Bottlenecks | References |
|---|---|---|---|---|---|---|---|
| Traditional Handcrafted | Block DCT + LBP + SVM Classifier | CASIA v1.0 | 97.50% | Not Reported | Moderate against basic rotation and scaling | Quadratic computational complexity; high false alarms on smooth patches | [2], [9] |
| Traditional Handcrafted | JPEG Ghost & Error Level Analysis (ELA) | CASIA v2.0, Columbia | 92.23% | Not Reported | Detects re-quantization anomalies | Fails on single-compressed images and non-JPEG formats | [10] |
| Transfer Learning CNN | Fine-Tuned VGG16 Architecture | CASIA v1.0 / CASIA v2.0 | 81.70%–97.14% | 95.25%–97.09% | High feature representation across standard forgeries | High parameter footprint; vulnerable to aggressive post-processing | [9] |
| Mobile & Lightweight CNN | MobileNet v1 / ResNet-50 | CASIA v1.0 | 94.37%–94.85% | 86.70%–91.01% | Fast convergence; suitable for resource-constrained setups | Lower recall on subtle, feathered boundary splices | [2], [9] |
| Dual-Channel Segmentation | Dual-Channel U-Net / Regularized U-Net | CASIA v1.0, MICC-F2000 | 97.52% | Not Reported | Generates fine-grained pixel-level boundary masks | High GPU memory demand; slow processing on ultra-high resolutions | [2], [8], [9] |
| Noise-Resilient Hybrid | DSNI-Net (Median Filter + CNN) | CASIA + Noise Perturbations | 96.80% | 95.90% | Robust under salt-and-pepper and Gaussian noise | Tailored primarily for splicing rather than composite copy-move | [1], [18] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Hierarchical Transformer | Multi-Scale Forensic Transformer | Multi-Domain Sets (Diffusion/GAN) | 86.80% | 76.00% | Effective global long-range dependency modeling | Severe accuracy drops on out-of-distribution web data | [11] |
| Relational Graph Learning | GNN + CNN Multi-Branch Fusion | CASIA v1.0 | 98.55% | 98.93% | Captures topological and spatial dependencies | Graph construction overhead on large pixel superpixels | [1], [9] |

## 5. Open Challenges and Conclusion

### 5.1 Open Challenges and Practical Bottlenecks

Despite the remarkable performance gains achieved by deep convolutional, segmentation, and graph-based architectures, substantial technical hurdles impede their widespread transition from controlled laboratory benchmarks to unconstrained, real-world forensic deployment [3], [17] .

**1**. **Cross-Dataset Generalization and Domain Shift:** A critical limitation of deep learning-based image forensic systems is their acute vulnerability to domain shifts [16] . Most models report state-of-the-art localization accuracy when trained and evaluated on closed, homogeneous datasets (such as CASIA or Columbia) [9] . However, when tested on out-of-distribution imagery characterized by diverse camera sensors, varied indoor/outdoor lighting conditions, and unconstrained resolutions detection accuracy drops substantially [16].Detectors frequently overfit to dataset-specific noise distributions or synthetic generation artifacts rather than learning universal tampering traces [16]. Developing self-supervised representation learning and domain-adaptation strategies is essential to ensure models generalize effectively to previously unseen manipulation techniques [16] .

**2. Robustness Against Social Media Transcoding and Image Degradation:** In practical forensic scenarios, fraudulent media rarely circulates in raw, uncompressed formats [10], . Instead, images shared across social media platforms undergo automated processing pipelines involving severe, lossy JPEG compression, spatial downscaling, metadata stripping, and filtering [3], [10], [19] . Such post-processing heavily degrades high-frequency forensic footprints, PRNU sensor noise, and fine boundary inconsistencies [10]. Furthermore, transmission across noisy channels introduces distortions such as Gaussian or salt-and-pepper noise, which disrupt learned feature representations [18], . While specialized noise-resilient architectures like DSNI-Net demonstrate improved robustness under controlled noise

injections [18],building detectors that consistently survive aggressive multi-platform transcoding remains an unresolved challenge [17].

**3. Anti-Forensic Techniques and Adversarial Vulnerabilities:** As detection algorithms grow more advanced, counterpart anti-forensic techniques are engineered to deliberately conceal tampering traces and deceive forensic classifiers [4], [7] . Attackers employ median filtering, contrast enhancement, noise dithering, and adversarial perturbations to fool automated systems [7], [19].Deep neural networks are particularly vulnerable to imperceptible adversarial noise (such as FGSM or PGD attacks) that can force a false-negative classification without altering the semantic appearance of the image . Establishing certified robustness through adversarial training and developing countermeasures against deliberate tampering concealment represent vital research frontiers [7] .

**4. The Emergence of Generative AI and Inpainting Hybrids:** The convergence of classical manipulation with modern generative artificial intelligence—including Generative Adversarial Networks (GANs) and latent diffusion models—has blurred the boundaries of digital forgery [3]. Rather than relying solely on simple cut-and-paste splicing or copy-move cloning, contemporary forgers utilize AI-driven inpainting and object synthesis [3] . These techniques synthesize photorealistic textures with seamless boundary blending, avoiding traditional geometric discontinuities and double-compression ghosts [3], [10]. Consequently, forensic frameworks must evolve beyond static binary assumptions to address hybrid manipulations that combine classical compositing with generative infilling [3].

**5. Explainability, Evidence Integrity, and Chain of Custody:** In judicial, investigative, and regulatory proceedings, black-box classification scores are legally insufficient [5].Forensic evidence must be explainable, transparent, and legally defensible [5].The absence of interpretability in deep neural networks limits their admissibility in court, where expert witnesses must substantiate why a particular region is deemed forged [5].In addition, preserving the integrity and provenance of visual evidence from acquisition to court presentation is vital [6], [19]. Recent studies highlight that integrating digital forensics with blockchain technology and ambient federated intelligence can secure an immutable chain of custody, preventing unauthorized tampering with evidentiary media during legal investigations [6], [19].

## 5.2 CONCLUSION

This review paper has systematically examined the evolution of digital image tampering detection, charting the technical paradigm shift from classical handcrafted features to state-

of-the-art deep learning architectures . Early forensic methodologies laid an indispensable foundation by exploiting physical camera characteristics (PRNU), transform-domain representations (DCT, DWT), and compression anomalies (JPEG ghosts, ELA) , . However, their inherent sensitivity to affine transformations, lossy re-compression, and prohibitive computational overhead necessitated the adoption of representation-learning frameworks .

The integration of convolutional neural networks (CNNs), transfer learning pipelines, dual-stream architectures, attention modules, and graph neural networks (GNNs) has elevated tampering detection into an automated, highly accurate, and pixel-precise discipline.By concurrently modeling spatial textures, noise residuals, and topological relationships, modern hybrid frameworks achieve detection accuracies exceeding 98% on established benchmarks .

Nevertheless, as digital manipulations grow increasingly sophisticated through generative AI and adversarial anti-forensics, static single-dataset classifiers remain inadequate for open-world cyberspace. The future of multimedia security relies on developing lightweight, cross-domain, and explainable models that fuse spatial, frequency, and semantic forensic cues while maintaining an immutable, verifiable chain of custody for digital evidence .

## REFERENCES


1. D. S. Hirde and S. S. Sherekar, "Performance analysis of deep and hybrid frameworks for automated image splicing and forgery detection," International Journal of Modern Science and Research Technology (IJMSRT), vol. 4, no. 10, pp. 81–88, Oct. 2026.
2. D. P. Timothy and A. K. Santra, "Detecting digital image forgeries with copy-move and splicing image analysis using deep learning techniques," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 15, no. 5, pp. 1299–1306, 2024.
3. M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake Detection: A Systematic Literature Review," IEEE Access, vol. 10, pp. 25494–25513, Mar. 2022.
4. Singh, S. K. Singh, N. Singh, and S. K. Nayak, "Cyber-Crime and Digital Forensics: Challenges Resolution," in Proc. 2023 Int. Conf. Computer Communication and Informatics (ICCCI), Coimbatore, India, Jan. 2023.
5. R. T. Yadav, "AI-driven digital forensics," International Journal of Scientific Research in Engineering and Trends (IJSRET), vol. 10, no. 4, Jul.–Aug. 2024.
6. B. J. Pathan, N. Nadaf, and S. Desai, "Blockchain-Enhanced Digital Forensics: Strengthening the Chain of Custody," Journal of Emerging Technologies and Innovative Research (JETIR), vol. 11, no. 6, pp. 348–356, Jun. 2024.

7. N. Asim, J. Osamor, F. Olajide, C. Iwendi, and N. Okeke, "Detecting and mitigating anti-forensic techniques: A comprehensive framework for digital investigators," in Proc. AI-Driven Smart Healthcare for Society 5.0, Feb. 14–15, 2025.
8. H. Ding, L. Chen, Q. Tao, Z. Fu, L. Dong, and X. Cui, "DCU-Net: A dual-channel U-shaped network for image splicing forgery detection," Neural Computing and Applications, vol. 35, no. 7, pp. 5015–5031, Mar. 2023.
9. D. P. Timothy and A. K. Santra, "Detecting Digital Image Forgeries with Copy-Move and Splicing Image Analysis using Deep Learning Techniques," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 15, no. 5, pp. 1137–1145, 2024.
10. D. Singh, P. Singh, R. Jena, and R. S. Chakraborty, "An image forensic technique based on JPEG ghosts," Multimedia Tools and Applications, vol. 82, no. 11, pp. 14153–14169, Sep. 2023.
11. E. U. H. Qazi, T. Zia, and A. Almorjan, "Deep Learning-Based Digital Image Forgery Detection System," Applied Sciences, 2022.
12. N. V. S. K. Vijayalakshmi K, J. Sasikala, and C. Shanmuganathan, "Deep learning techniques-based image forensics approaches - A Review," International Journal of Novel Research and Development (IJNRD), vol. 8, no. 10, Oct. 2023.
13. M. Trigka and E. Dritsas, "A comprehensive survey of deep learning approaches in image processing," Sensors, vol. 25, no. 2, Art. no. 531, Jan. 2025.
14. F. Niloy, K. K. Bhaumik, and S. S. Woo, "CFL-Net: Image Forgery Localization Using Contrastive Learning," arXiv preprint arXiv:2210.02182, 2022.
15. E. U. H. Qazi, T. Zia, M. Imran, and M. H. Faheem, "Deep Learning-Based Digital Image Forgery Detection Using Transfer Learning," Intelligent Automation & Soft Computing (IASC), vol. 38, no. 3, pp. 225–240, Feb. 2024.
16. H. Kwon, J. Choi, and H. Kim, "Predictive forensic framework for unseen manipulations using self-supervised anomaly detection," Pattern Recognition Letters, vol. 165, pp. 1–9, Jan. 2023.
17. W. Ding, M. Abdel-Basset, A. M. Ali, and N. Moustafa, "A survey of intelligent multimedia forensics for internet of things communications: Approaches, strategies, perspectives, and challenges for a sustainable future," Engineering Applications of Artificial Intelligence, vol. 138, p. 109451, Oct. 2024.

18. W. Shan, Y. Xie, P. Wang, D. Zou, J. Qiu, and J. Li, "DSNI-Net: Forensic Network for Detecting Splicing in Salt-and-Pepper Noise," arXiv preprint arXiv:2503.04123, Mar. 2025.
19. A. A. Khan, X. Zhang, F. Hajjej, J. Yang, C. S. Ku, and L. Y. Por, "ASMF: Ambient social media forensics chain of custody with an intelligent digital investigation process using federated learning," Heliyon, vol. 10, no. 2, p. e23254, Jan. 2024.