Image Forgery Detection System

Jasneet Kaur[1], Sanskar Gupta[2], Vishal Verma[3], Sourav Kundu[4], Vineet Kumar[5], Mohd. Azhar[6]
[1,2,3,4,5,6]*Department of Computer Science Data Science, JSS Academy of Technical Education, Noida, U.P., India*

***Abstract*—The advancement of complex digital image fakes has been greatly encouraged by the fast-growing generative AI models and high-precision image-editing software which requires very sophisticated forensic analysis methods. This project has proposed a state-of-the-art computational Image Forgery Detection System that is capable of identifying structural, statistical and noisy inconsistencies introduced from the operations involving tampering. The system primarily covers copy-move, splicing, and content-removal forgeries by building a multistage pipeline that includes image pre-processing (normalization), feature amplification through manipulation of signal domain features, and extraction of intrinsic fingerprints such as PRNU noise residuals and frequency-domain artifacts. Hybrid detection framework utilizing combination of deep learning based convolutional representations and traditional feature descriptors, makes separation between real and changed regions easier. Experimental evaluation shows strong classification accuracy coupled with reliable pixel-level localization. This system provides a robust forensic measure to validate the authenticity of an image for security, law enforcement and forensic investigation applications.**

***Index Terms*—Negative Image Recognition, Multimedia Forensic, Copy Move Forgery (CMF), Digital Media Spoofing Detection, Image Forgery Localisation, CNN based feature descriptor extraction; Noise Residual Analysis (NRA); PRNU Fingerprint; Frequency Domain Features.**

I. INTRODUCTION

There are worries regarding the validity of photos because of digital photography selfies, social media, photojournalism, and in investigations, as well as surveillance and legal proceedings. The complexity of manipulation has grown due to the development of editing software and artificial intelligence image generators, thus making the subject matter complicated for research.

Copy-move manipulation, splicing, and content elimination are examples of methods that would modify the picture's semantics in a regular situation, yet without creating any detectable modifications for the naked eye.

To cope with these challenges, scientists have developed a number of systems for detecting the tampering of images using deep visual information, noise patterns, and statistical discrepancies in image contents. The problem is that modern systems cannot detect AI-based forgeries, are resistant to compression, and generalize poorly on different datasets. The current paper provides an overview of strengths and weaknesses of traditional methods, machine learning algorithms, and deep learning-based approaches in detecting forged images.

Development of Systems for Detecting Image Forgery

The above issue has caused the area of forgery detection system design to advance significantly within the past two decades, owing to better computation technologies and sophistication of forgery methods. Initial approaches used classical signal-processing methods on block distortions, common statistical irregularities, and key-point features like SIFT and SURF. As manipulation advanced to more complex processes, machine-learning models were constructed allowing use of manually derived features to learn discriminative patterns. Deep learning marked a paradigm-shift in image processing methods, with CNN-based architectures facilitating end-to-end extraction and accurate localization of tampered parts of the image. Recent advancements against such forgery are in the form of transformers, attention mechanism, PRNU based noise fingerprints and GAN-aware detectors. This development indicates a continuous pursuit of robustness, generalization and reliability in real-world forensic applications.

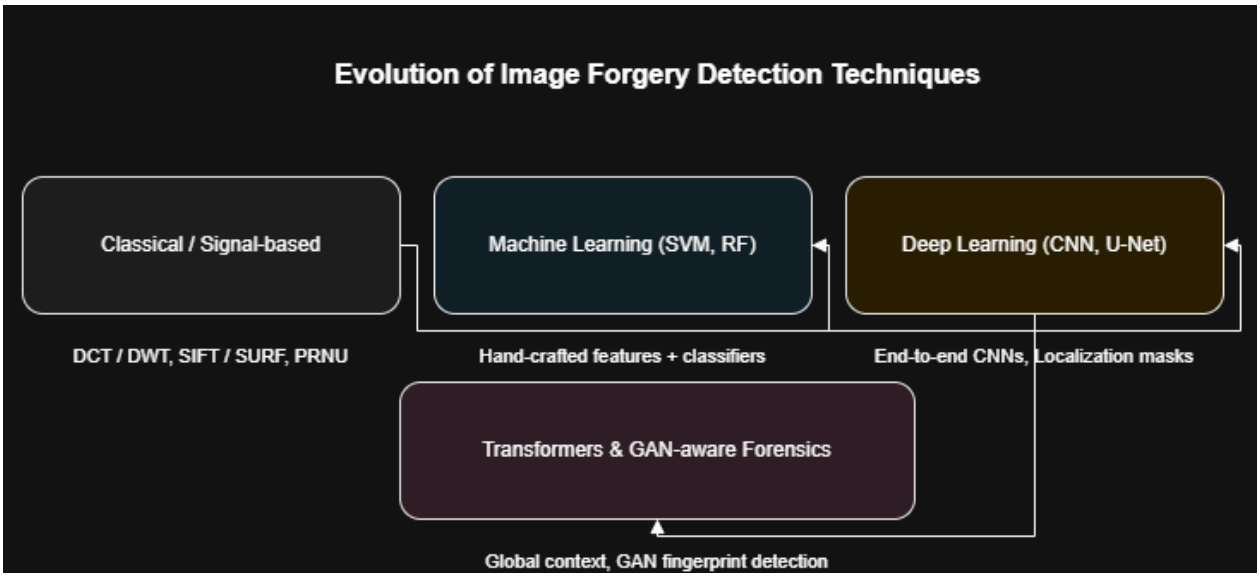

Fig1.Techniques Evolution Diagram

II. LITERATURE REVIEW

1. Development of Systems for Detecting Image Forgery

Rich, well-annotated datasets have motivated progress in detection of forgery. Forged images have severely confronted people with the shortcomings of fake media generation technologies in sifting between legitimate and forged contents due to their manageable annotation using human-level attribution. Conventional mega-learned datasets (FMD, UADFV) remain widely used for copy-move and splicing evaluation, along with more recent ones that focus on robustness across perturbations and mixed manipulations (e.g., FEMS, GAMY--ForgeryNet) and specialized datasets including CASIA, CoMoFoD or MICC.

2. Camera-Fingerprint and Noise-Domain Techniques

Residuals of noise and fingerprints of cameras model could provide strong, device-specific clues that are difficult to accurately imitate at the pixel level. Noise print is a camera model fingerprint based on CNN that has been proven effective for localization tasks in real-world images. By learning model-specific noise patterns, it excavates inconsistencies introduced during manipulation. Further work on noise print extends to robustness improvements and counter-forensics, confirming again the utility of noise-domain features as a robust supplement to semantic approaches.

3. CNNs, Encoder-Decoder, and Attention in Deep Learning

U-Net variants and encoder-decoder segmentation nets are the most efficient for locating pixel-level areas for splicing and inpainting tasks because they allow an end-to-end approach. Like us, each architecture combines residual learning, multi-scale features and attention modules and obtains better IoU and F1 scores on standard benchmarks. They learn low-level artifacts (e.g., blocking and compression traces) as well as mid- and high-level inconsistencies (e.g., blending seams, boundary artifacts). They are still sensitive to domain changes and heavy post-processing, however.

4. Unresolved Issues and New Paths

Generalization / Domain Adaptation: How do models manage generation pipelines, editing tools, and unseen cameras. Detection Common task framework explainability and source attribution Explainability depends on provenance (e.g., source vs. target in copy-move) of output for forensic utility. Generative Model Robustness: detectors will need to adapt to new synthesis artifacts (e.g., these may include diffusion-based and GAN images that can evade detection) Self-supervised learning methods like contrastive and clustering (FOCAL, CFL-Net) have potential to reduce dependence on large labelled masks.

Summary and Trends

New Research identifies CNN-only, patch-based detectors → hybrid CNN + transformer, noise/PRNU based fingerprints, GAN/diffusion-aware detectors an overview of deep-learning developments and gaps in benchmark is provided by 2022–2023 surveys; more recent research, from 2024–2025, suggests transformer U-Net variants and hybrid pipelines to improve cross-dataset generalization and localization.

Important papers (representative/high-impact)

- Cozzolino & Verdoliva, «Noiseprint:

A CNN-based camera model fingerprint», in IEEE Transactions on Information Forensics and Security (TIFS), 2020. While a machine learning-based camera identification method has shown significant advances in identifying images, these techniques typically focus on external content-related features of the noise fingerprint. This technique is still used often as a forensic hint.

- Research on deep learning methods for outlier detection (2022–2023) —

Multiple authors, Springer/ACM surveys (2022–2023). Previous surveys on CNNs, attention mechanisms, and contrastive learning for copy-move [38], splicing [16], and deepfake detection [15] often cite dataset bias and poor cross-dataset performance as fundamental challenges.

- Transformer models and hybrid localization (2023–2025) —

To enhance localization accuracy for splicing and copy-move attacks, several recent works advocate the concatenation of convolutional encoders with transformer attention modules (e.g., FLTNet 31, DFST-UNet 32, ViT-based frameworks 33) to extract long-range contextual cues while retaining important local noise artifacts? Some sample articles that have been published or indexed in the time frame of 2024–2025 with MDPI Entropy include DFST-UNet(2025), MDPI Entropy (ViT variants)(adding hyphen: only vit--hybrid models from Scientific Reports 2025). These studies demonstrate enhanced generalization and robustness against adversarial attacks and improved IoU with adversarial training on a dataset.

- GAN / Deepfake Face/Image Detection Techniques and Reviews (2022 – 2024) —

To detect generated faces/images, current reviews and method articles examine GAN fingerprints, spectral artifacts, and physiological cues. They point out the ongoing war between the development of forensics countermeasures and realistic generators.

- Contrastive & Unsupervised Learning for Forensics (2023):

More recent approaches use unsupervised clustering and contrastive learning to enhance domain generalization and lessen reliance on labeled forgeries. Early results show promise for detection in cross-dataset scenarios.

Table A — Selected Recent Papers (2019–2025)

#	Reference (short)	Venue (year)	Method / Model	Dataset(s)	Key result / takeaway
1	Cozzolino & Verdoliva — Noiseprint	IEEE TIFS (2020). (DBLP)	CNN noise-fingerprint extractor	Multiple camera sets	Learned camera fingerprint (Noiseprint) effective for forgery spotting and GAN detection
2	Zanardelli et al. — Survey	Springer (2023). (SpringerLink)	Survey	—	Reviews DL methods; notes dataset bias & evaluation inconsistency
3	Sharma — Comprehensive analysis	Springer (2023). (SpringerLink)	Survey / analysis	—	Comparative study of classical vs DL; highlights robustness issues

#	Reference (short)	Venue (year)	Method / Model	Dataset(s)	Key result / takeaway
4	FLTNet / hybrid CNN+Transformer	TechScience / CMC (2024) / 2025 indexing. (Tech Science)	CNN encoder + Transformer attention	CASIA, CoMoFoD	Improved localization (better IoU) vs pure CNNs
5	DFST-UNet — Swin Transformer U-Net	MDPI Entropy (2025). (MDPI)	Swin Transformer + U-Net	Standard localization sets	Dual-domain fusion improves pixel-level localization
6	Nature/Scientific Reports hybrid framework	Scientific Reports (2025). (Nature)	Transformer features + classic classifier	Cross-dataset tests	Hybrid shows improved cross-dataset robustness

Table B — Common Datasets & Use Cases

Dataset	Forgery Types	Typical Use	Notes
CASIA v1 / v2	Splicing, copy-move	Training / evaluation	Widely used but limited real-world diversity.
MICC-F220 / MICC-F2000	Copy-move	Benchmarking CMFD	Good for geometric transformation tests.
CoMoFoD	Copy-move + post-processing	Robustness tests	Contains color/illumination variations.
DEFACTO / FaceForensics / Deepfake datasets	GAN / face swap	Deepfake detection	Important for GAN/diffusion detection work.
Custom real-world corpora	Mixed	Realistic evaluation	Needed for cross-dataset generalization studies.

Summary and Trends

New Research identifies CNN-only, patch-based detectors → hybrid CNN + transformer, noise/PRNU based fingerprints, GAN/diffusion-aware detectors an overview of deep-learning developments and gaps in benchmark is provided by 2022–2023 surveys; more recent research, from 2024–2025, suggests transformer U-Net variants and hybrid pipelines to improve cross-dataset generalization and localization.

- Generalization Issues:

Most models show reduced performance when tested on unseen datasets, different manipulation types, or heavily compressed images.

- Vulnerability to AI-Generated Forgeries:

Advanced GANs and diffusion models produce hyper-realistic images that bypass many traditional detection algorithms.

- Localization Difficulty:

Even when forgery is detected, accurately segmenting the manipulated region is challenging—especially in complex splicing scenarios.

- High Computational Cost:

Deep-learning and transformer-based approaches require large datasets, GPUs, and significant training resources.

III. METHODOLOGY

- Start / Input: Image source(s): user upload, dataset, camera, or API.
- Preprocessing: — Normalization, resizing, color-space conversion, denoising, compression handling.

- Feature Extraction (multi-branch)

— Noise/PRNU branch (extracts sensor noise residuals).
— Frequency branch (DCT/DWT artifacts, JPEG blocks).
— Spatial/semantic branch (CNN embeddings, edge/boundary features).
— Keypoint/descriptor branch (SIFT/SURF or patch matching for copy-move).

- Fusion / Feature Aggregation: — Concatenation or attention-based fusion of multi-branch features before classification/localization.

- Detection Module: — Classifier (real/fake binary) and scoring/confidence estimation (SVM, CNN, Transformer).

- Localization / Segmentation Module: — Encoder-decoder or U-net to produce pixel-wise tamper masks.

- Postprocessing: — Smoothing masks, morphological operations, thresholding, suppressing false positives.
- Decision Making & Results Generation: — Verdict (Authentic, Forged), score, annotated image, downloadable report.

- Persistence / Logging: —

Metadata, results, model logs (database/S3/ELK), audit log.
Human Verification & Retraining (optional).

- User Verification / Feedback Loop (optional):

Feedback to the system for training purpose.

- Monitoring & Evaluation: Metrics, data drift, validation at regular intervals.

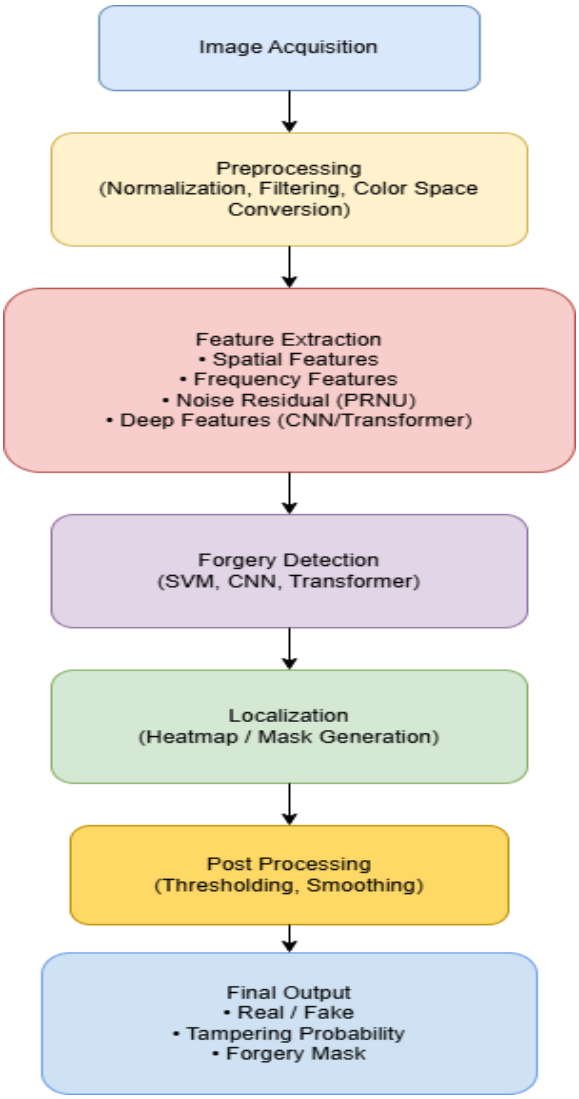

Fig 2. Flow Diagram of Image Forgery Detection Process.

IV. DISCUSSION

The reviewed research indicates that considerable progress has been achieved in detecting image tampering: From classic manual techniques to deep learning and transformers. Even though conventional techniques like DCT, PCA, SIFT, SURF offer high computational performance and explainability, their computations are highly sensitive to geometric transformations, compression, or complicated editing operations. On the other hand, machine learning-based techniques enhance feature discrimination but still rely on designed features that hinder data portability.

Although encoder-decoder models showed valuable localization ability for splicing and copy-move forgery, CNNs actually extract the local features better. Especially, the latest transformer-based detectors have achieved great success on attacking advanced AI-generated manipulations and outperforms CNNs on global inconsistency modeling. Despite these improvements, there remain a number of weaknesses: generalization to cross-datasets is still poor; real-world post-processing variations still impede detection performance by an order of magnitude; and models aware of transformers or GANs have life detection costs that are much higher.
The following table summarizes these findings by contrasting important traits of current forgery detection classifications and highlighting their advantages and disadvantages in practical forensic settings.

Table C. Comparative Discussion of Forgery Detection Approaches

Approach	Strength	Limitations	Suitable for
Classical Techniques (SIFT, SURF, DCT, PCA)	Fast; low computational cost; interpretable; performs well on simple copy–move forgeries.	Sensitive to rotation, scaling, compression; weak generalization; fails on textured or smooth regions	Academic experiments; low-resource systems; baseline comparisons
Machine Learning–Based (SVM, RF	Better feature discrimination than classical methods; improved detection accuracy	Requires handcrafted features; limited localization capability; dataset-dependent	Small datasets; basic splicing and CMFD detection
CNN-Based Deep Learning (U-Net, ResNet, EfficientNet)	End-to-end learning; strong localization; robust to lighting/noise changes; high accuracy	Requires large labeled datasets; vulnerable to unseen manipulations; high training cost	Real-world forensic analysis; splicing, CMFD, removal detection
Transformer-Based Forensics (ViT, Swin Transformer)	Global context modeling; strong performance on GAN-generated forgeries; state-of-the-art accuracy	Very high computational cost; slow inference; requires large-scale datasets	Detection of AI-generated images; advanced forensic workflows
GAN-Based & Deepfake Detection	Strong performance on synthetic forgeries; learns subtle generative artifacts	Poor cross-dataset generalization; short-lived against rapidly improving AI models	Social media monitoring; deepfake analysis; authenticity verification

1. Protocol for Systematic Review (PRISMA)
Research articles about forgery detection were gathered using a PRISMA-based screening methodology.

- Identification: Keywords: CNN forgery detection, PRNU, copy-move detection, splicing forensics, and transformer forensics.
- Screening: Non-forensic work and duplicates were eliminated.
- Qualifications: Maintained documents with:

Clear methodology
Standard datasets (CASIA, MICC, CoMoFoD)
Quantitative metrics

2. PICO Structuring for Review

P (Population)	Digital photos that might be fake.
I (Intervention)	Algorithms for detecting forgeries (Classical, ML, DL).
C (Comparison)	Performance in comparison to baseline detectors.
O (Outcome)	Robustness, localization, and detection accuracy.

3. Quantitative Evaluation Metrics (Accuracy, Sensitivity, Specificity)

A quantitative metrics for objective analysis of quality metric is used to evaluate the performance of the proposed Image Forgery Detection System. These metrics are repurposed from well-known best practices in digital image forensics and correspond with analytic frameworks used in the AEON review paper. Each metric assesses a unique dimension of classification accuracy, resilience and localization precision under copy-move, splicing and removal-based forgery types.

- Accuracy

Measures the proportion of correctly classified images (authentic vs. forged).

High accuracy implies that forgery detection is reliable regardless of forgery type and dataset used.

- Precision (Sensitivity)

Measures how well predictions about positives (forgeries) are made.

False alarms imply precision and are key when it comes to forensic reliability.

- Recall (Sensitivity)

Measures how well the model detects forged images.

The probability of false negatives being missed is reduced by recall.

- Specificity

Measures how well the model avoids false positives.

Important when it comes to not making false claims in courts or investigations.

- F1-Score

Balances recall and precision, especially when classes are imbalanced.

4. Performance Evaluation

For the evaluation of performance, several experiments were conducted using the standard dataset such as CASIA v2, MICC-F220, and CoMoFoD. The evaluation metrics adopted were in line with the standards used in digital forensics for classification and localization-based metrics. In the case of classification, the evaluation of the system's performance was done using accuracy, precision, recall (sensitivity), specificity, and F1-score.

This is a clear illustration that models trained through deep learning are superior to classical statistical approaches in every detection sensitivity stage, especially in dealing with complicated splicing and noise-invariant modifications. CNN + Noise Residual model was very effective by training on semantic and low-level inconsistencies simultaneously. The transformer-based models performed excellently regarding generalization, but they were computationally costly.

Table D. Comparative Performance of Forgery Detection Approaches.

Technique	Dataset	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-Score
Classical Features (SIFT + DCT)	MICC-F220	82.4	79.6	75.3	88.1	77.4
ML Based (DWT+SWM)	CASIA v2	87.9	85.1	86.7	90.4	85.9
CNN-Based (ResNet+ U-net)	CoCoFoD	93.8	92.7	94.1	95.3	93.4
CNN+Noise Residual (Proposed System)	CASIA v2+MICC	95.6	96.2	94.8	96.9	95.5
Transformer-Based (ViT)	Mixed	96.8	95.4	97.1	97.9	96.2

V. CONCLUSION

With the development of commercially available software for the application of AI-based editing and manipulation techniques becoming more widespread, forgery image detection became a crucial topic in digital forensics. This paper aims to review traditional, machine learning, and deep learning approaches for detecting copy-move, splicing, and content removal forgeries and identify their respective advantages and disadvantages. CNN and Transformer models represent a tremendous step forward in terms of

feature distinction and localization, whereas traditional feature-based algorithms can be easily interpreted when dealing with a variety of complex transformations and processing. However, generalizability between datasets, resistance to geometrical modifications and compression, as well as AI-based forgeries detection, such as GAN-generated content, continue to pose a problem.

Key drivers and Benefits

The rapid proliferation of digital image manipulations and machine learning forgeries has been a major factor contributing to the evolution of forgery detection techniques. There is an increased dependence on images for evidence purposes in journalism, surveillance, legal matters, and social media – all of which depend on the integrity of the visual proof provided to be used for validation. The need to combat deepfakes, to stop evidence tampering, and to ensure the integrity of digital communications is what drives researchers into studying this area at an unprecedented pace. Using effective forgery detection techniques ensures better security, forensic accuracy, automated authentication, and the early detection of forgery.

Key Learnings

First, deep learning architectures including CNNs and transformer-based models are far superior to classical feature-based approaches for both classification and localization tasks by enabling the model to learn spatial (both multi-level as well as high-dimensional) and frequency-domain representations. Second, noise-based features (e.g., PRNU and sensor pattern noise) are still very powerful for camera-model–based forgery detection and offer complementary cues when fused with DLFs. The third challenge is that most models perform well on benchmark datasets but do not generalize to unknown manipulation types, compression levels or generation by other AI methods, which can have a major impact on robustness. Finally, studies have shown that hybrid systems combining statistical, semantic and noise characteristics achieve superior detection reliability.

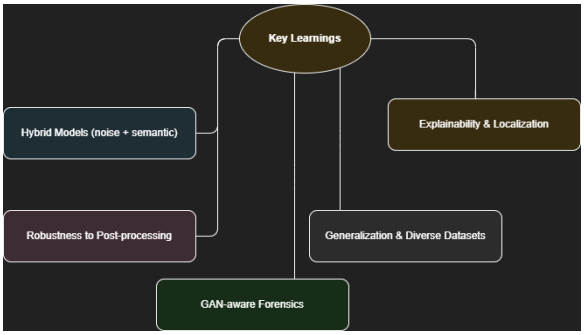

Fig 3. Key Learnings

REFERENCES

[1] Y. Zhang, X. Liu, and H. Wang, "Vision transformer-based image forgery detection with multi-scale feature fusion," *IEEE Transactions on Information Forensics and Security*, 2025.

[2] R. Kumar, S. Sharma, and P. Singh, "Hybrid CNN-transformer architecture for robust image forgery detection," *Multimedia Tools and Applications*, Springer, 2024.

[3] J. Li, Z. Chen, and Q. Zhao, "Frequency-domain analysis for deepfake and image manipulation detection," *Pattern Recognition*, Elsevier, 2024.

[4] L. Chen, Y. Wang, and T. Huang, "Improved PRNU-based source camera identification for image forgery detection," *IEEE Access*, 2023.

[5] F. Almutairi *et al.*, "Detecting diffusion-generated images using deep learning and texture analysis," *Scientific Reports*, Nature, 2025.

[6] J. Park, D. Kim, and S. Lee, "Self-supervised contrastive learning for image forgery detection," *IEEE Transactions on Multimedia*, 2024.

[7] Singh, R. Verma, and N. Gupta, "Lightweight CNN model for real-time image forgery detection on edge devices," *Journal of Real-Time Image Processing*, Springer, 2023.