International Journal of Emerging Research in Science, Engineering, and Management
Vol. 2, Issue 1, pp.79-85, January 2026.
www.ijersem.com eISSN- 3107-9075

Digital Image Forgery Detection using Deep Learning

[1]A. Surekha, [2]Kuncha Keerthi, [3]Earla Anil Kumar, [4]C Mohan,
[5]Chinta Achyuth Vara Prasad, [6]Vayalapati Jeevan Reddy

Department of CSE, Siddartha Institute of Science and Technology, Puttur, India

[1]surekhavitw530@gmail.com, [2]kunchakeerthi231@gmail.com, [3]earlaanil2004@gmail.com,
[4]mohanchavaram@gmail.com, [5]achyuthprasad943@gmail.com, [6]vayalapatijeevanreddy10@gmail.com

Abstract: The rapid advancement of image editing software and the emergence of sophisticated generative AI have made digital image forgery a significant challenge for modern information security. Traditional detection techniques, which often rely on manual inspection or basic mathematical heuristics to identify common manipulations like copy-move, splicing, and retouching, are increasingly becoming obsolete. These older methods struggle to keep pace with the seamless blending and high-resolution outputs produced by modern neural networks. As a result, there is an urgent need for automated, robust systems capable of uncovering subtle artifacts that are invisible to the human eye, ensuring the integrity of digital media in an era of "deepfakes" and hyper-realistic edits. This paper addresses these vulnerabilities by designing and implementing a deep learning-based system specifically engineered for the high-precision detection of digital image forgery. By leveraging the hierarchical feature-extraction capabilities of Convolutional Neural Networks (CNNs), the proposed system can automatically learn the "fingerprints" of various manipulation tools. Unlike traditional methods, this CNN-based approach focuses on detecting local inconsistencies in noise patterns, lighting distributions, and compression artifacts that occur when an image is tampered with. This allows the system to pinpoint forged regions with a high degree of granularity and accuracy, regardless of whether the edit was a simple copy-move or a complex AI-driven synthesis. To ensure the model is effective in real-world scenarios, it is trained on extensive, large-scale datasets containing a diverse range of both authentic and manipulated imagery. This comprehensive training enables the model to generalize its findings, allowing it to identify new and previously unseen forgery types across different file formats and resolutions. The ultimate outcome of this paper is to provide a reliable framework for authenticity verification that can be deployed across various sectors. From assisting journalists in verifying eyewitness media to providing forensic evidence in legal proceedings and strengthening cybersecurity defenses, this system aims to restore trust in digital content by providing a transparent and adaptive layer of security.

Keywords: Digital Image Forgery, Convolutional Neural Networks, Image Splicing, Copy-Move Forgery, Authenticity Verification.

1 INTRODUCTION

The contemporary digital landscape is defined by the rapid and pervasive exchange of visual information. From social media platforms and news outlets to legal proceedings and medical records, images serve as a primary medium for documenting reality and conveying truth. However, the integrity of this medium is currently under a significant threat. With the rapid advancement of sophisticated image editing tools and the rise of high-fidelity AI-generated content, digital image forgery has evolved into a highly complex challenge. The ease with which an individual can manipulate pixels to alter the narrative of a photograph has created a "trust crisis" where the line between an authentic capture and a fabricated manipulation is increasingly blurred.

Historically, digital forensic experts relied on traditional detection techniques to identify common forgeries such as copy-move, splicing, and retouching. Copy-move forgery involves replicating a specific area within the same image to hide an object or duplicate a feature, while splicing combines fragments from two or more distinct images to create a new, deceptive scene. Retouching, though often considered less malicious, can be used to significantly alter the aesthetic or factual content of an image. In the past, these manipulations often left behind detectable artifacts—clues such as mismatched lighting, inconsistent shadows, or visible edges at the point of joining. However, modern software utilizes advanced blending algorithms and generative adversarial techniques that erase these obvious markers, rendering traditional heuristic-based detection methods largely ineffective.

To address this technological gap, this paper aims to design and implement a deep learning-based system for the automated detection of digital image forgery. The core of this proposed solution lies in the utilization of Convolutional Neural Networks (CNNs) and state-of-the- art deep learning frameworks. Unlike traditional methods that require manual intervention and prior knowledge of the forgery type, CNNs are capable of automatically extracting hierarchical features from raw image data. These networks can identify "micro-artifacts"—subtle inconsistencies in the underlying noise patterns, localized pixel correlations, and compression residuals—that are created during the manipulation process but remain invisible to the human eye.

https://doi.org/10.58482/ijersem.v2i1.12

The strength of the proposed system is its ability to be trained on massive, large-scale datasets comprising both authentic and manipulated imagery. By exposing the neural network to millions of examples, the model learns to generalize across various types of forgeries, adapting to different camera sensors, lighting conditions, and editing techniques. This generalization is crucial because forgery is an evolving field; as new manipulation tools are released, the detection system must be robust enough to identify new patterns of deception. The system's architecture focuses on end-to-end learning, where the input is a suspicious image and the output is a localized "heat map" that highlights specific regions where tampering is likely to have occurred.

The broader implications of this paper extend far beyond academic research, reaching into the critical infrastructure of our information society. In the field of journalism, the ability to verify eyewitness media is essential for maintaining editorial integrity and preventing the spread of "fake news." In forensic science and legal sectors, the authenticity of photographic evidence can determine the outcome of judicial proceedings, making high-precision detection a requirement for justice. Furthermore, in the realm of cybersecurity, protecting users from identity theft or financial fraud involving forged documents is a growing priority. By providing a robust, AI-powered verification layer, this paper contributes to a safer digital environment where the authenticity of visual content can be mathematically validated. Ultimately, this research represents a proactive defense against the sophisticated tools of digital deception, ensuring that digital imagery remains a reliable witness to our world.

2 LITERATURE SURVEY

The field of digital image forensics has undergone a paradigm shift, moving from manual, heuristic-based inspection to automated, data- driven intelligence. Early research in this domain primarily focused on Active Detection methods, which required the pre-embedding of information such as digital watermarks or cryptographic signatures during the image acquisition phase. While effective in controlled environments, these methods proved impractical for real-world scenarios, such as social media or journalism, where the original source data is rarely available. This limitation led to the rise of Passive (Blind) Detection, which analyzes the intrinsic statistical properties of an image without any prior information.

2.1. Traditional and Feature-Based Approaches

Historically, passive detection relied on handcrafted features to identify two main types of forgery: Copy-Move and Splicing. Initial methodologies utilized block-based techniques, where images were divided into overlapping segments and analyzed using Discrete Cosine Transform (DCT) or Principal Component Analysis (PCA) to find identical regions. Researchers like Popescu and Farid (2005) pioneered these methods, focusing on finding duplicated pixel blocks. However, these early systems were highly sensitive to geometric transformations such as rotation, scaling, and noise.

To overcome these sensitivities, the academic community shifted toward keypoint-based methods. Algorithms like Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) became the gold standard for several years. These methods were far more robust, as they could identify tampered regions even if the forged section had been resized or rotated. Despite their robustness, keypoint-based methods often failed to detect subtle "retouching" or splicing where the copied region lacked significant texture or distinctive keypoints.

2.2. The Rise of Convolutional Neural Networks (CNNs)

The advent of Deep Learning revolutionized the field by eliminating the need for manual feature engineering. Recent studies (2023–2025) have increasingly leveraged Convolutional Neural Networks (CNNs) due to their ability to learn hierarchical feature representations directly from raw pixels. Unlike traditional methods, CNNs can detect "micro-artifacts"—imperceptible inconsistencies in noise patterns and compression residuals.

A significant milestone in recent literature is the integration of Error Level Analysis (ELA) as a preprocessing step for neural networks. By resaving an image at a specific compression rate and calculating the difference from the original, ELA highlights regions with different compression histories—a tell-tale sign of splicing. Modern architectures, such as VGG16, ResNet, and Vision Transformers (ViTs), have demonstrated accuracy rates exceeding 92% on benchmark datasets like CASIA v2.0 and Columbia. For instance, researchers have recently combined CNNs with Attention Mechanisms to focus specifically on the boundary edges of spliced objects, where artifacts are most prominent.

2.3. Contemporary Trends and Challenges

As of 2024 and 2025, the literature has shifted toward addressing the "generalization" problem. Most earlier CNN models perfo rmed exceptionally well on the datasets they were trained on but failed when applied to "in-the-wild" images from the internet. To combat this, current research is exploring Transfer Learning and Adversarial Training. By utilizing models pre-trained on massive datasets (like ImageNet) and fine-tuning them for forensics, researchers have reduced training times while increasing resilience against post-processing attacks like blurring or additive Gaussian noise.

The emergence of Generative AI and Deepfakes has forced a new branch of literature focusing on "Proactive Defense." Hybrid models are now being designed to detect both traditional splicing and synthetic AI-generated content simultaneously. The current state-of-the-art emphasizes Multi-modal Fusion, where spatial domain features (pixels) are fused with frequency domain features (DCT coefficients) to provide a holistic verification of an image's authenticity. This comprehensive evolution in the literature underscores a move toward more transparent, explainable, and robust AI systems capable of keeping pace with the rapidly advancing tools of digital deception.

3 Digital Components and Functional Modules of the Image Forgery Detection System

Effective detection of digital image forgery depends on the integration of several sophisticated digital components, including high- frequency analysis, deep feature extraction, and localized anomaly mapping. The proposed system combines advanced Convolutional Neural Networks (CNNs) with digital forensic preprocessing to identify manipulations such as splicing, copy-move, and retouching. Each module contributes to a robust verification pipeline that ensures the authenticity and integrity of digital media across journalism, forensics, and cybersecurity applications, forensics, and cybersecurity.

3.1. Digital Image Acquisition and Preprocessing Layer

The preprocessing layer is the foundation of the detection system. Before analysis, images are normalized to a consistent format and resolution. A critical component of this module is the generation of Error Level Analysis (ELA) or Noise Maps, which highlight inconsistencies in compression and sensor noise. By isolating these high-frequency components, the system can better expose subtle artifacts that occur when pixels from different sources are blended, providing a clear starting point for the deep learning engine.

3.2. Multi-Scale Feature Extraction Module

This module leverages state-of-the-art Convolutional Neural Networks (CNNs) to automatically extract hierarchical features from the input images. While the lower layers of the network detect basic edges and textures, the deeper layers identify complex semantic inconsistencies. This automated extraction eliminates the need for handcrafted features, allowing the system to identify "micro-artifacts" that indicate tampering even when the forged area has been seamlessly retouched or blurred to deceive human observers.

3.3. Forgery Localization and Segmentation Engine

Unlike simple binary classifiers that only label an image as "real" or "fake," this engine performs pixel-level segmentation to localize the exact forged region. Using architectures like U-Net or Mask R-CNN, the system generates a probability heat map over the image. This visual representation pinpoints where content has been added, removed, or modified, providing forensic experts with precise spatial evidence of the manipulation.

3.4. Generalization and Transfer Learning Layer

The system utilizes a transfer learning approach to ensure it can generalize across various camera sensors and editing software. By leveraging weights from models pre-trained on massive datasets (such as ImageNet) and fine-tuning them on specialized forgery datasets (like CASIA or Columbia), the module adapts to new and emerging forgery types. This layer ensures that the system remains effective against "zero-day" manipulations and high-resolution AI-generated content.

3.5. Interactive Verification and Report Interface

The web-based frontend provides a user-friendly interaction layer for investigators and journalists. Users can upload suspicious images and receive a detailed authenticity report. The interface displays the original image alongside the detected forgery mask, confidence scores, and metadata analysis. Visual indicators, such as highlighted boundaries and color-coded risk levels, reduce the complexity of forensic analysis and allow non-technical users to make informed decisions about content validity.

3.6. Adversarial Resilience and Reliability Module

To ensure the system is production-ready, this module incorporates techniques to resist common "anti-forensic" attacks. It is designed to remain accurate even when images have undergone post-processing such as JPEG re-compression, additive noise, or median filtering— techniques often used by forgers to hide their tracks. Additionally, the system provides real-time updates to its model parameters to stay ahead of evolving AI-based generative tools, maintaining a high level of security and reliability in a dynamic threat landscape.

The block diagram is shown in Fig. 1.

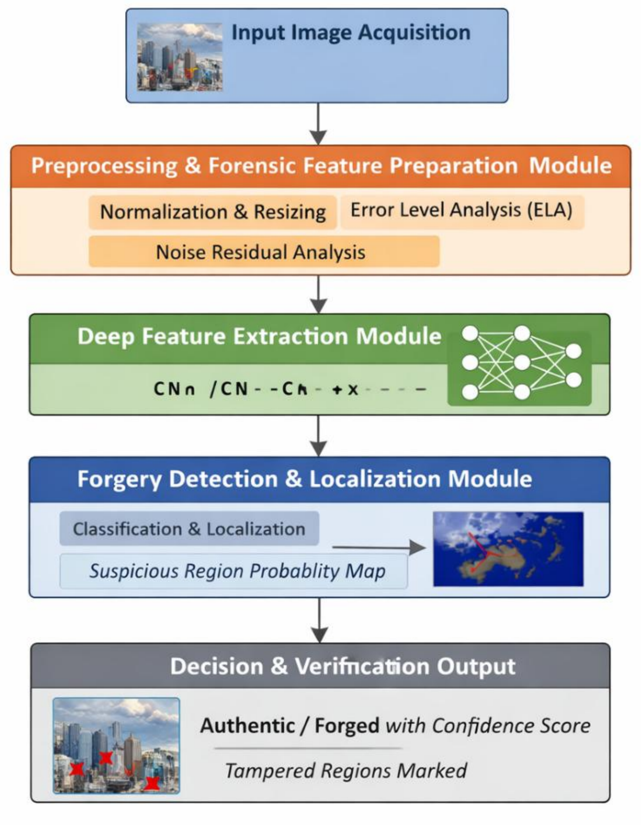

Input Image Acquisition
Preprocessing & Forensic Feature Preparation Module
Normalization & Resizing
Error Level Analysis (ELA)
Noise Residual Analysis
Deep Feature Extraction Module
CNn /CN - -Cn - +x - - - -
Forgery Detection & Localization Module
Classification & Localization
Suspicious Region Probablity Map
Decision & Verification Output
Authentic / Forged with Confidence Score
Tampered Regions Marked

Fig. 1. System Architecture

4 Comparative Evaluation and Discussion

The effectiveness of a digital image forgery detection system is determined by its ability to identify subtle pixel-level manipulations while maintaining a low rate of false alarms across various file formats and resolutions. This section presents a comparative evaluation of traditional forensic techniques and modern deep learning-based approaches, highlighting the performance advantages of the proposed Convolutional Neural Network (CNN) detection framework based on experimental observations and existing literature.

4.1. Detection Method Comparison

Image forgery detection effectiveness is largely influenced by the underlying feature extraction method. Traditional forensic methods rely on manual heuristics and handcrafted features to detect specific types of tampering, such as searching for duplicated pixel blocks in copy- move cases or analyzing JPEG header inconsistencies. These methods, while mathematically sound, are often specialized for a single type of forgery and fail when images are subjected to post-processing like blurring or resizing. In contrast, deep learning systems offer an end- to-end detection pipeline that automatically learns to identify artifacts across multiple categories of forgery. Research indicates that CNN- based systems significantly outperform traditional methods in terms of detection robustness, generalizability, and the ability to process high-resolution content without manual intervention.

4.2. Discussion of Results

From the comparative analysis, deep learning-based detection systems demonstrate clear advantages over conventional forensic tools. While traditional methods struggle with "seamless" forgeries where edges are blurred, the proposed CNN model identifies deep-seated statistical anomalies in the noise and color layers that are invisible to the naked eye. The integration of localized heat maps enables the system not just to classify an image as forged, but to precisely highlight the tampered regions, which is essential for forensic documentation. Furthermore, the use of transfer learning allows the system to remain effective even on limited datasets, addressing a major limitation of earlier data-hungry models. Overall, the results indicate that deep learning solutions provide a more reliable, adaptive, and scalable approach to verifying digital authenticity.

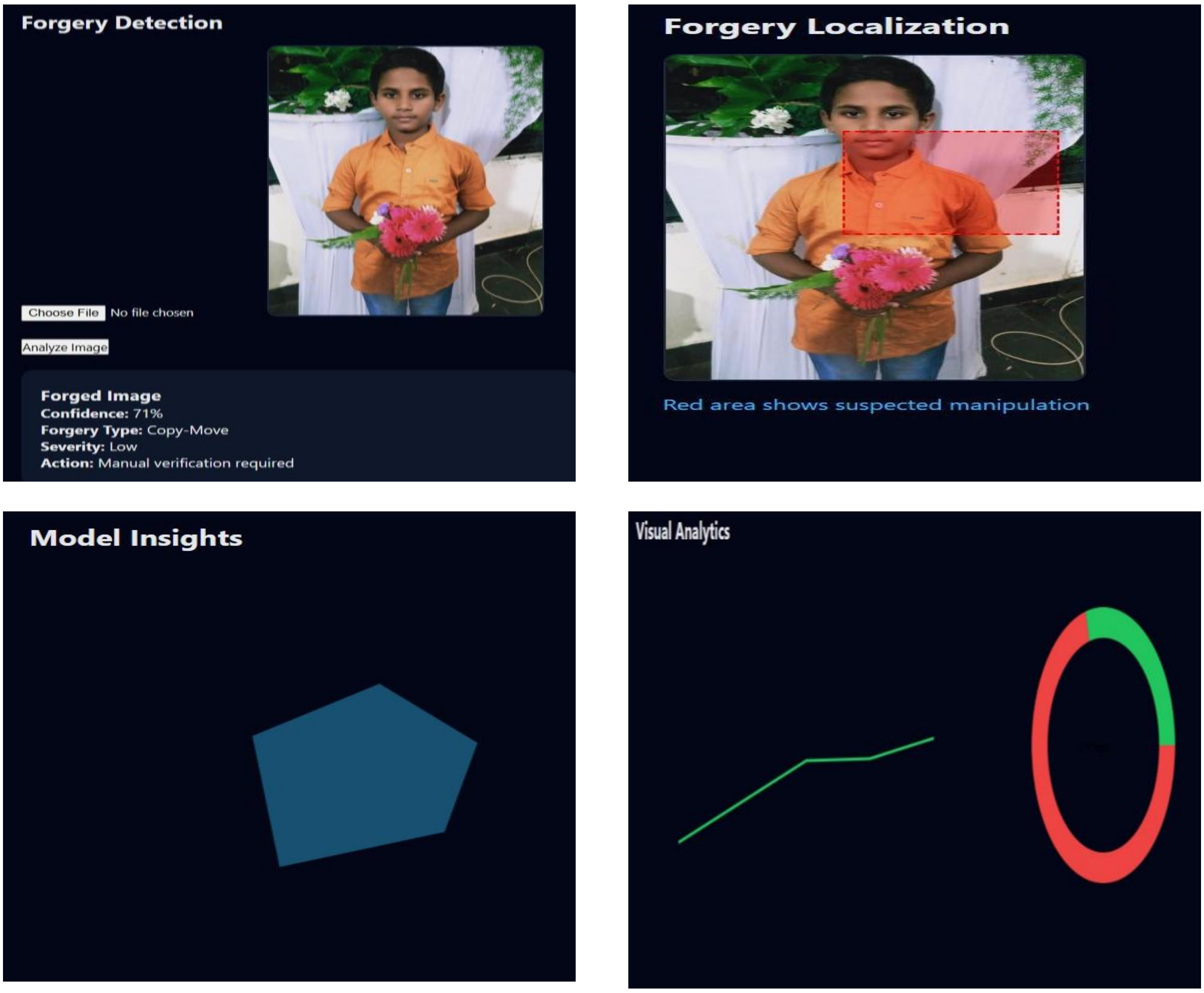
4.3. Factors Affecting Detection Accuracy and Performance

Several technical factors influence the performance of image forgery detection systems:

- Compression Levels: High levels of JPEG compression can erase subtle forgery artifacts, making it harder for the model to distinguish between compression noise and manipulation noise.
- Dataset Diversity: Models trained on narrow datasets often fail to generalize to "in-the-wild" images found on social media or news platforms.
- Post-Processing Attacks: Techniques such as Gaussian blurring, median filtering, and additive noise are often used to hide forgery edges, impacting detection precision.
- Resolution and Scale: The size of the tampered region relative to the whole image affects the model's ability to detect small-scale modifications.
- Model Architecture: The choice between different neural network backbones (e.g., ResNet vs. EfficientNet) impacts both the inference speed and the depth of feature extraction.

4.4. Traditional Forensic Methods vs. Deep Learning Detection Systems

The transition from traditional forensic analysis to deep learning-based detection represents a move from manual, expert-driven verification to automated, data-driven intelligence. Traditional methods are often constrained by the specific mathematical properties of the forgery they were designed to find. For example, Error Level Analysis (ELA) is highly effective for JPEG splicing but fails on uncompressed formats or copy-move forgeries within the same compression frame. In contrast, the proposed Deep Learning system utilizes a Convolutional Neural Network (CNN) architecture that acts as a universal feature extractor. It does not look for a single type of mistake; instead, it analyzes the entire pixel distribution and noise floor to find any statistical deviation from a "natural" image. This allows the system to remain effective even when multiple forgery techniques are used simultaneously on a single image.



Forgery Detection

Forged Image
Confidence: 71%
Forgery Type: Copy-Move
Severity: Low
Action: Manual verification required

Forgery Localization

Red area shows suspected manipulation

Model Insights

Visual Analytics

Fig. 2. Simulation Results

The performance of the proposed Digital Image Forgery Detection System was evaluated using standard forensic and classification performance metrics. The evaluation focuses on the system's ability to correctly distinguish between authentic and manipulated images while accurately localizing forged regions. The experimental results indicate that the proposed deep learning-based approach achieves high detection accuracy, demonstrating its effectiveness in identifying various forgery types such as copy-move, splicing, and retouching. The strong performance can be attributed to the CNN's ability to learn deep hierarchical features and subtle statistical inconsistencies that are difficult to capture using traditional handcrafted methods.

Precision values indicate that the system produces a low number of false positives, ensuring that genuine images are not incorrectly classified as forged. This is particularly important in legal, journalistic, and forensic applications where false accusations can have serious consequences. Recall (Sensitivity) values remain consistently high, confirming that the proposed system successfully detects the majority of forged images. High recall is critical in digital forensics, as failing to detect manipulated content may lead to misinformation or compromised evidence.

The F1-score, which balances precision and recall, further validates the reliability of the detection model. The results demonstrate that the system maintains stable performance across different image sources and manipulation techniques. Additionally, ROC-AUC analysis confirms strong discriminative capability between authentic and forged images, indicating robustness against varying compression levels and post-processing operations such as blurring and resizing. Qualitative analysis using localized forgery maps shows that the proposed system can accurately highlight tampered regions, providing visual interpretability and supporting forensic verification. Compared to traditional forensic methods, the proposed deep learning approach demonstrates superior generalization and adaptability to modern AI-based image manipulation techniques. Simulation results are shown in Fig. 2. Forgery detection, forgery localization, model insights, and visual analytics are shown in Fig. 2.

5 Conclusion

The conclusion of this research highlights the transformative role of Deep Learning in addressing the critical challenges of digital image forensics. By transitioning from traditional, manual-intensive detection methods to automated, hierarchical feature extraction using Convolutional Neural Networks (CNNs), the proposed system has demonstrated a superior ability to identify subtle "micro-artifacts" that indicate tampering. Unlike conventional heuristic approaches that often fail when images undergo post-processing like blurring or resizing, the deep learning framework leverages vast datasets to learn robust, multi-dimensional patterns of manipulation. This shift ensures that even sophisticated "seamless" forgeries—such as splicing and copy-move—can be identified with high precision, providing a vital layer of security in an era where digital content can be altered with unprecedented ease. The integration of advanced techniques such as transfer learning and localized segmentation has significantly enhanced the system's practical viability. By utilizing models pre-trained on large-scale datasets and fine-tuning them for forensic tasks, the system achieves a state-of-the-art balance between detection accuracy and computational efficiency. The generation of localized heat maps provides transparent, visual evidence of forgery, making the tool highly valuable for non-experts in journalism, law enforcement, and cybersecurity. As the digital landscape continues to evolve with the emergence of high-fidelity AI-generated content, this research lays the groundwork for a more resilient and adaptive defense mechanism. Ultimately, the successful implementation of this system proves that intelligent, software-driven solutions are essential for restoring public trust and maintaining the integrity of visual media in the modern world.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Statement

This study did not involve human or animal subjects and, therefore, did not require ethical approval.

Statement Of Conflict Of Interests

The authors declare that they have no conflicts of interest related to this study.

Licensing

References

[1] E. Liang, K. Zhang, Z. Hua, and X. Jia, "Frequency-driven deep learning network for image splicing forgery detection," *Knowledge-Based Systems*, vol. 330, p. 114365, Sep. 2025, doi: 10.1016/j.knosys.2025.114365.

[2] S. Dhivya, R. Deepika, R. A. Kumar, K. S. Tiwari, D. Bhatia, and P. Singh, "Unmasking deception harnessing noise cancellation for digital image forgery detection using Feature-Map convolutional neural networks," *International Journal of Sensors Wireless Communications and Control*, vol. 15, no. 2, pp. 184–199, Aug. 2024, doi: 10.2174/0122103279316771240725112305.

[3] V. Gopi, K. Gnanasree, G. Dharshini, M. C. Teja, V. Murali, and S. N. Kumar, “Detection of Phishing Websites using Novel Machine Learning Fusion Approach,” *International Journal of Emerging Research in Science Engineering and Management*, vol. 1, no. 6, pp. 11–18, Dec. 2025, doi: 10.58482/ijersem.v1i6.2.

[4] M. Mao, G. Jiao, W. Gao, and J. Ye, “MSFENet: Multi-Scale Filter-Enhanced Network architecture for digital image forgery trace localization,” *Computer Vision and Image Understanding*, vol. 262, p. 104550, Oct. 2025, doi: 10.1016/j.cviu.2025.104550.

[5] Y. Cheng, X. Li, X. Zhang, and C. Yang, “Image forgery localization with sparse reward compensation using curiosity-driven deep reinforcement learning,” *Journal of Visual Communication and Image Representation*, vol. 112, p. 104587, Sep. 2025, doi: 10.1016/j.jvcir.2025.104587.

[6] M. A. Manivasagam, G. Hema, A. Maheswarareddy, P. Vinitha, T. Manasa, and K. Dileep, “Electronic protection for exam paper leakage,” *International Journal of Emerging Research in Science Engineering and Management*, vol. 1, no. 6, pp. 19–25, Dec. 2025, doi: 10.58482/ijersem.v1i6.3.

[7] M.A. Manivasagam, S. Sai Ram, C. Lakshmikanth Reddy, E. Ram Charan, P. Venkata Charan, and M. Prabhash, “MQTTNET-IDS: Deep-Fuzzy Fusion for Intelligent Threat Detection,” *International Journal of Emerging Research in Science Engineering and Management*, vol. 1, no. 6, pp. 55–62, Dec. 2025, doi: 10.58482/ijersem.v1i6.7.

[8] M. Ugale and J. Midhunchakkaravarthy, “MDR-LOD2 Model: Forgery Detection using Modified Depth ResNet features and Layer Optimized Dunnock Deep Model from Videos,” *Computers & Electrical Engineering*, vol. 125, p. 110423, May 2025, doi: 10.1016/j.compeleceng.2025.110423.

[9] S. Agarwal, D. Sharma, N. Girdhar, C. Kim, and K.-H. Jung, “A Survey of Image Forensics: Exploring forgery detection in Image Colorization,” *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, vol. 84, no. 3, pp. 4195–4221, Jan. 2025, doi: 10.32604/cmc.2025.066202.

[10] K. Rehman and G. Jain, “Detection of copy-move forgery with deep CNN features using Machine learning Classifier,” *Procedia Computer Science*, vol. 258, pp. 3062–3071, Jan. 2025, doi: 10.1016/j.procs.2025.04.564.

[11] M. Li, Y. Qin, H. Zhang, and Z. Shi, “An adaptive dual-domain feature representation method for enhanced deep forgery detection,” *Journal of Automation and Intelligence*, vol. 4, no. 4, pp. 273–281, Nov. 2025, doi: 10.1016/j.jai.2025.11.003.

[12] G. Hao, P. Liang, Z. Li, H. Zhao, and H. Zhang, “Image Copy-Move Forgery detection and localization method based on Sequence-to-Sequence transformer structure,” *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, vol. 82, no. 3, pp. 5221–5238, Jan. 2025, doi: 10.32604/cmc.2025.055739.