iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76420

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

A Review on Digital Image Forgery Detection Using Deep Learning, Transfer Learning, and Hybrid Techniques

Ajna Ashraf[1], Arsha K B[2], Navami P M[3] , T S Rithuparna[4], Ms. Mahshiya Mishab[5]
[1,2,3,4]*B.Tech Student,* [5]*Asst. Professor, CSE Department, Universal Engineering College, Thrissur, Kerala*

Abstract: ***Digital images have become an essential part of everyday life, appearing in social media, news, medical, and legal contexts. However, as image-editing tools have advanced, it has become increasingly easy to alter or fabricate images. These manipulations, such as copy-move, splicing, and retouching, can spread misinformation or serve as false evidence, making image forgery detection a critical area of research. Traditional detection methods relied on manually crafted features like noise patterns or color inconsistencies, which often proved ineffective when images were compressed, resized, or modified in complex ways.In recent years, deep learning has revolutionized this field by enabling models to automatically learn useful features from data. Transfer learning, in particular, leverages pre-trained con volutional neural networks such as VGG, ResNet, and MobileNet to achieve higher accuracy, even with smaller datasets.This review examines the increasing use of transfer learning in digital image forgery detection.It discusses how techniques like Error Level Analysis (ELA) and recompression-based preprocessing help identify forgery clues, while Grad-CAM visualization aids in interpreting model decisions by highlighting manipulated regions. The paper concludes by emphasizing the need for future systems that balance accuracy, interpretability, and efficiency to provide reliable and explainable solutions for real-world image forgery detection.***

Keywords: Image Forgery Detection, Transfer Learning, Error Level Analysis, Deep Learning, Grad-CAM, MobileNetV2, CNN, Explainable AI.

I. INTRODUCTION

In today's digital world, images are one of the most effective and trusted means of communication. They are used across various domains such as social media, advertising, journalism, and education, influencing how people perceive and interpret information. However, with the availability of advanced editing tools such as Adobe Photoshop, GIMP, and AI-based image generators, it has become easier than ever to manipulate or fabricate images without leaving visible traces. This has raised serious concerns regarding the authenticity and credibility of digital content, especially when visual media is often perceived as reliable evidence.

Image forgery has emerged as a critical issue in the digital era, with significant social, political, and legal implications. Forged images are used to spread misinformation, damage reputations, fabricate evidence, or mislead the public. In some cases, they have even influenced political outcomes or judicial proceedings. The most common types of forgery include copy move, where a region of the image is duplicated within the same frame, and splicing, where components from multiple images are combined. These manipulations are often so seam less that human observers struggle to detect them manually.

Traditional forgery detection methods relied on low-level image analysis techniques such as color inconsistencies, illumination artifacts, and sensor noise estimation [3]. While such handcrafted methods provided initial success, they often failed under geometric transformations, compression, or complex editing scenarios. The need for more robust and intelligent detection systems led to the integration of deep learning into image forensics. In recent years, deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized image analysis and forgery detection. CNNs can automatically extract com plex spatial and contextual features from images, enabling high accuracy and adaptability. Among these, transfer learning has emerged as a powerful approach that leverages pre-trained models such as VGG, ResNet, DenseNet, and MobileNet for forgery detection tasks [1], [2], [5]. This allows the reuse of learned features from large-scale datasets, improving performance even with limited data.

Several researchers have combined preprocessing methods with deep models to further enhance performance. For instance, Error Level Analysis (ELA) and recompression-based preprocessing help highlight compression inconsistencies that reveal tampering traces . Furthermore, explainable AI techniques such as Grad-CAM provide visual interpretations of how CNNs detect forgery regions, improving model transparency and user trust.

This review paper examines various deep learning and transfer learning approaches used for digital image forgery detection. It highlights how modern architectures have significantly improved accuracy, generalization, and interpretability compared to traditional methods. The study also emphasizes how combining preprocessing, transfer learning, and explain able AI can lead to efficient and trustworthy systems for verifying the authenticity of digital images.

II. LITERATURE SURVEY

This section reviews significant research contributions in the field of digital image forgery detection, particularly focusing on deep learning and transfer learning approaches. The summarized works collectively demonstrate the evolution of the field—from early convolutional network applications to recent hybrid and transformer-based architectures.

The paper entitled "Image Forgery Detection Based on Deep Learning and Transfer Learning" pioneered one of the earliest explorations of transfer learning for image forgery detection. Their method utilized a Siamese network pre-trained to distinguish between cats and dogs, transferring the learned semantic dissimilarities to classify images as pristine or forged [1]. Despite being validated on a relatively small dataset of 1,348 images, the approach achieved a validation accuracy of 94.89%, proving that transferred knowledge could accelerate convergence and enhance classification efficiency. Similarly, the work "Transfer Learning Approach for Splicing and Copy-Move Image Tampering Detection" proposed a framework that effectively combined traditional preprocessing with deep learning. They employed Error Level Analysis (ELA) to reveal tampering traces, followed by six pre-trained CNN models—VGG16, VGG19, ResNet50, DenseNet121, DenseNet169, and DenseNet201—for classification. Experiments on the CASIA v2.0 dataset showed that ResNet50 achieved the highest accuracy of 97.58% [2], highlighting the advantages of deeper architectures with skip connections in preventing overfitting.

The study "Manipulation Classification for JPEG Images Using Multi-Domain Features" introduced MCNet, a frame work integrating spatial, frequency, and compression domain features to classify manipulation types in JPEG images. Their multi-stream architecture enabled the network to differentiate distortions caused by both tampering and compression artifacts. Tested on diverse manipulations such as blurring, morphing, and resampling, MCNet achieved a top-1 error of 15.21%, outperforming MISLNet and ManTraNet, and demonstrated fine-tuning potential for related tasks like Deep Fake detection [3]. Another work, "Image Forgery Detection Using Deep Learning by Recompressing Images," proposed a lightweight CNN-based approach leveraging recompression differences to highlight tampered regions.

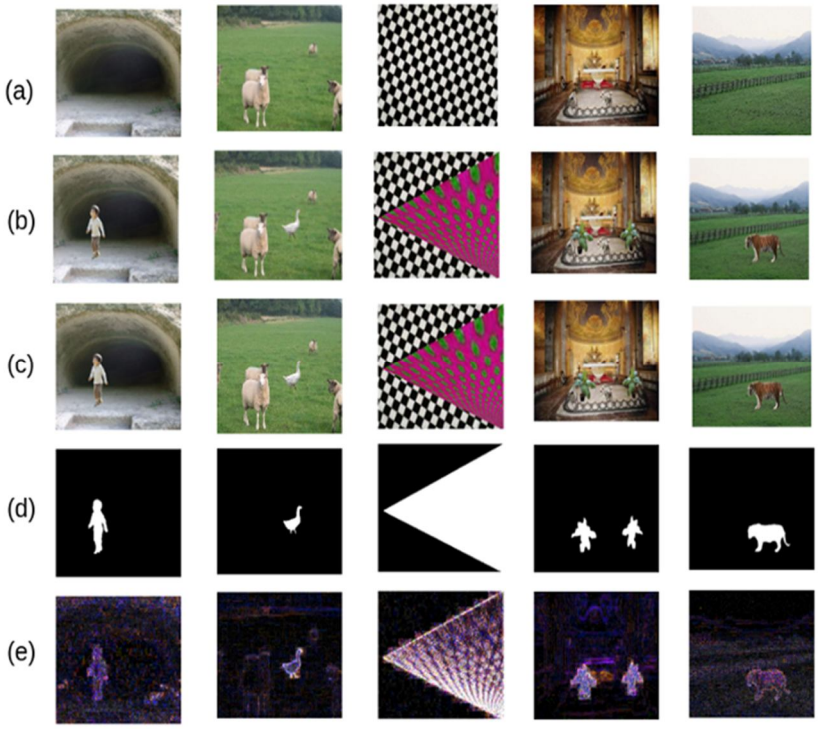


Fig. 1. Various sample images and their processed forms

By comparing an image with its recompressed version, the model effectively amplified subtle forgery artifacts. Evaluated on CASIA 2.0, it achieved 92.23% accuracy [4] and processed each image in just 34 milliseconds, outperforming CAT-Net and Buster-Net while maintaining computational efficiency. The paper "Enhancing Digital Image Forgery Detection Using Transfer Learning" presented a unified detection system for both splicing and copy-move forgeries [5]. Their approach integrated recompression-based preprocessing with transfer learning models such as VGG16, ResNet, and MobileNetV2. Among these, MobileNetV2 achieved around 95% accuracy while being lightweight, emphasizing its suitability for resource-limited devices. Similarly, "Image Splicing Forgery Detection Using Feature-Based of Sonine Functions and Deep Features" introduced a hybrid model combining handcrafted Sonine function-based texture features with CNN-derived deep features. Using the CASIA V2.0 dataset, the model achieved an outstanding accuracy of 98.93%, outperforming DCT-LBP, Haar wavelet, and U-Net models, demonstrating the effective ness of feature fusion for robust and precise splicing detection [6].

Further advancement was shown in "Detection of Tamper Forgery Image in Security Digital Image," which developed a hybrid system combining spatial and transform domain features with CNN classification. Achieving 96.50% accuracy [7], the method proved effective even under compression and noise, making it valuable for security-oriented image forensics. In "An Active Image Forgery Detection Approach Based on Edge Detection," an active forgery detection framework using edge detection and watermark embedding was proposed [8]. By integrating Canny edge features into chrominance channels via LSB watermarking, the method localized tampered regions with minimal false positives, maintaining image quality and robustness against compression and geometric transformations. Another study, "Document Image Forgery Detection and Localization in Desensitization Scenarios," addressed the challenge of detecting tampering in desensitized documents. Their dual-branch neural network fused spatial and frequency features using multiscale attention modules, achieving 98% detection accuracy even under blurring or masking [9]. The study extended forgery detection into privacy-preserving ap plications.

The work "Transfer Learning of Real Image Features with Soft Contrastive Loss for Fake Image Detection" presented the Natural Trace Forensics (NTF) framework, which focused on learning stable features of real images instead of identifying fake ones. Using a soft contrastive loss function, the model achieved 96.2% mAP across multiple generative models and exhibited strong generalization to unseen diffusion models, marking a major step forward in real-versus-fake image analysis [10]. In another development, "Deep Learning-Based Digital Image Forgery Detection System" integrated ResNet50v2 and YOLO architectures through transfer learning to detect splicing forgeries. Trained on CASIA datasets, the system achieved an impressive 99.3% accuracy [11], highlighting the power of fine-tuned pre-trained models for high-precision forensic detection.

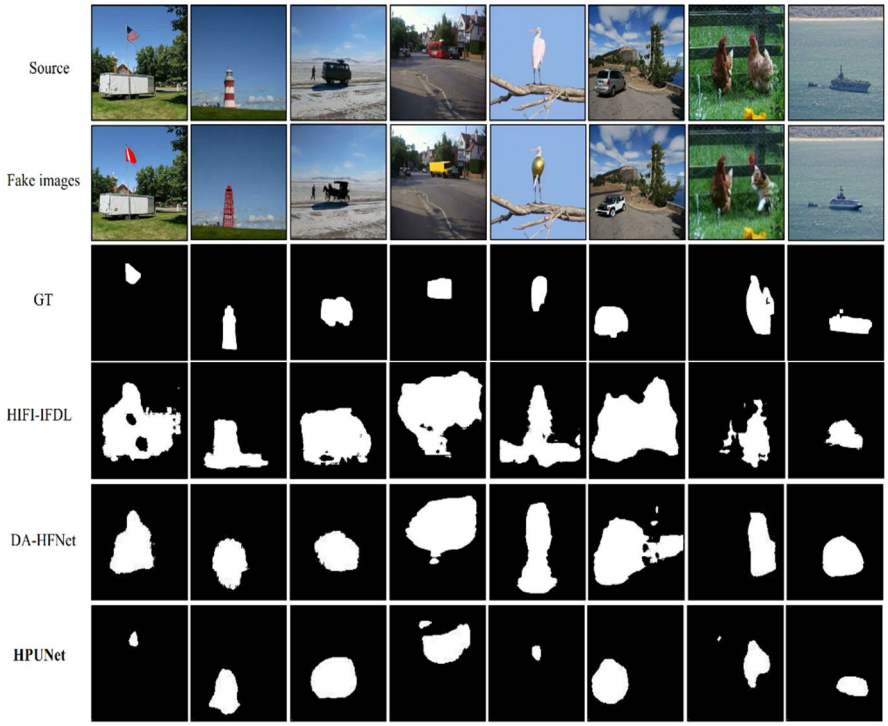


Fig. 2. Comparison of small-scale fake image localization results

The paper "An Image Forgery Detection Approach Based on Convolutional Neural Networks Using Transfer Learning" utilized a pre-trained VGG19 model fine-tuned for splicing and copy-move detection. Experiments on CASIA and Columbia datasets produced accuracies above 97% [12], demonstrating robust generalization and effective feature reuse via trans fer learning. "Image Tampering Detection Based on RDS YOLOv5 Feature Enhancement Transformation" introduced RDS-YOLOv5, combining multi-channel preprocessing with an enhanced YOLOv5 backbone. Their system achieved +6.46% F1-score and +5.13% mAP improvements over the baseline, confirming that multi-feature fusion significantly strengthens tampering detection [13]. Likewise, "Hierarchical Progressive Image Forgery Detection and Localization Method Based on UNet (HPUNet)" proposed a hierarchical UNet variant integrating spatial, frequency, and noise-domain features with dual-branch attention. Achieving 93.27% accuracy and 92.20% F1-score, HPUNet showed strong adaptability across datasets and superior localization capabilities [14].

A notable contribution, "A Two-Stage Detection Method of Copy-Move Forgery Based on Parallel Feature Fusion," developed a two-stage system combining SLIC-based segmentation with parallel SIFT-Hu moment fusion. Their approach achieved up to 99.01% accuracy on MICC-F220 and COMO FOD datasets, demonstrating exceptional robustness against compression and noise [15]. The introduction of "Grad CAM: Visual Explanations from Deep Networks via Gradient based Localization" provided an interpretability tool by producing class-discriminative heatmaps, making deep models more transparent and trustworthy—an important step towards explainable image forensics [16]. "Image Forgery Detection using VGG16-UNet and Error Level Analysis (ELA)" proposed a dual-stage framework combining ELA preprocessing with a VGG16-based UNet. Achieving 91.7% accuracy on the CASIA v2.0 dataset, the method highlighted the advantage of compression-based preprocessing coupled with transfer learning for forgery localization [17].

A comprehensive overview, "A Comprehensive Review of Deep Learning-Based Methods for Image Forensics," dis cussed the transition from handcrafted to automated feature learning, emphasizing challenges in dataset generalization and explainability, and identified transfer learning as a promising direction [18]. Similarly, "Copy-Move Forgery Detection using Deep Learning for Image and Video Forensics" analyzed CNN and Siamese architectures for detecting copy-move manipulations. Experiments on CoMoFoD and MICC-F220 datasets demonstrated that deep learning approaches outperform SIFT and SURF, though limited data and high computational requirements remain obstacles [19]. Finally, "The Effect of Error Level Analysis on Image Forgery Detection Using Deep Learning" investigated the impact of ELA preprocessing in CNN-based models. Their experiments on mixed datasets achieved up to 76% accuracy, showing a 2–3% improvement over non-ELA baselines and validating ELA as a simple yet effective preprocessing enhancement [20].

In summary, these studies collectively highlight the rapid evolution of digital image forgery detection. The integration of transfer learning, hybrid architectures, and feature fusion has led to impressive accuracy and robustness across datasets. Future research directions include improving interpretability, ensuring scalability to diverse image domains, and reducing computational overhead to enable real-time, trustworthy forgery detection systems.

III. ANALYSIS AND DISCUSSION

The evolution of fake image detection represents a re markable journey of technological advancement in the fields of computer vision and artificial intelligence. In the early stages, image forgery detection primarily relied on manually engineered features that examined various low-level inconsistencies such as color discrepancies, sensor noise patterns, edge artifacts, or compression irregularities. These traditional techniques, although effective for simple manipulations, often failed to handle complex editing operations like splicing, scaling, and recompression. As a result, their accuracy and generalization across diverse datasets were limited. The emergence of deep learning fundamentally transformed this landscape by introducing data-driven approaches that automatically learn hierarchical representations from raw images. Convolutional Neural Networks (CNNs) have demonstrated exceptional capability in extracting discriminative features, enabling them to detect subtle traces of image tampering that were previously overlooked by handcrafted methods. Transfer learning, in particular, has further accelerated progress in this domain. By leveraging pre-trained models such as VGG16, ResNet50, DenseNet121, and MobileNetV2—originally devel oped for large-scale image classification tasks—researchers have been able to fine-tune these architectures for specific forgery detection purposes. This not only reduces the dependence on large annotated datasets but also significantly decreases training time while maintaining high performance and stability. Several recent studies have also explored hybrid architectures that integrate deep learning with traditional machine learning classifiers such as Support Vector Machines (SVM) and Random Forests. These combinations are particularly beneficial in scenarios with limited computational resources or small datasets, offering a good balance between accuracy and efficiency. In parallel, classical preprocessing methods such as Error Level Analysis (ELA), Discrete Wavelet Transform (DWT), and recompression-based filtering continue to serve as valuable tools for highlighting tampered regions before feature extraction.

When coupled with deep models—such as VGG16-UNet or ResNet-UNet architectures—these pre processing techniques enhance both detection precision and tampering localization.

Moreover, the integration of handcrafted and deep features has proven to improve robustness against distortions introduced by compression, blurring, or occlusion. Lightweight CNN models such as MobileNetV2 and EfficientNet are particularly appealing for real-time and mobile-based implementations, where energy efficiency and low memory usage are crucial. Their deployment potential extends beyond academic research to real-world applications, including mobile forensics, social media integrity verification, and law enforcement investigations.

A notable emerging direction in this field is the use of explainable artificial intelligence (XAI) techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM) and Layer-wise Relevance Propagation (LRP). These visualization tools provide interpretable heatmaps that highlight which regions of the image most influenced the model's decision, thereby improving transparency, reliability, and human A notable emerging direction in this field is the use of explainable artificial intelligence (XAI) techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM) and Layer-wise Relevance Propagation (LRP). These visualization tools provide interpretable heatmaps that highlight which regions of the image most influenced the model's decision, thereby improving transparency, reliability, and human trust in automated systems. Such interpretability is essential in forensic and judicial contexts, where explainable evidence is as important as accuracy.

TABLE 1
Extended comparison of deep learning and transfer learning models discussed in this paper for image forgery detection.

Model	Features Extraction/Method Description	Accuracy(%)
VGG16	Fine-tuning on pre-trained layers; strong baseline with ELA preprocessing for forgery detection	90–96
ResNet50	Deep residual blocks with skip connections; high performance on CASIA v2.0 dataset	92-98
DenseNet121	Dense layer-wise feature reuse improves feature propagation and efficiency	91–97
MobileNetV2	Lightweight CNN architecture using depthwise separable convolutions; suitable for limited resource devices	88–94
Xception	Depthwise separable convolutions with residual learning for faster convergence and reduced computation	89–95
InceptionV3	Multi-scale convolution feature extraction for robust detection under varying image sizes	90–96
Hybrid CNN + SVM	Combines CNN-based feature extraction with SVM classification for better generalization on small datasets.	88–93
VGG16–UNet	Combines VGG16 encoder with UNet decoder for improved localization of manipulated image regions	91–94
RDS–YOLOv5	Enhanced YOLOv5 with multi-channel feature transformation; stronger tampering detection per formance.	94–98
HP–UNet	Hierarchical Progressive UNet integrating attention and multi domain features for precise forgery localization	93–96

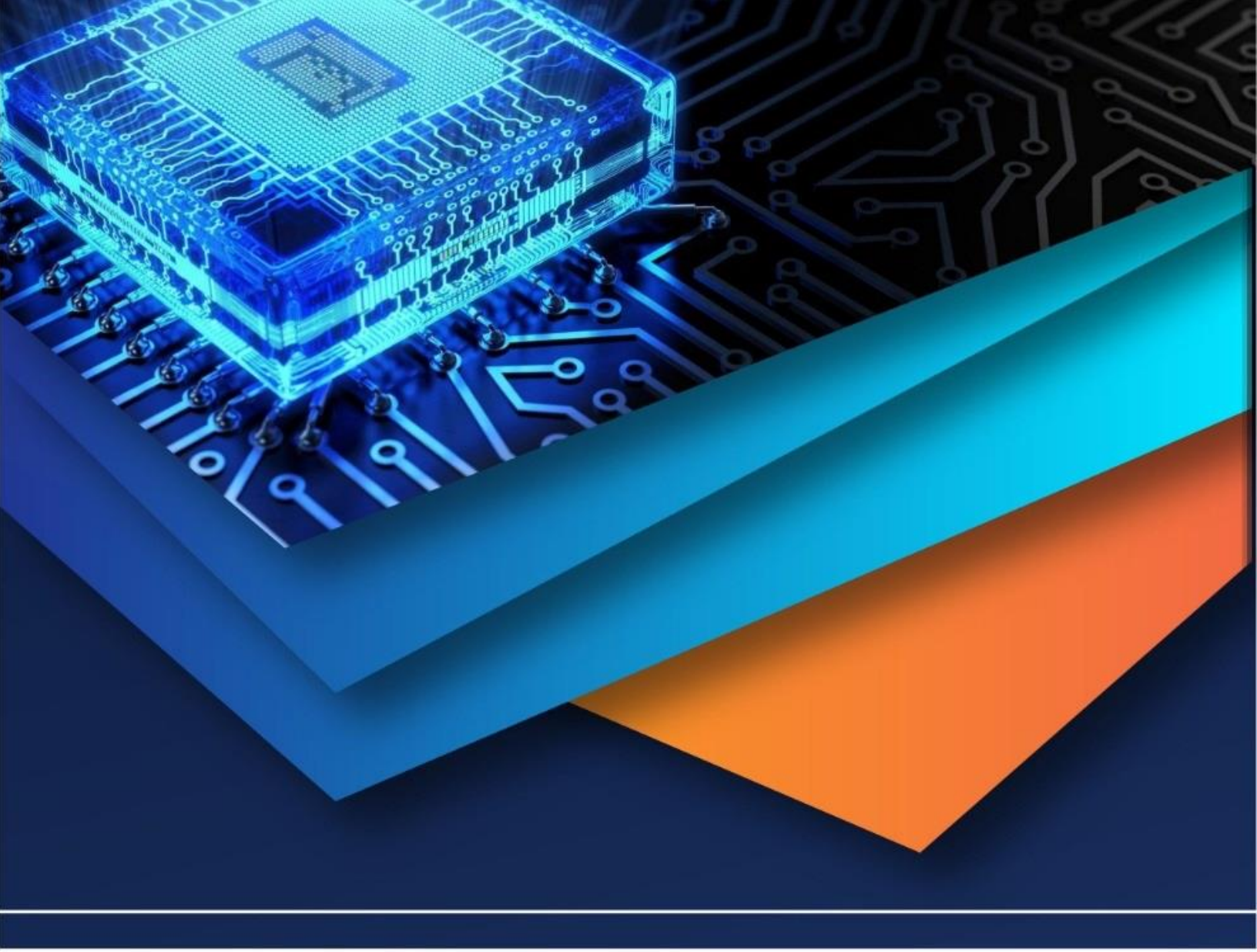
In summary, the most effective modern frameworks for digital image forgery detection integrate multiple complementary components deep learning, transfer learning, preprocessing, hybrid modelling, and explainable AI. Moving forward, research efforts should focus on enhancing model generalization across unseen datasets, reducing computational overhead for real-time applications, and ensuring fairness and transparency. These advancements will be vital for the trustworthy deployment of forgery detection systems in real-world forensic, cybersecurity, and media authentication domains.

IV. CONCLUSIONS

This paper reviewed several transfer-learning-based approaches for digital image forgery detection. Studies demonstrate that ELA preprocessing combined with lightweight CNN backbones like MobileNetV2 improves detection accuracy [2], [5]. However, interpretability and multi-forgery capability remain open research problems [18]. Future research should emphasize explainable AI models using Grad-CAM [16] and hybrid CNN architectures [13], [14] to enhance trust and usability in real-world applications.

REFERENCES

[1] Y. E. Abdalla, M. T. Iqbal, and M. Shehata, “Image forgery detection based on deep transfer learning,” European Journal of Electrical and Computer Engineering, vol. 3, no. 5, pp. 1–8, 2019.
[2] A. Hebbar and A. Kunte, “Transfer learning approach for splicing and copy-move image tampering detection,” in Proc. IEEE International Conference on Electronics, Computing and Communication Technologies, 2021.
[3] I.-J. Yu, S.-H. Nam, W. Ahn, M.-J. Kwon, and H.-K. Lee, “Manipulation classification for JPEG images using multi-domain features,” IEEE Access, vol. 8, pp. 210837–210854, 2020.
[4] S. S. Ali, I. I. Ganapathi, N.-S. Vu, S. D. Ali, N. Saxena, and N. Werghi, “Image forgery detection using deep learning by recompressing images,” Electronics, vol. 11, no. 3, p. 403, 2022.
[5] T. Khalil, S. A. Ali, and M. A. Khan, “Enhancing digital image forgery detection using transfer learning,” IEEE Access, vol. 11, pp. 91583 91594, 2023.
[6] A. R. Al-Shamasneh and R. W. Ibrahim, “Image splicing forgery detection using feature-based of Sonine functions and deep features,” Journal of King Saud University-Computer and Information Sciences, vol. 36, no. 1, p. 101925, 2024..
[7] M. F. Abdulqader, A. Y. Dawod, and A. Z. Ablahd, “Detection of tamper forgery image in security digital image,” Measurement: Sensors, vol. 27, p. 100746, 2023.
[8] H. B. Macit and A. Koyun, “An active image forgery detection approach based on edge detection,” in Proc. IEEE International Conference on Innovations in Intelligent Systems and Applications, 2023.
[9] W. Li, B. Li, K. Zheng, S. Li, and H. Li, “Document image forgery detection and localization in desensitization scenarios,” Engineering Applications of Artificial Intelligence, vol. 128, p. 107541, 2025.
[10] Z. Liang, W. Liu, R. Wang, M. Wu, B. Li, Y. Zhang, L. Wang, and X. Yang, “Transfer learning of real image features with soft contrastive loss for fake image detection,” Engineering Applications of Artificial Intelligence, vol. 133, p. 108542, 2024.
[11] E. U. H. Qazi, T. Zia, and A. Almorjan, “Deep learning-based digital image forgery detection system,” International Journal of Intelligent Automation and Soft Computing, vol. 34, no. 2, pp. 1125–1140, 2022.
[12] A. K. Jaiswal and S. Kumar, “An image forgery detection approach based on convolutional neural networks using transfer learning,” Journal of Intelligent Systems, vol. 32, no. 1, pp. 45–58, 2023.
[13] Y. Zhu, X. Wang, and L. Chen, “Image tampering detection based on RDS-YOLOv5 feature enhancement transformation,” Scientific Reports, vol. 14, no. 1, p. 5678, 2024.
[14] H. Liu, W. Zhang, and R. Yang, “Hierarchical progressive image forgery detection and localization method based on UNet (HPUNet),” Big Data and Cognitive Computing, vol. 8, no. 3, p. 234, 2024.
[15] W. Ye, J. Li, and H. Wang, “A two-stage detection method of copy-move forgery based on parallel feature fusion,” EURASIP Journal on Wireless Communications and Networking, vol. 2022, no. 1, p. 45, 2022.
[16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” in Proc. IEEE International Conference on Computer Vision (ICCV), pp. 618–626, 2017.
[17] M. Karthika, P. Saranya, R. Sudhakar, V. Uma, and R. Kiruthika, “Image Forgery Detection using VGG16-UNet and Error Level Analysis (ELA),” in Procedia Computer Science, vol. 232, pp. 1502–1510, 2024.
[18] S. Agarwal, P. Singh, and D. S. Chauhan, “A Comprehensive Review of Deep Learning-Based Methods for Image Forensics,” in Journal of Imaging, vol. 7, no. 11, pp. 1–20, 2021.
[19] R. Kaur, A. Kaur, and R. Vig, “Copy-Move Forgery Detection using Deep Learning for Image and Video Forensics,” in Journal of Imaging, vol. 7, no. 5, pp. 1–14, 2021..
[20] A. Khatoon, S. Khan, and A. Rehman, “The Effect of Error Level Analysis on Image Forgery Detection Using Deep Learning,” in Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, vol. 6, no. 2, pp. 195–202, 2021



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY