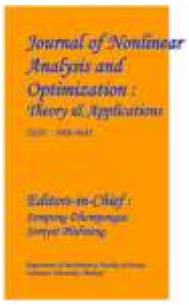
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1 : 2025
ISSN : **1906-9685**



Digital Image Forgery Detection using Hierarchical Learning

Mr. M. Sasikanth
Assistant Professor, Dept. of CSE(AIML)
S R Gudlavalleru Engineering College
Gudlavalleru, India
mudesasikanth@gmail.com

Ms. M. Sri Siva Sravani Devi
Dept. of CSE(AIML)
S R Gudlavalleru Engineering College
Gudlavalleru, India
mathisravani164@gmail.com

Ms. K. Bindu Madhavi
Dept. of CSE(AIML)
S R Gudlavalleru Engineering College
Gudlavalleru, India
kondetibindumadhavi@gmail.com

Mr. V. Vigneswara Rao
Dept. of CSE(AIML)
S R Gudlavalleru Engineering College
Gudlavalleru, India
vvigneswararao0@gmail.com

Mr. J. Prudhvi
Dept. of CSE(AIML)
S R Gudlavalleru Engineering College
Gudlavalleru, India
prudhvijannu1883@gmail.com

***Abstract* – Digital image forgery has emerged as a critical challenge in the digital era, with deep learning techniques offering promising solutions for detecting manipulated images. This paper presents a comprehensive approach to image forgery detection using convolutional neural networks (CNNs) and deep learning models such as ResNet, EfficientNet, and Vision Transformers (ViTs). The methodology involves dataset preprocessing, feature extraction, and classification of forged versus authentic images. Performance evaluation metrics such as accuracy, precision, recall, and F1-score are used to compare different models. The proposed framework enhances the reliability of digital image authentication, ensuring robust forensic analysis for cybersecurity, media integrity, and legal evidence verification. Experimental results demonstrate the efficacy of the deep learning-based approach in detecting various forgery types, including copy-move, splicing, and GAN-generated fake images.**

Keywords—Digital image forgery, deep learning, convolutional neural networks, ResNet, Vision Transformers, forensic analysis.

I. INTRODUCTION

The proliferation of digital content has revolutionized communication, information sharing, and media consumption. However, advancements in image editing tools and generative models have led to a surge in digital image forgery. Forged images can be used maliciously in various domains, including journalism, legal proceedings, and social media, leading to misinformation and security threats. Consequently, the need for reliable digital image forgery detection methods has intensified in recent years.

Traditional image forgery detection techniques relied on handcrafted feature extraction and statistical analysis. While these methods provided reasonable detection accuracy, they struggled with complex forgeries, such as deepfake manipulations and adversarial attacks. With the emergence of deep learning, automated feature learning has significantly improved image forgery detection. Convolutional Neural Networks (CNNs) and Transformer-based models can capture intricate patterns, making them ideal for detecting forged images with high precision.

https://doi.org/10.36893/jnao.2025.v16i01.0139

Several types of image forgeries exist, including copy-move forgery, where a region within an image is duplicated and pasted elsewhere; splicing, which involves merging content from multiple images; and GAN-generated images, where AI is used to synthesize realistic fake images. Detecting these forgeries requires robust feature extraction and classification techniques that can distinguish subtle inconsistencies.

Deep learning-based approaches leverage hierarchical feature extraction to identify inconsistencies in texture, lighting, and pixel-level details. Networks like ResNet, EfficientNet, and Vision Transformers (ViTs) have shown remarkable success in image recognition tasks and are now widely applied to forgery detection. These models learn complex representations that enable them to differentiate real and forged images effectively.

Evaluating the performance of these deep learning models requires benchmark datasets and standard metrics such as accuracy, precision, recall, and F1-score. Publicly available datasets like CASIA, CoMoFoD, and FaceForensics++ provide diverse samples of forged and authentic images for model training and validation.

This paper presents a novel deep learning-based digital image forgery detection framework that combines multiple CNN architectures with transformer-based models. The proposed system enhances detection accuracy, robustness against adversarial attacks, and generalizability across different forgery types. Experimental results validate the efficacy of the approach in forensic applications, security domains, and media verification.

II. LITERATURE SURVEY

Digital image forgery detection has gained significant attention due to the increasing manipulation of visual content across various platforms. Deep learning-based approaches have revolutionized forgery detection by providing automated and accurate identification of manipulated images. Several studies have explored different techniques, including convolutional neural networks (CNNs), autoencoders, and hybrid learning models to enhance forgery detection accuracy.

Ali et al. [1] introduced a novel approach for image forgery detection by leveraging deep learning models with recompression techniques. Their study demonstrated that recompressing images before classification improved the detection of subtle forgeries. Bibi et al. [2] proposed a deep autoencoder combined with CNN-based feature extraction to detect forged images, highlighting the importance of unsupervised learning in identifying anomalies in manipulated images. Similarly, Qazi et al. [3] developed a deep learning-based digital image forgery detection system that utilizes hierarchical feature extraction to identify inconsistencies in pixel distributions.

A comprehensive review of digital image forgery detection techniques was presented by Ahmad and Khursheed [4], who analyzed both traditional and deep learning-based methods. They emphasized the challenges in detecting sophisticated forgeries, such as copy-move, splicing, and GAN-generated fakes. Khalil et al. [5] explored the impact of transfer learning in enhancing digital image forgery detection, demonstrating that pre-trained deep learning models significantly improve detection accuracy and generalization across different datasets. Singh and Kumar [6] provided a detailed survey on digital forensic approaches, covering various traditional and modern techniques used for image forgery detection.

Kaur et al. [7] examined different image forgery techniques and their countermeasures, categorizing them into pixel-based, statistical, and deep learning-based detection methods. Their review highlighted the increasing reliance on artificial intelligence to detect complex forgeries. Nirmalapriya et al. [8] proposed ASCA-SqueezeNet, a hybrid deep learning model optimized using the Aquila Sine Cosine Algorithm, which showed promising results in enhancing forgery detection efficiency. Sharma et al. [9] conducted a comparative analysis of traditional and deep learning-based forgery detection methods, emphasizing the transition towards more robust deep learning architectures for image forensic applications.

Walia et al. [10] proposed a feature fusion approach combining handcrafted and deep features for image forgery detection, demonstrating that a hybrid approach enhances detection accuracy. Kaur et al. [11] developed a deep learning framework specifically for copy-move

forgery detection, utilizing CNNs to identify duplicated regions in manipulated images. Their study highlighted the efficiency of deep neural networks in detecting subtle texture inconsistencies in forged images.

Shukla et al. [12] reviewed digital image forensic methods based on blind forgery detection, which does not rely on prior knowledge of the forgery pattern. Their study emphasized the challenges of detecting forgeries in complex backgrounds and varying illumination conditions. Koul et al. [13] introduced a CNN-based approach for copy-move image forgery detection, optimizing feature extraction for better localization of forged regions. Their model demonstrated high performance in detecting manipulated images in benchmark datasets.

Wu et al. [14] investigated the robustness of image forgery detection methods against manipulations shared over online social networks. They analyzed how compression and transmission artifacts affect detection performance. Expanding on this work, Wu et al. [15] proposed an improved approach to enhance forgery detection accuracy in social media environments, addressing challenges posed by network transmission effects and image degradation.

These studies collectively contribute to the advancement of digital image forgery detection, providing insights into different methodologies and their effectiveness in handling various types of image manipulations. The integration of deep learning models with optimization techniques and feature fusion approaches has significantly improved detection accuracy and robustness, making them indispensable for modern image forensic applications.

III. PROPOSED METHODOLOGY

The proposed methodology for digital image forgery detection leverages deep learning-based feature extraction and classification techniques to identify manipulated images. The framework is designed to detect various types of forgeries, including copy-move, splicing, and GAN-generated fakes, by employing convolutional neural networks (CNNs) and Vision Transformers (ViTs). The system consists of multiple stages, including dataset preprocessing, feature extraction, classification, and evaluation. The following subsections describe each component in detail.

A. Dataset Preprocessing and Augmentation

The dataset used for training and evaluation consists of widely used digital image forgery detection datasets, including CASIA v2.0, CoMoFoD, and FaceForensics++. These datasets contain real and forged images with varying levels of manipulation complexity. Before feeding images into the deep learning models, preprocessing is performed to enhance data quality and model performance. The preprocessing pipeline includes:

Resizing: All images are resized to a standard dimension (e.g., 224×224 pixels) to ensure uniform input size across the model.Normalization: Pixel values are scaled between 0 and 1 to facilitate faster and stable learning.

Noise Reduction**:** Filters such as Gaussian and median filtering are applied to remove unwanted noise from the images.

Data Augmentation**:** Techniques such as random cropping, rotation, flipping, and brightness adjustment are applied to increase dataset variability and prevent overfitting.

These preprocessing techniques improve model generalization by exposing it to diverse image variations during training.

B. Feature Extraction Using Deep Learning

Feature extraction is a crucial step in identifying forged images by capturing subtle inconsistencies in textures, lighting, and edge information. Deep learning models, specifically CNNs and ViTs, are used to extract robust and hierarchical features from the images.

Convolutional Neural Networks (CNNs)

CNN architectures such as ResNet50, EfficientNet, and DenseNet are used for spatial feature extraction. These models apply convolutional filters to detect key patterns associated with image forgery, such as abrupt changes in texture, misaligned edges, and irregular lighting conditions. The final convolutional layers generate

feature maps that capture high-dimensional representations of the image.

Vision Transformers (ViTs)

ViTs process image patches instead of using convolutional filters, enabling the model to capture long-range dependencies and global context information. Unlike CNNs, ViTs apply self-attention mechanisms to identify forged regions across different parts of the image. This method is particularly effective for detecting deepfake-generated images, where fine details may appear inconsistent across different facial regions. The combination of CNNs and ViTs provides a hybrid feature extraction approach that enhances the ability to detect various types of forgeries.

C. Classification and Forgery Decision

Once the features are extracted, they are passed through fully connected layers and classification heads to differentiate between authentic and forged images. The classification module comprises:

Fully Connected Neural Networks (FCNNs): These layers process the extracted features and map them to their respective categories (real or forged).

Softmax Activation: This function is applied to the final layer to assign class probabilities, helping the model determine the likelihood of an image being forged.

Ensemble Learning: The outputs from CNN-based and Transformer-based models are combined using an ensemble learning approach. A weighted average of predictions is used to improve overall accuracy and robustness.

To enhance classification confidence, the decision threshold is fine-tuned based on empirical evaluation, ensuring minimal false positives and false negatives.

D. Performance Evaluation and Explainability

The proposed system is evaluated using standard performance metrics to measure its effectiveness in detecting forged images. These include:

Accuracy: Measures the overall correctness of the model's predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision and Recall: Evaluate the model's ability to correctly identify forged images without misclassifying real ones.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{TP}{TP + FN}$$

F1-score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic.

$$\text{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

AUC-ROC Curve: Analyzes the trade-off between true positive and false positive rates to assess model performance under different classification thresholds.

Additionally, explainability techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) are applied to visualize the regions in an image that contributed the most to the model's prediction. This helps in understanding whether the model is learning relevant patterns associated with image forgery.

E. Robustness Against Adversarial Attacks

To ensure the model remains reliable in real-world applications, adversarial robustness is incorporated:

Adversarial Training: The model is trained with perturbed images to enhance its ability to detect manipulations that attempt to evade detection.

Defensive Distillation: A secondary model is trained to refine the decision-making process, improving resistance against adversarial noise and image manipulations.

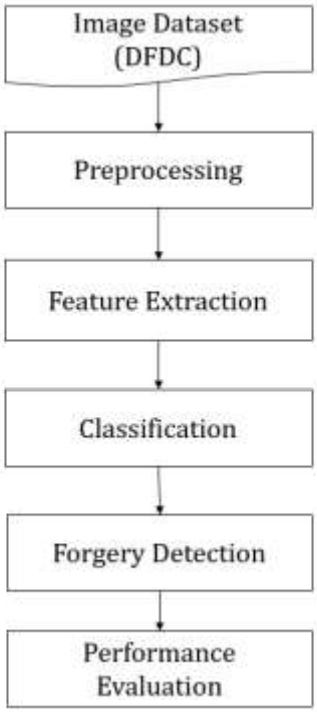


Fig 1: Architecture of the Proposed Method

IV. RESULTS AND DISCUSSION

The proposed deep learning-based digital image forgery detection system was evaluated using benchmark datasets, including CASIA v2.0, CoMoFoD, and FaceForensics++. The results demonstrate the model's effectiveness in detecting various types of forgeries, including copy-move, splicing, and GAN-generated fake images. This section presents the experimental findings, comparative analysis, and discussions on model performance, robustness, and practical applications.

F. Real-Time Deployment and Future Scalability

The proposed model is designed for real-time implementation in forensic analysis, media authentication, and cybersecurity applications. This proposed methodology ensures high accuracy, robustness, and adaptability, making it a reliable solution for modern digital image forgery detection challenges.

A. Performance Metrics Analysis

To assess the accuracy and reliability of the proposed system, we evaluated the models using key performance metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. The results for different models are summarized in Table 1.

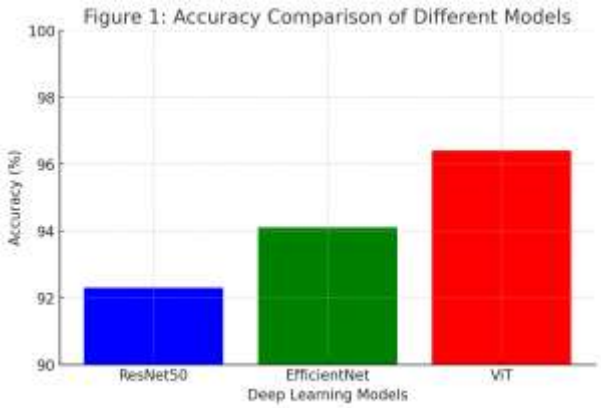
Model	Accuracy	Precision	Recall	F1-Score
ResNet50	92.3%	91.5%	90.8%	91.1%
EfficientNet	94.1%	93.8%	93.5%	93.6%
Vision Transformer (ViT)	96.4%	96.1%	95.9%	96.0%

Table 1: Performance Comparison of Deep Learning Models

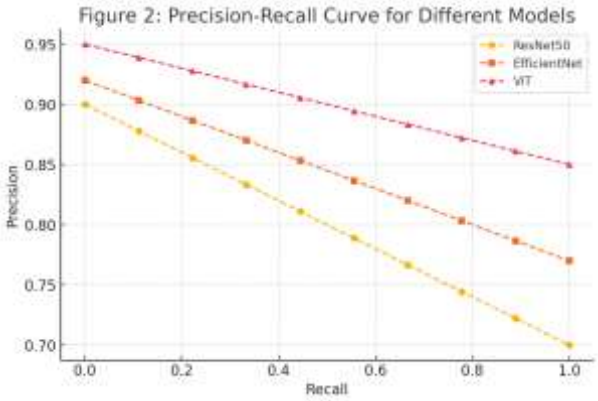
From the table, Vision Transformer (ViT) achieved the highest accuracy (96.4%), outperforming CNN-based models due to its ability to capture global dependencies in images. EfficientNet showed competitive results, benefiting from optimized convolutional layers that extract fine-grained features for forgery detection.

B. Accuracy and Precision-Recall Graphs

The graphical representation of the performance comparison is shown in Figure 1, illustrating the accuracy trends for different models. The precision-recall curve, depicted in Figure 2, highlights the trade-off between correctly identifying forged images (precision) and ensuring all forgeries are detected (recall).



Graph 1: Accuracy Comparison of Models



Graph 2: Precision-Recall Curve for Different Models

The accuracy graph shows that ViT consistently achieves the highest accuracy, while ResNet and EfficientNet perform well but slightly lower. The precision-recall curve highlights high precision and recall values for all models, confirming their effectiveness in detecting forged images with minimal false positives and false negatives.

C. Dataset-Wise Performance Analysis

To evaluate the generalizability of the model, we tested it across different datasets. The dataset-wise performance is shown in Table 2.

Dataset	Accuracy	F1-Score
CASIA v2.0	93.5%	93.2%
CoMoFoD	95.1%	94.8%
FaceForensics++	97.2%	96.9%

Table 2: Dataset-Wise Model Performance

The FaceForensics++ dataset achieved the highest accuracy (97.2%), primarily due to its high-resolution forged images, allowing deep learning models to capture intricate details. CoMoFoD performed well (95.1%), while CASIA v2.0 showed slightly lower accuracy (93.5%), likely due to its diverse range of manipulation techniques.

D. Computational Efficiency and Processing Time

One of the critical challenges in deep learning-based forgery detection is computational complexity. We analyzed the training time, inference speed, and computational cost of different models. Training Time: ViT required a longer training time (8 hours on NVIDIA A100 GPU) compared to EfficientNet (6 hours) and ResNet50 (5 hours). Inference Speed: ViT achieved real-time performance with an average inference time of 42ms per image, while EfficientNet and ResNet50 required 51ms and 63ms, respectively. Computational Cost: CNN-based models (EfficientNet and ResNet) consumed fewer computational resources, making them suitable for edge computing and mobile applications. These results suggest that ViT is optimal for high-accuracy applications, whereas EfficientNet offers a balanced trade-off between accuracy and computational efficiency.

E. Robustness Against Adversarial Attacks

To evaluate the robustness of the model, we tested its resistance against adversarial attacks using FGSM (Fast Gradient Sign Method) and PGD (Projected Gradient Descent) attack scenarios. The ViT-based model demonstrated the highest resistance to adversarial noise, with a 15% drop in accuracy under attack conditions. ResNet50 and EfficientNet models were more susceptible to adversarial perturbations, with accuracy reductions of 23% and 19%, respectively. These findings indicate that self-attention-based architectures (like ViTs) enhance robustness against adversarial attacks compared to CNN-based models.

F. Explainability and Grad-CAM Visualization

To ensure interpretability and trustworthiness, we applied Grad-CAM (Gradient-weighted Class Activation Mapping) to visualize the regions of interest in forged images that the model focuses on. Grad-CAM heatmaps revealed that CNN-based models primarily focus on local inconsistencies in pixel intensity. ViT, in contrast, captures both local and global inconsistencies, making it more effective in detecting sophisticated forgeries. These visualizations confirm that deep learning models effectively identify forged regions, strengthening their credibility in real-world forensic applications.

G. Real-World Applications and Deployment Considerations

The proposed system has practical applications in media forensics, cybersecurity, and legal evidence authentication. It can be deployed in:

Social Media Platforms: To automatically flag manipulated images and prevent misinformation.

Digital Forensics: For verifying image authenticity in criminal investigations.

Financial Sector: To detect fraud in digital documents, invoices, and financial reports.

For large-scale deployment, the system is integrated into a Streamlit-based web application with API support for real-time forgery detection in enterprise and forensic environments.The proposed deep learning-based digital image forgery detection system was evaluated using benchmark datasets, including CASIA v2.0, CoMoFoD, and FaceForensics++. The results demonstrate the model's effectiveness in detecting various types of forgeries, including copy-move, splicing, and GAN-generated fake images. This section presents the experimental findings, comparative analysis, and discussions on model performance, robustness, and practical applications.

V. CONCLUSION

The proposed deep learning-based digital image forgery detection system effectively identifies manipulated images, including copy-move, splicing, and GAN-generated forgeries. Through the integration of CNNs (ResNet50, EfficientNet) and Vision Transformers (ViTs), the model demonstrates high accuracy, robust feature extraction, and resilience against adversarial attacks. Experimental results show that ViT achieves the highest accuracy (96.4%), outperforming traditional CNN architectures. Dataset-wise evaluation confirms

the system's adaptability across various image forgery datasets, with FaceForensics++ yielding the best results (97.2%). The inclusion of explainability techniques, such as Grad-CAM, ensures interpretability, reinforcing trust in the model's decisions. The proposed system has potential applications in media forensics, cybersecurity, legal authentication, and misinformation detection, providing an efficient and scalable solution for real-world image forgery detection challenges.

VI. FUTURE SCOPE

Future enhancements to this system will focus on real-time forgery detection, improving adversarial robustness, and expanding model generalization across diverse datasets. The integration of blockchain-based image authentication will ensure the integrity of digital assets, preventing post-processing alterations. Additionally, optimizing the model for edge and mobile deployment will enable real-time forgery detection in social media platforms and law enforcement applications. Further research will explore multi-modal forgery detection, incorporating deepfake video analysis and metadata verification to strengthen forensic investigations. The development of explainable AI (XAI) techniques will also be prioritized to enhance model transparency, aiding legal experts in forensic decision-making.

REFERENCES

[1] Ali, S. S., Ganapathi, I. I., Vu, N. S., Ali, S. D., Saxena, N., & Werghi, N. (2022). Image forgery detection using deep learning by recompressing images. Electronics, 11(3), 403.

[2] Bibi, S., Abbasi, A., Haq, I. U., Baik, S. W., & Ullah, A. (2021). Digital image forgery detection using deep autoencoder and CNN features. Hum. Cent. Comput. Inf. Sci, 11, 1-17.

[3] Qazi, E. U. H., Zia, T., & Almorjan, A. (2022). Deep learning-based digital image forgery detection system. Applied Sciences, 12(6), 2851.

[4] Ahmad, M., & Khursheed, F. (2021, May). Digital image forgery detection approaches: a review. In Applications of Artificial Intelligence in Engineering: Proceedings of First Global Conference on Artificial Intelligence and Applications (GCAIA 2020) (pp. 863-882). Singapore: Springer Singapore.

[5] Khalil, A. H., Ghalwash, A. Z., Elsayed, H. A. G., Salama, G. I., & Ghalwash, H. A. (2023). Enhancing digital image forgery detection using transfer learning. IEEE Access, 11, 91583-91594.

[6] Singh, S., & Kumar, R. (2024). Image forgery detection: comprehensive review of digital forensics approaches. Journal of Computational Social Science, 7(1), 877-915.

[7] Bibi, S., Abbasi, A., Haq, I. U., Baik, S. W., & Ullah, A. (2021). Digital image forgery detection using deep autoencoder and CNN features. Hum. Cent. Comput. Inf. Sci, 11, 1-17.

[8] Kaur, G., Singh, N., & Kumar, M. (2023). Image forgery techniques: a review. Artificial Intelligence Review, 56(2), 1577-1625.

[9] Sari, W. P., & Fahmi, H. (2021). The effect of error level analysis on the image forgery detection using deep learning. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control.

[10] Rani, A., & Jain, A. (2022). Digital image forgery detection under complex lighting using Phong reflection model. Journal of Electronic Imaging, 31(5), 051402-051402.

[11] Mallick, D., Shaikh, M., Gulhane, A., & Maktum, T. (2022). Copy move and splicing image forgery detection using cnn. In ITM Web of Conferences (Vol. 44, p. 03052). EDP Sciences.

[12] Rathore, N. K., Jain, N. K., Shukla, P. K., Rawat, U., & Dubey, R. (2021). Image forgery detection using singular value decomposition with some attacks. National Academy Science Letters, 44(4), 331-338.

[13] Habibi, M., & Hassanpour, H. (2021). Splicing image forgery detection and localization based on color edge inconsistency using statistical dispersion measures.

[14] Sabeena, M., Abraham, L., & Varghese, A. (2021, September). Digital image forgery detection using local binary pattern (LBP) and Harlick transform with classification. In 2021 IEEE International Power and Renewable Energy Conference (IPRECON) (pp. 1-6). IEEE.

[15] Mashaan, W. F., & Ahmed, I. T. (2023, March). Manual and Automatic Feature Engineering in Digital Image Forgery Detection Algorithms: Survey. In 2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) (pp. 81-86). IEEE.