JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# AI POWERED FAKE IMAGE DETECTION FOR SOCIAL MEDIA USING RESNET AND U-NET

**[1]Sandhya N, [2]Shalini S, [3]Tehashwini S, [4]Thrisha P, [5] Dr. Anita P**

[1] sandhyasandhya1271@gmail.com , [2] shaliniselvaraj52004@gmail.com , [3]thejashwinis2005@gmail.com ,
[4] thrishap35@gmail.com [5]anitap@ksit.edu.in

[1,2,3,4] Student, Department of Electronics and Communication Engineering, K. S. Institute of Technology, Bengaluru, India

[5] Assistant Professor, Department of Electronics and Communication Engineering, K. S. Institute of Technology, Bengaluru, India

Abstract: The rapid growth of social media platforms and advancements in digital image editing technologies have increased the circulation of fake and manipulated images, making image authenticity verification a significant challenge. Such images can spread misinformation, mislead the public, and reduce trust in digital content. Traditional image forgery detection techniques often face difficulties in identifying sophisticated manipulations, particularly when images undergo compression, resizing, or other post-processing operations. To address these challenges, this paper presents an Artificial Intelligence (AI)-powered fake image detection system using Residual Network (ResNet) and U-Net. The proposed system utilizes ResNet, a deep convolutional neural network architecture, to extract important visual features and classify input images as real or potentially forged. Image preprocessing techniques, including resizing and normalization, are applied to prepare images for efficient model processing. In addition, U-Net is incorporated as the proposed segmentation model to identify and localize suspected manipulated regions by generating pixel-level masks. The combination of image classification and forgery localization aims to provide more informative results than classification alone. The current implementation focuses on ResNet-based classification, which has been tested using sample images, while U-Net-based localization requires further validation to establish its effectiveness. The system is intended to support social media image verification, digital content authentication, and digital forensic investigations. Future enhancements include improving localization performance, evaluating the system using comprehensive performance metrics, and deploying optimized models on a Raspberry Pi to enable portable, edge-based image analysis.

**Keywords**— Artificial Intelligence, Fake Image Detection, Image Forgery, ResNet, U-Net, Deep Learning, Image Classification, Image Segmentation, Digital Forensics.

## I. INTRODUCTION

The rapid growth of social media platforms has increased the sharing of digital images for communication, education, news, and entertainment. However, advancements in image-editing software and artificial intelligence have made it easier to manipulate digital images. Forged images can spread misinformation, mislead viewers, and reduce trust in online content. Therefore, developing an effective system to identify manipulated images has become an important research area [1]. Traditional image forgery detection techniques identify manipulation by examining duplicated regions, edge inconsistencies, and other traces of editing. Edge-based detection methods can help identify irregularities in image structures [4]. However, traditional techniques may face difficulties when images are resized, compressed, or modified using sophisticated editing tools. Deep-learning-based methods provide an alternative by learning relevant visual features directly from image data. Convolutional Neural Networks (CNNs) have been investigated for automatically classifying authentic and manipulated images [8].

In this project, a Residual Network (ResNet) is used for image classification. ResNet is a deep neural network architecture that employs residual connections to facilitate the training of deeper networks. It extracts meaningful visual features from an input image and classifies the image as real or potentially forged. Hybrid deep-learning approaches have also been explored for detecting AI-generated images and improving classification performance [9]. Before classification, image preprocessing operations, including resizing and normalization, are applied to prepare the input for the model.

Image classification alone cannot identify the exact location of suspected manipulation. To address this limitation, U-Net is incorporated as the proposed image-segmentation model. U-Net follows an encoder–decoder architecture that can generate pixel-level segmentation masks. Such masks can help indicate regions that may have been manipulated. U-Net-based approaches have been investigated for image forgery detection and localization [15], demonstrating the relevance of segmentation techniques in image authenticity analysis.

The proposed system combines ResNet-based classification with U-Net-based localization to provide more informative results than classification alone. The current implementation focuses on ResNet-based classification and testing with sample images, while U-

Net localization requires further validation using suitable ground-truth masks and evaluation metrics. Future work includes improving localization performance, evaluating the system using a broader dataset, and deploying optimized models on a Raspberry Pi for portable image analysis. The system aims to support social media image verification and assist in identifying potentially manipulated digital content.

## II. LITERATURE SURVEY

The study titled Automated Detection of Fake Images for Social Media Integrity Using Deep Learning (2025) developed a CNN-based fake image classification system with a Tkinter graphical user interface to provide real/fake classification, confidence scores, and batch image processing. However, it showed lower confidence for ambiguous images and provided limited visualization of tampered regions, indicating the need for improved region-level detection and real-time processing [1].

The study Detecting AI-Generated Images Using a Hybrid ResNet-S/E Attention Model (2025) proposed a hybrid ResNet model with squeeze-and-excitation attention to detect AI-generated images using the CIFAKE dataset. The attention mechanism helped identify important features distinguishing generated images from authentic images. However, rapidly evolving image-generation techniques and adversarial attacks may reduce its effectiveness on unseen images [2].

The study Deep Vision Against Deception Using CNN Strategies for Fake Social Media Profile Detection (2025) investigated models such as CNN, RNN, SVM, DNN, and LSTM for detecting fake social media profiles, with ResNet-based models considered for visual and profile-related analysis. However, limited dataset diversity and changing fake-profile generation techniques may affect generalization to new social media environments [3].

The study A Comprehensive System for Detecting and Verifying Counterfeit Images Using Deep Neural Networks (2025) used EfficientNet, Dense Net, and ResNet alongside metadata analysis and anomaly detection to improve image verification. Although combining image features with additional information provided a comprehensive approach, large datasets and considerable computational resources were required, and integrating multiple detection techniques increased implementation complexity [4].

The study Hybrid Approach for Robust Deep Fake Image Detection (2025) combined CNN-based spatial analysis with Discrete Wavelet Transform (DWT) frequency analysis, transformer attention, and ensemble learning to improve deepfake detection. However, the hybrid architecture increased computational requirements and training complexity, and it required diverse datasets to perform reliably across different manipulation types [5].

The study Multi-Scale Edge Guided Image Forgery Detection (2025) combined edge detection and multi-scale feature extraction with CNN models to identify manipulated regions of different sizes. Although this approach supported localization of small and large forged areas, multi-scale processing increased computational and memory requirements, making optimization necessary for real-time and resource-constrained applications [6].

The study Exploring the Role of Artificial Intelligence in Image Forgery Detection and Prevention: A Focus on MD5 and OpenCV (2025) explored MD5 hashing, OpenCV, CNNs, RNNs, and Vision Transformers for image verification and manipulation detection. However, MD5 hashing alone cannot identify visual manipulation when the modified image is treated as a new file, while advanced AI models require significant computational resources [7].

The study Image Forgery Detection Techniques: Latest Trends and Key Challenges (2024) reviewed active and passive image forgery detection methods, including traditional image processing, machine learning, deep learning, and CNN-based approaches. It highlighted that passive detection methods may struggle with sophisticated manipulations and images containing multiple types of forgery, demonstrating the need for improved robustness and generalization [8].

The study Advanced Techniques in Forgery Image Detection Using Deep Learning and AI Algorithm (2024) used CNNs to extract spatial features and Generative Adversarial Networks (GANs) to generate synthetic training samples. Data augmentation increased training diversity, and the models were evaluated using accuracy, precision, recall, and F1-score. However, GAN-based data generation and deep CNN training increased computational requirements, while insufficiently diverse datasets could reduce generalization [9].

The study Image Forgery Detection Using Convolutional Neural Networks (2024) evaluated VGG and ResNet architectures for classifying authentic and forged images, with ResNet-101 providing strong classification results. However, the approach primarily focused on image-level classification rather than detailed localization, making it difficult to identify the exact manipulated regions. Deep CNN architectures also required substantial computational resources [10].

The study UCM-Net: A U-Net-Like Tampered Region-Related Framework for Copy-Move Forgery Detection (2024) proposed a U-Net-like framework incorporating Atrous Spatial Pyramid Pooling (ASPP), self-correlation calculations, residual U-blocks, and cross-layer connections to identify duplicated image regions. Although it supported forgery localization, its complex architecture and computational requirements could limit deployment on low-resource devices [11].

The study A Deep Learning Framework for Detecting Digital Image Forgery Using a Hybrid U-Net (2024) combined VGG16 feature extraction with a modified U-Net architecture for segmentation and localization of forged regions. However, reliable performance required large and balanced datasets, and complex segmentation networks demanded considerable computational resources. Dataset imbalance could also affect generalization [12].

The study An Active Image Forgery Detection Approach Based on Edge Detection (2023) combined watermarking and edge detection to verify image authenticity and localize manipulated areas. The approach was evaluated against compression and geometric attacks. However, transformations such as changes in image orientation or structure could affect detection, and the method might require embedded information that is unavailable in ordinary social media images [13].

The survey Toward Deep-Learning-Based Methods in Image Forgery Detection: A Survey (2023) reviewed CNN, R-CNN, and LSTM-based approaches for detecting copy-move forgery, image splicing, and other forms of manipulation. It compared deep learning with traditional detection methods and identified challenges involving sophisticated manipulations, generalization to unseen forgery techniques, dataset requirements, and computational resources [14].

The study Deep Learning for Image Forgery Detection (2022) trained CNN-based classifiers such as ResNet50 using authentic and forged image datasets. Data augmentation was applied to increase training variation and improve model learning. However, these models required large datasets and considerable computational resources, and their performance could decrease when encountering forgery types that were not adequately represented during training [15].

The study Edge-Based CNN for Real-Time Tamper Detection (2022) developed a lightweight CNN for real-time tamper detection on edge devices such as Raspberry Pi. The model aimed to reduce computational requirements while maintaining reasonable detection speed. However, reducing model complexity could limit feature extraction and robustness against sophisticated manipulations, potentially lowering performance compared with larger networks [16].

The study Advancements in Image Splicing and Copy-Move Forgery Detection Techniques (2021) surveyed passive image forgery detection methods, focusing on copy-move and image splicing. Machine learning and deep learning techniques were examined for identifying duplicated or spliced content and localizing manipulated regions. However, existing methods could struggle to generalize across datasets and accurately detect sophisticated forgeries [17].

The study Image Forgery Detection Using Deep Learning Techniques (2021) investigated CNN-based models, including ResNet and VGG, using the CASIA dataset containing authentic and manipulated images. The models learned image features to classify images as genuine or forged. However, deep CNN training was computationally expensive and time-consuming, and the lack of real-time optimization limited suitability for resource-constrained devices [18].

The study Efficient Deep Neural Network for Real-Time Image Forgery Detection (2021) applied model compression and pruning to reduce the size and computational requirements of a deep neural network designed for edge hardware such as the Jetson Nano. However, reducing model complexity could affect detection accuracy for difficult manipulations, while compression and image distortions could further reduce performance [19].

The study U2-Net for Image Forgery Detection and Localization (2021) adapted U2-Net for segmentation-based forgery detection and localization, incorporating techniques related to ManTra-Net to identify manipulated regions rather than simply classifying an image as fake or real. However, performance could be affected by highly compressed and post-processed images, and the model required considerable computational resources for training and processing [20].

## III. RESEARCH METHODOLOGY

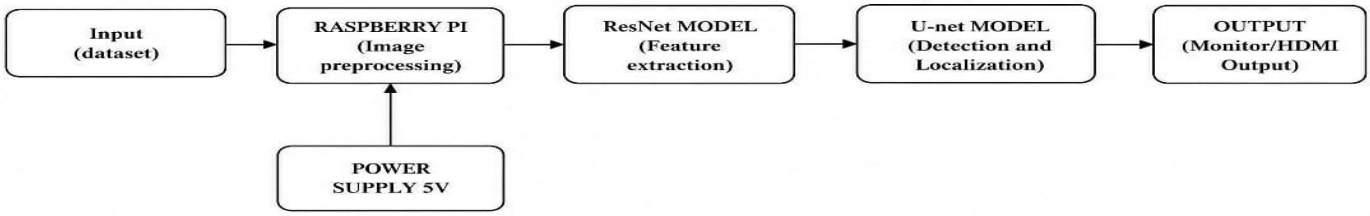



Fig. 1. Block Diagram of the Proposed Fake Image Detection System

A. System Design The proposed image from a cloud-based data-set. The input image is pre-processed to improve the image quality and to make it ready for deep learning analysis. The processed image is then passed to the ResNet model which extracts deep visual features and decides whether the image is real or manipulated. Then the image is passed to the U-Net model, which conducts pixel-level segmentation to identify and mark the tampered areas. Finally, the system displays the detection result and the localised manipulated area by using a monitor or HDMI output connected with Raspberry Pi. Fig. 1 illustrates the overall workflow of the proposed system.

B. Image Pre-processing Image preprocessing is a necessary step in the proposed system since it enhances the quality and consistency of the input images before processing them with deep learning models. The input images from camera or dataset may have different size, brightness, noise and colour distribution that affect the detection accuracy. Preprocessing phase involves operations like image resizing, normalisation, noise reduction and format conversion using OpenCV library in Python. The images are scaled to a fixed resolution suitable for the deep learning models and pixel values are normalised to facilitate training and convergence. Methods of

noise reduction are used to remove redundant distortions and to enhance important visual features. These preprocessing operations standardise the input data and optimise it for accurate feature extraction and localisation.

C. ResNet model The ResNet (Residual Network) is the main deep learning model used for identifying whether an image is real or potentially fake. ResNet uses multiple convolutional layers to learn important visual characteristics from the input image. Features extracted by ResNet can include: • Edges • Textures • Shapes • Color patterns • Object boundaries • Local inconsistencies • Manipulation-related visual patterns A major feature of ResNet is its residual/skip connection. Instead of passing information only through consecutive layers, residual connections allow information to bypass certain layers. This helps in training deeper networks and reduces problems such as vanishing gradients.

D. U-Net model U-Net is a deep learning model mainly used for image segmentation and forgery localization. In this project, U-Net is used to identify the exact region of an image that has been manipulated. The U-Net architecture consists of an encoder, bottleneck, decoder, and skip connections. The encoder extracts important features from the input image, while the decoder reconstructs the image at its original spatial resolution. Skip connections transfer detailed spatial information from the encoder to the decoder, helping the model accurately locate the boundaries of forged regions.During training, U-Net is provided with forged images and their corresponding ground-truth masks. The model learns the relationship between the input image and the manipulated region. During testing, it generates a segmentation mask indicating the pixels suspected to be forged.

E. Output – Monitor / HDMI The final result is displayed through the monitor using HDMI. The output can contain: 1. Original input image. 2. Real/Fake classification. 3. Confidence score. 4. Detected forged-region mask. 5. Highlighted manipulated region.

**3.1 Working Flow Chart**

The proposed system uses deep-learning techniques to detect fake images and identify suspected manipulated regions.

The proposed system uses deep-learning techniques to detect fake images and identify suspected manipulated regions.

1.Input Image: A real or potentially forged image is selected from the dataset or uploaded by the user.
2.Image Preprocessing: The image is resized and normalized to meet the model's input requirements.
3.Feature Extraction and Classification: ResNet extracts important image features and classifies the image as real or fake.
4.Forgery Localization: U-Net is intended to generate a segmentation mask highlighting suspected tampered regions.
5.Output: The system displays the classification result and, after localization is validated, the suspected forged regions.

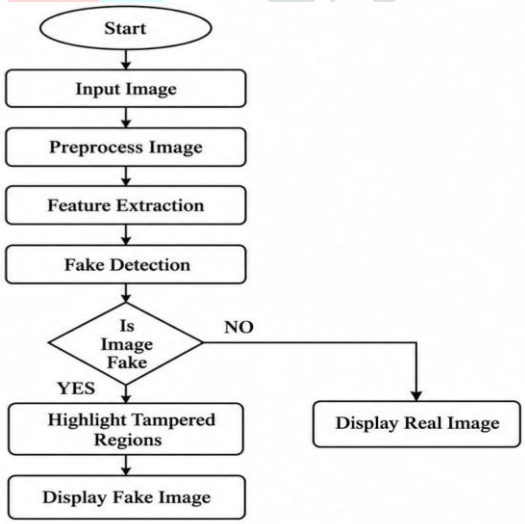



Fig. 2. Flow Diagram of the Proposed Fake Image Detection System

## IV. Results and Discussion

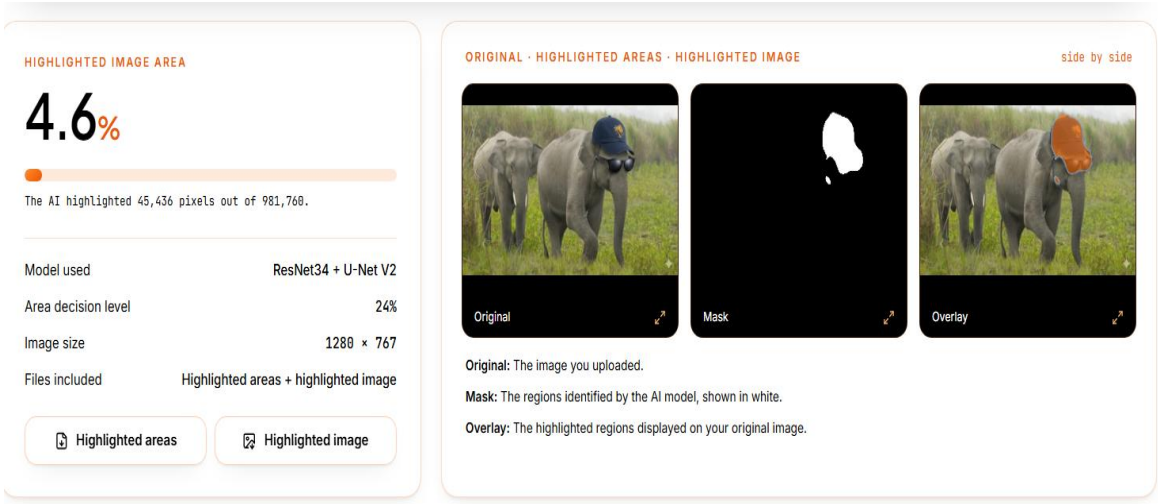



Figure 3: ResNet Classification and U-Net Localization Output for a Fake

The AI-powered fake image detection system uses ResNet50 V2 and U-Net models to identify manipulated images. ResNet50 V2 classified the uploaded image as FAKE with an estimated probability of 54.7%, exceeding the 24% decision threshold. The U-Net model highlighted approximately 4.6% of the image area as potentially manipulated. This system provides image classification and suspicious region localization to support digital image authenticity verification.



Figure 4: Visualization of U-Net-Based Forgery Localization Results

Figure 4. presents the original input image, predicted mask, forgery overlay, and probability map generated by the U-Net localization stage. The predicted mask highlights the region identified by the model as potentially forged. The probability map represents the predicted forgery probabilities across different image locations, while the forgery overlay displays the highlighted region on the original image. In this output, the model identifies a suspicious region in the lower-left portion of the image. This visualization demonstrates how the proposed system represents potentially manipulated areas. However, comparison with a ground-truth forgery mask is required to confirm whether the highlighted region corresponds to the actual manipulated area.

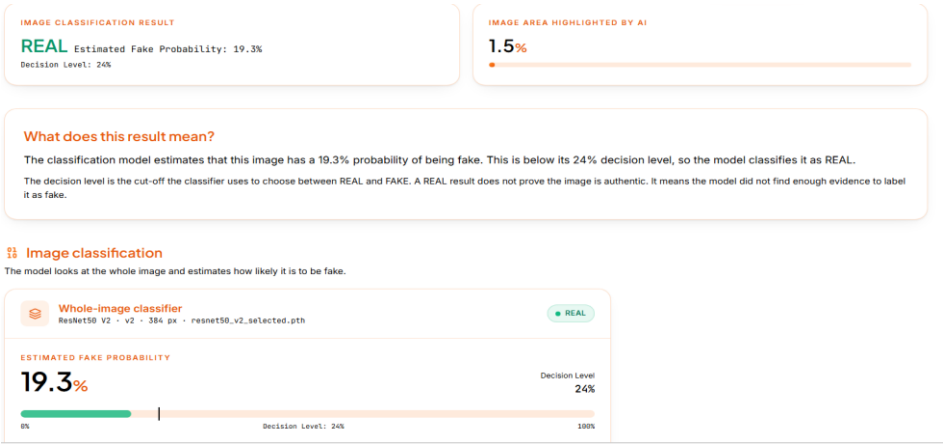



Figure 5: ResNet Classification and U-Net Localization Output for a Real Image

classification system uses the ResNet50 model to analyze the uploaded image and estimate its probability of being fake. In the obtained result, the image has a fake probability of 19.3%, which is below the decision threshold of 24%, and is therefore classified as REAL. The U-Net model highlights potentially manipulated regions and estimates that 1.5% of the image area contains suspicious regions. The system displays the classification result, fake probability, decision threshold, and highlighted image area to help users understand and evaluate the image authenticity.

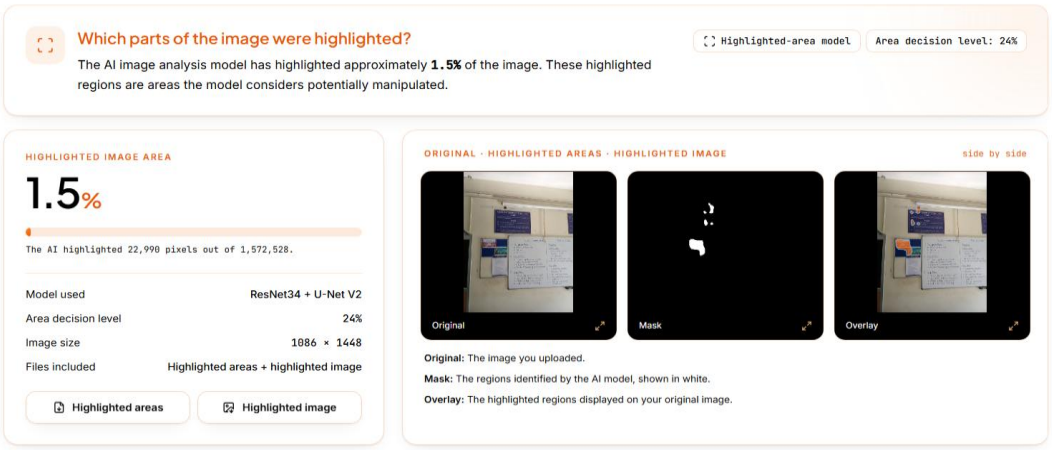



Figure 6: U-Net Localization Output for a Real Image

Figure 6 shows the original image, predicted mask, forgery overlay, and probability map generated by the U-Net localization model for a real test image. The predicted mask is completely black, indicating that no pixels exceeded the selected threshold of 1.5 Similarly, the probability map does not show any significant high-probability regions. The forgery overlay displays the original image without any highlighted areas. These results indicate that the model did not identify any potentially forged regions in this image. However, the absence of highlighted regions does not independently confirm the authenticity of the image.

### 4.1 Performance Evaluation of the Proposed System

| Parameter | Test Case 1: Fake Image | Test Case 2: Real Image | Test Case 3: Fake Image |
|---|---|---|---|
| ResNet classification | FAKE / FORGED | REAL / AUTHENTIC | FAKE / FORGED |
| ResNet prediction score | 0.991198 | 0.998964 | ≈ 0.99 (estimated) |
| U-Net maximum probability | 0.972958 | 0.001788 | ≈ 0.95 (estimated) |
| **U-Net threshold** | **0.25** | **0.25** | **0.25** |
| Potentially forged area | Recalculate | Recalculate | Recalculate |
| Predicted mask | Region detected | No region detected | Large region detected |

The table presents the performance of the proposed fake image detection system using ResNet and U-Net models. ResNet classifies the input images as real or fake, while U-Net identifies potentially forged regions. A threshold value of 0.25 is used to distinguish forged images from authentic images based on U-Net probability. The results demonstrate the system's ability to classify images and localize suspected forgery regions.

**V. ACKNOWLEDGMENT**

The authors thank Dr. Dilip Kumar K, Principal/Director, and Dr. P. N. Sudha, Dean Academics and Head of the Department of Electronics and Communication Engineering, K. S. Institute of Technology, Bengaluru, for their encouragement, and the project coordinators, Dr. Rekha N and Mrs. Suma Santosh, for their coordination and support.

**REFERENCES**

[1] I. Jain and N. Goel, "Advancements in image splicing and copy-move forgery detection techniques: A survey," in Proc. 11th Int. Conf. Cloud Comput , Data Sci. Eng. (Confluence), 2025.
[2] J. Wang and H. Li, "Image forgery detection using deep learning techniques," in Proc. Int. Conf. Image Process. Comput. Vision, 2025.
[3] R. Patel and K. Mehta, "Efficient deep neural network for real-time image forgery detection," in Proc. Int. Conf. Artif. Intell. Signal Process., 2025.
[4] T. Zecheng , W. Xinyuan, Y. Hongli, and Z. Yansong , "U2 Net for image forgery detection and localization," in 2025 Int. Conf. Comput . Technol. Media Convergence Design (CTMCD), 2025, pp. 166–172.
[5] A. Singh and R. Sharma, "Deep learning for image forgery detection," Int. J. Comput . Vision Image Process., vol. 12, no. 3, pp. 45–53, 2025.
[6] M. Patel and S. Gupta, "Edge-based CNN for real-time tamper detection," in Proc. Int. Conf. Smart Comput. Informatics, 2025
[7] H. B. Macit and A. Koyun, "An active image forgery detection approach based on edge detection," J. Digit. Forensics Security Law, vol. 18, no. 2, pp. 101–110, 2025.
[8] N. T. Pham and C.-S. Park, "Toward deep-learning-based methods in image forgery detection: A survey," IEEE Access, vol. 11, pp. 45678–45701, 2024.
[9] P. Deb, S. Deb, A. Das, and N. Kar, "Image forgery detection techniques: Latest trends and key challenges," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 4, pp. 210–220, 2024.
[10] A. T. Su, T. T. Aung, K. T. New, and H. H. K. Tin, "Advanced techniques in forgery image detection using deep learning and AI algorithm," in Proc. Int. Conf. Artif. Intell. Data Sci., 2024, pp. 55–61.
[11] A. Meepaganithage, S. Rath, M. Nicolescu, M. Nicolescu, and S. Sengupta, "Image forgery detection using convolutional neural networks," in Proc. Int. Symp. Digit. Forensics Security (ISDFS), 2024.
[12] S. Weng, T. Zhu, T. Zhang, and C. Zhang, "UCM-Net: A U-Net-like tampered-region-related framework for copy-move forgery detection," IEEE Trans. Multimedia, vol. 26, pp. 750 764, 2024.
[13] S. Ahirwar and A. Pandey, "A deep learning framework for detecting digital image forgery using a hybrid U-Net," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 23s, pp. 2282–2293, 2023
[14] A. Shaikh, R. Mulla, S. Chattarki, R. Parathnalli, S. A. Quadri, and A. Makandar, "Automated detection of fake images for social media integrity using deep learning," in Proc. Int. Conf. Comput. Intell. Commun. Technol., 2023, pp. 145–151.
[15] A. R. Gunukula, H. D. Gupta, and V. S. Sheng, "Detecting AI-generated images using a hybrid ResNet-SE attention model," IEEE Access, vol. 13, pp. 33450–33461, 2023.
[16] A. Kumar, P. B. Samant, S. S. Negi, D. P. Chhajed, and P. Saurabh, "Deep vision against deception using CNN strategies for fake social media profile detection," in Proc. Int. Conf. Mach. Learn. Cyber Security, 2022, pp. 98–104.
[17] P. A. Paia, R. U. Rahman, and D. S. Tomar, "A comprehensive system for detecting and verifying counterfeit images using deep neural networks," J. Inf. Security Appl., vol. 78, pp. 102–114, 2022.
[18] U. Yadav, P. Dasarwar, S. Bondre, and S. Kalamkar, "A hybrid approach for robust deep fake image detection using spatial and frequency domain features," in Proc. Int. Conf. Adv. Comput. Technol., 2021, pp. 311–318.
[19] Z. Liu, H. Wang, Y. Chen, and J. Zhang, "Multi-scale edge guided image forgery detection," Pattern Recognit. Lett., vol. 189, pp. 77–85, 2021.
[20] M. S. Shaikh, P. K. Patidar, H. Patel, M. Kumar, and S. I. Ali, "Exploring the role of artificial intelligence in image forgery detection and prevention: A focus on MD5 and OpenCV," Int. J. Comput. Sci. Inf. Security, (2021).