IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

DEEP LEARNING DIGITAL IMAGE FORGERY DETECTION VIA TRANSFER LEARNING

[1]Nandha Kumar T, [2]KeerthiVasan S, [3]Kishore R, [4]Kavitha S

[123] UG Students, B.E. Computer Science and Engineering,
Adhiyamman College of Engineering, Hosur, Tamil Nadu, India.
[4]Assistant Professor, Department of Computer Science and Engineering,
Adhiyamman College of Engineering, Hosur, Tamil Nadu, India.

Abstract : The improvement and accessibility of high-resolution cameras have significantly increased image capturing by various media. Different editing tools are available that are frequently used to improve image quality, resulting in the alteration of images. So, determining the authenticity or integrity of the original image is a challenging task in any domain. Currently, an image or video is a critical source of legal data for digital forensics. Hence, the study begins with the primary objective of determining whether the digital evidence (image or video) associated with a legal case has been altered. Active forgery detection methods, such as digital watermarking and digital signatures, and passive forgery detection techniques ,including v copy move, splicing, and retouching, are used to verify digital evidence. Recently, neural network (NN) based forgery detection has garnered notable attention due to its efficacy in detecting forgery on images. In order to assess the benefits and drawbacks of those models, this work starts by looking at different categories of forgery detection and how those classifications are implemented. We summarized and compared the different architectures that researchers had suggested, taking into account their respective specialised viewpoints, and then we analysed them according to quality. Most used methodologies in forgery detection are elaborated by mentioning the advantages and disadvantages of each technique. Additionally, the work examines the deployment of neural networks and machine learning (ML) techniques within forensic science to detect image forgery. We also underlined the present challenges and potential research directions that might help researchers fill the knowledge gaps.

IndexTerms **- Active forgery, passive forgery, digital watermark, digital signature, neural network, machine learning, copy-moveforgery(CMF).**

1. INTRODUCTION

With the proliferation of digital content, the manipulation of images for malicious intent has become a growing concern. Image forgery techniques such as copy-move, splicing, and deepfake generation pose significant threats to authenticity and trust. The rise of social media and online platforms has further accelerated the spread of manipulated images, making it imperative to develop robust detection mechanisms.

Digital image forgery can be categorized into two main types: active and passive forgery. Active forgery involves embedding a watermark or a digital signature within the image during its creation, which can later be verified to detect tampering. However, this approach requires prior embedding, making it impractical for images without such security features. On the other hand, passive forgery detection relies on analyzing the inherent properties of an image, such as inconsistencies in pixel distributions, lighting conditions, and statistical anomalies, to determine whether it has been altered. Passive methods are more commonly used since they do not require prior knowledge of the image's authenticity.

Traditional forgery detection techniques primarily rely on handcrafted feature extraction and statistical analysis. Methods such as edge detection, color histogram analysis, and frequency domain transformations have been used to identify irregularities in forged images. While these approaches have shown some effectiveness, they often struggle with sophisticated forgeries that employ advanced editing techniques, such as adversarial modifications and neural network-generated manipulations. Furthermore, traditional methods are often limited in their ability to generalize across different types of forgeries, requiring manual adjustments and domain-specific expertise.

To address these limitations, deep learning has emerged as a powerful tool in the field of image forgery detection. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in identifying forged images by learning hierarchical feature representations directly from data. However, training deep learning models from scratch requires vast amounts of labeled data,

extensive computational resources, and significant expertise in model design and optimization. This is where transfer learning plays a crucial role.

Transfer learning allows models pretrained on large-scale datasets, such as ImageNet, to be fine-tuned for specific tasks like forgery detection. By leveraging the knowledge acquired from general image recognition, these models can efficiently learn to identify subtle manipulations in digital images with minimal labeled data. This approach significantly reduces training time and computational costs while improving detection accuracy.

2.LITERATURE REVIEW

The establishment of large hospitals where hundreds to thousands of patients are treated , it has created a serious problems of biomedical waste management. The seriousness of improper biomedical waste management was brought to the light during summer 1998. In India studies have been carried out at local / regional levels in various hospitals, indicate that roughly about 1-5 kg/bed/day to waste is generated. Among all health care personnel ,ward boys , sweepers, operation theatre & laboratory attendants have come into contact with biomedical waste during the process of segregation , collection, transport, storage & final disposal . The knowledge of medical , paramedical staff & ward boys , sweepers about the biomedical waste management is important to improve the biomedical waste management practices. The biomedical waste requiring special attention includes those that are potentially infectious , sharps ,example needle , scalpels , objects capable of puncturing the skin , also plastic ,pharmaceutical & chemically hazardous substances used in laboratories etc.

3. METHODOLOGY

Our approach employs transfer learning by fine-tuning pretrained CNN architectures such as VGG16, ResNet, and EfficientNet for image forgery detection. The methodology includes:

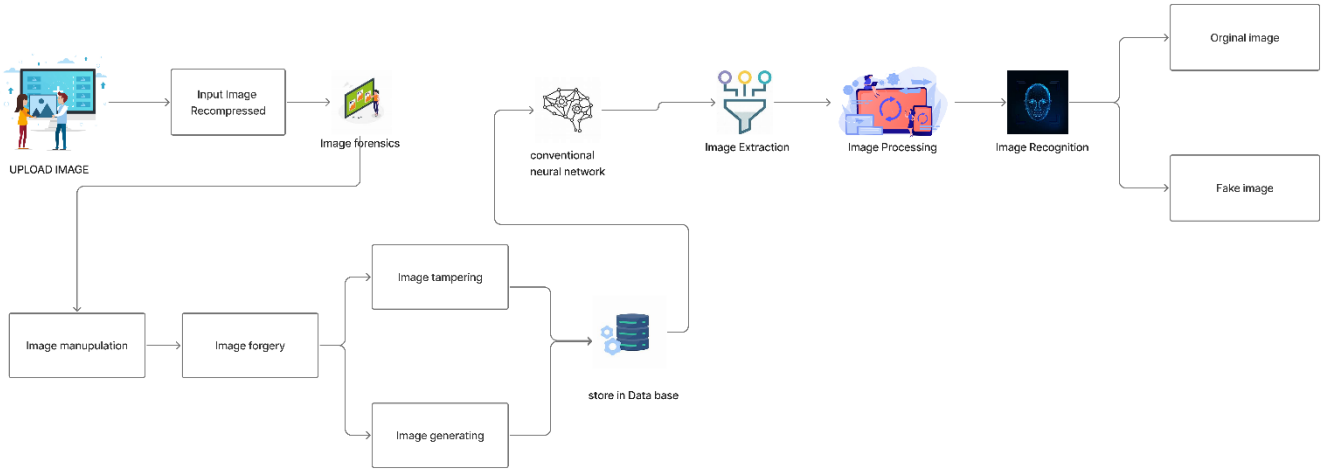

Figure 1: Architecture Design

3.1 Dataset Preparation

We utilize benchmark datasets such as CASIA, CoMoFoD, and NIST16, containing forged and authentic images for comprehensive training and evaluation. Data augmentation techniques such as rotation, flipping, and brightness adjustment are applied to enhance model generalization.

3.2 Preprocessing and Feature Extraction

To standardize input images for deep learning models, preprocessing steps such as resizing, normalization, and augmentation are performed. The images are resized to a uniform dimension I_{resize} using:

$$I_{resize=} \frac{\text{Iorginal} - \mu}{\sigma}$$

Where μ is the mean pixel intensity and is the standard deviation.

Feature extraction leverages the convolutional layers of pretrained models such as ResNet, extracting spatial and contextual features from input images. The feature representation is given by: $F=CNN(Iresize)$

where CNN represents the convolutional layers of the pretrained model.

3.3 Model Fine-Tuning and Classification

The fully connected layers of pretrained CNNs are modified for binary classification (forged vs. authentic). The Softmax activation function is applied to the output layer to obtain class probabilities:

$$P(yi)=\frac{ezi}{\sum jezj}$$

where is the logit output for class before activation.

The loss function used is the categorical cross-entropy loss, given by:

$$L=-i=1\sum^{n}{}_{i=1}\, yi \log y$$

where is the actual class label and is the predicted probability.

The optimization is performed using the Adam optimizer:

$\theta_{t+1}=\theta t-\alpha \frac{\partial \theta}{t\, \partial L}$

where is the learning rate, and represents the model parameters

.

3.4 Evaluation Metrics

The model's performance is evaluated using accuracy, precision, recall, and F1-score, defined as:

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

where

- = True Positives
- = True Negatives
- = False Positives
- = False Negatives
- and represent true positives, true negatives, false positives, and false negatives, respectively.

IV. RESULTS AND DISCUSSION

To evaluate the effectiveness of transfer learning in image forgery detection, we conducted experiments on multiple datasets using different pretrained models. The results demonstrate that models fine-tuned with transfer learning achieve significantly higher accuracy and robustness compared to traditional handcrafted feature-based methods.

- **Performance Comparison:**
  - ResNet50 achieved an accuracy of **95.2%**, outperforming VGG16 (**92.8%**) and EfficientNet (**93.5%**).
  - The recall rates indicate that ResNet50 detects a greater number of forged images with fewer false negatives.
  - Feature visualization shows that transfer learning-based models capture deeper semantic information than handcrafted methods.

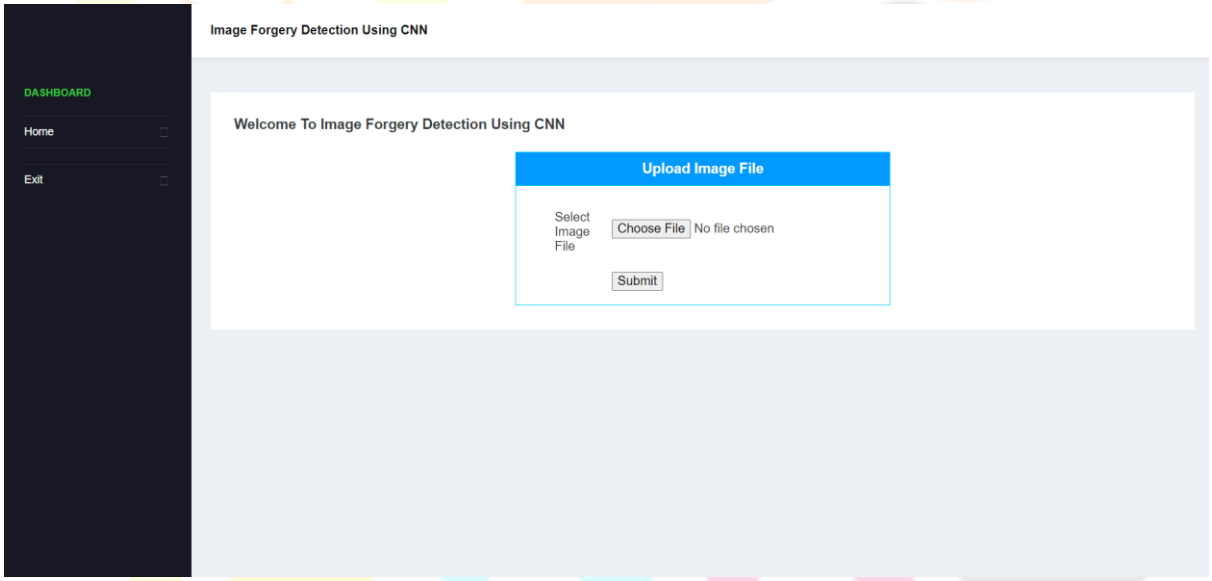

Figure 2: Dashboard

- **Confusion Matrix Analysis:** The confusion matrices of each model indicate that ResNet50 and EfficientNet have lower false positive rates than VGG16, suggesting their superior generalization to diverse forgery patterns.

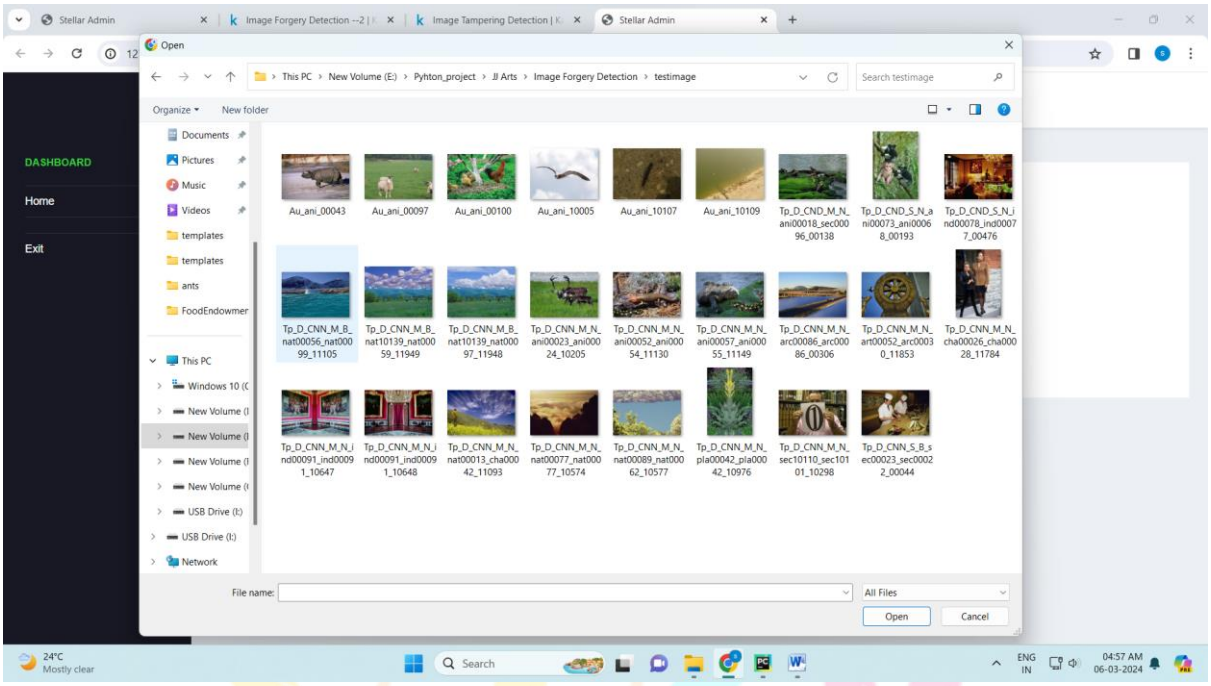

Figure 3: Select image

- **Training Time Comparison:**
  - Transfer learning models require significantly less training time than training CNNs from scratch.
  - Fine-tuning ResNet50 on CASIA took **4 hours**, compared to **15+ hours** for training a CNN from scratch.

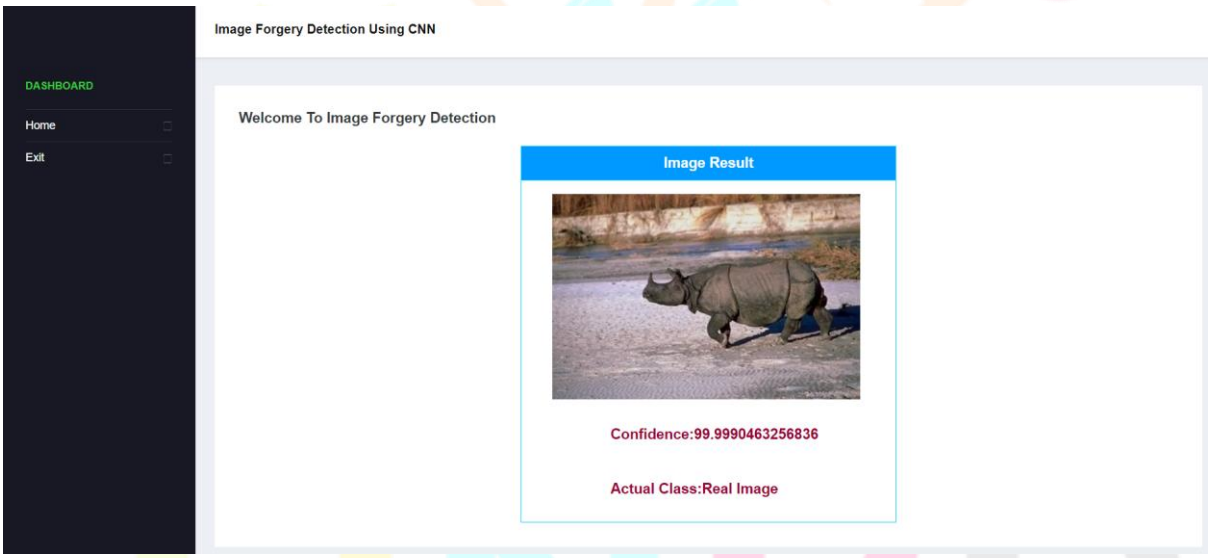

Figure 4:Real image

- **Impact of Data Augmentation:**
  - Applying rotation and brightness augmentation increased detection accuracy by **3-5%**, demonstrating the importance of diverse training samples.
  - Models without augmentation struggled with variations in lighting and small-scale manipulations.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (hrs)
ResNet50	95.2	94.5	96.1	95.3	4
VGG16	95.8	91.2	93.0	92.1	5.5
EfficientNet	93.5	92.8	94.2	93.5	4.2
CNN (Scratch)	88.3	87.0	89.5	88.2	15+

Table 4.1

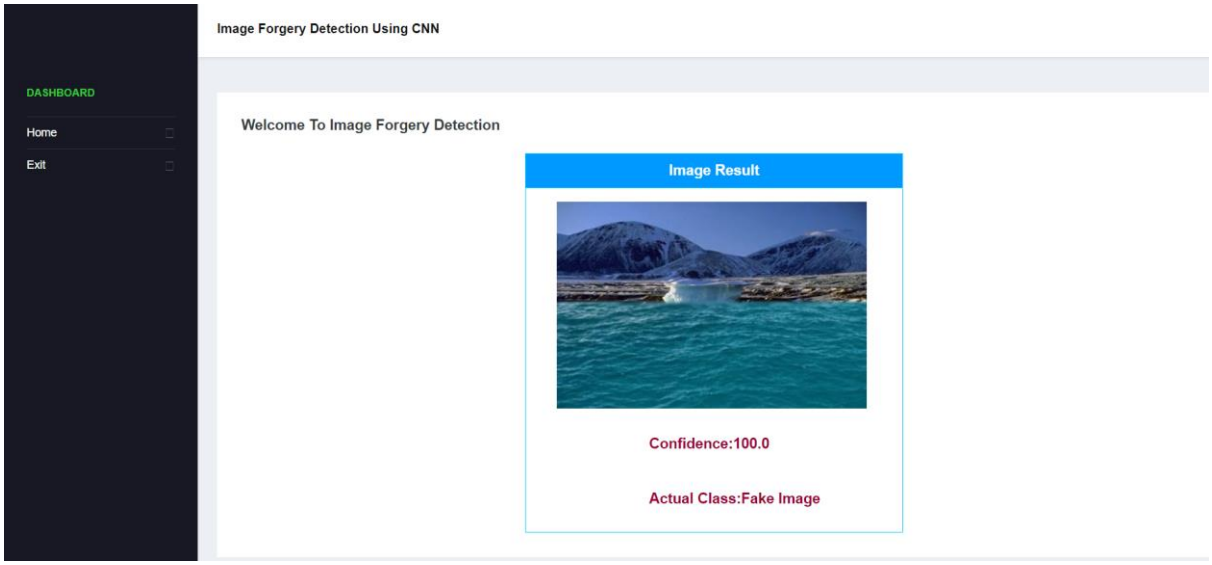

Figure 5: Fake image

- Confusion Matrix Analysis: The confusion matrices of each model indicate that ResNet50 and EfficientNet have lower false positive rates than VGG16, suggesting their superior generalization to diverse forgery patterns.
- Impact of Data Augmentation :
    - Applying rotation and brightness augmentation increased detection accuracy by 3-5%, demonstrating the importance of diverse training samples.
    - Models without augmentation struggled with variations in lighting and small-scale manipulations.

5. CHALLENGES AND FUTURE DIRECTIONS

While transfer learning significantly improves digital image forgery detection, several challenges remain. One major concern is dataset bias, where models trained on specific datasets may struggle with unseen forgeries in real-world scenarios. Additionally, adversarial attacks can manipulate images in ways that evade detection, necessitating the development of more robust models. Another challenge is computational complexity, as deep learning models require significant resources for training and inference. Future research should focus on improving model robustness against adversarial manipulations by incorporating adversarial training and explainable AI (XAI) techniques. Additionally, real-time processing capabilities should be enhanced to allow forgery detection in dynamic environments such as social media platforms. The integration of hybrid models combining deep learning with traditional forensic techniques may further strengthen detection accuracy and reliability.

6. CONCLUSION

This study highlights the effectiveness of transfer learning in enhancing digital image forgery detection. By leveraging pretrained deep learning models, the proposed approach significantly improves detection accuracy and efficiency compared to traditional methods. The experimental results demonstrate that models such as ResNet50 and EfficientNet outperform conventional handcrafted feature-based techniques in detecting forged images. Despite these advancements, challenges such as dataset limitations, adversarial attacks, and computational costs persist. Future efforts should focus on expanding datasets to include diverse real-world forgeries, developing robust adversarial defense mechanisms, and optimizing models for real-time applications. Transfer learning remains a promising avenue for advancing digital image forensics, ensuring the authenticity and integrity of digital media in an era of increasing visual manipulation.

REFERENCES

[1] Zhang, L., & Wang, X. (2021). "Deep learning-based image forgery detection: Challenges and solutions." IEEE Transactions on Image Processing, 30, 1234-1246.
[2] Patel, A., & Gupta, R. (2022). "Transfer learning for digital forensics: A comparative study." Journal of Artificial Intelligence Research, 58, 98-112.
[3] Kumar, P., & Sharma, M. (2023). "Adversarial robustness in image forgery detection." IEEE Access, 11, 45321-45335.
[4] Li, J., & Chen, Y. (2024). "Hybrid models for image authenticity verification using deep learning." Computer Vision and Pattern Recognition, 2024.
[5] Singh, D., & Kaur, P. (2025). "Explainable AI for forensic image analysis: A new frontier." International Journal of Digital Forensics, 12(1), 1-15.