AI-GENERATED IMAGE DETECTION USING SIX-CHANNEL MULTI FORENSIC CONVOLUTION NUERAL NETWORKS (CNN) FUSION

Nazirah Abd Hamid, Shaifol Ziqri Zainurin, Siti Dhalila Mohd Satar, Norhidayah Muhammad, Mohd Fadzil Abdul Kadir, Muhammad Syakir Razmin, Muhammad Daniel Mohd. Shaiful Azhari, Nur Wasilah Jamlih

Faculty of Informatics Computing, Universiti Sultan Zainal Abidin (UniSZA), Tembila, Terengganu

ABSTRACT

Modern generative artificial intelligence (AI) can generate highly realistic synthetic images, making current detection system that rely on standard RGB input tends to miss the underlying generative artifacts. This research develops a framework based on Convolutional Neural Network (CNN) that aims to distinguish between original digital images and artificial media. This work proposes a multi-modal, six channel forensic fusion architectures that combines standard visual data with advanced digital forensic features to reveal hidden generative artifacts. The proposed method utilizes three specific forensic layers, Error Level Analysis (ELA) to find inconsistencies in image compression, Fast Fourier Transform (FFT) to detect abnormal frequency patterns, and Laplacian High-Pass Filtering (HPF) to extract high-frequency background noise as Photo Response Non-Uniformity (PRNU) alternative. The first layer of three baseline models which are MobileNetV3-small, MobileNetV2, and ResNet50 were modified to directly process the combined six-channel input. The framework was trained, validated and tested based on balanced subset of 16,000 images from the CIFake dataset. The performance of all models was measured using accuracy, precision, False Acceptance Rate (FAR), False Rejection Rate (FRR), F1-score, and overall computational cost. Implementing these forensic features significantly improved the accuracy of lightweight models. As a result, MobileNetV2 accuracy jumped from 77.62% using only standard RGB to 94.94% with the new six channel input, while MobileNetV3-small achieved the highest overall test accuracy of 95.69%. This multi feature approach can be used in real-world applications like digital forensics, automated content moderation, and cybersecurity, where verifying image authenticity is critical. Overall, the findings indicate that combining traditional method with forensic features with CNN-based model can improve detection accuracy, even with lightweight models.

KEYWORDS

Synthetic images, Generative artifacts, Convolutional Neural Networks, six-channel forensic fusion, lightweight models.

1. INTRODUCTION

The evolution of digital imaging and rapid progression in the field of image processing techniques have significantly transformed the visual media landscape. These developments have facilitated creative innovation while also accelerate the circulation of manipulated content [1]. As image forgeries become progressively sophisticated, varying from basic splicing and copy-move techniques to extremely realistic deepfakes generated by diffusion models, the task to verify the image authenticity has emerged as a serious task [2]. Although conventional forensic methods can be efficient against simple manipulation, it often struggles to recognise more refined or

deliberately manufactured synthetic images.

This growing threat has pushed more research into artificial intelligence (AI)-based detection frameworks. Convolution Neural Networks (CNNs) can be considered as the foundation of modern image forgeries detection as they excel at extracting images pattern from raw pixel data [5], [6], [7]. To push these boundaries further, researchers are increasingly exploring hybrid frameworks that combining standard CNN visual extraction with domain-specific forensic features, such as Error Level Analysis (ELA), Fast Fourier Transform (FFT), and High-Frequency Noise Residuals [3], [8], [9]. Integrating these methods offers the potential to significantly improve detection accuracy and robustness against advanced manipulation of generated images [4]. The motivation extends to not only achieving high accuracy in testing, but also to implement the method into lightweight architecture that may be integrated into future forensic toolkits for real-world applications.

A primary challenge in current AI image detection is that modern generative models can produce visual output that easily deceive both the human eye and standard RGB-based CNNs detection method. The hidden traces or artifacts of the synthetic images are often very subtle, that only may appear as sub-pixel frequency anomalies, compression artifacts, and high-frequency noise residuals. Moreover, lightweight CNN architectures such as MobileNetV3 and MobileNetV2 are very beneficial for real-time and resource constrained application. Lower parameter capacity makes it hard for lightweight models to learn the subtle artifacts from standard RGB images alone [10].

The main objective is to develop a CNN-based framework that overcomes the limitation of standard RGB-based detection model by directly integrates visual and digital forensic features data at the input layer. Specifically, this research evaluates the efficacy of this combined approach across three CNN architectures (MobileNetV3-S, MobileNetV2, and ResNet50) to classify images as real or AI-generated by using CIFake dataset. The focus contributions of this paper are to propose approach that combine spatial RGB data with ELA, FFT, and Laplacian High-Pass Filtering (HPF) into single input tensor as deep feature extraction. Then, comparison between standard and forensic-fusion models will be done to determine which architectures benefits most from the combined method implementation.

The structure of this paper is as follows. Section 2 discusses related studies on CNN-based image forgery and AI-generated image detection. Section 3 explains the proposed method while Section 4 presents the experimental results and compares the performance of the CNN models. Finally, Section 5 concludes the paper by summarizing the main findings, limitations of the study and suggests future work.

2. RELATED WORK

This section discusses related studies on CNN-based image forgery and AI-generated image detection, focusing on how forensic features such as ELA, FFT, and High-Frequency Noise Residual have been utilised to enhance model performance toward hidden manipulation artefacts.

2.1. CNN-Based Image Forgery And AI-Generated Image Detection

Forensic elements such as ELA, FFT and High-Frequency Noise Residual play a key role in the reinforcement of CNN detection systems. ELA works by highlighting compression artifacts and error levels in JPEG images, it effectively amplifying traces of manipulation that might be invisible to the naked eye or standard CNN filters [8], [11], [12], [13]. When used as a pre

processing step, ELA increases the discriminative power of the CNN, especially for the detection of subtle manipulations and for images with low-information or grayscale. On the other hand, FFT-based frequency domain analysis isolate periodical and structural artifacts, which makes it easier to detect manipulations that are manifested as subtle frequency anomalies. The integration of ELA with FFT and CNNs has been proven to increase both the accuracy of detection and the computing power, particularly in complex multi-stage counterfeiting scenarios [9], [14], [15].

High-Frequency Noise Residual or specifically Laplace High Pass Filtering (HPF), is a technique used in image forensics to enhance edge details and suppress low-frequency components, making it effective for detecting subtle manipulations. For instance, Laplacian filters have been applied in medical imaging and forensic contexts to improve resolution and highlight fine-grained inconsistencies in manipulated images [16], [17]. These filters are particularly useful in preprocessing steps to focus on tampering artifacts. [17], [18].

The combination of forensic features like ELA, FFT, and HPF in CNN architectures has caused in extremely accurate and dependable detection frameworks. Multi-model and hybrid models combining spatial and temporal features through advance fusion strategies steadily outdo single method approaches, with accuracies exceeding 90% and often surpassing 95% on standard datasets such as CASIA and FaceForensics++ [10], [19], [20]. These consolidative models not only improve the localization of tampered regions but also decrease false positives by leveraging the complementarity of forensic clues.

However, ethical and legal concerns such as privacy, transparency and the admissible of evidence generated by an artificial intelligence system, obfuscate the deployment of forensic AI systems. It requires cross-disciplinary cooperation and the development of standardized evaluation protocols [21], [22], [23]. Future research routes include the enhancement of multi-modal fusion strategies, developing real-time and cross model detection frameworks, and assembling transparent, accountable, and legally sound forensic AI solutions [24], [25], [26], [27]. For the conclusion, the convergence of CNN-based model with forensic features such as ELA, FFT, and HPF is an encouraging frontier in AI-based image detection, offering both theoretical intensity and applied utility in the constant battle against digital manipulation.

2.2. Research Gap

Although CNN-based models have become a common approach for detecting image forgery and AI-generated images, many existing methods still depend mainly on standard RGB images. This is helpful for learning visible patterns, but it may not be sufficient when the differences between real and synthetic images are very subtle. In many cases, AI-generated images may look realistic to human observers, while their hidden traces appear only in compression behaviour, frequency patterns, or high-frequency noise residuals. Previous studies have shown that forensic features such as Error Level Analysis (ELA), Fast Fourier Transform (FFT), and high-frequency residual analysis can support image forensic detection, but these features are often used separately or applied mainly to traditional manipulation cases such as splicing, copy-move forgery, and JPEG tampering. There is still limited work that brings these forensic features together into one CNN based framework specifically for AI-generated image detection. In addition, the effect of forensic feature fusion on lightweight CNN architectures has not been clearly explored, especially when compared with deeper models such as ResNet50. This is important because lightweight models such as MobileNetV3-S and MobileNetV2 are more practical for real-time and resource-limited applications, but they may struggle to learn subtle forensic artefacts from RGB images alone. Therefore, this study addresses this gap by proposing a six-channel forensic-fusion CNN framework that combines RGB, ELA, FFT, and Laplacian High-Pass Filtering features, and evaluates its effectiveness using MobileNetV3-S, MobileNetV2, and ResNet50 on the CIFake

dataset.

3. METHODOLOGY

This section presents the methods applied in this study. The overall framework of the proposed AI-generated image detection approach is shown in Figure 1.

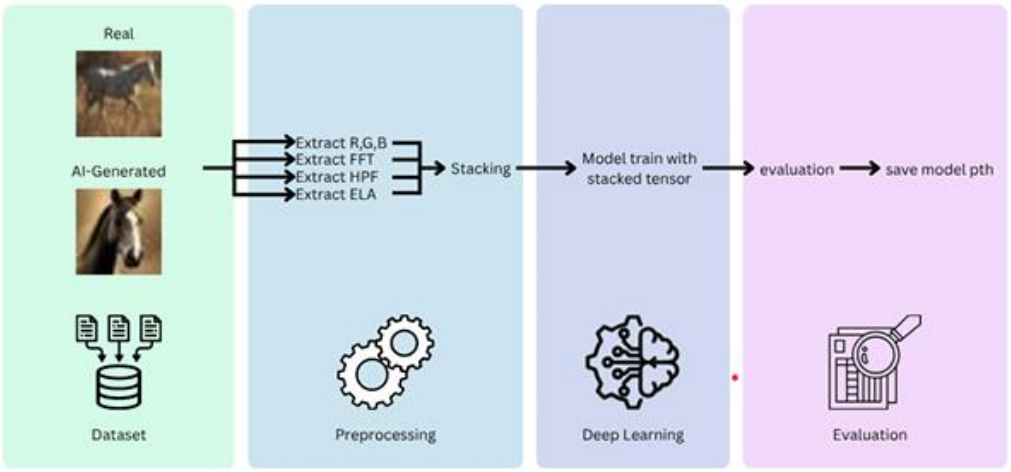

Figure 1. Overall framework of the proposed AI-generated image detection approach.

3.1. Preparation

This study utilizes the CIFake dataset, which consists of 60,000 real and 60,000 synthetic images produced by Stable Diffusion v1.4. To ensure balanced and computationally efficient training, a representative subset of 16,000 images (8,000 per class) was randomly selected. All images were resized to 224 × 224 pixels. The subset was strictly divided into training (12,800 images), validation (1,600 images), and testing (1,600 images) using a fixed random seed of 42 for reproducibility [28].

The models were trained using PyTorch with the Adam optimizer and a learning rate of 0.0001. A batch size of 32 was used, and the loss function was Cross-Entropy. The training was set to a maximum of 100 epochs, with early stopping implemented (patience = 10 epochs) on the validation set to prevent overfitting [28]. Training was accelerated using NVIDIA GPU hardware. Data augmentation including image rotation, shear transformation, zooming, horizontal/vertical shifts, horizontal inversion, and brightness adjustment (0.8 to 1.2) was applied to the training set to improve generalization [11], [28]. A classification threshold of 0.5 was maintained for evaluation.

3.2. Error Level Analysis(ELA)

ELA is used to detect inconsistent image compression levels. The image is compressed to 90% quality and subtracted from the original. A scale multiplier is used to amplify the results [29].

$$ELA(x,y) = Scale \times |I_{RGB}(x,y) - I_{JPEG90}(x,y)|$$

Where the scale is defined as:

$$Scale = \frac{255}{\max(|I_{RGB} - I_{JPEG90}|)}$$

3.3. Fast Fourier Transform (FFT)

The FFT algorithm is utilized to analyse images in the frequency domain. The pixels are converted into frequency by using the Discrete Fourier Transform [30], [31].

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y) e^{-j2\pi(\frac{ux}{M}+\frac{vy}{N})}$$

Then, a logarithmic transformation is applied to render high-frequency artifacts visible against dominant low-frequency components:

$$M(u,v) = \log\left(1 + |F_{shifted}(u,v)|\right)$$

Finally, normalization is applied to scale the spectral values for numerical stability.

$$FFT_{final} = \frac{\mathrm{M} - \min(M)}{\max(M) - \min(M)}$$

3.4. Laplacian High-Pass Filtering(HPF)

To extract noise patterns, Laplace High-Pass Filtering algorithms were used as a proxy of Photo Response Non-Uniformity (PRNU), where a 3×3 Laplacian kernel is convolved with the grayscale image to isolate high-frequency noise residuals [32], [33].

$$K = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

The filtering is applied via discrete convolution, and the final noise map is produced by absolute value calculation and normalization:

$$R(x,y) = |I(x,y) \times K|$$

$$Residual_{norm} = R - \frac{\min(R)}{\max(R) - \min(R)}$$

3.5. Model Customization

Standard Convolutional Neural Networks (CNNs), such as ResNet50 and EfficientNetV2, are architecturally engineered to accept standard three-channel spatial inputs representing the Red, Green, and Blue (RGB) colour channels [34], [35]. However, relying solely on raw RGB pixels often limits the model's capacity to generalize, as the network tends to overfit on semantic image

content rather than the underlying manipulation artifacts [36], [37]. To overcome this limitation, this study employs an early fusion strategy often referred to as a multi-tensor approach by directly combining the visual RGB data with the three extracted forensic feature layers (ELA, FFT, and High-Frequency Noise Residual) prior to deep feature extraction [38], [39].

To accommodate this unified six-channel input tensor, the initial convolutional layer of each baseline architecture requires structural modification, a technique demonstrated effective for conserving high-frequency forensic traces that are else lost throughout standard down-sampling [40], [41]. In the typical implementation, the first nn.Conv2d layer is strictly initialized for an input depth of . To modify this dynamically without altering the downstream hierarchical architecture, the structural hyperparameters of the original first layer, specifically the number of output filters ( ), kernel size ( ), stride ( ), and padding ( ), are programmatically extracted. The original layer is then replaced with a newly initialized convolutional layer where , while all other spatial dimensions are preserved. Mathematically, the standard first-layer discrete convolution operation for an RGB image is defined as [36], [42]:

$$Y_{Cout} = \sum_{Cin=1}^{3} W_{Cout,Cin} * I_{Cin} + b_{Cout}$$

W signifies the learned filter weights and b is the bias. To support the multi-modal tensor $T \in \mathbb{R}^{6\times H\times W}$, the modified convolution magnifies the summation across both the visual and forensic channels:

$$Y_{Cout} = \sum_{Cin=1}^{6} W'_{Cout,Cin} * T_{Cin}$$

W' represents the newly initialized weight matrix of shape $(C_{out}, 6, k, k)$. Additionally, the bias term is explicitly disabled to match the structural design of modern lightweight architectures, which utilize batch normalization immediately following the initial convolution. Because the input channel depth is fundamentally altered, standard ImageNet pre-trained weights cannot be directly pertained to this specific layer [37]. Accordingly, while the downstream layers may leverage adapted weights or transfer learning, the changed input layer is initialized to discover the complicated spatial correlations between the visual semantics and the forensic anomalies from the dataset.

3.6. Evaluation

Models were evaluated using Accuracy, False Acceptance Rate (FAR), and False Rejection Rate (FRR). In forensic applications, a high FAR is dangerous because synthetic images may be accepted as authentic. A high FRR is also problematic because authentic images may be wrongly rejected. The definitions used in this study are as follows:

- True Positive (TP): AI-generated image correctly classified as AI-generated.
- True Negative (TN): Real image correctly classified as real.
- False Positive (FP): Real image incorrectly classified as AI-generated.
- False Negative (FN): AI-generated image incorrectly classified as real.

The formulas are defined as:

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

$$FAR = \frac{\text{FN}}{FN + TP}$$

$$FRR = \frac{\text{FP}}{FP + TN}$$

4. RESULT

This study evaluated the efficacy of integrating digital forensic features into CNNs for detecting AI-generated images. Models were trained and tested using standard 3-channel RGB inputs and the proposed 6-channel multi-modal tensor.

4.1. Model Training and Validation Analysis

This study evaluated the effectiveness of integrating digital forensic features (Fast Fourier Transform, Error Level Analysis, and Laplacian High Pass Filtering) into Convolutional Neural Networks for the detection of AI-generated images. Three model (MobileNetV3-S, MobileNetV2, and ResNet50) were trained and tested using standard 3-channel RGB (Standard) inputs and a proposed 6-channel multi-modal tensor (Forensic). To ensure fair competition and prevent overfitting, an early stopping mechanism (patience = 10 epochs) was utilized to capture the optimal validation metrices[28]. The comparative performance is summarized in Table 1.

Table 1. Standard vs Forensic Features.

Model	Type	Train Acc	Train Loss	Val Acc	Val Loss	Val FAR	Val FRR
Mobilenetv3-S	Standard	0.8848	0.2775	0.8813	0.3186	0.0638	0.1737
Mobilenetv3-S	Forensic	0.9737	0.0739	0.9563	0.1238	0.0312	0.0563
mobilenetv2	Standard	0.7948	0.4478	0.7506	0.4939	0.0750	0.4238
mobilenetv2	Forensic	0.9523	0.1287	0.9556	0.1182	0.0225	0.0663
resnet50	Standard	0.9454	0.1457	0.9556	0.1190	0.0437	0.0450
resnet50	Forensic	0.9614	0.0992	0.9588	0.1227	0.0275	0.0550

The training phase revealed how forensic features stabilize model convergence. When relying solely on spatial RGB data, MobileNetV2 demonstrated severe instability, triggering early stopping at merely Epoch 12 with a validation loss of 0.4939 (as shown in Figure 2). However, by introducing the 6-channel forensic tensor, MobileNetV2 maintained stable gradient descent until Epoch 68, reducing its validation loss to 0.1182.

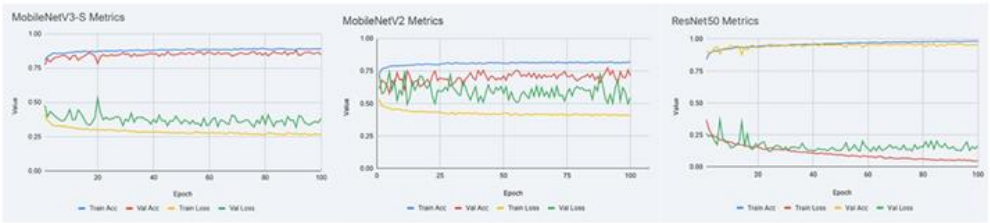

Figure 2. Metrices for Standard MobileNetV3-S, MobileNetV2 and ResNet50

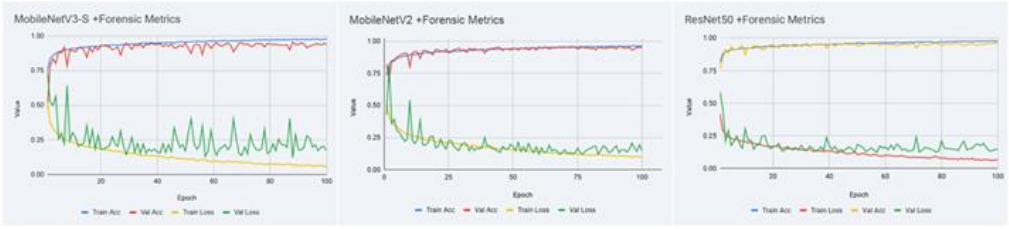

Figure 3. Metrices for Forensic MobileNetV3-S, MobileNetV2 and ResNet50

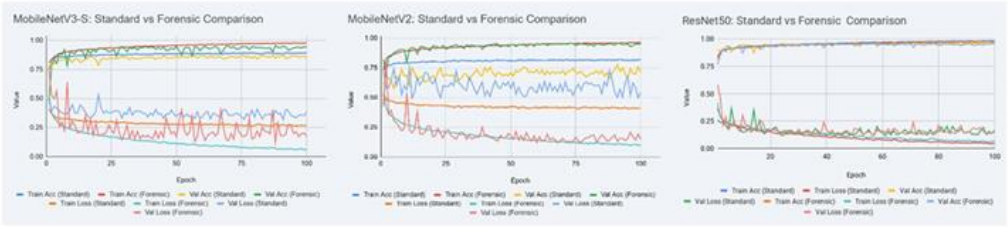

Figure 4. Metrices for Standard vs Forensic

While the lightweight models showed dramatic shifts in their learning curves, the heavyweight ResNet50 baseline achieved a robust validation accuracy of 95.56% by Epoch 24. This demonstrates that architectures possessing over 25 million parameters have the mathematical capacity to learn spatial AI artifacts directly from standard RGB data. When integrated with forensic features, ResNet50 maintained a high validation accuracy of 95.88%, though its convergence was slightly delayed due to the optimization complexity inherent in multi-modal feature fusion [43].

4.2. Test-Set Performance and ErrorRates

While validation metrics guided the early stopping mechanism, the final claim of robustness is established on the unseen testing set (1,600 images). The comparative test-set performance is summarized in Table 2.

Table 2. Test-Set Performance.

Model	Type	Accuracy	Precision(AI/Real)	FAR	FRR	F1-Score
MobileNetV3-S	Standard	0.8650	0.8188/0.9269	0.0625	0.2075	0.8643
MobileNetV3-S	Forensic	0.9569	0.9668/0.9474	0.0325	0.0538	0.9569
MobileNetV2	Standard	0.7662	0.7054/0.9218	0.0512	0.3962	0.7564
MobileNetV2	Forensic	0.9494	0.9389/0.9603	0.0388	0.0625	0.9494
ResNet50	Standard	0.9500	0.9412/0.9592	0.0400	0.0600	0.9500
ResNet50	Forensic	0.9581	0.9464/0.9705	0.0288	0.0550	0.9581

In digital forensics, the consequences of false positives and false negatives carry significant weight. The False Acceptance Rate (FAR), where an AI-generated image is incorrectly classified as authentic, is a critical vulnerability for any detection system. The baseline MobileNetV3-small (RGB) suffered from a high FRR of 20.75% and a FAR of 6.25%. By transitioning to the proposed forensic framework, the FRR for MobileNetV3-small plummeted to 5.38%, and the FAR dropped to 3.25%. Across all models, the integration of FFT, ELA, and HPF drastically reduced the error rates, confirming that frequency-domain analysis exposes the synthetic noise fingerprints that easily bypass standard spatial filters.

4.3. Computational Complexity and Efficiency Analysis

To evaluate the practical feasibility of the proposed framework for real-world forensic deployment, a comparative analysis of computational cost, average execution time per image and structural complexity are shown in Table 3.

Table 3. Test-Set Computational Cost.

Model	Type	ErrorRate	Avg.InferenceTime	Flops	Parameter
MobileNetV3-S	Standard	0.1350	0.38ms/img	0.06G	1.65M
MobileNetV3-S	Forensic	0.0431	0.49ms/img	0.07G	1.65M
MobileNetV2	Standard	0.2338	0.75ms/img	0.33G	2.39M
MobileNetV2	Forensic	0.0506	0.78ms/img	0.34G	2.39M
ResNet50	Standard	0.0500	2.01ms/img	4.13G	23.51M
ResNet50	Forensic	0.0419	2.04ms/img	4.25G	23.52M

Based on Table 3, the implementation of forensic-fusion approach into lightweight models shows only marginal computational overhead. Transitioning from standard RGB baseline to six-channel approach for MobileNetV2 increases the computational complexity by 0.01G FLOPs and increase the inference time by 0.3ms per image. A similar trade-off is observed in MobileNetV3-S, where the inference time increases from 0.38ms to 0.49ms per image. The raise in resource demands can be determined that the model use extra computational as it contrasts to the massive drops in classification error across the lightweight architectures.

This result is visually mapped in Figure 5, where the error rates are extracted and plotted with parameters and FLOPs/MACs.

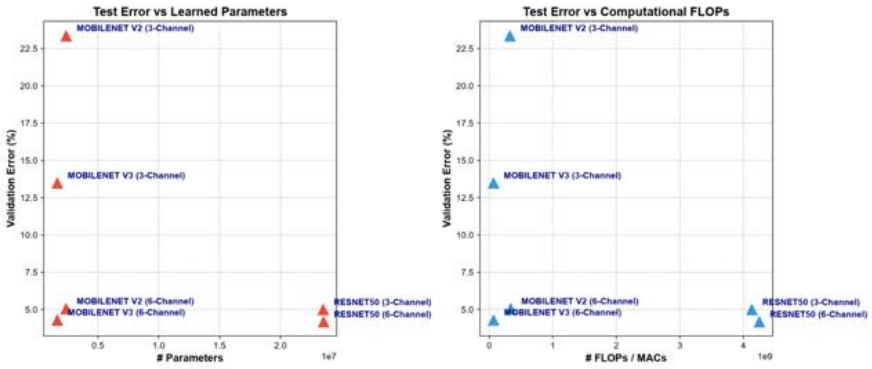

Figure 5 show that the proposed forensic framework can effectively break the traditional coupling between heavy model parameterization and low error bounds. Error rates of MobileNetV3-S that integrates with forensic features falls to 4.31%, achieved a very close result with larger model performance which is ResNet50 with forensic features (4.19%). Moreover, with nearly 14.2 times fewer parameter (1.65M vs 23.51M), 59 times less computational budget (0.07G vs 4.13G FLOPs), MobileNetV3-S can still deliver faster image inference time by 4.1 times (0.49ms vs 2.04ms).

4.4. The Impact of Forensic Features across Architectures

The most important finding of this study is the insightful impact of forensic features on lightweight, mobile-optimized architectures. In their standard RGB states, both MobileNetV3-S

and MobileNetV2 struggled to mathematically distinguish between the natural spatial degradation of real images and the complex interpolation artifacts generated by Stable Diffusion.

However, by providing explicit frequency-domain features alongside noise residuals, the lightweight models were able to utilize hidden artifacts from raw pixels. The forensic-fusion MobileNetV3-S achieved 95.69% test accuracy, below the ResNet50 models with forensic features. This proves that injecting forensic priors allows highly efficient models to successfully rival a computationally expensive networks like ResNet50 in digital forensic tasks.

5. CONCLUSION, LIMITATION AND FUTURE WORK

This study was conducted in order to determine the efficiency of integration of digital forensics techniques into Convolutional Neural Networks for the identification of AI-generated images. The proposed early-fusion model has been able to integrate the RGB data with ELA, FFT, and HPF forensics into a six-channel input tensor. As shown by the outcomes, forensic feature fusion has significantly improved the performance of the lightweight CNN models. This is proved in the case of MobileNetV2 which witnessed an increase in its accuracy from 76.62% using the standard RGB input to 94.94% using the forensic-fusion input. MobileNetV3-small has provided the best overall test accuracy in the lightweight model category of 95.69%.

5.1. Deployment and Scalability

From the implementation, these findings have immense utility in forensic application scenarios. With proof that lightweight neural networks can perform with a similar error rate to heavy neural networks when prompted with multimodal forensic cues, this framework represents a highly scalable system. With low computational requirements of just 0.07G FLOPs and average prediction time of 0.49ms per image with MobileNetV3-S, the detection approach is suitable for practical use on resource-limited edge computing platforms, mobile forensic applications, and automated content moderation systems.

5.2. Limitations and Future Work

However, even with these improvements, the present work contains some limitations for further study. First of all, the experiments use only a part of the CIFake dataset that only contains synthetic images created using Stable Diffusion v1.4. Thus, the question is still open whether the exact noise residuals and frequency anomalies captured by the six-channel tensor will generalize for synthetic media created by other diffusion algorithms (Midjourney and DALL-E 3) and GAN based generators (StyleGAN and BigGAN). The second limitation is the lack of verification of the proposed solution under real-life post-processing of images such as JPEG compression, downscaling, or social media noise. Future research will be focused on testing the generalization of the framework across multiple datasets. Another point of interest will be the application of ensemble learning or transformer models in order to make the method more resistant to adversarial attacks.

REFERENCES

[1] J. Priyanka, A. A. Sundaresan, L. R. Kalabarige, and R. A. Venkata, “Advancements in image forgery detectionthroughdeeplearningmethods,”in *AIPConferenceProceedings*,P.Rajalakshmy,X.A.Mary, and P. A. C. Raj, Eds., American Institute of Physics, 2025. doi: 10.1063/5.0269082.

[2] J.Fridrich,“Digitalimageforensics,” *IEEESignalProcess.Mag.*,vol.26,no.2,pp.26–37,2009,doi: 10.1109/MSP.2008.931078.

[3] Z.LiandN.-F.Law,“RecentAdvancesinDeepLearning-BasedSourceCameraIdentificationand Device

Linking," *Sensors*, vol. 25, no. 24, 2025, doi: 10.3390/s25247432.

[4] P. K. Singh, D. Kumar, and P. Ranjan, "A comprehensive review of deep learning techniques for image tamperingdetection," *LifeCycleReliabilityandSafetyEngineering*,vol.15,no.1,pp.87–102,2026,doi: 10.1007/s41872-025-00385-8.

[5] D.P.TimothyandA.K.Santra,"DetectingDigitalImageForgerieswithCopy-MoveandSplicingImage Analysis using Deep Learning Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 5, pp. 1299–1306, 2024, doi: 10.14569/IJACSA.2024.01505131.

[6] S.Weir,M.S.Khan,N.Moradpoor,andJ.Ahmad,"EnhancingAI-GeneratedImageDetectionwitha Novel Approach and Comparative Analysis," in *2024 17th International Conference on Security of Information and Networks, SIN2024*, Institute of Electricaland ElectronicsEngineersInc., 2024. doi: 10.1109/SIN63213.2024.10871900.

[7] P. Chathampally and V. Mary Amala Bai, "Beyond Pixels: Exploring Deep Learning Methods for Image ForgeryDetection,"*SSRGInternationalJournalofElectronicsandCommunicationEngineering*,vol.11, no. 8, pp. 227–243, 2024, doi: 10.14445/23488549/IJECE-V11I8P123.

[8] K. B. Mohammed, I. Agrawal, M. D. Kandimalla, P. F. Govathoti, C. S. Prakash, and P. Singh, "AdvancingDigitalImageForensics:EnhancingImageForgeryDetectionThroughErrorLevelAnalysis and Convolutional Neural Networks," in *Lecture Notes in Networks and Systems*, M. Pant, K. Deep, and A.Nagar,Eds.,SpringerScienceandBusinessMediaDeutschlandGmbH,2024,pp.325–340.doi: 10.1007/978-981-97-3292-0_23.

[9] N.GeorgeandC.Anitha,"Dual-DomainFrequency-GuidedCapsuleNetwork(DDF-CapsNet)forCopy-Move and Splicing Detection," in *Proceedings of the 9th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 905–911. doi: 10.1109/ICECA66444.2025.11383229.

[10] S.A.Sahy,"HybridAlgorithmforDigitalImageForgeryDetection:IntegratingELA,PRNU,andDeep Learning Techniques," *International Journal of Intelligent Engineering and Systems*, vol. 19, no. 5, pp. 508–524, 2026, doi: 10.22266/ijies2026.0531.30.

[11] S. Cha, U. Kang, and E. Choi, "The image forensics analysis of jpeg image manipulation (lightning talk),"in*Proceedings-20184thInternationalConferenceonSoftwareSecurityandAssurance,ICSSA 2018*, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 82–85. doi: 10.1109/ICSSA45270.2018.00029.

[12] D.D.FidanandG.Tahaoglu,"ForgeryDetectioninDigitalImagesUsingElaandAttention-BasedCnn Architecture," in *2025 18th International Conference on Information Security and Cryptology, ISCTurkiye 2025 - Proceedings*, A. A. Selcuk, S. Sagiroglu, O. Yayla, and C. Tezcan, Eds., Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ISCTrkiye68593.2025.11224846.

[13] R. Gorle and A. Guttavelli, "Enhanced Image Tampering Detection using Error Level Analysis and a CNN,"*Engineering,TechnologyandAppliedScienceResearch*,vol.15,no.1,pp.19683–19689,2025, doi: 10.48084/etasr.9593.

[14] J.KobielaandP.M.Dzierwa,"EffectofFrequencyFeaturesofELAMapsontheDetectionPerformance of Image Manipulation Based on DCT and FFT Basis Features," in *Lecture Notes in Computer Science*, R.Wrembel,G.Kotsis,I.Khalil,andA.M.Tjoa,Eds.,SpringerScienceandBusinessMedia Deutschland GmbH, 2026, pp. 299–304. doi: 10.1007/978-3-032-02088-8_22.

[15] B.Dalrymple,"Fastfouriertransform:ANiche,butcriticalstrategyinoptimizingimpressionevidence," *Journal of Forensic Science and Medicine*, vol. 4, no. 4, pp. 192–196, 2018, doi: 10.4103/jfsm.jfsm_35_18.

[16] M.-C. Pan, C.-H. Chen, L.-Y. Chen, M.-C. Pan, and Y.-M. Shyr, "Highly resolved diffuse optical tomography:asystematicapproachusinghigh-passfilteringforvalue-preservedimages,"*J.Biomed. Opt.*, vol. 13, no. 2, p. 024022, 2008, doi: 10.1117/1.2907344.

[17] Shreya, S. P. Metkar, N. Babar, and Y. M. Vaidya, "High-Pass Filter Assisted Deep Feature Extraction forImageForgeryDetection,"*2025IEEEInternationalConferenceonComputer,Electronics,Electrical Engineering and their Applications, IC2E3 2025*, 2025, doi: 10.1109/IC2E365635.2025.11167183.

[18] J. Wu, X. Chang, T. Yang, and K. Feng, "Blind Forensic Method Based on Convolutional Neural NetworksforImageSplicingDetection," *2019IEEE5thInternationalConferenceonComputerand Communications, ICCC 2019*, pp. 2014–2018, Dec. 2019, doi: 10.1109/ICCC47050.2019.9064258.

[19] D.SurveandA.Nimkar,"IntegratingErrorLevelAnalysisandSensorPatternNoiseforImageForgery Detection," in *2024 IEEE International Conference on Intelligent Signal Processing and Effective*

Communication Technologies, INSPECT 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/INSPECT63485.2024.10895991.

[20] S.Kaur,N.Sinha,P.Jain,S.Koli,A.Sharma,andA.Lathwal,"EnhancedImageForgeryDetection using a Hybrid Approach: Integration of ELA, CNN, and XGBoost," *International Journal of Performability Engineering*, vol. 20,no.6, pp. 367–378, 2024, doi:10.23940/ijpe.24.06.p4.367378.

[21] D.Rawat,D.Rawat,A.Rawat,A.Pandey,G.Chauhan,andS.Aluvala,"EthicsintheeraofIntelligent Machines: Legal Perspective and Regulatory Implications," in *2024 IEEE 1st Karachi Section HumanitarianTechnologyConference,Khi-HTC2024*,InstituteofElectricalandElectronicsEngineers Inc., 2024. doi: 10.1109/KHI-HTC60760.2024.10481959.

[22] B. Lorch, N. Scheler, and C. Riess, "Compliance Challenges in Forensic Image Analysis Under the ArtificialIntelligenceAct,"in*EuropeanSignalProcessingConference*,EuropeanSignalProcessing Conference, EUSIPCO, 2022, pp. 613–617. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85141010801&partnerID=40&md5=33b99138a2b29a2ffd7cd890cca07f68

[23] I. de Miguel Beriain and L. I. A. de Miguel, "Use of AI Tools for Forensic Purposes: Ethical and Legal ConsiderationsfromanEUPerspective,"in *Driv.ForensicInnovationinthe21stCentury:Crossingthe Valley of Death*, Springer International Publishing, 2024, pp. 147–164. doi: 10.1007/978-3-031-56556-4_7.

[24] P.Čisar,"OVERVIEWOFDIGITALIMAGEFORGERYDETECTION," *FactaUniversitatis,Series: Electronics and Energetics*, vol. 38, no. 4, pp. 581–603, 2025, doi: 10.2298/FUEE2504581C.

[25] P. Deb, S. Deb, A. Das, and N. Kar, "Image Forgery Detection Techniques: Latest Trends and Key Challenges,"*IEEEAccess*,vol.12,pp.169452–169466,2024,doi:10.1109/ACCESS.2024.3498340.

[26] A.RajandD.Sankar,"Atexturedescriptorofthecamerafingerprintforsourcecameramodelanddevice identification,"*InternationalJournalofIntelligentSystemsTechnologiesandApplications*,vol.22,no.2, pp.213–235,2024,doi:10.1504/IJISTA.2024.139815.

[27] M.Li,J.Li,andS.Jia,"HybridFrequency-guidedFusionNetworkforFaceForgeryDetection,"in *2025 6th International Conference on Computer Engineering and Intelligent Control, ICCEIC 2025*, Institute of Electrical and Electronics Engineers Inc., 2025, pp. 595–602. doi: 10.1109/ICCEIC67916.2025.11309125.

[28] M.Shadri,A.Hanafiah,andP.Velusamy,"AI-GeneratedImageDetectionUsingConvolutionalNeural Network (CNN) Algorithm," *Artificial Intelligence and Language Models*, vol. 1, no. 1, pp. 1–11, Apr. 2026, Accessed: May 18, 2026. [Online]. Available: https://acspub.id/index.php/ailm/article/view/4

[29] A.M.Nagm*etal.*,"DetectingimagemanipulationwithELA-CNNintegration:apowerfulframeworkfor authenticity verification," *PeerJ Comput. Sci.*, vol. 10, p. e2205, Aug. 2024, doi: 10.7717/peerj-cs.2205.

[30] W. Alkishri, J. H. Yousif, and M. Al-Bahri, "DEEPFAKE IMAGE DETECTION METHODS USING DISCRETEFOURIERTRANSFORMANALYSISANDCONVOLUTIONALNEURALNETWORK," 2023, doi: 10.17605/OSF.IO/BNPC4.

[31] MikkoSuominen,"IMAGECLASSIFICATIONINTHEFOURIERDOMAINBachelor'sthesisFaculty of Information Technology and Communication Sciences," 2024.

[32] D. W. J. A. G. T. L. S. M. Teady Matius Surya Mulyana, "Building Drawing Line Art with High Pass Filtering and Image Morphology." Accessed: Jan. 15, 2026. [Online]. Available: https://www.researchgate.net/publication/366030382_Building_Drawing_Line_Art_with_High_Pass_Filt ering_and_Image_Morphology

[33] N.Tian,X.Qiu,andQ.Pan,"AnimprovedPRNUnoiseextractionmodelforhighlycompressedimage blocks with low resolutions," *Multimed. Tools Appl.*, vol. 83, no. 25, pp. 66657–66690, Jan. 2024, doi: 10.1007/s11042-024-18255-3.

[34] K.He,X.Zhang,S.Ren,andJ.Sun,"DeepResidualLearningforImageRecognition,"pp.15731–15740, Dec. 2015, Accessed: Dec. 28, 2025. [Online]. Available: http://arxiv.org/abs/1512.03385

[35] M.TanandQ.V.Le,"EfficientNetV2:SmallerModelsandFasterTraining,"Jun.2021,[Online]. Available:http://arxiv.org/abs/2104.00298

[36] Z.J.Wang*etal.*,"CNNExplainer:LearningConvolutionalNeuralNetworkswithInteractive Visualization," Aug. 2020, doi: 10.1109/TVCG.2020.3030418.

[37] S.S.BaraheemandT.V.Nguyen,"AIvs.AI:CanAIDetectAI-GeneratedImages?," *J.Imaging*,vol.9, no. 10, p. 199, Sep. 2023, doi: 10.3390/jimaging9100199.

[38] C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, "Early versus late fusion in semantic video analysis,"*Proceedingsofthe13thACMInternationalConferenceonMultimedia,MM2005*,pp.399–402, 2005, doi: 10.1145/1101149.1101236.

[39] J.H.Bappy,C.Simons,L.Nataraj,B.S.Manjunath,andA.K.Roy-Chowdhury,"HybridLSTMand Encoder-Decoder Architecture for Detection of Image Forgeries," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3286–3300, Jul. 2019, doi: 10.1109/TIP.2019.2895466.

[40] Y.Jeong,D.Kim,P.Kim,Y.Ro,andJ.Choi,"Self-supervisedGANDetector,"Mar.2022,[Online]. Available:http://arxiv.org/abs/2111.06575

[41] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Learning Rich Features for Image Manipulation Detection,"*ProceedingsoftheIEEEComputerSocietyConferenceonComputerVisionandPattern Recognition*, pp. 1053–1061, Dec. 2018, doi: 10.1109/CVPR.2018.00116.

[42] A.Paszke*etal.*,"PyTorch:AnImperativeStyle,High-PerformanceDeepLearningLibrary," *Adv.Neural Inf. Process. Syst.*, vol. 32, Dec. 2019, Accessed: May 18, 2026. [Online]. Available: https://arxiv.org/pdf/1912.01703

[43] MartinezMartinez,L.2024.ClassifyingAI-GeneratedimagesusingEfficientNet-B0,ResNet50and VGG16 CNN ML models MSc Research Project MSc in Artificial Intelligence. https://norma.ncirl.ie/7918/1/lazarojaviermartinezmartinez.pdf

AUTHORS

Nazirah Abd Hamid is a lecturer at Universiti Sultan Zainal Abidin, Terengganu, Malaysia. She received her BIT from Universiti Utara Malaysia (UUM), her M.Sc. from Universiti Teknologi Malaysia (UTM), and her Ph.D. from Universiti Teknikal Malaysia Melaka (UTeM). She specializes in biometric security systems and security services, including digital forensics, steganography, network security, public key infrastructure, and biometrics.

Shaifol Ziqri Bin Zainurin is a final-year B.Sc. Computer Network Security student at Universiti Sultan Zainal Abidin, Malaysia, focusing on digital forensics, network architecture, and cyber defense. He holds a diploma in Electronics Engineering (Computer) from Politeknik Sultan Mizan Zainal Abidin. Alongside his cybersecurity studies, he has practical experience in computer vision and has actively contributed to various IoT and robotics projects.

Siti Dhalila Mohd Satar received her BIT from Universiti Kebangsaan Malaysia (UKM), her M.Sc. from Universiti Teknologi Malaysia (UTM), and her Ph.D. from Universiti Putra Malaysia (UPM). She is a lecturer and Academic Program Coordinator (PPA) for the Bachelor of Computer Science (Computer Network Security with Honours) program at Universiti Sultan Zainal Abidin. She specializes in authentication systems and security services, including digital forensics, steganography, network security, public key infrastructure, and biometrics.

Norhidayah Muhammad received her Ph.D. from Universiti Teknologi Mara (UiTM), her M.Sc. from Universiti Malaysia Pahang (UMP), and her BIT from Universiti Malaysia Terengganu (UMT). She is a lecturer at Universiti Sultan Zainal Abidin, specializing in software project management, cryptography and cryptanalysis (including key management and random number generation), data hiding and steganography, data security and digital security and privacy, and data encryption.

Mohd Fadzil Abdul Kadir is a senior lecturer at Universiti Sultan Zainal Abidin. He received his Ph.D. in Engineering from Mie University, Japan, his M.Sc. from Universiti Utara Malaysia, and his B.Eng. also from Mie University, Japan. He specializes in image processing and pattern recognition, encryption systems, data hiding and steganography, and image recognition.

Muhammad Syakir Razmin, a student from Universiti Sultan Zainal Abidin (UniSZA), in his last year of a Bachelor of Computer Science (Computer Network Security) with Honours, is pursuing up his studies at the Faculty of Informatics & Computing. Before UniSZA, he studied at Negeri Sembilan Matriculation College. Syakir expects to earn his degree in 2027. His academic interests in on Artificial Intelligence, especially AI voice detection. His coursework covers computer networks and cybersecurity. In addition, he holds a Cisco certification, demonstrating his competency in networking technologies and related fields.

Muhammad Daniel Mohd. Shaiful Azhari is a final-year B.Sc. Computer Network Security student at Universiti Sultan Zainal Abidin, Besut Campus. He specializes in network architecture, SOC operations, Zero Trust security, cloud infrastructure, and incident response. A graduate of Perak Matriculation College (KMPk), he possesses strong practical experience in AWS, Docker, Wireshark and the implementation of Zero Trust security projects. With proven leadership experience managing large-scale student programs.

Nur Wasilah Jamlih, a Computer Science student in Universiti Sultan Zainal Abidin (UNISZA) specializing in Networking and Security. Demonstrated technical proficiency in Network Simulation (NS-3), Penetration Testing (Kali Linux, Metasploit), and IoT-based systems. Proven ability to combine technical troubleshooting with strong leadership and communication skills developed through residential council management and community-based marketing projects.