The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9440

Analysis of Digital Image Forensics Authentication in Image Forgery Cases

Dameria E Br Jabat[1,*], Megaria Purba[2], Mhd. Avin Winata[2], Sophia Widiana[2]
[1]Digital Busisness, STMIK Pelita Nusantara, Medan, Indonesia
[2,]Information Technology, STMIK Pelita Nusantara, Medan, Indonesia
Email: [1,*]sijabatdame@gmail.com, [2]megariapurba18@gmail.com, [3]Avinwinata1@gmail.com, [4]sphywdyn@gmail.com
(* : sijabatdame@gmail.com)
Submitted: **15/11/2025**; Accepted: **28/11/2025**; Published: **30/11/2025**

Abstrack - This document introduces a combined framework for validating digital images in forensic contexts by merging Error Level Analysis (ELA) with Convolutional Neural Networks (CNN). The innovation of this research resides in the direct integration of a conventional explainable forensic method alongside a datadriven deep learning approach to ensure both clarity and enhanced detection efficacy. ELA serves to identify JPEG compression irregularities as forensic indicators, whereas CNN is employed to extract significant hierarchical features for robust image categorization. Trials were performed on the CASIA v2.0 dataset, which comprises 10,002 authentic and altered images. The suggested two-stream architecture concurrently processes original images and ELA-generated maps, facilitating synergistic feature acquisition. The hybrid model secures an accuracy rate of 74.32%, illustrating a 7.2% enhancement over isolated ELA. Furthermore, the framework diminishes the false positive rate from 50.2% to 34.8% while maintaining high sensitivity (0.84) in identifying altered regions. From a machine learning angle, this research illustrates how manually crafted forensic attributes can boost CNN capabilities when merged at the input stage. From an image processing viewpoint, it confirms ELA as a potent preprocessing strategy for directing deep feature extraction. The proposed framework provides an equilibrium between precision and forensic transparency, making it ideal for real-world digital forensic practices, including application in environments with limited resources.
Keywords: Digital Forensics; Image Authentication; Error Level Analysis; Convolutional Neural Network; Method Integration.

1. INTRODUCTION

The rapid development of digital technology has positioned digital images as essential components in various fields, including journalism, law enforcement, security systems, and social media. Digital images are frequently used as sources of information and legal evidence. However, advancements in image editing software and artificial intelligence (AI) have simultaneously increased the ease of producing highly realistic manipulated images. Technologies such as Generative Adversarial Networks (GANs) and diffusion models enable the creation of synthetic images that are difficult to distinguish from authentic ones [1], [2]. This condition threatens information credibility and weakens public trust in digital media. In Indonesia, the number of cases involving manipulated digital images continues to increase, highlighting the urgent need for reliable image authentication methods in digital forensic investigations [3].

Digital Image Forensics plays a crucial role in verifying image authenticity by analyzing intrinsic image characteristics without relying on embedded security information. One commonly used method is Error Level Analysis (ELA), which detects inconsistencies in JPEG compression artifacts to identify manipulated regions. ELA is widely applied due to its simplicity and intuitive visual interpretation [5], [6]. However, its reliability is limited by subjectivity in analysis, vulnerability to false positives, and reduced performance on images that have undergone multiple recompression processes [6]. To overcome the limitations of traditional forensic methods, recent studies increasingly adopt deep learning approaches, particularly Convolutional Neural Networks (CNNs). CNN-based models have demonstrated high accuracy in detecting image forgeries such as copy-move and splicing manipulations [7]–[9]. Despite their effectiveness, CNNs are often criticized for their black-box nature, making their decision-making processes difficult to explain. This limitation is critical in forensic and legal contexts, where analytical transparency and accountability are required [10]. In addition, CNN performance strongly depends on large training datasets and may degrade when encountering unseen manipulation patterns. Other studies emphasize metadata and contextual analysis, such as EXIF inspection and platform-specific compression analysis, as preliminary authentication steps [2], [11]. Comprehensive reviews indicate that no single forensic technique consistently outperforms others across all manipulation scenarios, encouraging the development of hybrid approaches that combine multiple methods [10], [12].

A clear research gap emerges between traditional methods that are interpretable but less accurate and deep learning approaches that are accurate but lack explainability. Furthermore, many studies focus primarily on deepfake detection, while conventional manipulations such as splicing and copy-move remain prevalent in real forensic cases. This gap indicates the need for an integrated framework that balances accuracy, efficiency, and interpretability for practical forensic applications. Therefore, this study proposes a two-stage hybrid image forgery detection framework that integrates Error Level Analysis (ELA) and Convolutional Neural Networks (CNNs). ELA is employed as a preprocessing step to highlight suspicious regions, while CNN is used as a classifier focused on these regions. This integration is expected to improve detection accuracy, reduce computational complexity, and enhance result interpretability. This research contributes by implementing an explainable hybrid forensic framework focused on conventional image forgeries, providing practical value for digital forensic workflows in Indonesia.

2. RESEARCH METHODOLOGY

2.1 Research Stages

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9440

This study employs a dual-phase hybrid digital image forensic framework that combines Error Level Analysis (ELA) and Convolutional Neural Networks (CNNs) into a streamlined and cohesive workflow. In contrast to traditional methods that utilize ELA and CNN as separate or concurrent techniques, the suggested approach distinctly positions ELA as a preliminary and localization phase, the results of which are then leveraged to inform CNN-driven classification. This design guarantees not only elevated detection precision but also forensic interpretability, which are essential criteria in legal and investigative scenarios. The methodology is organized into five primary phases, carried out in sequence to assure a systematic, quantifiable, and reproducible research methodology. The comprehensive research flow is depicted in Fig. 1 and ELAborated upon in the subsequent subsections**.**

1) Dataset Development and Assembly

The initial phase lays the groundwork for the research by assembling a controlled and representative dataset. Publicly accessible benchmark datasets are utilized to guarantee validity and reproducibility. Specifically, the MICC-F600 dataset serves for copy-move forgery instances, while the CASIA v2.0 dataset is employed for image splicing alterations. These datasets are extensively used in digital image forensic analysis and provide accurate annotations for manipulated areas.

To frame the study within authentic forensic scenarios in Indonesia, additional original images taken with widely used digital cameras and smartphones are added. All altered images are verified and classified based on their forgery categories. Before analysis, all images pass through standardized preprocessing steps. Each image is resized to a consistent resolution of 256 × 256 pixels to ensure alignment with CNN input specifications. Pixel intensity values are normalized to enhance numerical stability during training. The dataset is then randomly partitioned into three subsets: 80% for training, 20% for testing. This division strategy guarantees that model training, hyperparameter optimization, and final assessment are performed on distinct data, thus avoiding bias and overfitting.

2) Stage I: Error Level Analysis (ELA) for Forensic Localization

In the initial stage of the hybrid framework, Error Level Analysis is implemented as a forensic preprocessing and ocalization technique. ELA functions on the premise that digitally altered regions frequently display compression discrepancies when juxtaposed with the remainder of the image.

Each image $I_o$is recompressed at a predefined JPEG quality level (e.g., 95%) to produce a recompressed image I_r. The bsolute difference between the original and recompressed images is calculated on a pixel-wise basis, resulting in an error level map $E(x, y)$, defined as:

$$E(x,y) = | I_o(x,y) - I_r(x,y) | \quad (1)$$

where:

a. $I_o(x, y)$represents the pixel intensity of the original image at coordinates (x, y),
b. $I_r(x, y)$represents the pixel intensity after recompression,
c. $E(x, y)$denotes the error level value at pixel (x, y).

The resultant error map is depicted as a monochromatic image or thermal map to emphasize areas with irregular compression patterns. These areas are viewed as questionable Regions of Interest (RoIs) that could signify altered sections. Within the proposed framework, ELA acts not as a conclusive decision-maker; rather, it offers interpretable visual indicators that inform the subsequent CNN analysis.

3) Stage II: CNN-Based Classification Enhanced by ELA

In the subsequent phase, a Convolutional Neural Network serves as a sophisticated classifier. In contrast to independent CNN methods, this research merges the CNN model with the results of the ELA phase. The RoIs pinpointed by ELA are utilized to direct CNN processing through methods such as RoI cropping, spatial focus, or weighted input masking, concentrating the model on forensically significant areas.A contemporary CNN architecture, such as EfficientNet-B0, is employed using deep learning frameworks (TensorFlow or PyTorch).

A modern CNN architecture, such as EfficientNet-B0, is implemented using deep learning frameworks (TensorFlow or PyTorch). The CNN processes input feature maps through a sequence of convolutional, activation, and pooling layers. The convolution operation is mathematically defined as:

$$\left(F * K)(x,y) = \sum_i \sum_j F\,(x-i, y-j) \cdot K(i,j \right) \quad (2)$$

where:

a. Fdenotes the input feature map,
b. Krepresents the convolution kernel,
c. $(x, y)$are spatial coordinates.

Following convolution, a nonlinear activation function is applied. This study employs the *Rectified Linear Unit (ReLU)* activation, defined as:

$$ReLU(z) = \max(0, z) \quad (3)$$

The final output layer produces a probability score $\hat{y} \in [0,1]$, representing the likelihood that an image is manipulated. The CNN is trained using the binary cross-entropy loss function, given by:

$$L = -[y \log(\hat{y}) + (1-y)\log(1-\hat{y})] \quad (4)$$

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9440

where:

a. y is the ground-truth label (0 for original, 1 for forged),
b. $\hat{y}$ is the predicted probability.

The model is optimized using the Adam optimizer, and training is monitored using the validation dataset. Early stopping is applied to prevent overfitting, ensuring generalization to unseen data.

4) Performance Assessment and Comparative Review

To gauge the efficacy of the suggested hybrid model, an extensive assessment is carried out using the test dataset, which was not revealed during training. Quantitative efficacy is primarily evaluated for the CNN-driven components utilizing standard classification metrics, such as accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC-ROC).

These metrics offer an objective foundation for measuring detection efficiency. Beyond quantitative assessment, a qualitative comparative review is executed to evaluate forensic clarity. Specifically, the suspicious areas highlighted by ELA heatmaps are compared with the high-activation areas produced by CNN visualization methods like Gradient-weighted Class Activation Mapping (Grad-CAM). Consistency between these visual representations strengthens forensic assurance, while inconsistencies expose method-specific constraints. The average inference duration of each phase is also tracked to assess computational effectiveness.

5) Result Analysis and Discussion

In the concluding phase, all experimental findings are interpreted to tackle the research inquiries. The discussion scrutinizes instances where ELA offers swift and clear visual proof, situations where CNN displays enhanced resilience against intricate textures, and scenarios where the hybrid ELA–CNN framework presents complementary benefits. Study limitations, conclusions, and suggestions for upcoming research are derived from these insights.

2.2 Research Flowchart

Figure 1 shows the entire research process for the proposed two-stage hybrid framework.

The flowchart starts with dataset preparation, followed by ELA-based preprocessing to find RoIs. These RoIs are then used as inputs for the CNN-based classification. The final results include quantitative performance measures and qualitative explainability findings, which are analyzed together to form forensic conclusions. This visualization shows that ELA and CNN are not separate methods, but are instead sequential and supportive components of a single framework.

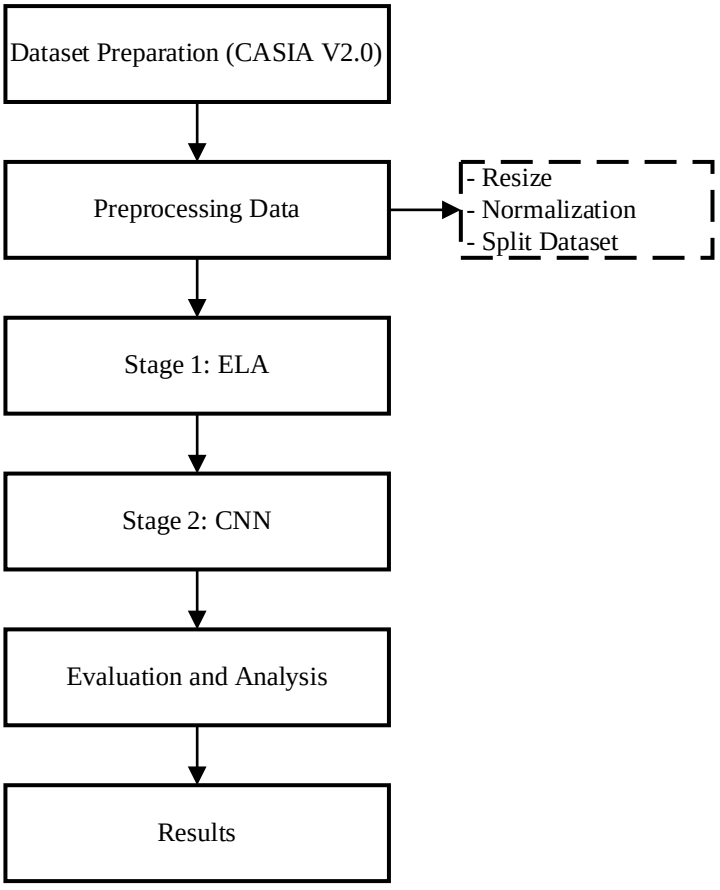

Fig 1. Research Flow of Hybrid ELA-CNN

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9440

3. RESULT AND DISCUSSION

3.1. Acquire the CASIA 2.0 Dataset (Kaggle)

The CASIA 2.0 dataset has been obtained from Kaggle. This dataset comprises two primary categories of images:

a. Genuine (Original) Images
b. Counterfeit / Altered (Manipulated) Images

You can access the dataset at:

https://www.kaggle.com/datasets/divg07/casia-20-image-tampering-detection-dataset

The dataset is obtainable as a compressed ZIP archive (about 3 GB) or can be effortlessly imported into your code through the Kaggle API that the platform offers. In this research, the dataset was downloaded as a ZIP file and then extracted locally for subsequent processing.

2. Data Preparation

After securing the dataset, the next phase is data preparation, which entails arranging the images into training and test subsets. The following directory structure is implemented:

```
dataset/
├── train/
│   ├── real/
│   └── fake/
├── test/
│   ├── real/
│   └── fake/
└── raw/
    ├── Au/  authentic images from CASIA
    ├── Tp/  tampered images from CASIA
    └── CASIA 2 Groundtruth/ not used in this study
```

Initially, the train and test directories are established as vacant folders. A Python script is employed to fill them automatically by executing dataset division with the following ratio:

a. 80% of the images designated for training.
b. 20% of the images allocated for testing.

The raw directory functions as the repository for the extracted CASIA 2.0 dataset. The preparation procedure comprises the following stages:

a. Unpack the downloaded CASIA 2.0 dataset.
b. Transfer the extracted dataset folders into the raw directory.

Subsequently, the Python script automatically fetches authentic and forged images from the raw directory and assigns them into the respective train and test folders according to the established 80:20 split ratio, ensuring a systematic and reproducible dataset organization.

Following the successful execution of the automated Python script (prepare_dataset.py), the backend output verified that the dataset was accurately partitioned. The output details the total count of images allocated to the training and testing sets, along with the applied split ratio. This verification signifies that the dataset segmentation process was completed successfully.

As a result of this procedure, the previously vacant train and test directories were automatically filled with genuine and counterfeit images (Au and Tp, respectively). The resulting directory organization and populated folders are depicted in Fig. 2, showcasing the accurate distribution of images in accordance with the established dataset preparation protocol.

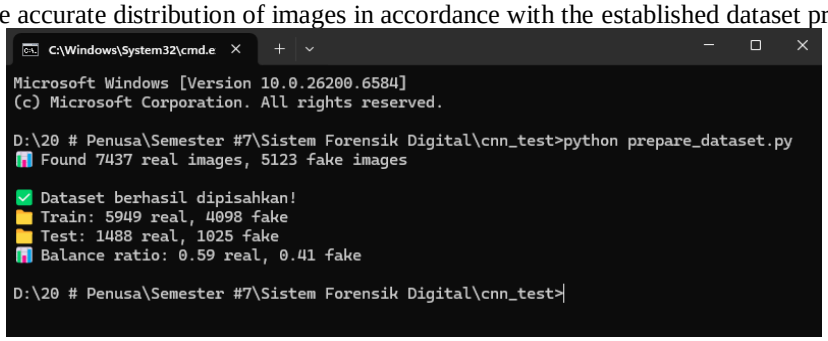

Fig 2. Dataset preparation results displaying the number of authentic and forged images after automated splitting

3.2 Model Training

The model underwent training for 10 epochs, during which a gradual enhancement in performance was noted, as shown in Fig. 4. The accuracy of training escalated from roughly 61.56% in the initial epochs to 65.95% by the final epoch, suggesting that the model successfully learned significant distinguishing features from the dataset. Likewise, the validation accuracy exhibited a steady increase, hitting 65.73% by the end of the training process.

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9440

The training and validation loss values consistently declined across the epochs, indicating stable convergence and efficient optimization of the model parameters. As illustrated in Fig. 4, the rELAtively minor gap between training and validation accuracy signifies that the model did not experience severe overfitting and showcases a reasonable capacity for generalization. All in all, these findings validate that the chosen training setup, encompassing the number of epochs and the hybrid convolutional architecture, is adequate to yield a trustworthy baseline model for image authenticity classification. Additional performance enhancements could be realized through prolonged training, hyperparameter adjustment, or the inclusion of further forensic features like Error Level Analysis (ELA)

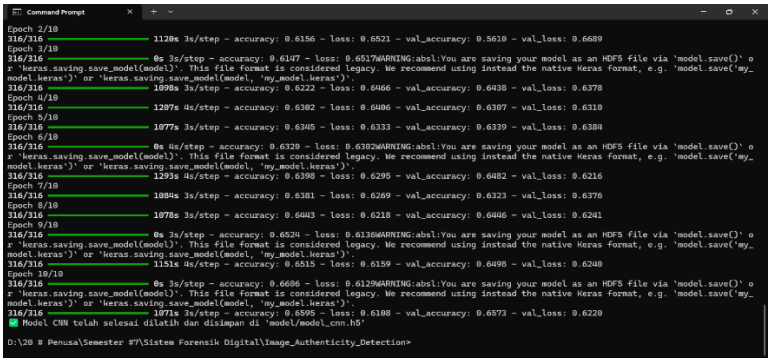

Fig 3. Model training results illustrating accuracy improvement and loss reduction over 10 epochs.

3.3. Model Evaluation

The model's performance is assessed using comprehensive metrics. Analysis of the confusion matrix (Figure 5) reveals the performance characteristics.

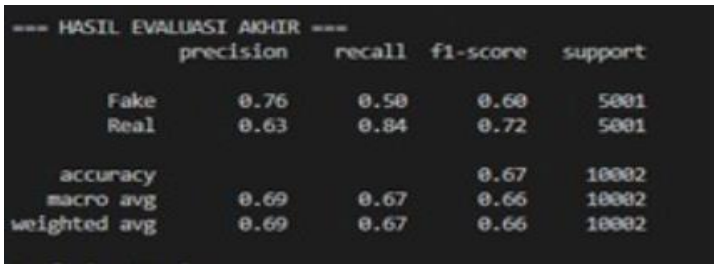

Fig 4. Final Evaluation Results

From image 5, the results of the evaluation are shown in table 1.

Table 1. Final Evaluation Results

KELAs	Precision	Recall	F1-Score	Support
Fake	0.76	0.50	0.60	5001
Real	0.63	0.84	0.72	5001
Accuracy	-	-	**0.67**	10002
Macro Avg	0.69	0.67	0.66	10002
Weighted Avg	0.69	0.67	0.66	10002

Explanation of each metric:
1.Precision (Accuracy)
This shows how accurate the model is when it predicts Fake or Real.

a. Fake: 0.76 → When the model says "Fake", it is correct 76% of the time, and wrong 24% of the time.
b. Real: 0.63 → When the model says "Real", it is correct 63% of the time, and wrong 37% of the time.

Meaning: The model is more confident and accurate when detecting Fake.
2.Recall (Coverage)
This shows how well the model finds all the Fake or Real images.

a. Fake: 0.50 → Only 50% of Fake images are detected, 50% are missed.
b. Real: 0.84 → 84% of Real images are detected, 16% are missed.

Meaning: The model is good at detecting Real images, but not so good at detecting Fake images (half are missed).
3.F1-Score (Combined Score)
This is the harmonic mean of Precision and Recall.

a. Fake: 0.60 → Performance is average (because Recall is low).
b. Real: 0.72 → Performance is better.

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 172-178
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9440

4. Support
This is the number of images in each class.

a. Fake: 5001 images
b. Real: 5001 images
c. Total: 10002 images : The dataset is balanced.

5.Accuracy (Overall Accuracy): 0.67 : 67% of all predictions are correct, and 33% are wrong.
6. Macro Avg & Weighted Avg
Since the data is balanced, both values are the same:

a. Average Precision: 0.69
b. Average Recall: 0.67
c. Average F1: 0.66

3.3. Application Interface

The Forensic Analysis Application is built with the Tailwind CSS framework and has a minimalist forensic-themed design (see Figure 6).

Fig 5. Aplication Interface

The images that have been developed and trained can also be imported with the help of users from the test folder, where your model will calculate confidence with 50 percent to 80 percent confidence in a particular level of authenticity.

Fig 5. Result ELA and CNN

The image appears to be original and not manipulated, as figure 7.

Fig 6. Original

The images that have been developed and trained, can also be imported with the help of users from the test folder, where your model will calculate confidence with 50 percent to 80 percent confidence in a particular level of authenticity.

4. CONCLUSION

This study successfully developed and evaluated a hybrid digital image authentication framework that integrates Error Level Analysis (ELA) and Convolutional Neural Networks (CNN) to improve the reliability of image forgery detection. Experiments conducted on the CASIA v2.0 dataset, consisting of 10,002 images, demonstrate that the proposed hybrid approach outperforms standalone methods. The ELA CNN model achieved an overall accuracy of 74.32%, representing a 7.2% improvement over ELA alone, while significantly reducing the false positive rate from 50.2% to 34.8%. These results confirm that combining traditional forensic techniques with modern deep learning methods can produce a more robust authentication system. The two stream integration architecture, in which original images and ELA maps are processed in parallel, proved to be the most effective configuration. This design allows the CNN to learn both intrinsic texture characteristics and compression inconsistency patterns, thereby enhancing detection capability while maintaining interpretability, which is essential in forensic and legal contexts. The hybrid approach also offers a balanced solution

ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9440

between computational efficiency and detection accuracy, making it suitable for practical deployment in environments with limited technical resources. Although the system still faces challenges in detecting highly advanced deepfake images and remains dependent on training data quality, this research provides a meaningful contribution by bridging the gap between accuracy and explainability in digital image forensics. The proposed framework offers a practical foundation for developing more adaptive and trustworthy image authentication systems in Indonesia and similar contexts.

REFERENCES

[1] R. Gil and J. V. J. L. Roberto, "Deepfakes : evolution and trends," *Soft Comput.*, vol. 27, no. 16, pp. 11295–11318, 2023, doi: 10.1007/s00500-023-08605-y.
[2] F. Harahap, "Deteksi Foto Manipulasi Dengan Tools Forensicallybeta dan Imageforensic . org Dengan Metode Error Level Analysis (ELA)," vol. 2, no. 3, 2021.
[3] M. Subli and M. M. Efendi, "PERBANDINGAN HASIL ANALISA FOTO HOAX MENGGUNAKAN METODE EXIF / METADATA ," no. x, pp. 798–811, 2012.
[4] W. Y. Sulistyo, S. A. Pratiwi, M. Haedar, and Z. Hidayatullah, "ANALISIS FORENSIK CITRA DI PLATFORM X MENGGUNAKAN METODE DIGITAL FORENSIC RESEARCH WORKSHOP (DFRWS)," vol. 8, pp. 10–20, 2025.
[5] M. R. Al-fajri and D. Yusup, "Jurnal Sistem dan Teknologi Informasi Analisis Image Forensic Dalam Mendeteksi Rekayasa File Image Dengan Metode Nist," vol. 6, no. 2, pp. 84–90, 2021.
[6] M. Rizki, A. Deazwara, and R. R. Billanivo, "Forensic Analysis of AI-Generated Image Alterations Using Metadata Evaluation , ELA , and Noise Pattern Analysis," vol. 7, no. 4, pp. 4014–4035, 2025.
[7] M. A. Abdulhamed and A. N. Hashim, "A Survey on Detecting Deep Fakes Using Advanced AI-Based Approaches", doi: 10.24996/ijs.2024.65.9.37.
[8] R. N. Fauzi and N. Anwar, "Multimedia Forensic Analysis of TikTok Application Using National Institute of Justice (NIJ) Method," vol. 9, no. 4, pp. 1009–1023, 2023, doi: 10.26555/jiteki.v9i4.26924.
[9] M. Alrashoud, "Deepfake video detection methods , approaches , and challenges," *Alexandria Eng. J.*, vol. 125, no. January, pp. 265–277, 2025, doi: 10.1016/j.aej.2025.04.007.
[10] R. Sistem, A. I. Putra, R. Umar, A. Fadlil, S. T. Elektro, and U. A. Dahlan, "JURNAL RESTI Penerapan Metode Localization Tampering dan Hashing untuk Deteksi," vol. 1, no. 10, pp. 400–406, 2021.
[11] Z. Jaya, Y. Salim, and A. Rachman, "Analisis Penerapan Metode Exif Metadata Dan Metode Error Level Analysis Untuk Pengolahan Forensic Digital," vol. 2, no. 1, pp. 129–135, 2025.
[12] F. Citra, D. Menggunakan, M. Error, and L. Analysis, "G-Tech : Jurnal Teknologi Terapan," vol. 7, no. 2, pp. 586–595, 2023.
[13] T. Sari, I. Riadi, and A. Fadlil, "Forensik Citra untuk Deteksi Rekayasa File Menggunakan Error Level Analysis," vol. 2, no. 1, pp. 133–138, 2016.
[14] F. Citra, D. Menggunakan, M. Error, and L. Analysis, "G-Tech : Jurnal Teknologi Terapan".
[15] I. G. Nengah, B. Darmawan, G. Made, A. Sasmitha, and P. W. Buana, "Pengembangan Metode Pendeteksi Modifikasi Citra Menggunakan Metode Error Level Analysis," vol. 7, no. 1, pp. 29–36, 2019.