International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

Digital Forensic Approaches to Identifying and Analyzing Deepfake Media

Divyajeetsinh Tripalsinh Rayjada[1], Aditya More[2] Bhumika Doshi[3], Dr. Kashyap Joshi[4], Prof (Dr) Kapil Kumar[5]
[1]*M.Sc. Cyber Security & Forensics*
[2]*PhD Research Scholar, Cyber Security and Digital Forensics*
[3,4]*Assistant Professor, Cyber Security and Digital Forensics*
[5]*Professor and Coordinator*
Department of Biochemistry & Forensic Science, Gujarat University, Ahmedabad, India

***Abstract*-- The rapid advancement of artificial intelligence has brought a troubling side effect the proliferation of deepfake media. These AI-generated images and videos, which can convincingly mimic real people and events, pose serious threats to public trust, democratic institutions, legal proceedings, and individual privacy. Traditional content verification methods are no longer sufficient to address this challenge. This research article presents a layered digital forensic framework combining Error Level Analysis (ELA), metadata examination, cryptographic hash verification using HashCalc, and an AI-based image classifier built on a 7-layer Convolutional Neural Network (CNN). Tested on real and manipulated images from the Kaggle dataset, the framework achieved a classification accuracy of 96.51%. The findings show that no single tool is adequate alone, but when these methods work together, they provide a reliable, multi-dimensional basis for distinguishing authentic media from AI-generated fakes.**

***Keywords*-- Deepfake detection, digital forensics, Error Level Analysis, metadata analysis, hash verification, CNN, AI-generated media, image authentication.**

I. Introduction

We live in a time when seeing something is no longer enough to believe it. A video of a political leader making a damaging statement, a photograph of someone at the wrong place at the wrong time, or an audio clip of a person's voice saying things they never said any of these could be the product of artificial intelligence rather than reality. This is the world that deepfake technology has created.

The term 'deepfake' is a combination of 'deep learning' and 'fake,' and the technology relies on Generative Adversarial Networks (GANs), first described by Ian Goodfellow and colleagues in 2014. In a GAN system, two neural networks compete: a generator creates synthetic images that look convincingly real, while a discriminator tries to spot the fakes. Through iterative training, the generator improves until its outputs are nearly indistinguishable from genuine content.

Open-source frameworks such as TensorFlow, Keras, and PyTorch, along with consumer applications like DeepFaceLab and FaceSwap, have lowered technical barriers significantly. Today, deepfakes are used in political disinformation, non-consensual imagery financial fraud, and fabricated legal evidence.

This research addresses this gap by developing a layered forensic framework that combines Error Level Analysis (ELA), metadata analysis, hash-based integrity verification using HashCalc, and an AI-based image classifier trained on a labelled dataset. Each method contributes a different analytical perspective, and together they provide a more reliable basis for media authentication.

II. Problem Statement

The core problem is straightforward: deepfake images have become sufficiently convincing that neither human observers nor most automated systems can reliably identify them. The challenge operates on multiple levels simultaneously. At the technical level, modern deepfake generation tools including StyleGAN3, Face Swap-GAN, and audio driven talking head synthesis systems produce outputs that replicate subtle facial movements, adjust lighting, synchronise lip movements, and simulate micro-expressions. Visual artefacts that previously exposed deepfakes have largely been eliminated by newer generation methods.

Existing detection systems each have significant weaknesses. Conventional forensic tools like ELA and metadata inspection were not designed for AI-generated content and are unreliable alone. Machine learning detection models generalise poorly to techniques outside their training distribution. Many powerful systems require GPU clusters, making them impractical for real-time or resource-limited deployment. Additionally, deepfake creators actively work to defeat detection through adversarial noise injection and forensic signal suppression.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

What is needed is a comprehensive multi-method approach where the absence of any single signal does not cause the overall analysis to fail. This is the fundamental motivation for the hybrid framework developed here.

III. Research Objectives

This study was guided by three primary objectives:

First, to conduct a thorough review of existing methods for detecting manipulated images and videos, assessing both traditional forensic approaches (ELA, metadata analysis) and recent machine learning-based systems (CNN classifiers, hybrid models).

Second, to test a hybrid detection approach combining four complementary methods: Error Level Analysis, metadata examination, hash-based integrity verification, and AI-based image classification. The rationale is that each method examines a different dimension of an image compression behaviour, provenance data, file-level identity, and visual feature patterns.

Third, to evaluate the combined approach in practice and assess implications for real-world applications, including forensic investigation, journalism, cybersecurity, and legal proceedings.

IV. Literature Review

A. Existing Detection Approaches

Research into deepfake detection has expanded significantly since the problem entered public awareness around 2017–2018. Convolutional Neural Networks have become dominant in machine learning-based detection, with systems such as the DeepFake Detection (DFD) model trained on large, labelled datasets including FaceForensics++, Celeb-DF, and the DeepFake Detection Challenge (DFDC) dataset.

Conventional forensic methods such as ELA and EXIF metadata inspection predate the deepfake era. ELA was originally developed to detect Photoshop-style edits and has since been repurposed for AI-generated content. Tools like Deepware Scanner and Reality Defender combine neural network analysis with metadata inspection for combined authenticity assessments.

B. Known Weaknesses and Gaps

Generalisation remains the most widely discussed limitation. A detection model trained on one generation pipeline will often fail on deepfakes created with a different technique. Dataset bias is closely related many benchmark datasets underrepresent certain demographic groups, raising both technical and ethical concerns in law enforcement applications.

Computational cost is a recurring theme; the most accurate systems require GPU clusters that are impractical for real-time or resource-limited deployment. The adversarial problem is also well-documented: deepfake creators use noise injection, manipulated compression signatures, and forensic signal stripping to defeat detection systems. This literature review confirms the need for a multi-method approach that compensates for the weaknesses of each individual method.

V. Methodology

A. Scope and Data

This study focused on still images rather than video, as still image manipulation is prevalent and more amenable to systematic framework validation. The dataset was sourced from Kaggle (deepfake-and-real-images, compiled by Manjil Karki), containing both authentic photographs and deepfake images created using various AI-based techniques. Images were obtained in original formats to preserve forensic information and labelled into real and manipulated categories.

B. Forensic Tools Used

PhotoForensics (FotoForensics) is a web-based platform providing Error Level Analysis. ELA re-saves an image at a known compression level and examines the differences from the original. Unedited JPEG images show uniform compression artefacts; deepfake regions exhibit different compression characteristics, visible as brightness inconsistencies in the ELA heatmap. The tool also extracts EXIF metadata revealing camera model, lens settings, timestamps, and GPS coordinates.

HashCalc is a Windows application computing cryptographic hash values (MD5, SHA-1, SHA-256, SHA-512, RIPEMD-160, CRC32). Hash values serve as fixed-length fingerprints; even a single changed byte produces a dramatically different hash, enabling reliable detection of modifications that leave no visual trace.

C. AI-Based Image Classifier

A seven-layer Convolutional Neural Network was built using Python and TensorFlow, with Conda for environment management. The architecture comprises convolutional layers, pooling layers, and fully connected (dense) layers, designed to progressively extract visual features. The model was trained on the Kaggle dataset over ten epochs (1,310 total training steps at 131 steps/epoch) using the Adam optimiser. Training framework: TensorFlow, selected for scalability, broad architecture support, and GPU acceleration.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

D. Forensic Workflow

Each image followed a consistent analytical sequence: (1) original acquisition and secure storage without modification; (2) ELA analysis via FotoForensics with examination of the resulting heatmap; (3) metadata extraction and review for anomalies; (4) HashCalc processing to compute cryptographic fingerprints; (5) AI classifier probability assessment. Results from all four analyses were considered collectively before any conclusion was reached.

VI. Results

A. Error Level Analysis Findings

ELA results showed clear and consistent differences between authentic and deepfaked images. Real images displayed largely uniform ELA colouring error levels distributed evenly throughout consistent with unmodified JPEG compression. Deepfaked images showed notable brightness variations, particularly around facial features (eyes, mouth, contours), indicating compression inconsistency from AI reconstruction. Additional ELA observations included unnaturally smooth textures in AI-synthesised regions and sharply defined edges around inserted elements.

Figure 1The FotoForensics submission interface utilized for importing image URLs or local files to initiate Error Level Analysis (ELA) and metadata extraction.

Table I: Error Level Analysis Results

ELA for Real Image	ELA for Deepfaked / Morphed Image
Consistent ELA colours indicating uniform compression and even distribution of error levels.	Uneven error levels: brightness varies across image regions, suggesting tampering.
Uniform lighting, textures, and compression artefacts throughout the image.	High-error areas around AI-generated features (eyes, mouth, facial margins).
Edges and objects appear natural; no unexpected brightness in different regions.	Blurry or patchy textures visible where AI-synthesised pixels show uneven ELA colouration.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

Figure 2FotoForensics interface illustrating the side-by-side comparison of the source image and its generated Error Level Analysis (ELA) compression heatmap.

B. Metadata Analysis Findings

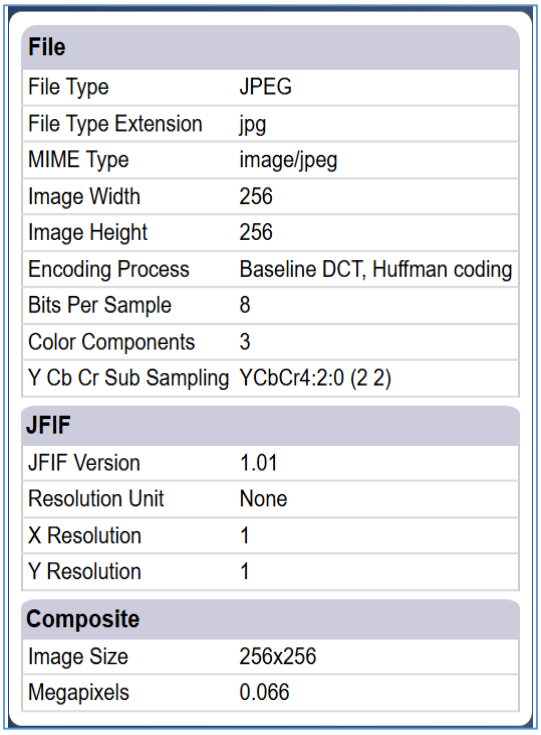
Authentic images consistently showed complete and internally consistent EXIF data: camera model, manufacturer, lens specifications, encoding process, resolution, timestamps, and (where applicable) GPS data.

Deepfaked images most lacked all camera-origin metadata, as deepfake generation tools do not write realistic camera metadata by default. Some deepfaked images also showed resolution values inconsistent with real camera specifications and encoding patterns suggesting multiple processing steps.

TABLE II: METADATA ANALYSIS RESULTS

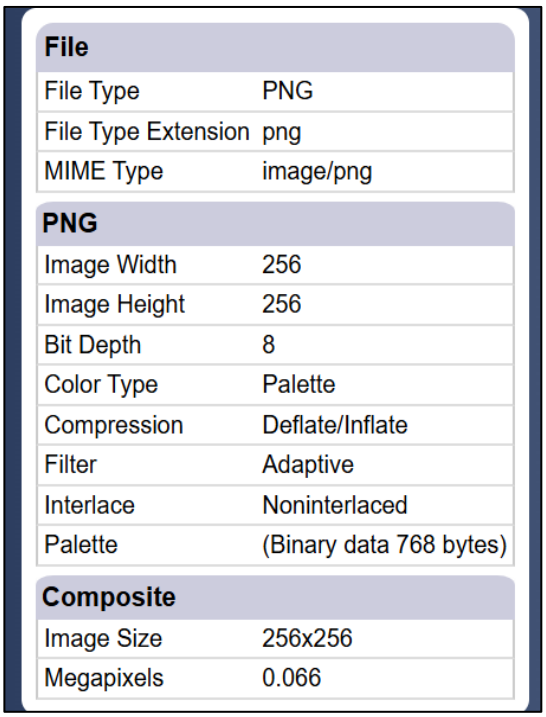
Metadata for Real Image	Metadata for Deepfaked / Morphed Image
Real photographs display authentic camera information (model, manufacturer, lens).	AI-generated images frequently display unknown or absent metadata in place of real camera data.
Timestamps match the expected capture timeline accurately.	Capture date may be absent, inconsistent, or altered to support a fake narrative.
Standard metadata fields (shutter speed, aperture, ISO) are present and consistent.	Resolutions may be incompatible with real camera specifications.
Embedded colour profile complies with camera standards.	Uneven compression patterns in metadata indicate fabricated changes.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)



File	
File Type	JPEG
File Type Extension	jpg
MIME Type	image/jpeg
Image Width	256
Image Height	256
Encoding Process	Baseline DCT, Huffman coding
Bits Per Sample	8
Color Components	3
Y Cb Cr Sub Sampling	YCbCr4:2:0 (2 2)
JFIF	
JFIF Version	1.01
Resolution Unit	None
X Resolution	1
Y Resolution	1
Composite	
Image Size	256x256
Megapixels	0.066

Figure 4 FotoForensics composite panel showing the extracted EXIF metadata profile for an authentic real image sample.



File	
File Type	PNG
File Type Extension	png
MIME Type	image/png
PNG	
Image Width	256
Image Height	256
Bit Depth	8
Color Type	Palette
Compression	Deflate/Inflate
Filter	Adaptive
Interlace	Noninterlaced
Palette	(Binary data 768 bytes)
Composite	
Image Size	256x256
Megapixels	0.066

Figure 3 FotoForensics composite panel showing the extracted, altered or missing metadata profile for a deepfaked image sample.

C. Hash Verification Findings

HashCalc analysis revealed reliable structural differences between authentic and deepfaked files. Real images were predominantly JPEG/JFIF format with a clean single-layer compression structure and no unusual binary string patterns.

Deepfake images were frequently PNG format (preferred by AI generation pipelines) and exhibited fragmented encoding structures with multiple IDAT chunks. String extraction from deepfake binary data revealed random character patterns and unusual symbols absent from genuine photographic files. EXIF metadata was absent or altered in most deepfake cases.

TABLE III: HASHCALC TOOL RESULTS

Feature	Deepfake Image	Real Image
File Format	PNG (IHDR, IDAT, IEND)	JPEG (JFIF)
Encoding Structure	Fragmented IDAT chunks	Standard JPEG quantization
Random String Patterns	Present (e.g., ziiqqjri, NfY5)	Absent
Special Characters	Unusual symbols (!Q{p:zxD*8)	None
EXIF Metadata	Often missing or altered	Typically present
Compression Artifacts	Multiple encoding layers detected	Single compression layer

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

D. AI Classifier Performance

The seven-layer CNN achieved a classification accuracy of 96.51% with a loss value of 0.0921 after training for ten epochs on the Kaggle dataset using the Adam optimiser.

This result demonstrates that a compact neural network architecture can reliably distinguish real from deepfaked images when given sufficient labelled training data.

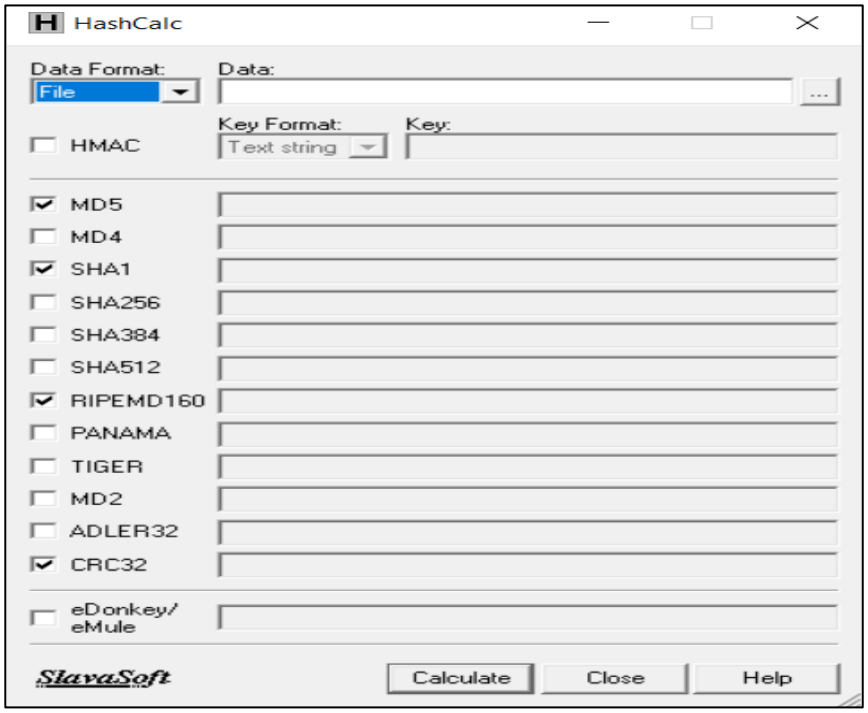


Figure 5 HashCalc home interface displaying selectable cryptographic hash algorithms for file integrity verification.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

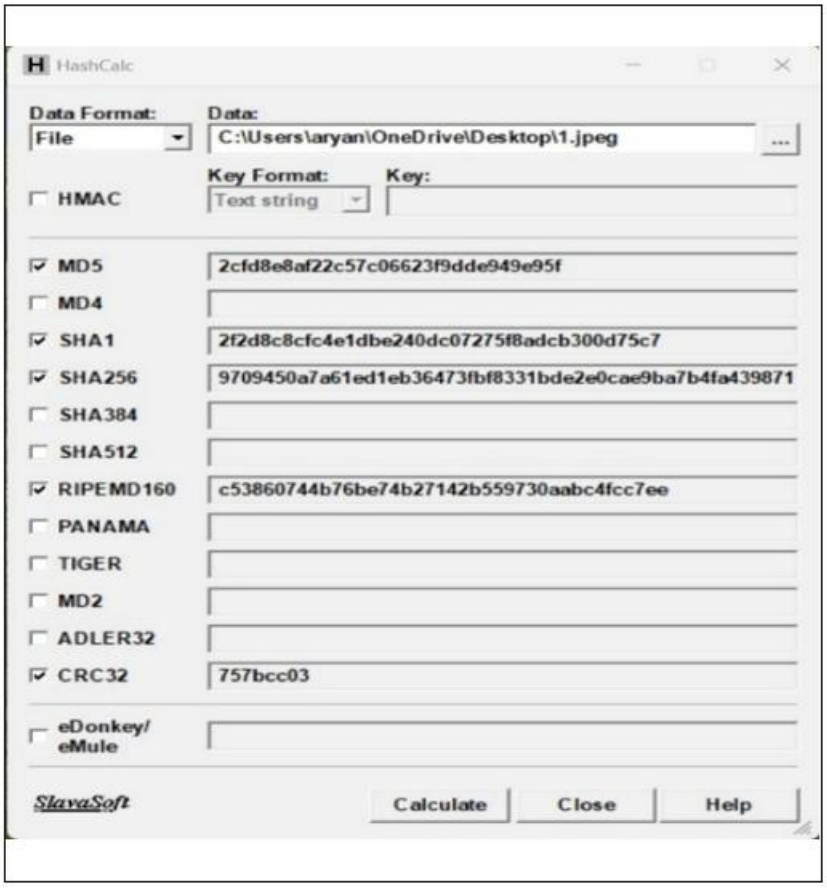


Figure 6 HashCalc interface displaying the calculated cryptographic hash values for an authentic, unmodified real image sample.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

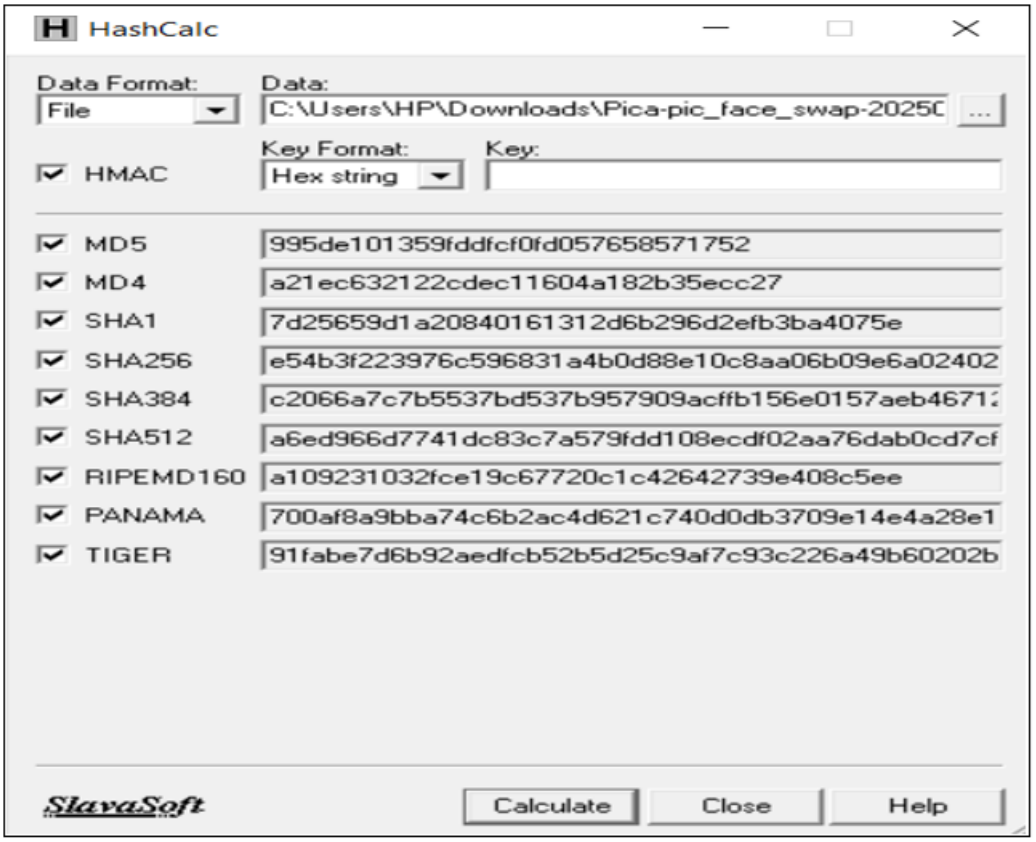


Figure 7 HashCalc interface displaying the calculated cryptographic hash values for a manipulated deepfake image sample.

TABLE IV: AI-BASED CLASSIFIER RESULTS

Component	Tool/Model	Optimizer	Accuracy (%)	Loss	Key Characteristics
Framework	Tensor Flow	Adam	96.51	0.0921	Deep learning, GPU support, scalable
AI Model	7-layer CNN	Adam	96.51	0.0921	Learns deep visual patterns; high dataset performance

VII. DISCUSSION

The results demonstrate that deepfake detection cannot be solved by any single tool or technique. Each of the four methods examined contributes something distinct and irreplaceable.

ELA is strongest at detecting visual manipulations compression inconsistencies arising from AI reconstruction. However, images re-saved multiple times can develop unusual compression patterns even without manipulation, and some modern deepfake methods have reduced ELA signal strength.

Metadata analysis excels at establishing provenance identifying missing camera data, implausible timestamps, or encoding inconsistencies but metadata can be stripped or fabricated. Hash verification is most valuable for integrity assurance, detecting any modification including those leaving no visual trace, though it ideally requires a trusted reference hash. The AI classifier provides statistical pattern recognition at scale, achieving 96.51% accuracy, but this figure reflects performance on the specific training dataset and likely diminishes on out-of-distribution deepfakes.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

Collectively, these four methods create a far more robust framework than any provides individually. When all four analyses point in the same direction, the conclusion carries considerably greater evidentiary weight. In practical terms, this framework offers clear value to forensic investigators, journalists, cybersecurity professionals, and legal practitioners seeking systematic, documentable media authentication.

VIII. Limitations

The most significant scope limitation is that this work focused exclusively on still images. Video deepfakes introduce additional complexity temporal analysis across frames, motion consistency, lip synchronisation, and physiological signals. Extension to video is an important direction for future work.

The study also relied on a specific Kaggle dataset, meaning the trained model reflects that dataset's distribution. Deepfakes produced by techniques not well represented in training data may be detected less reliably. The 96.51% accuracy figure should be understood as dataset-specific, not as a universal guarantee. Metadata and ELA components are also subject to known circumvention strategies including re-saving, metadata fabrication, and forensic signal stripping. Finally, the current framework is not designed for real-time or large-scale automated screening; each image requires individual multi-tool analysis.

IX. Future Scope

The most immediate extension would apply the multi-method framework to video deepfake detection, incorporating temporal analysis for consistency across frames, facial landmark trajectories, physiological plausibility checks, and audio-visual mismatch detection.

Larger and more diverse training datasets are needed to address the generalisation problem. More demographically inclusive data would produce more robust and equitable detection models. Incorporating Explainable AI (XAI) components into the classifier enabling explanation of why an image was classified as fake would significantly enhance usefulness in legal and journalistic settings.

Blockchain-based provenance systems represent a complementary mechanism for authenticating media at point of creation. At the policy level, standardised forensic protocols, certification standards for detection tools, and clear legal frameworks governing synthetic media are urgently needed to ensure consistent, cross-jurisdictional deployment.

X. Conclusion

Deepfake technology represents one of the most practically significant challenges to digital trust in recent years. This research has demonstrated that detecting deepfake images is achievable with reasonable reliability when forensic analysis uses multiple complementary methods. ELA reveals compression inconsistencies from AI facial manipulation. Metadata examination exposes absent authentic camera provenance. Hash-based verification through HashCalc identifies structural anomalies in file encoding. An AI-based CNN classifier, trained on labelled real and manipulated images, achieves 96.51% classification accuracy.

The central finding is that none of these methods is sufficient alone, but that combining them as a layered forensic framework substantially compensates for their individual weaknesses. This matters because deepfake detection has real consequences: incorrect forensic judgements can alter legal outcomes, damage innocent reputations, or allow disinformation to spread unchallenged. Sustained investment in forensic research, diverse training datasets, explainable AI tools, standardised protocols, and appropriate legal frameworks is essential. This study contributes a practically grounded, multi-method forensic approach toward maintaining the ability to authenticate digital media in an era of increasingly convincing synthetic content.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., "Generative adversarial nets," Advances in Neural Information Processing Systems, vol. 27, 2014.

[2] A. Rossler, D. Cozzolino, L. Verdoliva, et al., "FaceForensics++: Learning to detect manipulated facial images," Proc. IEEE ICCV, 2019.

[3] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A large-scale challenging dataset for deepfake forensics," Proc. IEEE/CVF CVPR, 2020.

[4] B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, "The deepfake detection challenge (DFDC) preview dataset," arXiv:1910.08854, 2019.

[5] M. Westerlund, "The emergence of deepfake technology: A review," Technology Innovation Management Review, vol. 9, no. 11, pp. 39-52, 2019.

[6] T. Nagarhalli, "A comprehensive review of deepfake and its detection techniques," Journal of Advances in Information Technology, 2024.

[7] A. Godulla, "Dealing with deepfakes – An interdisciplinary examination of the state of research," Studies in Communication and Media, 2021.

[8] G. Gupta et al., "A comprehensive review of deepfake detection using advanced machine learning and fusion methods," Electronics, vol. 12, no. 1, p. 95, 2023.

[9] S. Mohan, "Review on deepfake detection," International Journal of Innovative Research in Technology, vol. 10, no. 11, 2024.

[10] S. Alanazi and S. Asif, "Exploring deepfake technology: creation, consequences and countermeasures," Journal of Computer Science and Technology Studies, vol. 6, no. 3, pp. 49-60, 2024.

[11] D. Sarkar, "Combatting deep-fakes in India," Indian Law Review, 2024.

[12] P. N. Vasist, "Deepfakes: An integrative review of the literature and an agenda," Telematics and Informatics Reports, vol. 9, 2022.

[13] Z. Geradts, "Interpol review of forensic video analysis, 2019-2022," Forensic Science International: Synergy, 2023.

[14] L. Y. Gong et al., "A contemporary survey on deepfake detection: Datasets, algorithms, and challenges," IEEE Access, vol. 12, 2024.

[15] M. M. Sharma, "Deepfake pornography: Examining the impact on women's digital privacy and consent," International Journal for Multidisciplinary Research, 2024.

[16] B. U. Mahmud, "Deep insights of deepfake technology: A review," International Journal of Advanced Computer Science and Applications, 2020.

[17] S. Srivastava, "The danger of deepfakes, Indian laws and platform responsibility," Journal of Intellectual Property Rights and Media Law, 2025.