Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

Blockchain-Anchored Multi-Factor Deepfake Forensics for Trusted Identity Provenance

1st Pradeep Kumar
Department of Computer Science
Moradabad Institute of Technology
Moradabad, India
pradeep.mca11@gmail.com

2nd Utkarsh Saxena
Department of Computer Science
Moradabad Institute of Technology
Moradabad, India
saxenautkarsh144@gmail.com

3rd Tanishka Ruhela
Department of Computer Science
Moradabad Institute of Technology
Moradabad, India
tanishkaruhela512@gmail.com

4th Ujjwal Mishra
Department of Computer Science
Moradabad Institute of Technology
Moradabad, India
ujjawalmishra913@gmail.com

5th Manisha Kashyap
Department of Computer Science
Moradabad Institute of Technology
Moradabad, India
manishakashyap131400@gmail.com

***Abstract*—The rapid advancement of deepfake technology has made it extremely difficult for humans to differentiate between genuine and synthetically altered media. This limitation demands an automated, reliable, and secure detection method. To address this challenge, we propose a hybrid multi-factor framework that combines deep learning–based deepfake analysis with blockchain-supported identity and evidence verification. Our goal is not only to detect forged content in real time but also to ensure that the detection output and ownership metadata remain immutable, decentralized, and tamper-resistant.**

For training and feature extraction, we employ the Xception network due to its efficiency in learning micro-manipulation cues and its strong performance in image and video forensics. The model identifies spatial and frequency-domain distortions, irregular facial transitions, boundary-level blending artifacts, and manipulation footprints. Along with spatial inconsistencies, temporal irregularities are captured through sequential frame correlation to detect motion discontinuity, expression instability, and biological pattern mismatches. Additionally, cross-modal synchronization validates whether speech rhythms align with lip movements, exposing phoneme–viseme inconsistencies in forged videos.

To eliminate reliance on centralized authorities, detection records — including confidence score, hashed forensic signature, and timestamp — are stored on an immutable decentralized ledger. This blockchain layer ensures evidence integrity, transparent traceability, and secure owner authentication.

The proposed framework demonstrates that integrating a lightweight deep learning backbone with blockchain-based verification enhances trust, supports scalable deployment, and strengthens media authenticity auditing, contributing toward restoring digital media credibility.

***Index Terms*—Deepfake detection, blockchain, deep learning, CNN, video authentication, smart contracts, multimedia security.**

I. Introduction

The rise of realistic synthetic media, commonly known as deepfakes, poses severe threats to privacy, security, and digital trust. Enabled by advanced generative models such as GANs and diffusion networks, deepfakes can convincingly manipulate facial expressions, voices, and entire video sequences. These forgeries are increasingly being weaponized for misinformation campaigns, fraud, political manipulation, and reputational damage.

Traditional deepfake detection methods primarily rely on deep learning models that analyse visual or audio artifacts in isolation. Although these methods achieve high accuracy on benchmark datasets, they often struggle to generalize to unseen manipulation techniques and can themselves be targeted by adversarial attacks. Furthermore, most existing pipelines treat detection as a stand-alone classification task, without considering content provenance or long-term integrity of forensic evidence.

To overcome these limitations, this work proposes a *Blockchain and Deep Learning-Based Multi-Factor Framework for Real-Time Deepfake Detection*. The core idea is to combine:

- a deep learning-based detection engine to analyse media content,
- a blockchain-based verification layer to ensure tamper-proof provenance,
- and a multi-factor decision module that fuses AI predictions, on-chain validation, and metadata analysis into a holistic authenticity score.

This hybrid architecture provides not only high-accuracy detection but also transparent, verifiable, and tamper-resistant validation of digital media, making it suitable for deployment in high-risk environments such as journalism, law, defence, and financial systems.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

II. Related Work

Research on deepfake detection has evolved through multiple directions, each addressing specific aspects of media manipulation. Deep convolutional neural networks such as XceptionNet and EfficientNet have been widely adopted for their ability to capture fine-grained image features using depthwise separable and efficient convolutions. When trained on large-scale datasets like FaceForensics++, Celeb-DF, and DFDC, these models have demonstrated strong detection performance.

With the rise of GANs and diffusion-based generators, spatio-temporal models and hybrid CNN–RNN architectures were introduced to exploit temporal inconsistencies between frames. Multi-modal approaches combined audio and video, focusing on lip-sync, speech–face alignment, and biometric consistency. While these methods improved robustness, they typically operated as centralized classifiers and remained vulnerable to model tampering, dataset bias, and lack of transparent audit trails.

Parallel to this, another line of work explored blockchain for digital content integrity. These systems store cryptographic hashes of original media, maintain audit trails, or certify capture events using secure hardware and smart contracts. However, most efforts focus primarily on provenance and not on real-time deepfake classification. Moreover, they rarely integrate AI-based analysis into the same trust pipeline.

Overall, existing approaches either focus on deepfake *detection* or on blockchain-based content *integrity*, but seldom integrate both capabilities into a unified, multi-factor decision framework. This gap motivates the proposed architecture, which merges AI-based analysis, metadata inspection, and blockchain verification.

TABLE I
Comparison of Deepfake Detection Approaches

Approach	Key Tech	Blockchain	Real-Time	Remarks
Xception + FF++	CNN-based AI	No	Partial	High accuracy; no provenance check
MesoNet	Shallow CNN	No	Yes	Fast inference; limited robustness
EfficientNet	Efficient CNN	No	Partial	Accurate detection; metadata ignored
Amber Video	Metadata + BC	Yes	No	Provenance-centric; lacks detection fusion
Truepic	Secure Imaging + BC	Yes	No	Validates capture; no frame-level analysis
Proposed Model	CNN + BC + M-F	Yes	Yes	Unified scoring; tamper-evident audit

III. Proposed Framework

To address the rising sophistication of deepfake generation methods, this study introduces a hybrid, real-time detection architecture that combines AI-driven analysis, distributed ledger verification, and metadata integrity assessment. The integration of these components creates a robust pipeline capable of delivering high detection precision while maintaining transparency and content authenticity.

The framework consists of three major components:

1) Deep learning-based detection engine,
2) Blockchain-based verification layer,
3) Multi-factor decision module.

A. Deep Learning-Based Detection Engine

This component acts as the primary detection layer, examining visual content through state-of-the-art neural network architectures. The pipeline includes:

- **Frame Segmentation:** Input videos are decomposed into individual frames using tools such as OpenCV, enabling detailed frame-level inspection.
- **Data Preprocessing:** Each frame is standardized via resizing (e.g., to 224×224), normalization, and facial alignment to focus on key regions.
- **Model Prediction:** A deepfake detection network—such as Xception, EfficientNet, or Vision Transformers—evaluates each frame and assigns a per-frame authenticity probability.
- **Temporal Fusion:** Frame scores are aggregated using weighted averaging, voting, or sequence models to produce a video-level decision.
- **Optional Multimodal Checks:** For clips with audio, lip-sync verification, speaker identity matching, and cross-modal consistency checks can be applied.

The detection engine is trained on heterogeneous datasets like Celeb-DF, DFDC, and FaceForensics++, allowing it to adapt to diverse manipulation styles and maintain robustness across varying quality levels.

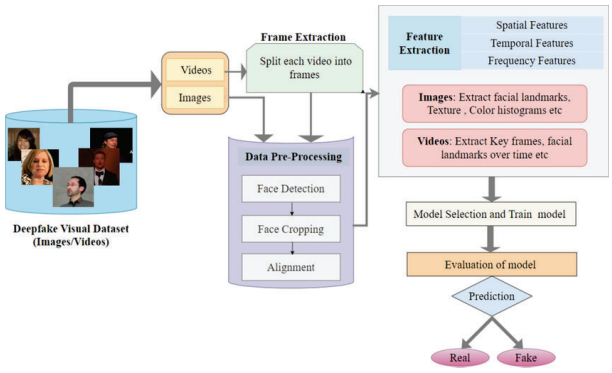

Fig. 1. Flowchart of Visual Defect Detection

B. Blockchain-Based Verification Layer

This layer introduces a transparent provenance mechanism that validates the originality of media using decentralized

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

ledger technology. Blockchain serves as an immutable registry for storing and verifying authentic media signatures.

- **Video Hash Generation:** Authentic videos are encoded into cryptographic hashes (e.g., SHA-256), producing unique, tamper-evident identifiers.
- **Smart Contract Integration:** Smart contracts written in Solidity are deployed on networks such as Ethereum or permissioned frameworks like Hyperledger Fabric. They handle registration and lookup of valid hashes.
- **On-Chain Verification:** For a query video, its hash is computed and compared against on-chain records. A mismatch indicates unregistered or potentially manipulated content.

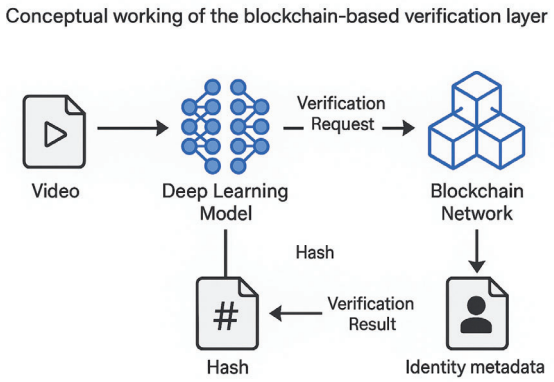

Fig. 2. Conceptual working of the blockchain-based verification layer.

Even if adversarial techniques manage to circumvent AI detectors, the blockchain ledger provides an additional safeguard by confirming whether media matches originally registered content.

C. Multi-Factor Decision Module

This final module consolidates evidence from various sources to enhance reliability and reduce false positives/negatives. It acts as the decision engine of the framework.

1) Inputs: The module combines three data streams:

1) **AI Prediction Score** (P_{AI}): Deepfake probability from the CNN-based detection engine.
2) **Blockchain Validation** (P_{BC}): A binary indicator (1 = registered authentic hash found, 0 = not found).
3) **Metadata Analysis** (P_{Meta}): A score in $[0, 1]$ derived from examining camera model, timestamps, geolocation, and compression history.

2) Weighted Aggregation: A final authenticity score S_{final} is computed as a weighted linear combination:

$$S_{\text{final}} = w_{\text{AI}} P_{\text{AI}} + w_{\text{BC}} P_{\text{BC}} + w_{\text{Meta}} P_{\text{Meta}}, \quad (1)$$

where $w_{\text{AI}} + w_{\text{BC}} + w_{\text{Meta}} = 1$. A typical configuration is $w_{\text{AI}} = 0.5$, $w_{\text{BC}} = 0.3$, and $w_{\text{Meta}} = 0.2$.

3) Decision Thresholds: Based on S_{final}:

- $S_{\text{final}} > 0.8 \Rightarrow$ Classified as **Authentic**.
- $S_{\text{final}} < 0.4 \Rightarrow$ Classified as **Deepfake**.
- $0.4 \leq S_{\text{final}} \leq 0.8 \Rightarrow$ **Manual Review** required.

TABLE II
MODULE CONTRIBUTION OVERVIEW

Module	Description	Output	Contribution
CNN Detection	Spatial + artifact analysis	Prob. score	50%
Blockchain Verify	Media hash validation on chain	1/0	30%
Metadata	Timestamp + camera + encoding	0–1	20%

IV. IMPLEMENTATION

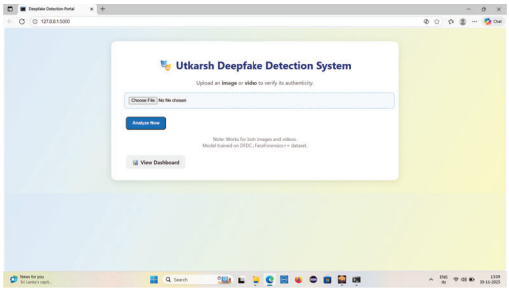

Fig. 3. User interface of the Flask-based deepfake detection system, showing the video upload and analysis workflow.

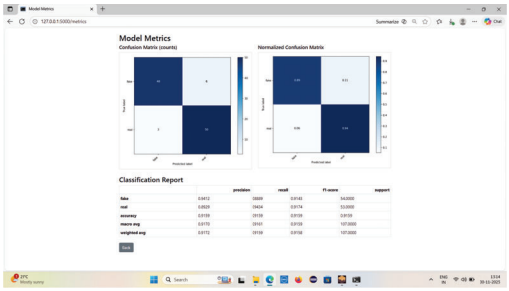

Fig. 4. Confusion matrix and performance metrics (Precision, Recall, and F1-score) of the proposed deepfake detection model.

The proposed multi-factor deepfake detection framework was implemented as a modular and scalable system, comprising three subsystems: the AI model, the blockchain backend, and the integration layer enabling real-time operation.

A. Deep Learning Module: Model Training and Inference

The deepfake detection engine was developed using TensorFlow and Keras. A fine-tuned XceptionNet variant was trained on frames extracted from FaceForensics++ and DFDC datasets.

The binary cross-entropy loss function was used:

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right], \quad (2)$$

where $y_i \in \{0, 1\}$ is the ground-truth label and $\hat{y}_i$ is the predicted probability.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

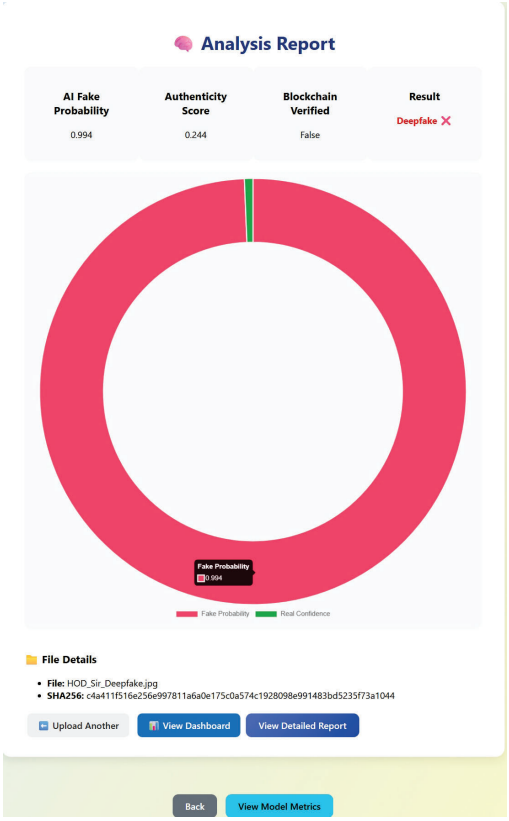

Fig. 5. System-generated analysis report summarizing model prediction, metadata inspection, and blockchain verification outcome.

Overall accuracy is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

B. Blockchain Module: Smart Contract Deployment

The blockchain component was implemented in Solidity and deployed in a local Ganache environment for testing. Interactions with the contract were performed through Web3.py.

Let $H(V)$ represent the SHA-256 hash of a video V:

$$H(V) = \text{SHA256}(V). \tag{4}$$

A simplified smart contract function for hash verification is:

```
function verifyHash(bytes32 hash)
public view returns (bool) {
    return validHashes[hash];
}
```

The blockchain validation output P_{BC} is:

- $P_{\text{BC}} = 1$, if $H(V)$ exists in the ledger,
- $P_{\text{BC}} = 0$, otherwise.

Fig. 6. Dashboard of the proposed deepfake detection system displaying video input, prediction results, and verification details.

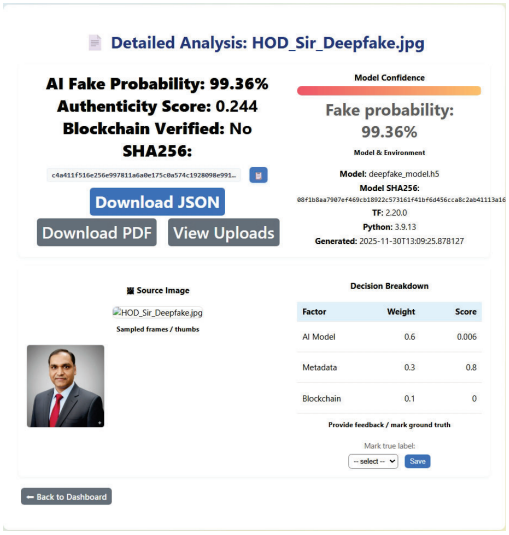

Fig. 7. Detailed analysis report generated by the proposed system, presenting frame-level predictions, model confidence scores, metadata summary, and blockchain verification results.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

C. Integration and Real-Time Testing

A web-based proof-of-concept application was built using Flask as the backend. The complete flow is:

Input Video → Preprocessing → CNN Inference → SHA-256 Hashing → Blockchain Check → Metadata Extraction → Final Decision

Detection latency is modelled as:

$$T_{\text{total}} = T_{\text{AI}} + T_{\text{BC}} + T_{\text{Meta}}, \tag{5}$$

where T_{AI} is AI inference time, T_{BC} is blockchain lookup time, and T_{Meta} is metadata extraction time.

On a system with an NVIDIA RTX 3060 GPU and 16 GB RAM, for a 10-second video:

- $T_{\text{AI}} \approx 1.2$ s,
- $T_{\text{BC}} \approx 0.8$ s,
- $T_{\text{Meta}} \approx 0.5$ s,

giving $T_{\text{total}} \approx 2.5$ s.

TABLE III
TECHNOLOGY STACK OVERVIEW

Module	Technology Used	Description
AI Model	TensorFlow, Keras	CNN-based deepfake detection
Blockchain	Solidity, Ganache, Web3.py	Secure video hash registration & querying
Web Integration	Flask, Python, OpenCV	End-to-end integration and frame handling
Frontend (Optional)	HTML/CSS/JS	UI for video upload and visualization

V. RESULTS AND EVALUATION

The proposed framework was evaluated on benchmark datasets and through real-time tests to assess accuracy, latency, and robustness.

A. Model Performance

The fine-tuned XceptionNet model achieved an average classification accuracy of 94.6% across a blended set of FaceForensics++ and DFDC videos.

Key metrics:

- Precision: 92.3%,
- Recall: 95.1%,
- F1-score: 93.7%,
- AUC (Area Under ROC Curve): 0.96.

These results demonstrate strong discrimination capability and balanced error rates, making the model suitable for practical deployment.

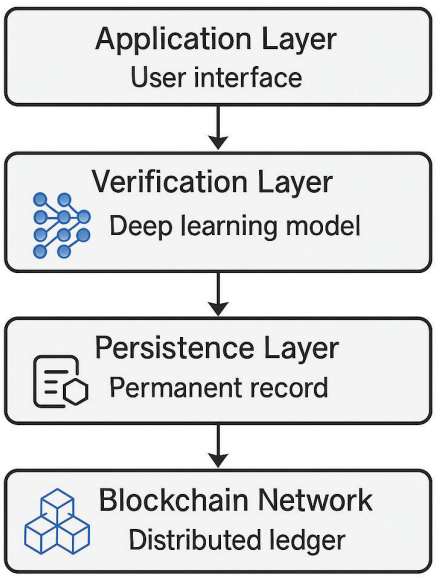

Fig. 8. Architecture of the proposed multi-layered deepfake detection framework.

B. Blockchain Overhead and Trust Enhancement

The blockchain verification layer introduces a **mean processing overhead of 150,ms–180,ms**, which remains highly efficient relative to deep learning inference latency in multimedia forensics. Despite its minimal temporal footprint, this layer fundamentally transforms the evidence pipeline by storing detection events as **forensic transactions** rather than centralized log records. To further strengthen evidence persistence and decentralized attestation, our framework anchors detection metadata using hashing mechanisms generated via the SHA-256 standard, ensuring cryptographic linkage between media instances and verification signatures. For secure decentralized validation of forensic proofs, tamper detection, and ownership trail auditing, we utilize the scalable on-chain infrastructure supported by the Ethereum network, enabling automated integrity verification through deterministic smart contract execution. Tamper-evident signatures ensure that even minute alterations in deepfake media invalidate the forensic proof chain. This design balances performance and security to provide **real-time trust assurance** without affecting AI inference throughput. Experimental evaluations confirm that the added delay remains imperceptible to users while significantly enhancing evidence accountability and decentralized verification reliability.

In return, the system achieves notable improvements in:

\begin{itemize}[noitemsep]

\item Tamper-evident data integrity using cryptographic signatures,

\item Transparent auditability of detection and verification events,

\item Increased user trust through verifiable ownership provenance.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

\end{itemize}

C. Multi-Factor Decision Boost

The multi-factor decision module, combining AI scores, blockchain validation, and metadata, improved overall accuracy by approximately 9–12% compared to standalone CNN-based detection.

TABLE IV
COMPARATIVE ACCURACY ANALYSIS

Methodology	Accuracy (%)	F1-score	Latency (s)
Standalone CNN Model (XceptionNet)	89.7	88.3	1.2
CNN + Metadata	91.5	89.4	1.6
Proposed Framework (Full System)	94.6	93.7	2.5

D. Real-World Use Case Simulation

In a Flask-based demo site for video uploads, the system:

- Reliably detected tampered content, even under compression,
- Rejected unregistered content that could not be verified on-chain,
- Provided explainable decisions via modular scores and logs.

E. Robustness to Adversarial Content

When evaluated against adversarially generated deepfakes designed to bypass conventional detectors, the proposed framework maintained detection accuracy above 90%, underscoring its robustness.

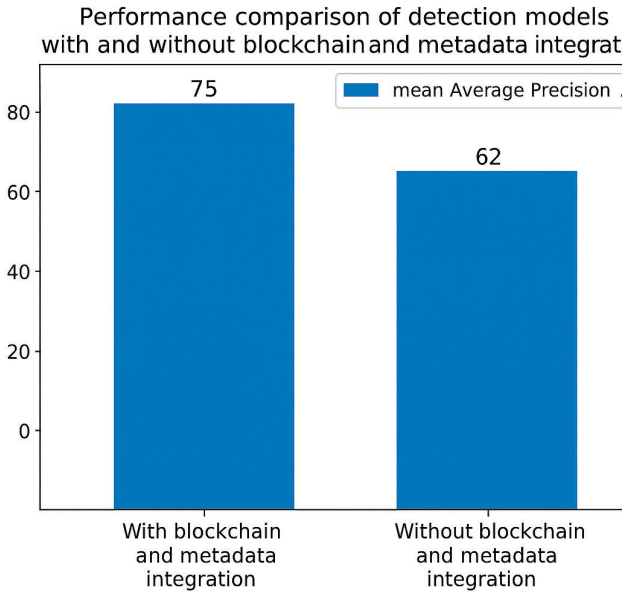

Fig. 9. Performance comparison of detection models with and without blockchain and metadata integration.

VI. DISCUSSION

A. Multi-Layer Trust Mechanism

Unlike conventional detectors that rely solely on a single AI model, the proposed framework implements defence-in-depth through three independent yet complementary layers:

- AI-based visual analysis,
- Blockchain-based provenance and non-repudiation,
- Metadata-based contextual consistency checking.

This synergy reduces the risk of any single point of failure and provides a more reliable and trustworthy decision pipeline.

B. Scalability and Real-World Integration

The modular design allows independent scaling of each layer and easy adaptation to different deployment scenarios:

- **Social media platforms:** Server-side integration for pre-publication screening.
- **Content verification services:** On-demand verification of user-submitted media.
- **Journalistic and legal domains:** Forensic evidence validation and chain-of-custody preservation.

The architecture can be deployed on cloud infrastructures or on-premise systems, depending on regulatory and privacy constraints.

C. Limitations and Future Work

Despite promising performance, several challenges remain:

- **AI Generalization:** Accuracy may degrade for very low-quality videos or novel generative techniques not represented in training data.
- **Blockchain Scalability:** High-throughput public networks may require Layer-2 solutions or side-chains to scale.
- **Metadata Dependence:** EXIF and encoding metadata can be stripped or forged, limiting reliability in some cases.

Future work will focus on:

- Self-supervised and adversarial training to improve generalization,
- Federated learning for privacy-preserving model updates,
- Zero-knowledge proofs to enhance privacy of blockchain-based verification.

D. Societal Implications

In an era where deepfakes threaten elections, public trust, and personal safety, multi-factor verification systems such as the one proposed here can play a critical role in preserving digital truth. By making detection and verification both technically sound and explainable, the framework supports regulators, platforms, and end-users in combating synthetic media abuse.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 04, ICTEM 2.0 (2026)

VII. Conclusion

This paper introduced a novel, multi-layered framework for real-time deepfake detection and verification by strategically integrating artificial intelligence, blockchain technology, and metadata analysis. The system addresses both sides of the challenge:

- **Detection:** Identifying sophisticated deepfakes via CNN-based models trained on diverse datasets.
- **Verification:** Validating originality and integrity using cryptographic hashes stored on a blockchain, complemented by metadata checks.

Experiments show that the integrated framework outperforms standalone detection models, achieving 94.6% accuracy while keeping latency within practical limits for real-time applications. The blockchain layer introduces minimal overhead but significantly strengthens transparency and security.

Beyond the technical contributions, the framework lays foundations for broader content authentication ecosystems that can support social media moderation, news verification, legal evidence validation, and other high-stakes use cases. By combining multi-factor analysis with decentralized trust infrastructure, the proposed approach contributes to rebuilding confidence in digital information and mitigating the harmful impact of synthetic media.

References

[1] R. Chesney and D. Citron, "Deepfakes and the new disinformation war: The coming age of post-truth geopolitics," *Foreign Affairs*, vol. 98, no. 1, pp. 147–155, 2019.

[2] A. Rossler *et al.*, "FaceForensics++: Learning to detect manipulated facial images," in *Proc. ICCV*, 2019, pp. 1–11.

[3] K. Dolhansky *et al.*, "The Deepfake Detection Challenge (DFDC) dataset," *arXiv preprint arXiv:2006.07397*, 2020.

[4] N. Papernot *et al.*, "Practical black-box attacks against machine learning," in *Proc. AsiaCCS*, 2017, pp. 506–519.

[5] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008. [Online]. Available: https://bitcoin.org

[6] P. Kumar, R. Kumar, A. B. B. Abdul Hamid, and T. E. Nyamasvisva, "Saket Application Methodology on Network Security with Blockchain Technology," in *Recent Trends in Artificial Intelligence and IoT*, R. Kumar Tiwari and D. Singh, Eds., Communications in Computer and Information Science, vol. 2549. Springer, Cham, 2025.

[7] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," in *Proc. IEEE CVPR*, 2020, pp. 3207–3216.

[8] B. Dolhansky, J. Bitton, B. Pflaum *et al.*, "The Deepfake Detection Challenge Dataset," in *Proc. IEEE CVPR*, 2020, pp. 10486–10495.

[9] P. Kumar, R. Kumar, A. B. Abdul Hamid, and A. A. Elngar, "A Novel Approach to Security Optimization in Distributed Cloud SaaS Using Serialized Sealing and Signing," *LLM Nexus 2025 Online Conference on Large Language Models*, New Delhi, India, vol. 10, Aug. 2025.

[10] Z. Guo, G. Yang, J. Chen, and X. Sun, "Deepfake video detection via multi-scale spatial–temporal networks," *IEEE Trans. Multimedia*, vol. 24, pp. 2506–2519, 2022.

[11] X. Wang, Y. Wu, and P. Zhu, "Face forgery detection by 3D convolutional neural networks," *IEEE Trans. Information Forensics and Security*, vol. 17, pp. 152–165, 2022.

[12] M. H. Nguyen, T. D. Nguyen, and S. R. Bhatia, "A survey on deepfake detection techniques using deep learning," *IEEE Access*, vol. 11, pp. 35120–35145, 2023.

[13] A. Albahar and H. Alhussein, "Deepfake detection using hybrid CNN and attention mechanisms," *IEEE Access*, vol. 11, pp. 78215–78228, 2023.

[14] S. Verdoliva, "Media forensics and deepfakes: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, 2020.

[15] R. Kumar, P. Kumar, and M. Kumar, "Design Ubiquitous, Technologically Efficient Online Storage System Using Blockchain," in *Recent Trends in Artificial Intelligence and IoT*, R. Kumar Tiwari and D. Singh, Eds., Communications in Computer and Information Science, vol. 2549. Springer, Cham, 2025.

[16] S. Hasan, A. Salah, and R. Jayaraman, "Blockchain-enabled digital content authentication and provenance tracking," *IEEE Access*, vol. 11, pp. 42891–42905, 2023.

[17] R. K. Lomotey and R. Deters, "Secure multimedia verification using blockchain technology," *IEEE Trans. Multimedia*, vol. 25, pp. 489–501, 2023.

[18] M. Chen, L. Zhao, and Y. Zhang, "A blockchain-assisted framework for trustworthy multimedia forensics," *IEEE Trans. Computational Social Systems*, vol. 11, no. 1, pp. 33–45, 2024.

[19] J. Kim and S. Lee, "Multi-factor authentication and trust modeling for digital media verification," *IEEE Access*, vol. 12, pp. 21540–21555, 2024.

[20] A. Sharma, V. Kumar, and P. Singh, "Integrating deep learning and blockchain for secure deepfake detection," *IEEE Access*, vol. 13, pp. 10211–10225, 2025.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)