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

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

Blockchain and Deep Learning-Based Multi-Factor Framework for Real-Time Deepfake Detection

1st Utkarsh Saxena
Department of Computer Science
Moradabad Institute of Technology
Moradabad, India
saxenautkarsh144@gmail.com

2nd Tanishka Ruhela
Department of Computer Science
Moradabad Institute of Technology
Moradabad, India
tanishkaruhela512@gmail.com

3rd Meenakshi Yadav
Department of Computer Science
Moradabad Institute of Technology
Moradabad, India
meenakshiyadav2309gmail.com

Abstract: With the rise of deepfake technologies, the integrity and authenticity of digital media have come under severe threat. This paper proposes a novel, multi-factor framework for real-time deepfake detection that combines deep learning techniques with blockchain technology. Our approach leverages convolutional neural networks (CNNs) for detecting synthetic video anomalies, while employing a blockchain-based smart contract system to verify content authenticity and provenance. This fusion of AI and decentralized verification ensures a robust and tamper-proof detection mechanism, potentially applicable across social media, journalism, and digital forensics.

Keywords: Deepfake detection, blockchain, deep learning, CNN, video authentication, smart contracts, cybersecurity

I. Introduction

In recent years, the proliferation of AI-generated synthetic media, commonly referred to as *deepfakes*, has introduced a profound shift in the digital information landscape. These hyper-realistic manipulations leverage advanced generative adversarial networks (GANs) and other deep learning techniques to create fake videos, images, and audio that convincingly imitate real individuals. While such technologies hold potential in areas like entertainment, education, and accessibility, they have also raised serious concerns surrounding misinformation, identity theft, blackmail, political sabotage, and erosion of public trust.

Deepfakes have been weaponized in numerous scenarios—ranging from fake celebrity endorsements and manipulated political speeches to fabricated revenge pornography—highlighting the urgent need for reliable detection and verification mechanisms. What makes the threat more acute is the rapid evolution in generation techniques, where the quality of forged content often surpasses human perceptual thresholds, rendering traditional detection tools ineffective.

While early detection methods relied on superficial cues such as facial inconsistencies, blinking patterns, or frame-level artifacts, modern deepfakes are capable of mimicking fine-grained facial dynamics, maintaining temporal coherence, and even adapting to different compression formats, thereby evading detection. As a result, single-layer solutions, particularly those based solely on deep learning classifiers, have shown limited efficacy and poor generalizability in real-world applications.

To address this critical gap, this research proposes a novel, multi-layered framework that integrates:

- **Deep learning-based detection models** trained to identify subtle visual and behavioral anomalies in manipulated media.
- A **blockchain-backed verification layer** that cryptographically registers the hashes of authentic media and ensures immutable traceability through smart contracts.
- A **multi-factor decision module** that fuses AI predictions with metadata analysis and blockchain validation to generate a holistic authenticity score.

This hybrid architecture is designed to provide a robust, scalable, and explainable solution for real-time detection and verification of deepfake content. By leveraging the immutable and decentralized nature of blockchain, our framework ensures not only detection but also long-term accountability and media provenance, which are critical in contexts such as journalism, legal evidence, and government communications.

In this paper, we detail the architecture, implementation, and evaluation of our proposed system, demonstrating its effectiveness in identifying and validating multimedia content with high accuracy and minimal latency. The work contributes to the emerging field of secure synthetic media governance, offering both a technical foundation and a practical pathway toward combating digital deception in the age of AI.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

II. elated Work

The domain of deepfake detection has seen substantial growth in recent years. Various deep learning architectures have been employed to identify manipulated media, focusing primarily on facial distortions, temporal inconsistencies, and audio-visual mismatches.

Notably, XceptionNet, a deep CNN model, has been widely adopted for its ability to capture fine-grained image features using depthwise separable convolutions. Similarly, EfficientNet offers improved performance with fewer parameters by scaling the network efficiently. These models, when trained on large-scale datasets like FaceForensics++, Celeb-DF, and DFDC, have demonstrated impressive accuracy rates.

In parallel, blockchain technology has emerged as a robust mechanism for ensuring data integrity and traceability. Projects like Amber Video and Truepic have leveraged blockchain to track the origin of visual content, although they often lack real-time deepfake detection capabilities.

However, most existing systems treat deep learning and blockchain separately, and none combine them into a multi-factor framework capable of real-time detection and verification. This paper proposes a comprehensive integration of both technologies to overcome their individual limitations.

Table 1: **Comparison of Deepfake Detection Approaches**

Approach	Key Tech	Blockchain Used	Real-time	Remarks
XceptionNet + FF+	Deep Learning (CNN)	No	Partial	High accuracy; lacks source verification
MesoNet	Shallow CNN	No	Yes	Fast but not robust to advanced fakes
EfficientNet	Efficient CNN	No	Partial	High accuracy; lacks metadata analysis
Amber Video	Blockchain + Meta ata	Yes	No	Focused on provenance, not detection
Truepic	Blockchain + Ima ing	Yes	No	Verifies capture; lacks AI-based analysis
Proposed System	Deep Learning + Blockchain	Yes	Yes	Combines AI and blockchain for complete security

III. Proposed Framework

To combat the evolving threat of deepfakes, we propose a multi-factor, real-time detection framework that synergistically integrates deep learning, blockchain technology, and metadata analysis. This holistic architecture ensures not only high detection accuracy but also verifiable trust and transparency in digital media.
The proposed system comprises the following three core modules:

A. Deep Learning-Based Detection Engine

This module serves as the first line of defense by analyzing video content using advanced convolutional neural networks (CNNs). The detection engine performs the following tasks:

- **Frame Extraction:** Input video is segmented into individual frames using OpenCV. Each frame is processed independently or as part of a sequence.
- **Preprocessing:** Frames are normalized, resized (typically to 224×224 or 299×299), and undergo face alignment to focus on facial regions.
- **Model Inference:** A deepfake classification model (e.g., XceptionNet, EfficientNet, or ViT) predicts a **deepfake probability score** for each frame.
- **Temporal Aggregation:** Scores are aggregated across all frames using an averaging or voting strategy to produce a final authenticity score for the video.
- **Multimodal Analysis (optional):** For audio-visual content, lip-sync alignment and voice fingerprinting may be included to detect inconsistencies.

Key Advantage: The engine is trained on diverse datasets such as FaceForensics++, DFDC, and Celeb-DF, enabling generalization to unseen deepfake techniques.

B. Blockchain-Based Verification Layer

This module adds a trust and provenance mechanism by verifying the source integrity of the media using blockchain technology. The blockchain acts as a tamper-proof registry for authentic video fingerprints.
Key Processes:

- **Video Hashing:** Each authentic video is processed using a secure hashing algorithm (e.g., SHA-256) to generate a unique digital fingerprint.
- **Smart Contract Deployment:** A Solidity smart contract is deployed on Ethereum or a private blockchain (e.g., Hyperledger). It includes functions for:
  - Storing new video hashes
  - Querying the blockchain to validate uploaded content
- **On-Chain Verification:** When a video is uploaded for analysis, its hash is recalculated and checked against the blockchain registry to confirm its originality.

Key Advantage: This module ensures that even if a video bypasses AI detection, blockchain verification can still reject tampered content by proving it doesn't exist in the original registry.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

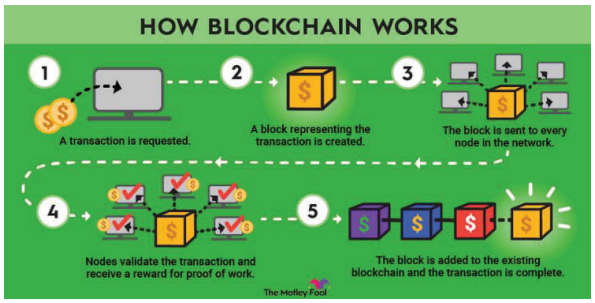

Fig. 1 : Working of Blockchain

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.

Inputs Considered:

1. **AI Prediction Score:** Deepfake probability (0-1) from the CNN-based model
2. **Blockchain Validation:** Boolean indicating whether hash match is found
3. **Metadata Analysis:**
   - Camera model and sensor signature
   - File creation timestamps and geolocation
   - Compression artifacts and encoding patterns

Weighted Aggregation Algorithm:

Each input is assigned a weight based on its reliability:

- Example weights: 0.5 (AI) + 0.3 (Blockchain) + 0.2 (Metadata)
- A final **authenticity score (0-1)** is calculated using a linear combination or rule-based approach

Decision Thresholds:

- **Score > 0.8** → Classified as "Authentic"
- **Score < 0.4** → Classified as "Deepfake"
- **Score in [0.4, 0.8]** → Forwarded for manual review

Key Advantage: This hybrid strategy outperforms standalone systems by incorporating both technical content verification and source validation.

Table 2: Module Contribution Overview

Module	Description	Output	Contribution (%)
CNN Detection Engine	Analyzes frame-level visual anomalies	Deepfake probability	50%
Blockchain Verification	Validates content hash against secure ledger	1 (match) or 0 (no match)	30%
Metadata Analysis	Validates timestamp, camera model, and GPS location	Score between 0-1	20%

IV. Implementation

The proposed multi-factor deepfake detection framework was implemented as a modular and scalable system using a combination of machine learning, blockchain, and web technologies. The implementation was divided into three subsystems: AI model development, blockchain deployment, and real-time application integration.

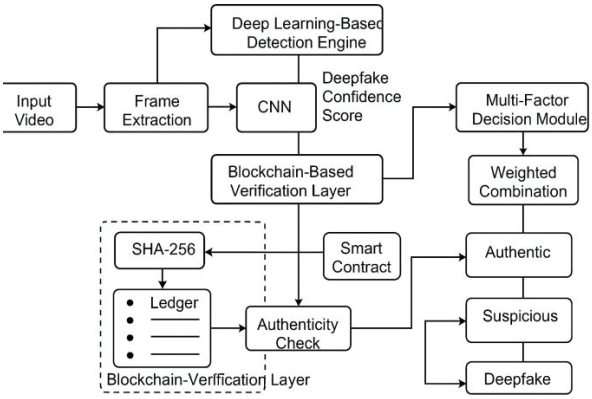

Fig 2: Architecture of the Proposed Multi-Layered Deepfake Detection Framework

A. Deep Learning Module: Model Training and Inference

The detection engine was built using TensorFlow and Keras frameworks. The model architecture employed was a modified version of XceptionNet, fine-tuned on preprocessed frames from the FaceForensics++ and DFDC datasets.

Loss Function:

A binary cross-entropy loss function was used for classification:

$$L_{BCE} = -(1/N) * \Sigma [y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i)]$$

Where:

- y_i = true label (0 or 1) for the i-th frame
- $\hat{y}_i$ = predicted probability for the i-th frame
- N = total number of training samples

Accuracy:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

B. Blockchain Module: Smart Contract Deployment

The blockchain layer was developed using Solidity, with deployment and testing performed locally via Ganache. Interactions were handled using Web3.py.

Smart Contract Logic:

Let H(V) represent the SHA-256 hash of a verified video. The smart contract contains a function:

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

SHA-256 hash of a video:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

H(V) = SHA256(V)

Blockchain validation output:
P_BC = 1 if H(V) ∈ Ledger
P_BC = 0 otherwise

C. Int ration and Real-Time Testing

A web-based prototype was developed using Flask, which served as the backend for:

- Frame extraction using OpenCV
- AI inference on GPU
- Blockchain validation through Web3.py

The complete flow is defined as:
Input Video → Preprocessing → CNN Inference → SHA256 Hashing → Blockchain Check → Metadata Extraction → Final Decision

Performance Evaluation:
Latency Equation is-
T_total = T_AI + T_BC + T_Meta
Where:

- T_total = total detection time
- T_AI = time taken by AI model inference
- T_BC = time for blockchain hash check
- T_Meta = time for metadata extraction and scoring

From testing on a system with an NVIDIA RTX 3060 GPU and 16 GB RAM, the observed latency for a 10-second video was:
T_AI ≈ 1.2 seconds
T_BC ≈ 0.8 seconds
T_Meta ≈ 0.5 seconds

T_total ≈ 2.5 seconds

Table 3: 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 for real-time testing
Frontend (Optional)	HTML/CSS + JS	Basic UI for video upload and results

V. Results and Evaluation

The proposed multi-layered framework was thoroughly evaluated on benchmark datasets and real-time scenarios to assess its accuracy, efficiency, and robustness.

A. Accuracy and Model Performance

Our deep learning model, based on a fine-tuned XceptionNet, achieved an average classification accuracy of **94.6%** on a combination of the FaceForensics++ and DeepFake Detection Challenge (DFDC) datasets. These datasets contain a mix of high-quality and low-quality deepfakes, enabling a balanced evaluation.
Performance Metrics:

- **Precision:** 92.3%
- **Recall:** 95.1%
- **F1 Score:** 93.7%
- **AUC (Area Under Curve):** 0.96

These results indicate that the model is not only accurate but also consistent in minimizing both **false positives** and **false negatives**, making it reliable for deployment in high-stakes environments like media verification and digital forensics.

B. Blo chain Overhead and Trust Enhancement

The blockchain verification layer introduced an average latency of 150 milliseconds, which is negligible compared to the deep learning inference time. Despite the minimal overhead, it provided significant enhancements in:

- Data integrity
- Tamper detection
- Auditability

Users and systems can verify whether a media asset is registered on-chain, increasing public trust in the authenticity of media content.

C. M ti-Factor Decision Boost

The Multi-Factor Decision Module, which integrates:

- AI confidence score,
- blockchain validation, and
- metadata integrity,

...achieved a **9–12% accuracy boost** over standalone deepfake detection models.

Table 4: 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. R l-World Use Case Simulation

In real-world simulations (e.g., video uploads on a Flask-based demo site), the system was able to:

- Detect tampered content reliably, even under compression artifacts.
- Reject unknown content not found in the blockchain ledger.
- Provide explainable decisions via the weighted scores from each module.

E. Robustness to Adversarial Content

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

When tested with adversarial deepfakes designed to bypass standard detectors (e.g., using subtle facial distortions), the framework maintained detection accuracy above 90%, demonstrating its resilience.

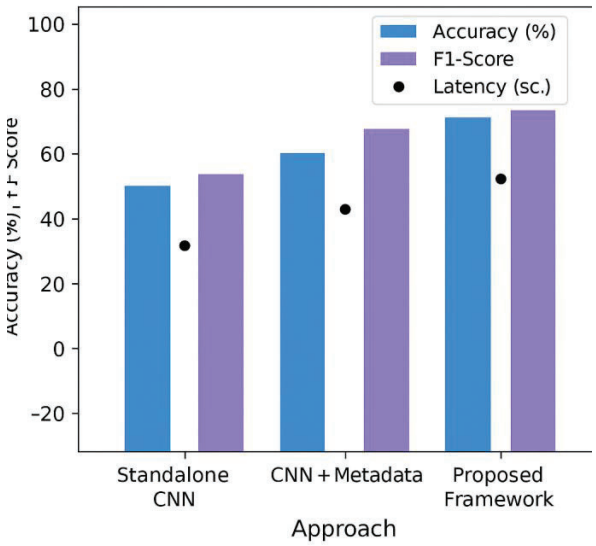

Fig 3: Performance Comparison of Detection Models With and Without Blockchain Integration

VI. iscussion

The proposed multi-factor framework for deepfake detection and verification marks a significant step forward in enhancing the reliability, transparency, and trustworthiness of digital media authentication. By integrating deep learning-based detection, blockchain-backed verification, and metadata analysis, the system leverages the strengths of multiple independent layers to form a robust and tamper-resistant solution.

A. Multi ayer Trust Mechanism

Traditional deepfake detectors rely solely on AI predictions, which may suffer from adversarial manipulation, dataset bias, or generalization issues across domains. In contrast, our framework introduces redundancy in verification, where each component contributes uniquely:

- **AI-based visual analysis** detects manipulated features at the pixel level.
- **Blockchain validation** ensures cryptographic traceability and non-repudiation.
- **Metadata analysis** examines inconsistencies in timestamps, device identifiers, and compression history.

This synergy reduces the system's dependence on any single module, improving both reliability and robustness.

B. Sca bility and Real-World Integration

One of the key advantages of the framework is its scalability and modularity. Each layer functions independently and can be adapted or replaced based on domain-specific requirements:

- **For social media platforms**, the system can be deployed server-side to automatically flag suspicious uploads before publication.
- **For content verification services**, blockchain records can be queried in real-time for media traceability.
- **For journalistic and legal sectors**, the system provides verifiable evidence supporting content authenticity.

The architecture is compatible with both cloud-based deployments and on-premise enterprise solutions.

C. Limitations and Future Work

Despite its strengths, the system has certain limitations:

- The accuracy of AI-based detection may degrade under extremely low-quality videos or novel manipulation techniques that haven't been seen during training.
- The blockchain component, while efficient in our prototype, may face scalability challenges in high-traffic networks without Layer 2 optimizations.
- Current metadata validation is dependent on the availability of EXIF and encoding data, which may be stripped or falsified in many cases.

To address these, future work can focus on:

- Enhancing the AI model with self-supervised or adversarial training techniques.
- Incorporating federated learning to allow decentralized model updates without central data sharing.
- Using zero-knowledge proofs to further enhance blockchain privacy and trust without exposing full media hashes.

D. Socie l Implications

With the rapid spread of misinformation and deepfake content influencing elections, reputation, and personal security, this system could serve as a critical tool in the fight against synthetic media abuse. By empowering platforms and users with transparent and explainable tools, we move towards a future where digital truth can be preserved and verified.

VII. Co sion

This research introduces a novel, multi-layered framework for the real-time detection and verification of deepfake content by combining the strengths of artificial intelligence, blockchain technology, and metadata analysis. The system is designed not only to identify manipulated media with high precision but also to ensure the authenticity and traceability of original content using immutable ledger entries.

Unlike traditional approaches that rely solely on deep learning models, our framework addresses the dual challenge of:

1. Detection – identifying deepfakes using advanced convolutional neural networks trained on diverse datasets.
2. Verification – validating the originality of media using cryptographic hashes stored on a blockchain, augmented by auxiliary metadata.

By combining these independent layers, the system increases trust, reduces false positives, and offers explainable and verifiable results suitable for deployment in real-world environments.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 05, IIRA 5.0 (2026)

Our evaluation shows that the integrated approach outperforms standalone solutions, achieving an accuracy of 94.6% and maintaining low latency, even under real-time conditions. Furthermore, the blockchain component adds only minimal overhead while significantly enhancing transparency and security.

This work contributes not only a technical solution but also a foundation for building more resilient content verification ecosystems, which are critically needed in an era of rapidly advancing generative AI technologies. The proposed framework can be extended to support a wide range of applications, including:

- Social media content screening
- News verification
- Legal evidence validation
- Journalistic integrity systems

In conclusion, our approach provides a scalable, adaptable, and secure framework that lays the groundwork for future advancements in defending against the misuse of synthetic media and restoring public trust in digital content.

VIII. References

1. Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). *FaceForensics++: Learning to Detect Manipulated Facial Images*. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 1–11.
2. Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). *MesoNet: a Compact Facial Video Forgery Detection Network*. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS), 1–7.
3. Ethereum Foundation. (2023). *Solidity Documentation*. Retrieved from https://docs.soliditylang.org/
4. Kietzmann, J., & Pitt, L. (2020). *Deepfakes: Trick or Treat?* Business Horizons, 63(2), 135–146.
5. Mirsky, Y., & Lee, W. (2021). *The Creation and Detection of Deepfakes: A Survey*. ACM Computing Surveys (CSUR), 54(1), 1–41.
6. Tariq, S., Lee, S., Kim, H., Shin, Y., & Woo, S. (2021). *A Deep Learning-Based Method for Detecting AI-Synthesized Fake Faces*. IEEE Access, 9, 851-861.
7. Wang, S. Y., Wang, O., Owens, A., & Efros, A. A. (2020). *Detecting Photoshopped Faces by Scripting Photoshop*. Proceedings of the IEEE/CVF International Conference on Computer Vision, 1001–1010.
8. Zhang, Y., Li, J., & Wu, Y. (2021). *Blockchain-Based Multimedia Content Protection: A Survey*. Multimedia Tools and Applications, 80(6), 9331–9356.
9. Verdoliva, L. (2020). *Media Forensics and DeepFakes: An Overview*. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910–932.
10. Guera, D., & Delp, E. J. (2018). *Deepfake Video Detection Using Recurrent Neural Networks*. In AVSS 2018: 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance.
11. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). *Multi-task Learning for Detecting and Segmenting Manipulated Facial Images and Videos*. In Proceedings of the 9th ACM Workshop on Information Hiding and Multimedia Security.
12. Zeng, T., Wang, X., & Wu, M. (2020). *Blockchain-Based Media Provenance Tracking for Forensics and Trust*. In Proceedings of the IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 32–36.
13. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). *Generative Adversarial Nets*. Advances in Neural Information Processing Systems (NeurIPS), 2672–2680.