© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1719388

Deepfake Detection Using Hybrid CNN-LSTM Architecture with Blockchain-Based Media Authentication

T PRIYA[1], S SANDHIYA[2]
[1]*Assistant Professor, Computer Science and Engineering, AVS College of Technology, Attur Main Rd, Chinnagoundapuram, Salem.*
[2]*Student (M.E Computer Science and Engineering), AVS College of Technology, Attur Main Rd, Chinnagoundapuram, Salem.*

Abstract- ***The rapid advancement of artificial intelligence has enabled the creation of highly realistic deepfake images and videos, posing significant threats to digital media authenticity, cybersecurity, and public trust. Existing deepfake detection techniques often rely on either spatial or temporal information, resulting in limited robustness against sophisticated manipulations. This paper proposes a hybrid CNN-LSTM framework integrated with blockchain-based media authentication for reliable deepfake detection and secure media verification. The Convolutional Neural Network (CNN) extracts discriminative spatial features from facial frames, while the Long Short-Term Memory (LSTM) network captures temporal inconsistencies across video sequences. To ensure media integrity, SHA-256 hashing is combined with blockchain technology to provide tamper-proof authentication and provenance tracking. The proposed model is evaluated using FaceForensics++, DFDC, and Celeb-DF datasets. Experimental results demonstrate high detection accuracy, improved robustness, and secure verification compared with existing approaches. The proposed framework offers an efficient solution for social media platforms, journalism, digital forensics, and cybersecurity applications.***

Keywords—Deepfake Detection, CNN, LSTM, Blockchain, Artificial Intelligence, Computer Vision, Digital Forensics.

I. INTRODUCTION

Deepfake technology has become increasingly accessible due to advances in deep learning and generative models such as Generative Adversarial Networks (GANs). These technologies can generate highly convincing fake videos and images that are difficult to distinguish from authentic media. Although deepfakes have useful applications in entertainment and education, they also introduce serious risks including misinformation, identity theft, financial fraud, and political manipulation.

Traditional deepfake detection techniques mainly focus on spatial image features or handcrafted artifacts. However, modern deepfake generation methods successfully remove many visible artifacts, making single-frame detection less effective. Furthermore, existing methods rarely verify whether original media has been altered after detection.

To overcome these limitations, this work proposes a Hybrid CNN-LSTM model that combines spatial and temporal learning with blockchain-based authentication. The CNN extracts facial features while the LSTM analyzes temporal inconsistencies between frames. Blockchain securely stores media hashes, enabling immutable verification of genuine content.

II. RELATED WORK

Several deep learning approaches have been proposed for deepfake detection. CNN-based methods such as XceptionNet and EfficientNet achieve high spatial detection accuracy but often fail when temporal consistency is required. Recurrent Neural Networks and LSTM-based approaches improve sequential analysis by modeling frame dependencies in videos.

Recent research also explores transformer-based detectors and frequency-domain analysis using FFT

and DCT features. Although these methods improve detection performance, they generally lack mechanisms for authenticating original media. Blockchain technology has recently emerged as a secure solution for media provenance by storing cryptographic hashes in decentralized ledgers. However, limited research combines hybrid deep learning with blockchain authentication into a unified framework.

III. PROPOSED METHODOLOGY

The proposed framework consists of four major stages:

1. Video preprocessing and frame extraction.
2. CNN-based spatial feature extraction.
3. LSTM-based temporal sequence analysis.
4. Blockchain-based media authentication.

Initially, videos are converted into frames and facial regions are detected using MTCNN. Images are normalized and enhanced before being processed by the CNN backbone. The extracted feature vectors are passed into the LSTM network to analyze temporal relationships between consecutive frames. The final classifier determines whether the video is genuine or manipulated.

For authentication, the SHA-256 hash of every verified media file is generated and stored on a blockchain network using smart contracts. During verification, the computed hash is compared with the blockchain record. Any modification changes the hash, immediately identifying tampered content.

IV. EXPERIMENTAL SETUP

The proposed model was trained and evaluated using publicly available datasets including FaceForensics++, DeepFake Detection Challenge (DFDC), and Celeb-DF.

The implementation was carried out using Python, TensorFlow, OpenCV, and Keras. Model training was performed using the Adam optimizer with binary cross-entropy loss. Performance was evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics.

The blockchain module was implemented using Ethereum smart contracts and SHA-256 cryptographic hashing to provide secure and immutable media authentication.

V. RESULTS AND DISCUSSION

The proposed Hybrid CNN-LSTM model achieved superior performance compared to existing CNN-only approaches. Experimental evaluation demonstrated an overall detection accuracy of approximately 97%, with improvements in Precision, Recall, and F1-Score.

The CNN effectively extracted spatial forgery artifacts while the LSTM captured temporal inconsistencies across video frames. The integration of blockchain introduced secure verification without significantly affecting computational performance.

Compared with existing methods, the proposed framework provides higher robustness against compression, face-swapping, and video manipulation while ensuring trustworthy media authentication.

VI. CONCLUSION

This paper presented a Hybrid CNN-LSTM architecture integrated with blockchain technology for deepfake detection and secure media authentication. The proposed framework successfully combines spatial feature extraction, temporal sequence learning, and decentralized verification to improve both detection accuracy and media integrity. Experimental results demonstrate that the proposed system outperforms several existing methods while providing reliable blockchain-based authentication.
Future work includes supporting diffusion-model deepfakes, lightweight deployment for mobile devices, transformer-based feature extraction, and real-time implementation for social media platforms.

REFERENCES

[1] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in Proc.

IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1–11.

[2] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[3] B. Dolhansky et al., "The DeepFake Detection Challenge (DFDC) Dataset," arXiv preprint arXiv:2006.07397, 2020.

[4] D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in Proc. IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 2018.

[5] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[6] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. International Conference on Machine Learning (ICML), 2019.

[7] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multi-task Cascaded Convolutional Networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016.

[8] T. Karras, M. Aittala, J. Hellsten, et al., "Alias-Free Generative Adversarial Networks (StyleGAN3)," Advances in Neural Information Processing Systems (NeurIPS), 2021.

[9] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[10] M. Hasan and K. Salah, "Combating Deepfake Videos Using Blockchain and Smart Contracts," IEEE Access, vol. 7, pp. 41596–41606, 2019.

[11] Q. Wang, W. Luo, and Y. Qiu, "Detecting GAN-Generated Images by Learning Their Spectral Representations," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020.

[12] J. Frank, T. Eisenhofer, L. Schönherr, et al., "Leveraging Frequency Analysis for Deep Fake Image Recognition," in Proc. International Conference on Machine Learning (ICML), 2020.

[13] Z. Zhao, S. Wang, and Y. Li, "Multi-Attentional Deepfake Detection," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[14] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," 2016.

[15] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.