Indiana Journal of Multidisciplinary Research
Abbreviated Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.21610284

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Review Articles

Volume-06|Issue-04|2026

Deepfake Video Detection: A Survey of Techniques, Limitations, and Future Perspectives

Bhavya S[1] & Swetha Rani L[2]

[1]Assistant Professor, Department of ECE, AMCEC Bangalore, India.
[2]Associate Professor, Department of ECE, AMCEC Bangalore, India.

Article History
Received: 05.05.2026
Accepted: 22.06.2026
Published: 25.07.2026

Citation
Bhavya, S. & Rani, S. L. (2026) Deepfake Video Detection: A Survey of Techniques, Limitations, and Future Perspectives. *Indiana Journal of Multidisciplinary Research, 6*(4), 388-392.

Abstract: The quick growth of Artificial Intelligence (AI) and Deep Learning (DL) advancements in technology has resulted in the emergence of deepfake videos, which are altered content that can convincingly imitate real people and their actions. Although these technologies show promise for creativity and business, they raise significant concerns about misinformation, identity theft, and loss of public trust. This literature survey offers a detailed review of recent methods and approaches for detecting deepfake videos. There are many techniques available for detection, which include traditional methods, deep learning-based methods, and hybrid models.

Keywords: Deep Learning, Artificial Intelligence

INTRODUCTION

Deepfakes are a class of AI-generated media—including videos, audio, and images—that are manipulated to appear authentic [1]. Enabled by advancements in deep learning, these synthetic contents are often used to alter or superimpose identities, frequently targeting public figures such as politicians and celebrities. While deepfakes can be leveraged for creative or entertainment purposes, their malicious use poses severe risks, including misleading audiences, inciting political or social conflicts, manipulating elections, and even influencing financial markets.

At the core of deepfake generation lies the Generative Adversarial Network (GAN), a deep learning framework composed of two competing neural networks: a generator (G), which synthesizes realistic media, and a discriminator (D), which evaluates its authenticity. This adversarial training process is expressed mathematically as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \text{ ---------(1)}$$

In this formulation, the generator improves by minimizing the discriminator's ability to detect fake samples $G(z)$, while the discriminator strengthens its capacity to distinguish between real data $x$and synthetic outputs. Through this iterative process, GANs evolve to produce highly realistic media that are increasingly difficult to differentiate from authentic content.

The growing sophistication of deepfakes has intensified the challenge of detection. Modern detection systems attempt to identify subtle anomalies, such as unnatural facial movements, inconsistent eye blinking, irregular lip synchronization, or imperceptible visual artifacts, that may not be noticeable to the human eye. Deep learning plays a dual role in this landscape: it powers the creation of deepfakes while simultaneously serving as the foundation for detection frameworks that learn discriminative patterns from large-scale labeled datasets.

Detection strategies can be broadly categorized into three groups: deep learning-based approaches, artifact-driven methods, and hybrid frameworks. These approaches can be analyzed across multiple dimensions, including their failure modes, signal provenance, and temporal scale. By systematically studying these methods, researchers can better understand current limitations and identify future directions for developing more robust and generalizable deepfake detection systems [2].

Deep Fake Video Detection

Deepfake videos, which use advanced artificial intelligence to construct seemingly real but false visuals, have become a significant concern due to their potential for misuse in spreading deceptive content, data misuse, and other unethical behaviors. Assessing the presence of such videos is crucial to maintaining the integrity of digital media and ensuring public trust.

***Corresponding Author:** Bhavya S

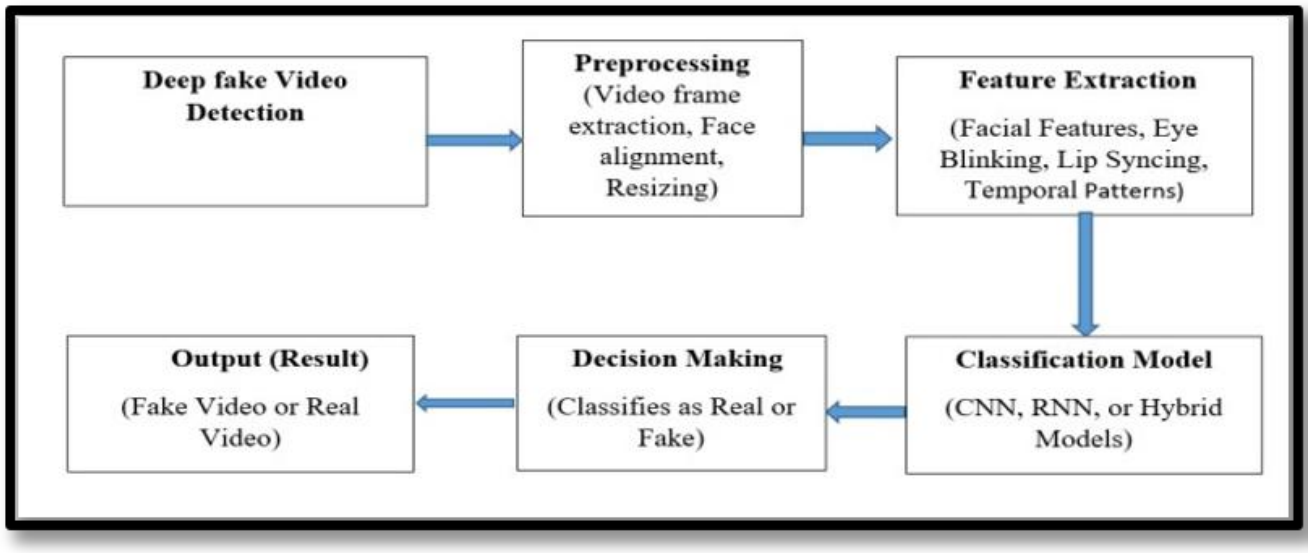

Figure 1: General Workflow for Deepfake Detection

Input: Initially, the system is provided with a video, which can be either genuine or synthetically produced using deepfake algorithms.

Pre-processing: The video is first subjected to a pre-processing phase, which enhances its visual quality and formats the data for effective analysis. This process may include operations like scaling, normalization, or other modifications to make the input suitable for deep learning techniques.

Feature Extraction: In the process of detecting deepfakes, feature extraction plays a vital role by isolating visual, spatial, and occasionally audio elements from video frames. The objective is to detect minute irregularities or artifacts that may have been embedded during the synthesis of fake content.

Create a Model: A deep learning detection model is developed using techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which are commonly utilized for feature extraction and pattern recognition.

Model Training: Training the detection model involves using datasets that contain annotated examples of both genuine and deepfake videos. Through this exposure, the model learns to recognize the key differences between authentic and manipulated content.

Model Testing: After the training is complete, the model is evaluated using unseen data to assess how well it generalizes its learned knowledge to unfamiliar instances.

Result Determination: In the final phase, the output from the trained model is used to decide if a video is real or artificially created. This process effectively assesses the video's authenticity.

Techniques for Detecting Deepfake Videos

Deep Learning-Based Approaches

CNNs extract spatial features from frames, while RNNs, particularly Long Short-Term Memory (LSTM) networks, capture temporal dependencies across sequences.

A CNN maps video frame $X$to feature representation F:

$F = f_\theta(X)$…… **(2)**

An LSTM aggregates sequential information:

$h_t = \sigma(Wx_t + Uh_{t-1} + b)$…… **(3)**

Recent models combine CNN backbones (e.g., ResNext, ResNet) with LSTMs for spatio-temporal learning. For example, Pathan *et al.* [4] proposed a ResNext50 + LSTM pipeline that achieved high precision on Celeb-DF (v2). These models are proficient in detecting unique patterns and characteristics within videos that help differentiate authentic footage from manipulated media. This capability underscores their strength in learning complex representations essential for robust detection. This approach takes advantage of pretrained models like ResNext50 to extract hierarchical features from video data. By combining these with sequential feature extractors such as LSTM, the system effectively captures temporal dependencies and crucial patterns for deepfake detection.

Similarly, Anjani *et al.* [5] applied ResNext-based CNNs for facial manipulation detection, followed by LSTM classification. It presents a revolutionary deep learning-based approach designed to reliably identify the difference between genuine and AI-generated fake videos. It briefly outlines the methodology, highlighting that the approach employs a ResNext-based Convolutional Neural Network (CNN) to perform feature extraction at the individual frame level, focusing on detecting deepfakes involving facial replacement and

***Corresponding Author:** Bhavya S

re-enactment. For the final classification task, a Recurrent Neural Network (RNN) enhanced with Long Short-Term Memory (LSTM) units is utilized to differentiate genuine videos from those that have been tampered with.

Deep learning has been one of the most effective methods for detecting deepfake videos. These methodologies employ architectures like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their hybrid combinations to examine both spatial and temporal elements within video content. CNNs are particularly effective at extracting spatial details from individual video frames, while RNNs—especially Long Short-Term Memory (LSTM) networks—excel at recognizing patterns across sequences of frames, capturing the temporal dynamics essential for deepfake detection [3].

Artifact-Based Methods

Artifact-driven methods detect forensic traces left during manipulation, such as:

- Blending boundaries (face edge mismatches).
- Frequency spikes in Discrete Fourier Transform (DFT) or Discrete Cosine Transform (DCT) coefficients.
- Camera pipeline traces (e.g., Color Filter Array noise).

A simple artifact detector can be formalized as:

$$A(f) = \sum_{u,v} | F(u,v) |^2 \cdot M(u,v) \ldots\ldots \textbf{(4)}$$

where $F(u,v)$represents frequency-domain coefficients and $M(u,v)$is a mask for high-frequency anomaly detection.

Liu and Wang [6] proposed an Artifact-Focused Attention Network (AFAN) to enhance spatial inconsistencies, which is designed to detect high-quality deepfake videos, particularly in short-frame scenarios where spatial information is limited. By utilizing explicit attention mechanisms and artifact enhancement strategies, the proposed framework effectively identifies subtle spatial inconsistencies and local artifacts that generative models introduce. Experimental results on the FF++ dataset and other datasets show that AFAN achieves better results in terms of accuracy and robustness compared to existing methods.

Li *et al.* [7] introduced Artifacts-Disentangled Adversarial Learning (ADAL), which isolates artifacts through a multi-scale feature separator and is robust against irrelevant noise. This method aims to effectively convey the features of the artifact and enhance the connection between the encoder and decoder. The Artifacts-Disentangled Adversarial Learning (ADAL) framework is designed to improve deepfake detection accuracy by separating artifacts from irrelevant information. This method improves detection accuracy while also offering visual cues by accurately identifying artifact patterns. Central to the framework is the Multi-scale Feature Separator (MFS), which is integrated into the disentanglement generator to isolate and highlight these features effectively. To optimize the link between the encoder and decoder, these components are designed to transmit artifact features accurately.

Hybrid Models

Hybrid approaches integrate multiple deep learning frameworks to capitalize on their respective advantages. A common example is the combination of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, which has proven effective in detecting deepfakes by simultaneously analyzing spatial details and temporal dynamics. Graph Neural Networks (GNNs) [8] capture facial region relationships, employing fusion strategies such as additive or multiplicative operations and implementing a two-phase detection process with innovative fusion techniques, achieving high accuracy and addressing existing gaps in deepfake detection methodologies.

The proposed technique divides the detection task into two distinct stages: one utilizing a mini-batch Graph Convolutional Network and the other involving a four-layer CNN pipeline consisting of convolutional operations, batch normalization, and activation functions. A flattening layer is then applied to link the convolutional outputs with fully connected layers. The outputs from both streams are combined using one of three fusion strategies—FuNet-A (additive), FuNet-M (multiplicative), and FuNet-C (concatenation-based). This dual-phase approach achieves a notable training and validation accuracy of 99.3% after 30 epochs [8].

As digitization increases, so do the threats to our data, escalating at a rapid pace. Creating fake videos no longer requires specialized knowledge, advanced hardware, or significant computational resources. However, detecting these fakes remains a challenge. Although numerous approaches have been proposed to solve this challenge, most still suffer from high computational demands, and a truly efficient model remains elusive. To overcome these limitations, a novel architecture named Deep Fake Network (DFN) has been introduced.

DFN is constructed using core elements of the MobileNet architecture, featuring a streamlined sequence of depthwise separable convolutions and max-pooling operations complemented by the Swish activation function. For the DFDC (Deep Fake Detection Challenge) dataset, the model achieved an accuracy of 93.28% and a precision of 91.03% [9].

A cutting-edge method for deepfake video detection combines Ant Colony Optimization and Particle Swarm Optimization (ACO-PSO) with deep learning frameworks. By extracting spatial and temporal features

***Corresponding Author:** Bhavya S

through ACO-PSO, the model captures fine-grained patterns indicative of video manipulation.

The extracted features are then used to train a deep learning classifier that distinguishes real video content from fabricated footage. Experimental results demonstrate that this approach achieves an accuracy of 98.91% and an F1-score of 99.12%, indicating superior performance in identifying deepfakes [10].

Datasets for Deepfake Detection

Performance strongly depends on dataset design.

Table 1: Summarizes Widely Used Benchmarks.

Dataset	Year	Videos	Manipulation Types	Compression
FaceForensics++ (FF++)	2019	1,000+	FaceSwap, Face2Face, NeuralTextures	Raw, HQ, LQ
Celeb-DF (v2)	2020	590	GAN-based facial swaps	Compressed
DFDC	2020	>100k	Mixed methods (GANs, swaps)	Heavy compression
FakeAVCeleb	2021	20k clips	Audio-visual manipulation	Compressed

Table 2: Comparative Analysis of Detection Methods

Reference	Approach	Dataset	Year of Publication	Results
[11]	CNN, TCNs, BiLSTM, Dynamic Time Warping	FakeAVCeleb, AVDeepfake1M, TAVIL, LAV-DF	2025	Accuracy: 0.9987 Accuracy: 0.9825 Accuracy: 0.9915 Accuracy: 0.9812
[9]	ResNet-Swish-BiLSTM	FF++, DFDC	2024	Accuracy: 99.13% Accuracy: 98.6%
[12]	Long Short-Term Memory (LSTM), ResNeXt	Celeb-DF (v2)	2024	Accuracy: 95.33%
[10]	Ant Colony Optimization–Particle Swarm Optimization (ACO-PSO)	DFDC	2024	Accuracy: 97.32%
[13]	Hybrid CNN-LSTM Model	DFDC, FF++, Celeb-DF	2022	Accuracy: 66.26% Accuracy: 91.21% Accuracy: 79.49%

DISCUSSION AND CHALLENGES

Despite significant progress, challenges remain:

- **Failure Modes:** CNN-LSTM hybrids degrade under heavy compression and short clips.
- **Signal Provenance:** Artifact-based models fail when re-encoding smooths traces.
- **Temporal Scale:** Frame-only detectors ignore sequence consistency, while temporal models require high-quality video.
- **Multimodality:** Few methods robustly exploit audio-visual synchronization; dataset imbalance (visual vs. audio) limits generalization.

CONCLUSION AND FUTURE DIRECTIONS

The increasing prevalence of deepfake videos poses a serious risk to the authenticity of digital media, affecting areas such as misinformation, political manipulation, cybercrime, and personal privacy. This review examines the latest techniques for detecting deepfake videos, including deep learning approaches, artifact-based methods, and hybrid models. Recent advancements, particularly approaches leveraging Convolutional Neural Networks (CNNs) [8], Recurrent Neural Networks (RNNs), and transformer-based architectures, have shown impressive performance on standard benchmark datasets.

Despite this, their ability to generalize and remain resilient in real-world scenarios remains limited. A key obstacle lies in the fast-paced advancement of generative technologies, which continue to enhance the realism of synthetic media while adopting increasingly sophisticated methods to evade detection.

As a result, detection models need to be continually adapted to remain effective. Additionally, the lack of large-scale, diverse, and well-annotated datasets poses a barrier to the training and evaluation of reliable detection systems. Future research should aim to design detection methods capable of generalizing across previously unseen generative methods, remaining robust against compression and post-processing artifacts, and providing explainable insights into their decision-making processes. Furthermore, combining detection tools with blockchain verification, media forensics, and policy frameworks can offer a more comprehensive defense against harmful deepfakes.

***Corresponding Author:** Bhavya S

In summary, **deepfake** identification is a dynamic and rapidly evolving field. Ongoing interdisciplinary collaboration among AI researchers, forensic experts, policymakers, and media organizations is crucial to tackling the complex challenges presented by synthetic media and ensuring the integrity of digital information. Future work should focus on:

- Cross-dataset generalization to unseen manipulations.
- Robustness to compression and re-uploads.
- Multimodal synchronization modeling for audio-visual fakes.
- Explainable detection frameworks to build trust.
- Integration with **blockchain** and forensic tools for scalable authenticity verification.

REFERENCE

1. Abbas, F., & Taeihagh, A. (2024). Unmasking deep fakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. *Expert Systems with Applications, 252*(Part B), Article 124260. https://doi.org/10.1016/j.eswa.2024.124260
2. Khan, R., Sohail, M., Usman, I., Sandhu, M., Raza, M., Yaqub, M. A., & Liotta, A. (2024). Comparative study of deep learning techniques for DeepFake video detection. *ICT Express, 10*(6), 1226–1239. https://doi.org/10.1016/j.icte.2024.05.010
3. Saikia, P., Dholaria, D., Yadav, P., Patel, V., & Roy, M. (2022). A hybrid CNN-LSTM model for video deepfake detection by leveraging optical flow features. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)* (pp. 1–7). IEEE. https://doi.org/10.1109/IJCNN55064.2022.9892884
4. Pathan, S., Khairnar, S., Joshi, S., Malshikare, H., Kharkar, A., & Suryawanshi, D. (2024). Deepfake detection using deep learning: ResNeXt and LSTM. In *Proceedings of the International Conference on Computing, Communication and Green Engineering (ICCubeA)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICCubeA61740.2024.10775025
5. Anjani, S. D. D., Sai, K. T., Venkata, S. S., Subhan, S., & Kumar, V. (2024). Detection of deep fakes using deep learning. *i-manager's Journal on Image Processing, 11*(1), 1. https://doi.org/10.26634/jip.11.1.20719
6. Liu, Y., & Wang, X. (2024). Artifact-focused attention networks for robust detection of high-quality deepfake videos in short-frame scenarios. In *Proceedings of the International Conference on Smart Technologies (ISCTech)* (pp. 1–5). IEEE. https://doi.org/10.1109/ISCTech63666.2024.10845508
7. Li, X., Ni, R., Yang, P., Fu, Z., & Zhao, Y. (2023). Artifacts-disentangled adversarial learning for deepfake detection. *IEEE Transactions on Circuits and Systems for Video Technology, 33*, 1658–1670. https://doi.org/10.1109/TCSVT.2022.3217950
8. El-Gayar, M. M., Abouhawwash, M., Askar, S. S., & Sweidan, S. (2024). A novel approach for detecting deep fake videos using graph neural network. *Journal of Big Data, 11*, 1–27. https://doi.org/10.1186/s40537-024-00884-y
9. Bansal, N., *et al.* (2023). Real-time advanced computational intelligence for deep fake video detection. *Applied Sciences, 13*(5), Article 3095. https://doi.org/10.3390/app13053095
10. Alhaji, H. S., Çelik, Y., & Goel, S. (2024). An approach to deepfake video detection based on ACO-PSO features and deep learning. *Electronics, 13*(12), Article 2398. https://doi.org/10.3390/electronics13122398

***Corresponding Author:** Bhavya S