International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5, Sep-Oct-2026, ISSN (Online): 2395-566X

AI-Based Deepfake Detection: A Review of Spatial Artifact, Temporal Consistency and Biological Signal Approaches for Real-Time Media Verification

Gadge Anuja Harishchandra[1] Prof. N. G. Rokade[2] Prof. M. C. Kolse[3] Prof. S. K. Said[4]

[1]Department Of Artificial Intelligence And Data Science Jaihind College Of Engineering Pune, Maharashtra, India

Abstract—Deepfakes, realistic manipulations of faces, voices and videos produced with generative adversarial networks and diffusion models, threaten digital media integrity, personal privacy and public security. Traditional verification by visual inspection or simple forensic tools struggles against these sophisticated alterations. This paper reviews AI-based deepfake detection with emphasis on approaches suited to a real-time framework. It explains how deepfakes are generated and then surveys three families of detection cues: spatial artifacts within individual frames, such as warping, blending boundaries and texture inconsistencies captured by convolutional networks; temporal inconsistencies across frames, including unnatural blinking and lip motion detected with recurrent and spatio-temporal models; and biological signals such as subtle colour changes caused by blood flow that synthetic faces fail to reproduce. Benchmark datasets including FaceForensics++, Celeb-DF and the DeepFake Detection Challenge are discussed. A five-layer real-time detection architecture covering media ingestion, face processing, multi-cue analysis, decision fusion and reporting is proposed, together with research gaps, objectives, a comparison of methods and challenges such as generalization to unseen generators, compression and adversarial attacks. The review concludes that fusing complementary cues offers the most reliable automated defence.

Index Terms—Deepfake detection, generative adversarial networks, face forensics, CNN, temporal consistency, biological signals, media integrity

I. INTRODUCTION

A few minutes of video of a person are now enough for widely available tools to create a convincing fake in which that person appears to say or do something they never did. Such deepfakes have been used for fraud, harassment, political misinformation and identity theft. Because they can be shared instantly, detection needs to be fast and automatic.Deepfakes are produced mainly with generative adversarial networks [1], in which a generator and a discriminator compete, and more recently with diffusion models [2]. Each generation of tools removes visible flaws, but subtle traces usually remain in the pixels, the motion or the physiology of the synthesized face. Tolosana *et al.* [3] reviewed face manipulation types and the corresponding detection methods This paper reviews AI techniques that exploit these traces and proposes a real-time detection framework.

II. BACKGROUND AND RELATED WORK

2.1 Types Of Facial Manipulation

Fig. 1 distinguishes entire-face synthesis, identity swap, attribute editing and expression or lip-sync reenactment, along with the cue families used for detection.

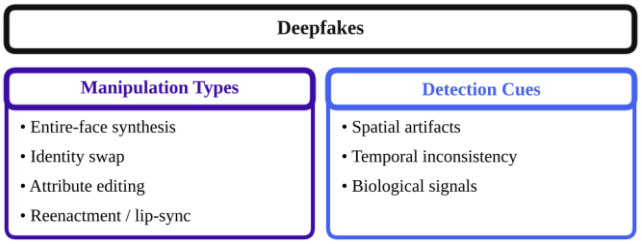


Fig. 1. Facial manipulation types and detection cue families.

2.2 Why Detection Is Possible

Face-swap pipelines usually generate a face at limited resolution, warp it to fit the target and blend it in, leaving resolution mismatches and boundary artifacts. Frame-by-frame generation can cause flicker and unnatural motion, and synthesized skin does not reproduce the subtle periodic colour changes caused by the heartbeat.

2.3 Review Of Related Studies

Rössler *et al.* [4] released FaceForensics++, a large benchmark of videos manipulated with several methods at different compression levels, and showed that a deep Xception network [5] trained on face crops outperformed classical forensic features. Afchar *et al.* [6] proposed MesoNet, a compact network that focuses on mesoscopic image properties, offering fast detection suitable for real-time use.

© 2026 IJSRET

Li and Lyu [7] observed that face-swap algorithms must warp faces to match pose and trained CNNs to detect the resulting warping artifacts. Li *et al.* [8] noticed that early deepfakes blinked rarely and used an LRCN model to detect abnormal blinking. Güera and Delp [9] combined a CNN feature extractor with an LSTM to capture temporal inconsistencies between frames.

Ciftci *et al.* [10] introduced FakeCatcher, which extracts photoplethysmography-like biological signals from facial regions and found that their spatial and temporal coherence distinguishes real from synthetic portraits.

Larger, harder datasets followed: Celeb-DF [11] contains high-quality fakes with fewer visible artifacts, and the DeepFake Detection Challenge dataset [12] provides thousands of videos with diverse actors and manipulation methods, revealing that detectors struggle to generalize. Table I summarizes these works.

TABLE I Summary of Literature Survey

Ref.	Author(s) / Year	Technique / Approach	Key Contribution	Limitation / Gap
[4]	Rössler et al., 2019	FaceForensics++ with XceptionNet	Large benchmark; strong CNN baseline	Known methods only
[6]	Afchar et al., 2018	MesoNet compact CNN	Fast mesoscopic detection	Lower accuracy on HQ fakes
[7]	Li and Lyu, 2019	Face-warping artifact detection	Exploits resolution mismatch	Newer methods reduce warping
[8]	Li et al., 2018	Eye-blink analysis (LRCN)	Detected abnormal blinking	Fakes now blink normally
[9]	Güera and Delp, 2018	CNN + LSTM temporal model	Captures frame inconsistencies	Small dataset
[10]	Ciftci et al., 2020	FakeCatcher biological signals	Generalizes across generators	Needs visible skin regions
[11]	Li et al., 2020	Celeb-DF dataset	High-quality challenging fakes	Celebrity faces only
[12]	Dolhansky et al., 2020	DFDC dataset	Diverse large-scale benchmark	Generalization gap exposed

III. RESEARCH GAP AND OBJECTIVES

3.1 Identified Research Gaps

- Detectors trained on known manipulation methods often fail on new generators.
- Social-media compression removes many subtle artifacts that detectors rely on.
- Real-time detection on long, high-resolution videos remains computationally demanding.
- Audio–visual deepfakes and Indian-language content are under-represented in datasets.
- Detection results are rarely explained in a form useful to journalists or courts.

3.2 Objectives Of The Review

- To review deepfake generation methods and the traces they leave.
- To examine spatial, temporal and biological detection approaches.
- To propose a real-time, multi-cue deepfake detection architecture.
- To compare detection methods by accuracy, robustness and speed.
- To identify challenges and future directions.

IV. DESIGN OF THE PROPOSED SYSTEM

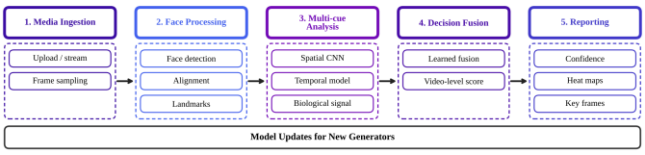


Fig. 2 shows the proposed architecture.

Fig. 2. Proposed real-time multi-cue deepfake detection architecture.

4.1 Media Ingestion Layer

Videos and images are uploaded or captured from streams; frames are sampled at a configurable rate.

© 2026 IJSRET

4.2 Face Processing Layer

Faces are detected, aligned and cropped; facial landmarks identify eyes, mouth and skin regions.

4.3 Multi-Cue Analysis Layer

Three analysers run in parallel: a spatial CNN scores each face crop for artifacts, a temporal model scores blink and lip-motion consistency across frames, and a biological-signal module estimates heartbeat-related colour variation.

4.4 Decision Fusion Layer

Scores are combined by a learned fusion model into a video-level probability of manipulation.

4.5 Reporting Layer

Results are presented with a confidence score, heat maps of suspicious regions and the frames that contributed most to the decision.

4.6 Methodology

Fig. 3 outlines the detection process. Frame-level probabilities p_t are aggregated into a video score, for example by averaging the top-k most suspicious frames, and the cue scores are fused as

$$P_fake = \sigma(w_s \cdot S_spatial + w_t \cdot S_temporal + w_b \cdot S_bio + b) \quad (1)$$

Performance is reported as accuracy and area under the ROC curve on in-dataset and cross-dataset tests, the latter indicating generalization.

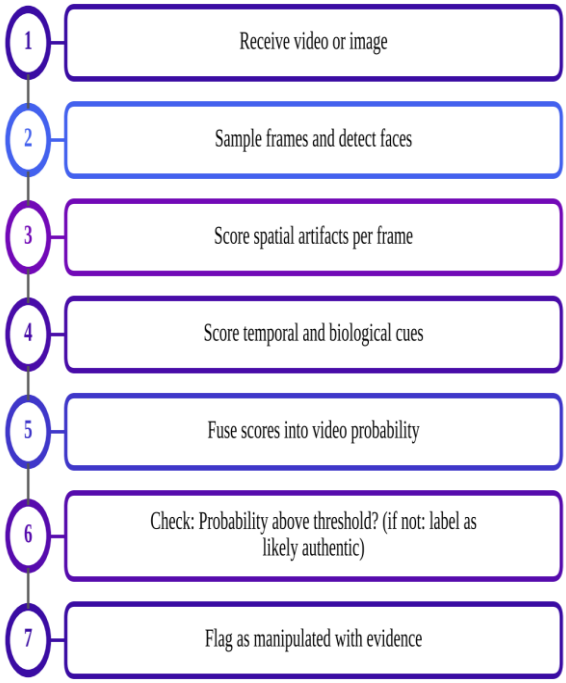


Fig. 3. Deepfake detection workflow.

V. DISCUSSION AND COMPARISON

Table II compares detection approaches. Spatial CNNs are accurate on known manipulations, temporal models capture motion inconsistencies, biological signals generalize well but need clear skin regions, and fusion combines their strengths.

TABLE II Comparison of Detection Approaches

Approach	Accuracy	Generalization	Speed
Spatial CNN (Xception)	High (known)	Low–Medium	Medium
Compact CNN (MesoNet)	Medium	Low	Very fast
Temporal CNN-LSTM	High	Medium	Medium
Biological signals	High	High	Slow
Multi-cue fusion	Very high	Medium–High	Medium

VI. CHALLENGES AND FUTURE DIRECTIONS

© 2026 IJSRET

Table III summarizes challenges. Future work includes detectors trained to generalize across generators, robustness to compression, joint audio–visual analysis, content provenance standards that sign authentic media at capture, and lightweight models for phones.

TABLE III Challenges and Mitigation

Challenge	Mitigation
Unseen generators	Diverse training; generalizable cues
Compression	Train on compressed data
Adversarial attacks	Robust training; ensembles
Real-time cost	Frame sampling; compact models
Explainability	Heat maps and key-frame evidence

VII. CONCLUSIONS

Deepfake detection is an ongoing contest between generation and detection technologies. The reviewed literature shows that spatial artifacts, temporal inconsistencies and missing biological signals each provide useful evidence, and that large benchmarks expose the difficulty of generalizing to new manipulation methods. A real-time framework that fuses these complementary cues and explains its decisions offers a reliable automated defence against misinformation, identity theft and fraud.

7.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

REFERENCES

1. I. Goodfellow et al., “Generative adversarial nets,” in Proc. NeurIPS, vol. 27, 2014.
2. J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in Proc. NeurIPS, vol. 33, 2020, pp. 6840–6851.
3. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “DeepFakes and beyond: A survey of face manipulation and fake detection,” Inf. Fusion, vol. 64, pp. 131–148, 2020.
4. A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “FaceForensics++: Learning to detect manipulated facial images,” in Proc. IEEE/CVF ICCV, 2019, pp. 1–11.
5. F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in Proc. IEEE CVPR, 2017, pp. 1251–1258.
6. D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “MesoNet: A compact facial video forgery detection network,” in Proc. IEEE WIFS, 2018.
7. Y. Li and S. Lyu, “Exposing DeepFake videos by detecting face warping artifacts,” in Proc. IEEE/CVF CVPR Workshops, 2019, pp. 46–52.
8. Y. Li, M.-C. Chang, and S. Lyu, “In ictu oculi: Exposing AI created fake videos by detecting eye blinking,” in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), 2018.
9. D. Güera and E. J. Delp, “Deepfake video detection using recurrent neural networks,” in Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), 2018.
10. U. A. Ciftci, I. Demir, and L. Yin, “FakeCatcher: Detection of synthetic portrait videos using biological signals,” IEEE Trans. Pattern Anal. Mach. Intell., early access, 2020.
11. 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 CVPR, 2020, pp. 3207–3216.
12. B. Dolhansky et al., “The DeepFake Detection Challenge (DFDC) dataset,” arXiv:2006.07397, 2020.

© 2026 IJSRET