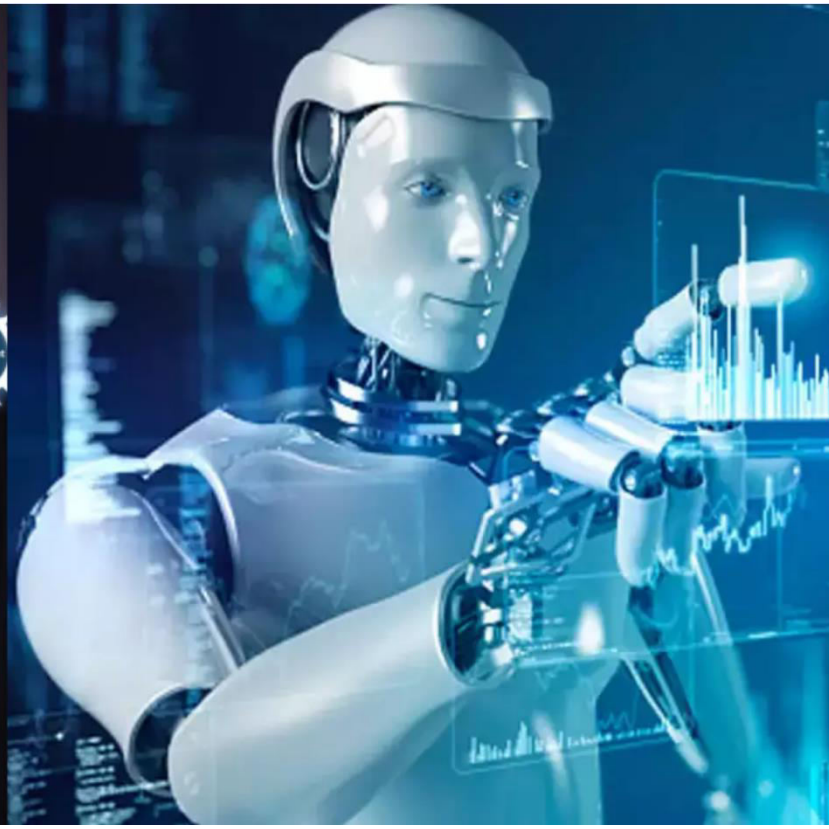
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 2, February 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502027|

A Comprehensive Survey on Fake News and Deepfake Media Detection: From Linguistic Analysis to Multimodal Fusion

Vaishnavi Sonawane[1], Nehal Chaudhari[2], Parth Dhumal[3], Krishna Palange [4], Parth Kolekar[5]

Mentor, Department of Artificial Intelligence and Machine Learning, AISSMS Polytechnic, Pune, Maharashtra, India[1]

Students, Diploma in Artificial Intelligence and Machine Learning, AISSMS Polytechnic, Pune, Maharashtra, India[2-5]

ABSTRACT: The rapid evolution of generative Artificial Intelligence (AI) has facilitated the creation of hyper-realistic synthetic media, posing significant threats to social stability and information integrity. This paper provides a systematic survey of state-of-the-art techniques for detecting fake news and deepfake video/image content. By synthesizing recent breakthroughs in linguistic propagation modeling, statistical trace learning in synthetic videos, and multimodal feature integration using Vision Transformers (ViTs) and Diffusion Models, we categorize the current landscape of digital forensics. We evaluate the efficacy of these methods against modern generators such as Sora and Pika, identify persistent challenges in zero-shot transferability, and outline future research directions in explainable AI.

KEYWORDS: Fake news detection, Deepfake video detection, Multimodal fusion, Generative AI, Synthetic media forensics, H.264 re-compression resilience.

I. INTRODUCTION

The rapid evolution of generative Artificial Intelligence (AI) has fundamentally altered the landscape of digital information, ushering in an era where the boundary between authentic media and synthetically engineered content is increasingly blurred. This "post-truth" environment is characterized by a dual threat: the strategic dissemination of text-based fake news and the creation of hyper-realistic "Deepfakes" in image and video formats. While early misinformation efforts were limited by technical barriers, modern frameworks such as Generative Adversarial Networks (GANs), Diffusion Models, and Large Language Models (LLMs) have democratized the ability to create deceptive content at scale, threatening social stability, individual privacy, and democratic integrity.

Misinformation is no longer a static classification problem but a dynamic, process-oriented challenge involving intentional creation, heteromorphic transmission, and controversial reception. Recent research emphasizes that fake news is often designed with specific stylistic markers and propagation patterns to maximize its impact. Simultaneously, the technological leap from synthetic images to videos has introduced a critical "trace gap." Evidence suggests that detectors specifically optimized for synthetic images often fail when applied to synthetic videos because modern video generators—such as Sora, Pika, and Stable Video Diffusion—introduce unique statistical fingerprints that differ fundamentally from those found in static forgeries.

Furthermore, modern forgeries have evolved into complex multimodal threats. Traditional detection methods relying solely on local feature extraction via Convolutional Neural Networks are no longer sufficient to counter high-fidelity fakes. Modern strategies now propose integrating local artifacts with global contextual relationships using Vision Transformers to track biological inconsistencies, such as mismatched eye movements and lip-sync errors. This survey provides a comprehensive analysis of these emerging methodologies, outlining a technical roadmap for identifying misinformation by synthesizing linguistic analysis, statistical video tracing, and multimodal global-local feature integration to stay ahead of evolving generative technologies.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502027|

II. PROBLEM STATEMENT

The fundamental challenge in modern digital forensics is the widening "asymmetry of information" between generative models and detection frameworks. While creators can leverage tools like Sora and Pika to produce hyper-realistic misinformation with minimal effort, detectors must combat an infinite variety of linguistic styles and visual artifacts. Current systems face a critical "trace gap" where detectors optimized for synthetic images fail to identify AI-generated videos because video generators leave distinct statistical fingerprints that differ from static content [1]. Furthermore, fake news has evolved into complex "heteromorphic transmission" across social networks, making it difficult to verify veracity through text alone as the misinformation is strategically designed to exploit public polarization [2]. Traditional detection methods relying solely on local feature extraction also frequently miss global biometric inconsistencies, such as eye-movement anomalies and audio-visual desynchronization [3]. Standardized preprocessing and facial alignment remain difficult in varied environmental conditions, further complicating the reliability of these forensic tools [4].

2.1 Project Objectives

To address these challenges, this research establishes the following objectives:

- **Bridge the Synthetic Media Trace Gap:** To develop specialized detection methodologies that identify the unique pixel-level statistical traces left by modern video generators [1].
- **Analyze the Misinformation Lifecycle:** To move beyond content-only verification by investigating the stages of intentional creation, transmission patterns, and controversial reception [2].
- **Implement Local-Global Feature Integration:** To enhance detection accuracy by combining local facial artifacts with global contextual signals using Vision Transformers (ViTs) [3].
- **Enhance Multimodal Synchronization:** To model cross-modal dependencies between audio and visual signals to detect subtle temporal inconsistencies [3].

III. LITERATURE SURVEY

The landscape of digital forensics has undergone a significant paradigm shift, moving from simple unimodal analysis to complex, process-oriented frameworks that address the multi-layered nature of modern misinformation. Research in this field now focuses on the intersection of linguistic propagation, statistical signal processing, and multimodal fusion to counter the sophistication of generative AI.

In the realm of linguistic analysis, recent studies have redefined the detection of fake news by analyzing its entire diffusion lifecycle rather than treating it as a static textual classification task. This research identifies three intrinsic stages: intentional creation, where specific inflammatory styles are used; heteromorphic transmission, which tracks the varied paths news takes across social platforms; and controversial reception, which utilizes user stance and polarization as a key indicator of veracity [2]. This perspective proves that the social context and propagation patterns of a news item are often more indicative of its truthfulness than the content itself.

Regarding visual media, a critical "trace gap" has been identified between synthetic images and videos. While image-based detectors are highly mature, they frequently fail when applied to modern video generators like Sora or Pika. This is because AI-generated videos leave unique pixel-level statistical fingerprints that differ fundamentally from those found in static GAN-generated images [1]. Evidence shows that these video-specific traces can be learned to perform reliable source attribution and detection, even when the media is subjected to H.264 re-compression common on social media platforms.

Finally, to address high-fidelity forgeries, researchers are now employing multimodal frameworks that integrate local and global features. While traditional methods rely on local facial artifacts captured by CNNs, modern approaches utilize Vision Transformers (ViTs) to model global contextual relationships, such as the synchronization between eye movements and audio signals [3]. By incorporating diffusion models as pre-processors to refine data and standardized alignment techniques to normalize facial regions [4], these systems achieve superior robustness against complex, cross-modal deepfakes.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502027|

IV. METHODOLOGY

The proposed methodology employs a multi-layered detection pipeline designed to address the unique artifacts of synthetic media. The process begins with **Multimodal Data Preprocessing**, where video streams are decomposed into constituent frames and audio tracks. To ensure spatial consistency, we utilize an efficient facial alignment and landmarking protocol [4] to crop and normalize the Region of Interest (ROI). This stage is critical for mitigating the impact of environmental noise and varied lighting conditions.

For **Feature Extraction**, the framework bifurcates into local and global analysis. Local features, such as microscopic blending artifacts and pixel-level inconsistencies, are captured via Convolutional Neural Networks (CNNs). Simultaneously, **Vision Transformers (ViTs)** are deployed to capture global contextual relationships and long-range temporal dependencies [3]. This is specifically targeted at identifying "Statistical Fingerprints" or traces unique to the generative architecture of the source model [1].

To detect linguistic misinformation, we implement a **Process-Oriented GCN**, which models the news diffusion lifecycle [2]. Finally, these features are integrated through a **Multimodal Fusion Layer**. We utilize diffusion models as pre-processors to refine noisy signals and employ a cross-modal consistency check to identify desynchronization between audio-visual cues [3], leading to the final classification and source attribution.

V. TECHNICAL ARCHITECTURE

The proposed system integrates four distinct analytical layers to identify multi-source misinformation. As shown in the diagram, the process begins with the Input Layer, which handles text, video, and audio streams.

The **Linguistic Layer** addresses text-based fake news by utilizing Graph Convolutional Networks (GCN) to model "Heteromorphic Transmission" [2]. Rather than scanning words alone, it analyzes the structural dispersion of the news and the sentiment of user responses (Controversial Reception) to detect artificial amplification.

The **Visual and Trace Layer** tackles the "trace gap" by processing video frames through two parallel paths. First, it identifies "Statistical Fingerprints"—microscopic noise patterns in pixel distributions that are unique to video generators like Sora or Pika [1]. Simultaneously, it performs facial alignment and landmarking to normalize the data for biometric analysis [4].

The **Multimodal Fusion Layer** acts as the core decision engine. It employs **Vision Transformers (ViTs)** to capture global contextual relationships, such as the temporal consistency between frames and the synchronization between audio phonemes and lip movements [3]. To improve robustness, a **Diffusion Model** is utilized for data refinement, "cleaning" artifacts from real-world compression to ensure high-fidelity feature extraction.

Finally, the **Decision Network** concatenates these cross-modal features to provide a final classification. Beyond a simple binary "Real or Fake" output, the system provides **Source Attribution**, identifying the specific generative model used, thereby offering a comprehensive forensic trail for the detected media. comprehensive forensic trail for the detected media.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502027|

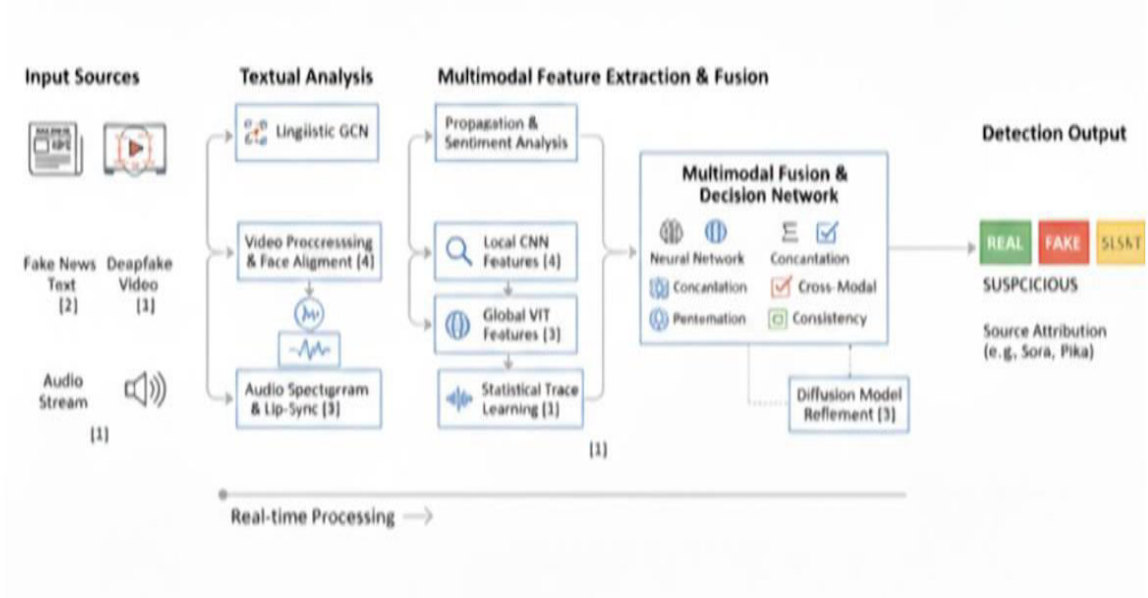


Fig.1. Proposed Multimodal Detection Framework for Synthetic Media and Fake News Analysis

VI. KEY FEATURES

The detection framework leverages several critical features to distinguish synthetic media from authentic content:

- **Statistical Trace Learning:** Identifying "digital fingerprints" or microscopic noise patterns in pixel distributions unique to specific video generators like Sora or Pika [1].
- **Structural Dispersion:** Analyzing the "heteromorphic transmission" patterns of news, where fake content often displays distinct, star-like propagation graphs compared to organic growth [2].
- **Controversial Reception:** Utilizing user stances and polarized sentiment in response layers as a proxy for veracity [2].
- **Local-Global Consistency:** Integrating local facial artifacts with global biometric signals, such as eye-movement anomalies and gaze patterns, through Vision Transformers [3].
- **Audio-Visual Synchronization:** Measuring the precise temporal alignment between phonemes and visemes to detect lip-sync errors [3].
- **Geometric Alignment:** Utilizing precise facial landmarking to ensure consistent feature extraction across varying environmental conditions [4]

VII. FUTURE SCOPE

The future of digital forensics lies in the development of more adaptive and interpretable systems that can keep pace with the accelerating "arms race" between generative models and detection frameworks. One of the most critical areas for growth is the implementation of **Explainable AI (XAI)**. Future systems must move beyond binary classifications to provide detailed heatmaps and natural language explanations that highlight exactly why a piece of media is flagged as synthetic, which is essential for legal and journalistic applications.

Furthermore, research will shift toward **Zero-Shot and Few-Shot Learning** capabilities. As new generators are released weekly, detectors must be able to identify novel synthetic "traces" without requiring thousands of updated training samples. Another vital direction is the enhancement of **Cross-Platform Resilience**. Future models will need to maintain high accuracy even when content has been heavily degraded by multiple rounds of social media compression or adversarial attacks designed to bypass filters. Finally, integrating **Blockchain-based Verification** with AI detectors could create a hybrid "trust-but-verify" ecosystem, where original content is digitally signed at the source, leaving AI to handle the forensic analysis of unsigned, suspicious media.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502027|

VIII. CHALLENGES AND RESEARCH GAPS

Despite significant advancements, several critical bottlenecks hinder the deployment of robust detection systems. A primary challenge is the **"Generalization Gap"**; models trained on existing datasets often experience a drastic drop in accuracy when exposed to "in-the-wild" content from emerging generators [1]. This is compounded by the **Resilience Deficit**, where common social media operations—such as aggressive H.264 compression, resizing, and noise—effectively "wash away" the microscopic statistical traces detectors rely on [2].

From a linguistic perspective, the **"Adversarial Evolution"** of fake news allows creators to bypass Graph Convolutional Networks by subtly altering propagation structures [2]. Furthermore, current multimodal systems struggle with **Temporal Scalability**; while they can detect inconsistencies in short clips, they often fail to maintain synchronization checks over longer video durations where generative errors may be less frequent [3]. Finally, there is a persistent lack of **Standardized Benchmarks** for multimodal forgeries that include diverse lighting, extreme angles, and low-resolution environments, which frequently lead to false positives in facial alignment modules [4].

IX. CONCLUSION

The rapid proliferation of AI-generated misinformation necessitates a multifaceted forensic response that transcends traditional detection boundaries. This survey has demonstrated that a robust defense must integrate the analysis of social propagation patterns, the identification of microscopic statistical traces in synthetic videos, and the monitoring of global biometric synchronization. By combining linguistic insights with multimodal feature fusion, researchers can better address the "trace gap" and the evolving sophistication of generative models. As the digital arms race continues, prioritizing explainability and cross-platform resilience will be essential to maintaining information integrity and restoring public trust in digital media ecosystems.

REFERENCES

[1] D. S. Vahdati, T. D. Nguyen, A. Azizpour, and M. C. Stamm, "Beyond Deepfake Images: Detecting AI-Generated Videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2024.

[2] L. Hu, S. Ji, T. Zhao, and H. Liu, "Process-Oriented Fake News Detection: Challenges and New Directions," *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[3] M. Javed, Z. Zhang, F. H. Dahri, and T. Kumar, "Enhancing multimodal deepfake detection with local–global feature integration and diffusion models," *Signal, Image and Video Processing*, vol. 19, no. 5, pp. 400-409, 2025.

[4] S. A. Qadir, et al., "An efficient deepfake video detection using robust deep learning," *Heliyon*, vol. 10, no. 3, p. e25757, 2024.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details