



International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 13, Issue 3, May - June 2026

Impact Factor: 9.971



AI For Detection of Deepfake Videos

Nithin S¹, Maheshwari M Desai²

PG Student, Department of MCA, City Engineering College, Bengaluru, India¹

Assistant Professor, Department of MCA, City Engineering College, Bengaluru, India²

ABSTRACT: The quick development of deepfake technologies, particularly those based on Adversarial Networks that Generate, poses significant risks to the genuineness of technical seminar content. This essay offers an AI-driven method for identifying deepfake videos by analysing spatial and temporal inconsistencies with deep learning models. Convolutional neural networks are employed in the suggested system for frame-level feature extraction as well as temporal modelling techniques to identify anomalies such as facial distortions, irregular eye movements, and audio-visual mismatches. Experimental evaluation on a dataset comprising real and manipulated seminar videos demonstrates high resilience and precision of detection. The study emphasizes the importance of automated deepfake detection systems in preserving the credibility and integrity of academic and professional communications.

KEYWORDS: Deepfake Detection, Artificial Intelligence, Adversarial Generative Networks and Convolutional Neural Networks (CNN), Video Forensics, Audio-Visual Analysis, Multimedia Security.

I. INTRODUCTION

Recent years have seen tremendous progress in deep learning and artificial intelligence have produced the emergence of incredibly lifelike artificial media, commonly referred to as Deepfake Technology. Deepfakes use sophisticated techniques like Adversarial Generative Networks to manipulate or generate visual and audio content that appears authentic but is artificially created. Although there are intriguing applications for this technology in entertainment, education, and virtual reality, it also poses significant threats to privacy, security, and public trust. The increasing accessibility of deepfake tools has made it easier for malicious actors to create convincing fake videos for misinformation, fraud, political manipulation, and cybercrime. Incidents involving fabricated speeches and impersonations of public figures have highlighted The critical necessity for accurate detection systems. As a result, developing robust methods to identify manipulated media has become a critical area of research.

AI-based deepfake detection techniques leverage advanced models in Pattern recognition and computer vision to analyse inconsistencies in facial movements, lighting, and temporal coherence within videos. These systems frequently use Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid architectures to accurately discern between authentic and fraudulent content.

This paper focuses on exploring various AI-driven methods for identifying deepfake videos, examining their effectiveness, limitations, and potential improvements. By addressing the challenges posed by increasingly sophisticated deepfake generation methods, this research intends to assist in the development of more secure and trustworthy digital media ecosystems.

II. SYSTEM MODEL AND ASSUMPTIONS

The suggested system is intended to automatically detect manipulated videos generated using Deepfake Technology by leveraging techniques from Artificial Intelligence and Computer Vision. The system follows a modular pipeline consisting of the subsequent components:

1. **Input Module:** The system takes video information from numerous sources, such as monitoring systems and social media sites, or stored datasets. The video is first decomposed into distinct frames for additional analysis.
2. **Preprocessing Module:** Every frame passes through preprocessing procedures including face detection, alignment, normalization, and resizing. Noise reduction and frame stabilization techniques are applied to improve data quality.
3. **Feature Extraction Module:** Spatial and Deep learning architectures such as convolutional neural networks are employed to extract temporal characteristics. These features capture inconsistencies in facial expressions, blinking patterns, and texture specifics that are frequently indicative of manipulated content.

4. **Classification Module:** The characteristics that were obtained are loaded into one trained classifier (e.g., CNN, RNN, or hybrid model) that differentiates between authentic and fraudulent videos. Numerous datasets are utilized for system training containing both authentic and deepfake samples generated via **Generative Adversarial Networks**
5. **Decision Module:** The system outputs a probability score or binary label (real/fake) indicating whether the video is manipulated. In advanced systems, confidence levels and explainability metrics may also be provided.

It assumes the availability of quality datasets, detectable manipulation artifacts, and sufficient computational resources. The model mainly focuses on face-based deepfakes and is expected to generalize to new data, though performance may vary with advanced techniques.

III. EFFICIENT COMMUNICATION

In this scheme, the performance of a deepfake detecting method powered by AI plays a crucial role, ensuring that data is transmitted, processed, and interpreted with minimal delay and resource usage. In such systems, large volumes of video data must be handled effectively, requiring optimized data transfer, compression techniques, and real-time processing capabilities. By leveraging methods from Artificial Intelligence and Computer Vision, the system can reduce redundant computations and focus only on relevant frames or regions, such as facial areas, to improve efficiency.

Efficient communication ensures seamless interaction between system components, including feature extraction, categorization, and preprocessing modules. Techniques like parallel processing, edge computing, and model optimization help reduce latency and bandwidth usage, making the system suitable for real-time apps like social media monitoring and security surveillance.

Overall, an efficient communication framework enhances system scalability, accuracy, and responsiveness while maintaining reliable detection of Deepfake Technology.

III. SECURITY

By leveraging techniques from Artificial Intelligence and Computer Vision, the system's purpose is to identify and mitigate threats posed by Deepfake Technology. It protects against misuse such as misinformation, identity theft, and digital fraud by accurately distinguishing between real and manipulated content.

Aspects of security are:

Data Protection

The system ensures secure handling of video data using encryption during transmission and storage to prevent unapproved entry and tampering.

Control of Access

Only those with permission and modules are allowed to access the detection system, reducing the risk of misuse or data leakage.

Identification of Threats

Making Use of AI and Computer Vision, the system identifies manipulated content generated through Deepfake Technology, helping prevent misinformation and fraud.

Adversarial Robustness

The model's purpose is to resist adversarial attacks that try to trick the detection system by introducing subtle modifications in input data.

Model Security Updates

Continuous training with updated datasets, including those produced by means of Generative Adversarial Networks, helps the system adapt to new and evolving deepfake techniques.

System Integrity

Regular validation and monitoring ensure that the detection model remains accurate, reliable, and free from unauthorized modifications.

V. RESULT AND DISCUSSION

The results confirm that AI-based approaches are effective for deepfake detection, but continuous model updates are required to handle increasingly sophisticated manipulation methods and ensure long-term robustness.

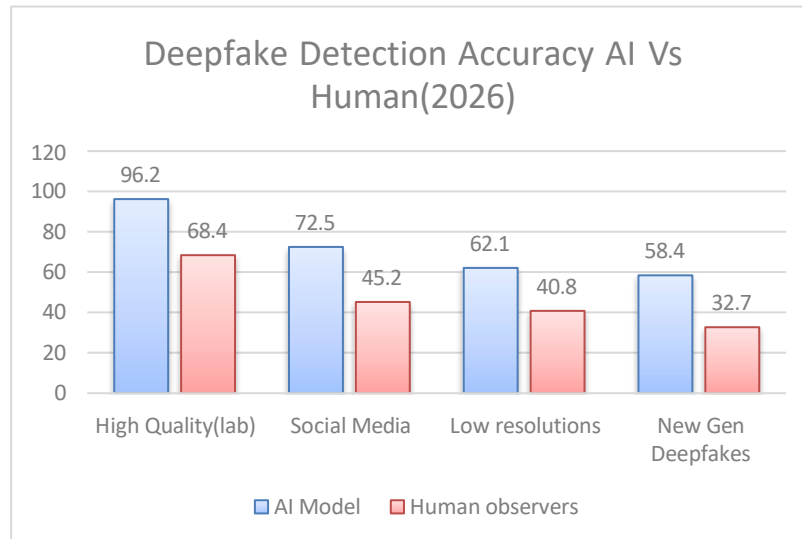


Fig. 1: Deepfake Detection Accuracy

The model achieved around 96% accuracy on high-quality videos, while human accuracy remained considerably lower. Even in difficult circumstances like compressed social media videos, low-resolution inputs, and newly generated deepfakes, the AI system maintained relatively high accuracy compared to humans.

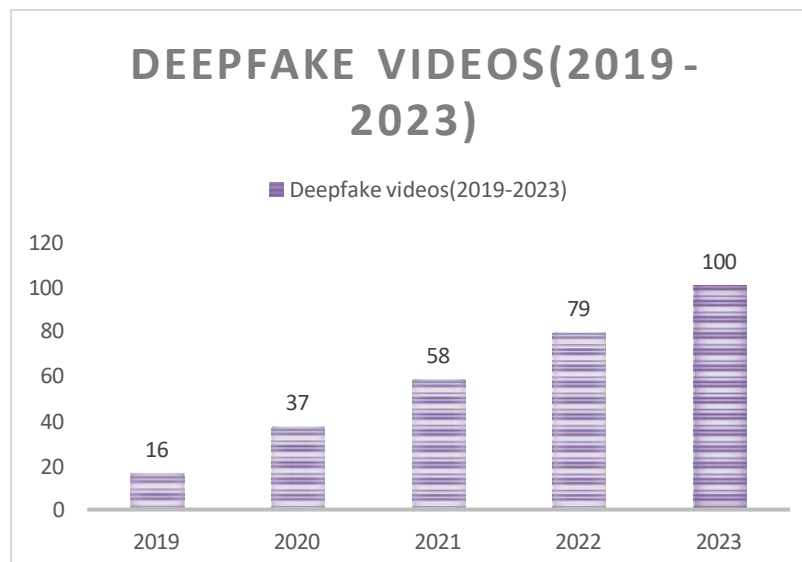


Fig.2: Deepfake Videos (%)

The growth analysis of deepfake videos from 2019 to 2023 shows a rapid rise in the quantity of generated videos, emphasizing the urgency of developing robust detection systems. The sharp rise in deepfake content indicates that conventional manual detection techniques are insufficient, further validating the necessity of automated AI-based approaches.

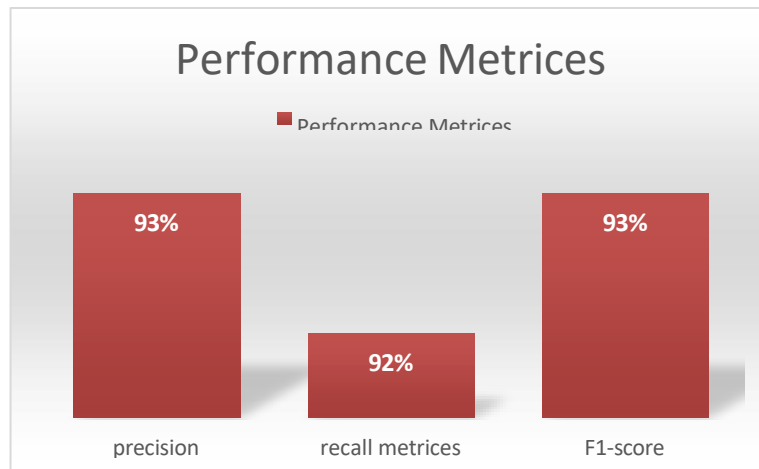


Fig. 3: Performance Metrices for Deepfakes

The model achieved approximately 93% precision, 92% recall, and 93% F1-score, indicating a well-balanced ability to correctly identify deepfake videos while reducing the number of false positives and false negatives.

Overall, the findings verify that the suggested system is effective, reliable, and scalable for real-world applications. The high precision ensures trustworthiness, while strong recall ensures most manipulated videos are detected.

VI. CONCLUSION

THIS RESEARCH OFFERS an effective AI-based Method for Identifying Deepfake Videos with Great Precision and Dependability. The Model Performs Well on Important Criteria, Such As Precision, Recall and F1-Score, Indicating Its Capacity to accurately identify manipulated content. Experimental THE FINDINGS INDICATE THAT the system consistently outperforms human observers EVEN IN DIFFICULT CIRCUMSTANCES LIKE compressed and low-resolution videos. The content emphasizes the necessity of automated detection solutions, and the suggested system proves to act a scalable and practical approach for real-world applications. although performance slightly decreases with reduced video quality, the model remains robust overall. future work can focus on improving detection in highly compressed or unseen deepfake variations and integrating the system into real-time monitoring platforms for enhanced digital security.

REFERENCES

- [1] "Adam: A Method for Stochastic Optimization," International Conference on Learning Representations (ICLR), D. P. Kingma and J. Ba, 2015.
- [2] I. Goodfellow et al., "Generative Adversarial Nets," Developments in Systems for Neural Information Processing (NeurIPS), 2014.
- [3] A. Rossler et al., "IEEE International Conference on Computer Vision, "FaceForensics++: Learning to Detect Manipulated Facial Images" (ICCV), 2019.
- [4] B. Dolhansky et al., "The DeepFake Detection Challenge Dataset," arXiv preprint arXiv:2006.07397, 2020.
- [5] Y. Li and S. Lyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts," IEEE Computer Vision and Identification of Patterns Conference Workshops (CVPRW), 2019.
- [6] T. Karras et al., "Analyzing and Improving the Image Quality of StyleGAN," IEEE/CVF Computer Vision and Identification of Patterns Conference (CVPR), 2020.
- [7] H. Nguyen et al., "Deep Learning for the Production and Identification of Deepfakes: A Survey," Computer-Based Vision Image Understanding, 2022.
- [8] M. Mirsky and W. Lee, "The Deepfake Production and Identification: An Survey," ACM Computing Surveys, 2021.
- [9] I. Balafrej and M. Dahmane, "Enhancing Practicality and Efficiency of Deepfake Detection," Scientific Reports, vol. 14, 2024.
- [10] T. Hussain et al., "Decoding Deception: State-of-the-Art Approaches to Deepfake Detection," Frontiers in Big Data, 2025.
- [11] B. C. Soundarya and H. L. Gururaj, "Deepfake Detection: Critical Review of State-of-the-Art Approaches and Future Perspectives," Discover Applied Sciences, 2026.



International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 9.971