



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 14    **Issue:** II    **Month of publication:** February 2026

**DOI:** <https://doi.org/10.22214/ijraset.2026.77607>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Deepdive into Deepfaked Detection Technology

Prof. Poonam Dharpawar, Prerna Pal, Shivanshi Gupta, Smruti Choudhary

Data Science Department, Usha Mittal Institute of Technology, Shreemati Nathibai Damodar Thackersey Women's University, Mumbai

**Abstract:** Concerns regarding disinformation, identity theft, and dwindling trust in digital material have been raised by the rise in deepfake video production brought on by the quick development of machine learning and artificial intelligence. With the help of this project, users may upload movies and get immediate authenticity analysis using an intuitive web interface. The system uses sophisticated algorithms to categorize videos as "real" or "fake," guaranteeing usability and accessibility. It seeks to inform consumers about the consequences of deepfake technology in addition to detection. The results highlight the need for strong detection systems in the digital age to protect privacy and maintain data integrity.

**Keywords:** Deepfake, Video Detection, Artificial Intelligence, User Interface, Mis information.

## I. INTRODUCTION

Deepfake technology creates incredibly lifelike synthetic audio-visual content by applying sophisticated deep learning algorithms. The main component of this technology is Generative Adversarial Networks (GANs), a kind of neural network architecture in which two models—the discriminator and the generator—compete with one another to generate output that are more and more convincing. While the discriminator tries to distinguish between authentic and artificially generated media, the generator produces synthetic content. This repeated technique eventually produces deepfakes that are almost identical to authentic recordings.

Deepfake technology has valid uses in industries like virtual reality, entertainment, and filmmaking, but when used improperly, it poses serious threats to information security, privacy, and public confidence. Deepfakes can be used by malicious actors to propagate political individuals or corporations through identity fraud and financial scams. Moreover, deepfakes have been used to fabricate speeches, impersonate public figures, and create misleading content that can damage reputations or influence elections.

Detecting deepfakes is an ongoing challenge, as generative models continue to evolve, producing increasingly sophisticated results that blur the line between reality and fabrication. Conventional detection techniques include frame-by-frame analysis to identify visual inconsistencies, examination of temporal discrepancies in facial expressions and speech synchronization, and deep learning-based classifiers that assess patterns within the generated content. However, as deepfake algorithms improve, detection methods must also advance to keep pace with new threats.

This paper provides an in-depth exploration of the growing challenges associated with deepfake detection, reviews existing techniques used to identify synthetic media, and highlights the urgent need for more robust and adaptive solutions. By understanding the capabilities and limitations of current detection frameworks, researchers and policymakers can work toward developing effective countermeasures to mitigate the risks posed by deepfake technology.

## II. LITERATURE REVIEW

Deepfake detection has been an active area of research, with various approaches proposed to identify manipulated media. Existing methods primarily rely on deep learning, computer vision techniques, and forensic analysis to detect inconsistencies in deep fake content.

With the help of Multi-Task Cascaded Convolutional Neural Networks (MTCNN) for facial recognition, VGG19 and EfficientNet for feature extraction, and Capsule Networks (CapsuleNet and ArCapsNet) for enhanced classification accuracy, the Golden Ratio Based Deep Fake Video Detection System is able to detect even the smallest manipulations [1].

The AUFF-NET framework integrates Bi-directional Long Short-Term Memory (Bi-LSTM) networks for temporal pattern analysis with Inception-Swish ResNet-v2 for spatial feature extraction. The detection of FaceSwap and face-reenactment deep fakes is improved by this combination, which achieves excellent accuracy in differentiating between modified and authentic video [2].

The Effective Method for Identifying Deepfake Videos To enhance detection against cyberattacks such as compression, noise, blurring, translation, and rotation alterations, the Robust Deep Learning approach makes use of ResNet-Swish and Bi-LSTM. Bi-LSTM improves classification resilience against adversarial manipulations by capturing temporal inconsistencies, whereas ResNet-Swish improves feature propagation [3].

The Dual Descriptor with Frequency Domain Reconstruction Learning approach combines spatial and frequency domain analysis using EfficientNet-NS-B3 and Generative Adversarial Networks (GANs). This method enhances deepfake detection by reconstructing manipulated regions in the frequency domain, improving classification performance [4].

The Graph Neural Networks (GNN)-based Approach enhances deepfake detection by treating images as graphs, improving model generalization across datasets. It leverages t-SNE for feature visualization, SCM for correlation detection, SAM for feature refinement, and ANDM for better classification. This method improves interpretability and effectively detects manipulated regions [5].

### III. IMPLEMENTATION OF THE PROPOSED DEEPFAKE DETECTION SYSTEM

Our proposed deepfake detection system follows a multi-step approach to analyze and classify video content. The methodology is divided into the following stages:

#### 1) Data Collection and Preprocessing

Video files are first preprocessed by extracting individual frames for analysis.

#### 2) Feature Extraction

Features are extracted from each frame using an advanced model: Inception V3 for extracting high-level features and efficient feature representation. This model helps capture the essential characteristics of the facial structure and movements.

#### 3) Classification Using Neural Networks

The system uses neural networks (CNN and RNN) to classify real vs. fake content. Capsule Networks are employed because they can analyze spatial and temporal inconsistencies between facial features, which is crucial for identifying subtle manipulations in deepfake videos.

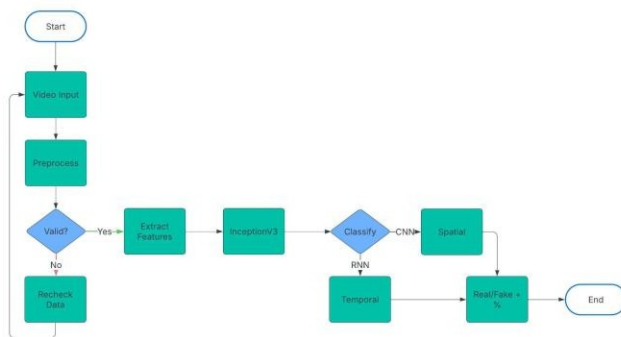
#### 4) Temporal Analysis

Temporal patterns are analyzed by evaluating facial movements across frames, including blinking, lip-sync, and head movements. This is achieved through RNN, which processes sequential data to detect inconsistencies in facial dynamics.

#### 5) Final Classification

The system classifies each video as either "real" or "fake" based on spatial and temporal features. Fusing the spatial and temporal analysis results enhances the overall detection accuracy.

This multi-faceted approach, combining spatial feature extraction and temporal analysis, improves the system's robustness against deepfake manipulations.



Flowchart

### IV. PERFORMANCE EVALUATION AND EXPERIMENTAL ANALYSIS

The deepfake detection project achieved high accuracy, demonstrating strong performance in identifying manipulated videos. It excelled in detecting lower-quality deepfakes, with promising precision, recall, and F1 scores. However, several challenges were observed that highlight areas for improvement. Some authentic videos were misclassified as deepfakes (false positives), while certain deepfakes went undetected (false negatives), reducing reliability. The model struggled against high-resolution deepfakes that exhibited natural facial expressions, realistic lip-syncing, and seamless blending, making detection more difficult.

While effective on known deepfake styles, the model failed to generalize well across newer manipulation techniques and cross-modal deepfakes (audio+video synthesis).

Current deepfake detection methods require high computational power, limiting their feasibility for live streaming platforms and real-time content monitoring. Additionally, attackers employ subtle perturbations, adversarial noise, and blending techniques to bypass detection models, making deepfakes increasingly harder to identify.

To improve detection, future advancements should focus on hybrid approaches that combine audio-visual cues, biological signals, and motion inconsistencies. Transfer learning and adaptive models can enhance generalization by fine-tuning on diverse datasets and enabling continuous learning from new deepfake variations. Explainable AI techniques, such as attention maps and feature visualization, can improve transparency and trust in deepfake classification.

Developing lightweight AI models capable of efficiently processing deepfakes in real-time applications, such as social media monitoring, is essential. Additionally, strengthening detection models through robust adversarial training and blockchain-based media verification can ensure content authenticity.

The project demonstrates the potential of AI in combating misinformation, but continuous advancements are essential to address evolving deepfake techniques. Future research should focus on enhancing model robustness, explainability, and real-time detection capabilities to ensure a more secure digital landscape.

Fig. 1 Represents the initial interface of the site

- Several videos have been analyzed, and their classification results have been displayed.
- The system allows users to upload and analyze deep fake videos.

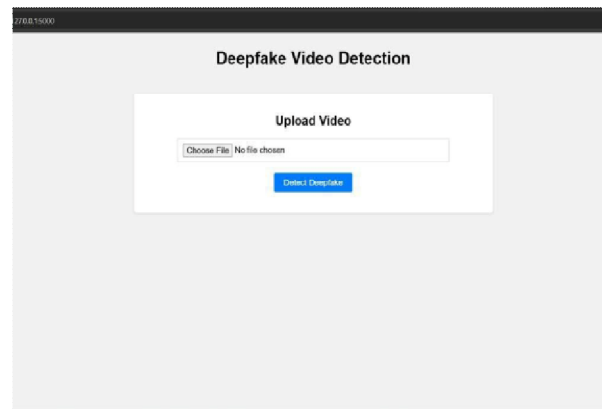


Fig. 1 Interface

Fig. 2 Preview of the uploaded video and its result

- A sample video named "Real6.mp4" was analyzed.
- The model predicted 55.34% REAL and 44.66% FAKE, indicating uncertainty in classification.

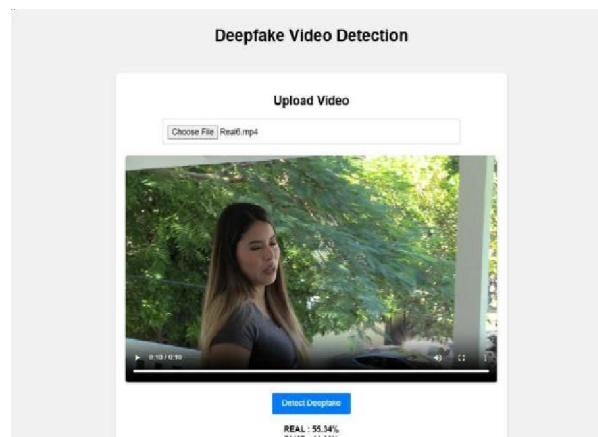


Fig. 2 Result

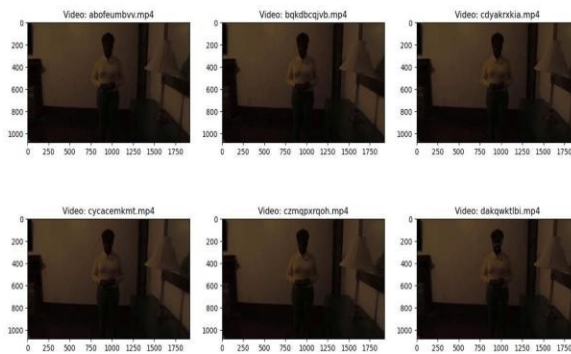


Fig.3 Breakdown of a sample fake video into frames      Fig.4 Distribution of labels in training set

- The dataset contains a distribution of labels (REAL/FAKE) across different videos.
- The number of samples for each class appears to be well-balanced.

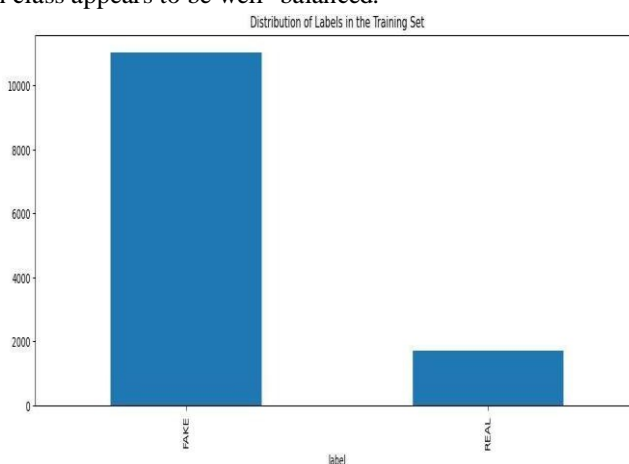


Fig.4.1 Used training set

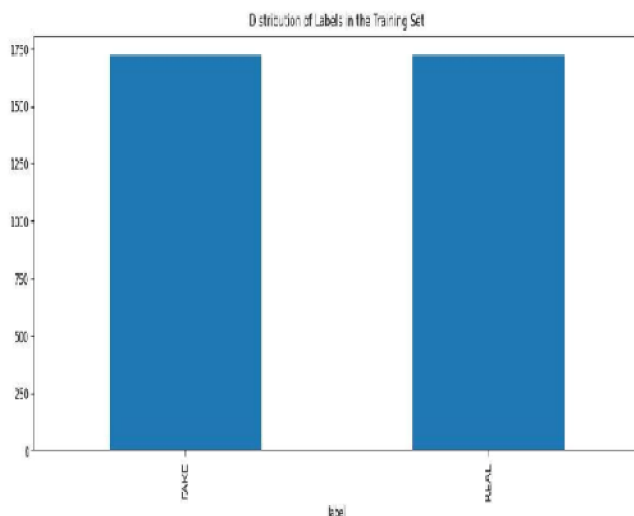


Fig.4.2 Original training set      Fig.5 Model accuracy and loss

- The training and validation accuracy are plotted, showing values around 0.810 to 0.780.
- The training and validation loss are also plotted, with values decreasing over epochs, indicating learning progress.

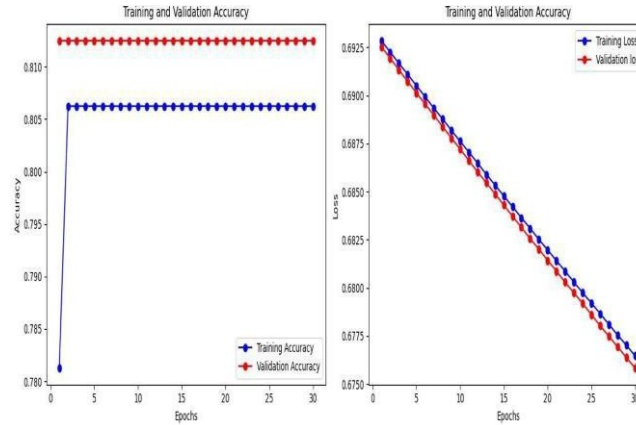


Fig.5 Model accuracy and loss

## V. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

The deepfake video detection model demonstrates promising accuracy, achieving around 81% training accuracy and 78% validation accuracy, though some uncertainty remains in classification. While it effectively differentiates real and fake videos, occasional misclassifications highlight the need for improvements in model architecture and dataset diversity. Future enhancements, such as advanced deep learning techniques, better data augmentation, and real-world testing, can further refine its performance. Overall, the project establishes a strong foundation for deepfake detection, with the potential for increased reliability through further optimization.

## VI. ACKNOWLEDGMENT

We thank Prof. Poonam Dharpawar for her guidance throughout this proposed work. Special thanks to our Principal Dr. Yogesh P. Nerkar, Head of Department Dr. Mohan Bonde, and the teaching staff for their support.

## REFERENCES

- [1] Samet Dincer, Guzin Ulutas, Beste Ustubioglu, Gul Tahaoglu, Nicolas Sklavos, "Golden ratio based deepfake video detection system with a fusion of capsule networks", *https://doi.org/10.1016/j.compeleceng.2024.109234*, ISSN: 2352-4845, 2024.
- [2] Shraddha Suratkar, Faruk Kazi, "Deep Fake Video Detection Using Transfer Learning https Approach", *Arab J Sci: //doi.org/10.1007/s13369-022-07321-3*, Eng 48, 9727-9737 (2023).
- [3] Xin Jin, Nan Wu, Qian Jiang, Yuru Kou, Hanxian Duan, Puming Wang, Shaowen Yao, "A dual descriptor combined with frequency domain reconstruction learning for face forgery detection in deepfake videos", *Forensic Science International: Digital Investigation* 49(2024)301747, <https://doi.org/10.1016/j.fsidi.2024.301747>
- [4] Abdul Qadir, Rabbia Mahum, Mohammed A. El-Meligy, Adham E. Ragab, Abdulmalik AlSalman, Muhammad Awais, "An efficient deepfake video detection using robust deep learning", *Heliyon*, Volume 10, Issue 5, e25757, March 15, 2024, <https://doi.org/10.1016/j.heliyon.2024.e25757>
- [5] H. She, Y. Hu, B. Liu, J. Li, and C.-T. Li, "Using graph neural networks to improve generalization capability of the models for deepfake detection," *IEEE*, vol. 19, ISSN: 1556-6021, 2024, <https://ieeexplore.ieee.org/abstract/document/10654318/>
- [6] A. Qadir, R. Mahum, M. A. El-Meligy, A. E. Ragab, A. AlSalman, and M. Awais, "An efficient deepfake video detection using robust deep learning," *Heliyon*, vol. 10, no. 5, 2024, [https://www.cell.com/heliyon/fulltext/S2405-8440\(24\)01788-2](https://www.cell.com/heliyon/fulltext/S2405-8440(24)01788-2)
- [7] L. Ma, P. Yang, Y. Xu, Z. Yang, P. Li, and H. Huang, "Deep learning technology for face forgery detection: A survey," *Neurocomputing*, vol. 618, 2025, <https://www.sciencedirect.com/science/article/pii/S0925231224018265>
- [8] Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Thien Huynh-The, Saeid Nahavandi, Thanh Tam Nguyen, Quoc-Viet Pham, Cuong M. Nguyen, "Deep learning for deep fakes creation and detection: A survey", <https://www.sciencedirect.com/science/article/pii/S1077314222001114>
- [9] A. Golda, K. Mekonen, A. Pandey, A. Singh, V. Hassija, V. Chamola, and B. Sikdar, "Privacy and security concerns in generative AI: A comprehensive survey," *IEEE Access*, vol. 12, 2024, <https://ieeexplore.ieee.org/abstract/document/10478883/>
- [10] T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, D. T. Nguyen, T. Huynh-The, S. Nahavandi, Q.-V. Pham, and C. M. Nguyen, "Deep learning for deepfakes creation and detection: A survey," *Computer Vision and Image Understanding*, vol. 223, 2022, <https://www.sciencedirect.com/science/article/pii/S1077314222001114>



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)