International Journal of Innovative Research in Engineering and Management (IJIREM)
ISSN (Online): 2350-0557, Volume 12, Issue 5, October 2025
DOI: https://doi.org/10.55524/ijirem.2025.12.5.7
Article ID IJIRE-1412, Pages 38-44
www.ijirem.org

Explore the Evolution of Deepfake Detection Techniques, focusing on the Transition from Traditional Methods to Hybrid Deep Learning Approaches

Sahana Kumari B[1], Dr. Thyagaraju G S[2] and Pradeep Rao K B[3]

[1, 3] Assistant Professor ,Department of Computer Science and Engineering, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire, Karnataka, India

[2] Professor and Head, Department of Computer Science and Engineering, Sri Dharmasthala Manjunatheshwara Institute of Technology, Ujire, Karnataka, India

Correspondence should be addressed to Sahana Kumari B ; sahana@sdmit.in

Received: 1 October 2025 Revised: 15 October 2025 Accepted: 29 October 2025

ABSTRACT- Deepfake videos, generated using advanced techniques like GANs and autoencoders, pose serious challenges to media authenticity, security, and public trust. These synthetic videos can convincingly alter facial expressions, speech, and identity, making detection increasingly difficult. Traditional unimodal detection methods—focused on either visual or audio cues—often fall short in handling the complexity and diversity of modern deepfakes.This review explores a hybrid deep learning and multimodal approach to deepfake video detection, which combines different learning paradigms (e.g., CNNs, RNNs, transformers) and fuses multiple data modalities such as visual, auditory, and physiological signals. By examining state-of-the-art models and their performance on datasets like FaceForensics++, DFDC, and Celeb-DF, we highlight how these integrated methods offer improved accuracy, robustness, and generalization.The review also addresses key challenges such as real-time detection, explainability, and adversarial robustness. Finally, it outlines future directions, emphasizing the development of lightweight, interpretable, and scalable detection systems. This work serves as a critical resource for advancing reliable deepfake detection technologies in an era of rapidly evolving synthetic media.

KEYWORDS- Deepfake Detection, Hybrid Deep Learning, CNN-RNN, Multimodal approach.

I. INTRODUCTION

Deepfakes are highly realistic fabricated media created using AI, leading to societal impacts such as misinformation, erosion of trust, harassment, and privacy violations. Their misuse can undermine democratic processes and facilitate identity theft, posing significant risks to individuals and society.[1]

The motivation for developing robust detection methods in deepfake video detection stems from the increasing sophistication of deepfake technologies and their potential for misuse in disinformation campaigns. As deepfake videos become more realistic, traditional detection methods, which often focus on single-frame analysis, are proving inadequate. This has led to a pressing need for advanced techniques that can effectively identify deepfakes across multiple frames and in diverse scenarios.

Sequence-based models leverage temporal information, making them more effective than single-frame detectors[2]. These models can analyze changes over time, enhancing detection accuracy against adversarial attacks, which have shown high success rates against simpler models[3].

The limitations of unimodal approaches in deepfake video detection are significant, primarily due to their inability to effectively analyze and interpret the complexities of manipulated media. Unimodal detection methods, which rely on a single type of data (e.g., visual or audio), often struggle to identify inconsistencies that may arise when both modalities are manipulated. This limitation can lead to a higher rate of false negatives and reduced overall detection accuracy. Unimodal detectors focus solely on one type of data, missing critical discrepancies between audio and visual components.For example, a deepfake may have altered visuals that do not match the corresponding audio, which unimodal systems cannot detect[4].

Hybrid models like CNN+RNN and ResNet+LSTM-CNN are effective in capturing both spatial and temporal features of deepfakes. CNNs are adept at extracting spatial features, while RNNs and LSTMs handle temporal dependencies, crucial for detecting dynamic facial expressions and movements across frames[5]. The integration of machine learning and deep learning models enhances detection accuracy by combining the strengths of both approaches. For instance, using advanced feature extraction techniques followed by ML and DL models can improve the detection of subtle manipulations in deepfake content[6].

The paper is organized as follows: Section II discusses the various steps involved in deepfake detection techniques. Section III discusses about various research works carried out on deepfake detection techniques. Section IV traces the progress of deepfake detection methodologies. Section V identifies research gaps and outlines future research directions. Finally, section VI concludes the paper by summarizing key findings and emphasizing the importance of research in this field.

II. PROCESS INVOLVED IN DEEPFAKE VIDEO DETECTION TECHNIQUES

The process of deepfake video detection using a hybrid deep learning and multimodal approach typically follows a structured pipeline that combines multiple data modalities (e.g., visual, audio, physiological) and leverages different deep learning models. Below are the key steps involved in detecting deepfake videos.

- Data Acquisition
- Preprocessing
- Feature Extraction
- Multimodal Fusion
- Classification
- Post-Processing and Interpretation
- Evaluation
- Deployment (Optional)

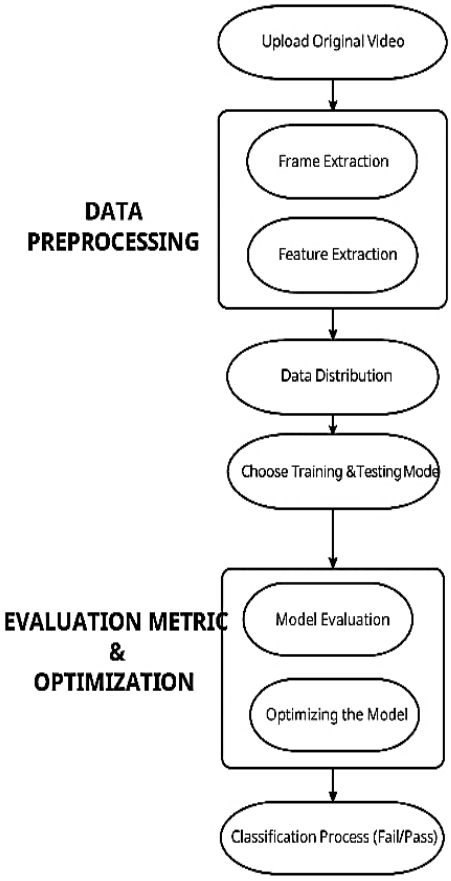

Figure 1: Process involved in detecting Deepfake videos

A. Data Acquisition

Common datasets used include UADFV, FaceForensics++, Celeb-DF, and DFDC, which provide a range of deepfake videos for training and testing[7]. The success of detection models is heavily reliant on the quantity and quality of training data, necessitating extensive data collection efforts to ensure balanced representation across different scenarios[8].

B. Preprocessing

Video Encoding and Frame Extraction: Initial steps often involve encoding videos, renaming, trimming, and extracting frames to create a manageable dataset for analysis[9]. Detecting and cropping faces from frames is a common preprocessing step, which helps focus the model on relevant features. This is often done using libraries like DLIB or MTCNN[10].

C. Feature Extraction

The RGB features extraction layer identifies forgery signs within the spatial domain of video frames, while the GAN features extraction layer detects fingerprints left by Generative Adversarial Networks (GANs) in the high-frequency region[11]. Hybrid models utilizing deep learning algorithms like Xception and ResNet50 have shown high accuracy rates, with ResNet50 achieving 98% accuracy and an AUC of 99.65%[12]. Additionally, combining facial landmarks detection with frequency analysis has proven effective, yielding over 95% accuracy[13]. The integration of ant colony optimization and particle swarm optimization (ACO-PSO) features with deep learning has resulted in a detection accuracy of 98.91%[14].

D. Multimodal Fusion

Incorporating attention allows models to focus on salient features within each modality, enhancing the detection of discrepancies in deepfake content[15].Utilizing Graph Neural Networks (GNNs) and attention networks helps capture temporal inconsistencies and spatial features, improving the robustness of detection systems[16].

E. Classification

A study utilized these algorithms alongside a Triplet Loss approach, achieving an accuracy of 84% with a precision of 90.45%[17]. Several papers employed CNN architectures, such as VGG-16 and EfficientNet, achieving high accuracy rates above 98% for detecting deepfake images[18].

F. Post processing and Interpretation

Post-processing methods, such as augmentations and transformations, can significantly alter the visual characteristics of deepfake videos, complicating detection efforts. Research indicates that detectors trained without considering these transformations may underperform in real-world scenarios[19].Comprehensive multilayer frameworks have been proposed, which include RGB feature extraction, GAN feature detection, and intra-frame inconsistency analysis. These layers work together to identify signs of forgery, even when post-processing is applied[20].

G. Evaluation

Measure performance using metrics like:Accuracy, Precision, Recall, F1-score, AUC.Test across multiple datasets and manipulation types for generalizability.

H. Deployment

Optimize model for real-time inference (e.g., pruning, quantization). Deploy in moderation pipelines, forensic tools, or social media monitoring systems.

III. RELATED WORK

Different Approaches used to detect Deepfakes based on different modalities are:

- Visual (frame-based)
- Audio based

- Text based
- Physiological (heart rate-based)

A. Visual (frame-based):

Hu J Liao et. al introduces FInfer, a novel frame inference-based framework aimed at detecting high-visual-quality Deepfake videos—a growing challenge due to increasingly realistic generative models. Traditional Deepfake detection methods falter as visual artifacts diminish; FInfer overcomes this by predicting future frames' facial representations using autoregressive modeling and comparing them with actual representations. The core innovation lies in using representation-prediction loss to distinguish real and fake content based on temporal coherence and mutual information. FInfer comprises four components: face preprocessing (via Gaussian-Laplace pyramids), representative learning (using convolutional encoders), predictive learning (via GRUs), and correlation-based learning for final classification. Experimental evaluations on multiple benchmarks (Celeb-DF, WildDeepfake, DFDC) reveal that FInfer offers competitive accuracy (up to 90.47%) and superior efficiency (lowest Mult-Adds at 96.75×10^6) compared to existing models.The paper's strengths include theoretical grounding using information theory and strong performance in both in-dataset and cross-dataset scenarios. However, limitations persist: FInfer's performance declines when long-range predictions are required, and its accuracy, while efficient, trails slightly behind some fusion-based deep models in robustness. Future work could integrate more dynamic temporal modeling and explore attention mechanisms for improved scalability[21].

B. Audio based

Sharma et. al proposes a novel multimodal framework for deepfake detection by integrating both audio and visual cues, addressing the limitations of monomodal systems. Traditional visual- or audio-only methods often fail against sophisticated deepfakes such as voice cloning and lip-syncing. The authors utilize mel-spectrograms and CNNs for audio analysis, and facial landmark detection with CNNs for visual analysis. These features are fused through a late-fusion strategy, enabling the system to exploit complementary strengths from both domains. Experimental evaluation on the DeepFakeTIMIT and DFDC datasets yielded high performance, achieving a precision of 0.78, F1-score of 0.82, and accuracy of 0.93, surpassing several state-of-the-art models. The study significantly contributes to robust deepfake detection, with implications for media integrity and security. However, limitations exist, including reliance on late fusion which may not fully exploit cross-modal relationships. Additionally, the approach may face challenges adapting to evolving deepfake techniques or real-world deployment conditions. Future directions include exploring attention-based fusion strategies, incorporating additional modalities like physiological signals or textual data, and improving model generalizability using transfer learning or meta-learning. This research provides a strong foundation for developing scalable, trustworthy deepfake detection systems in an era of increasingly synthetic content[22].

C. Text based

Singh V et. al presents SAVANA, an integrated framework for dual detection of AI-generated media: deepfake videos and machine-generated text. For video detection, it combines BlazeFace for face localization and EfficientNetB4 for classification, leveraging CNN-based feature extraction and regression to distinguish real from manipulated content. For AI text detection, the authors implement a hybrid system that merges double paraphrasing consistency checks (inspired by the SAVANA method) with probabilistic analysis using TF-IDF vectorization and logistic regression. This dual-structured model achieves remarkable accuracy—96% for video and 95% for text, validated on large, balanced datasets.The paper's strength lies in its bimodal architecture, enabling robust detection across formats (PDF, DOCX, multimedia). It innovatively captures both spatial-temporal inconsistencies in video and stylistic coherence in text, adapting well to real-world conditions. However, limitations include its sensitivity to compression artifacts in low-quality media and limited multilingual capabilities for text detection. Furthermore, late fusion of audio-video streams is not addressed, and live-stream detection remains an open challenge. The study offers a scalable, adaptable tool for content authenticity verification and proposes future improvements such as multilingual expansion and real-time deployment[23].

D. Physiological (heart rate) based:

Hernandez-Ortega et.al presents *DeepFakesON-Phys*, a novel DeepFake detection framework leveraging physiological cues, specifically heart rate signals estimated using remote photoplethysmography (rPPG). The method builds upon the DeepPhys architecture and introduces a Convolutional Attention Network (CAN) that captures spatial and temporal inconsistencies in facial video frames. By fine-tuning a heart rate estimation model, the approach exploits subtle blood flow changes—typically absent in synthetic videos—for distinguishing between real and fake content. Evaluations conducted on challenging datasets, Celeb-DF v2 and DFDC Preview, demonstrate impressive performance, achieving 99.9% and 98.2% AUC respectively, surpassing several state-of-the-art visual and physiological detectors. The model's strength lies in its robustness to diverse facial attributes and minimal preprocessing requirements.However, the approach exhibits limitations under external lighting conditions that mimic physiological color changes, potentially degrading detection accuracy. Additionally, the frame-level evaluation may miss temporal cues vital for robust classification, indicating a need for improved temporal integration. Future work should explore cross-database generalization and resilience against novel DeepFake techniques. Overall, this study highlights the untapped potential of physiological signals in advancing DeepFake forensics[24].

Table 1 summarizes recent studies on deepfake video detection techniques, highlighting strengths and challenges of each method.

Table 1: Comparative Analysis of Different Studies

Methodology	Techniques	Strengths	Challenges
CNN and feature extraction[25]	Multi-input Convolutional Neural Network (CNN) and facial features	Improved accuracy	Memory storage
Dynamic convolutional neural network (CNN)[26]	Dense Dynamic Convolutional Neural Network (CNN), Dynamic Dense Blocks, Attention Mechanism	Better Results Across Compression Rates and Datasets	Performance on Complex Compression Formats, Diverse Forgery Methods
Detector training and online detection[27]	Universal detection method	Good Detection Ability, High Generalizability, Good Transferability	Generalizability of dataset & Target Model Architecture
Multi-definition video deepfake detection-High-Level Semantics Reduction, Cross-Domain Training[28]	Facial Structure Destruction, Adversarial Jigsaw Loss, Domain Generalization	Performance on Low and Cross-Definition Videos, High-Level Semantics Reduction: Cross-Domain Generalization	Ineffectiveness of Existing Approaches on Low and Cross-Definition Videos
Feature Extraction Techniques, Classification Techniques (Capsule Networks)[29]	Feature Extraction Techniques, Classification Techniques (Capsule Networks)	Novel Frame Selection Method, Combination of Robust Feature Extraction and Classification, Model Fusion for Enhanced Performance, Strong Performance on Datasets	Data Scarcity and Quality for Training, Generalization to Unknown Deepfake Techniques, Real-time Detection Requirements, Ethical and Privacy Concerns

Figure 2 illustrates the accuracy achieved by different models reported in related studies. The graph highlights analysis of different Approaches used to detect Deepfakes based on different modalities.

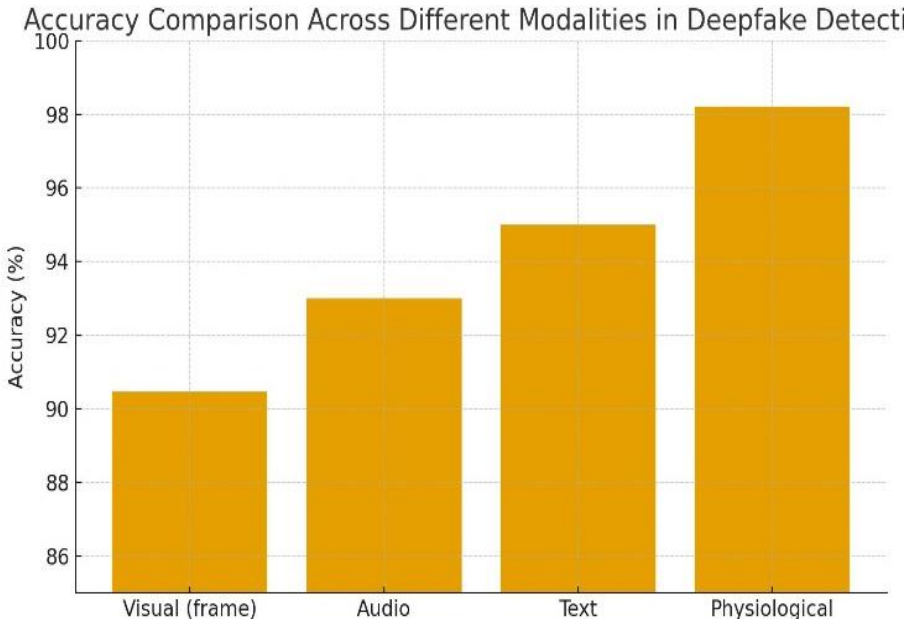

Figure 2: Accuracy reported in surveyed research works

IV. EVOLUTION OF DEEPFAKE VIDEO DETECTION TECHNIQUES

The evolution of descriptive answer script evaluation techniques reflects a dynamic interplay of advancements in natural language processing, machine learning, and ethical frameworks. Early studies concentrated on keyword matching, syntactic and semantic similarity measures, and foundational supervised learning models to automate grading. With the advent of deep learning and transformer-based architectures, research shifted towards enhancing accuracy, addressing biases, and improving efficiency via model explainability and human-in-the-loop systems. Recently, large language models and multimodal approaches have further refined evaluation methodologies, emphasizing transparency, fairness, and integration of ethical considerations in automated assessment systems.

A. Traditional Detection Techniques

Traditional methods often rely on extracting specific features from videos, such as facial landmarks and motion patterns, to identify inconsistencies that indicate manipulation[30]. The paper focuses on temporal and spatial detection techniques, utilizing 68 facial landmarks for alignment and feature extraction and detection was carried out on four datasets: UADFV, FaceForensics++, Celeb-DF, and DFDC.

Common algorithms include Support Vector Machines (SVM), Random Forests, and Decision Trees, which classify media based on the extracted features[31]. These

methods may struggle with the increasing sophistication of deepfake generation techniques, leading to reduced accuracy in detection.

B. CNN and Machine Learning Models

Combination of CNN with Random Forest and XGBoost: This approach utilizes CNNs for feature extraction, followed by machine learning classifiers like Random Forest, which outperformed XGBoost in accuracy[32].
CNN-SVM Hybrid: This model integrates Support Vector Machines (SVM) with CNNs to improve classification accuracy for deepfake images and videos[32].

C. Deep Learning Combinations

CNN and LSTM: This hybrid model captures spatial features through CNNs and temporal relationships using Long Short-Term Memory networks (LSTMs), achieving high precision and recall rates[34].
RNN, GAN, and CNN: This approach combines Recurrent Neural Networks (RNNs) for temporal analysis, GANs for adversarial training, and CNNs for feature extraction, demonstrating versatility and effectiveness in real-time detection[35].

V. GAPS AND FUTURE RESEARCH DIRECTIONS

Despite significant advancements in deepfake detection, several critical research gaps remain. Addressing these challenges is essential for building robust, scalable, and trustworthy detection systems.

A. Multimodal Pretraining Strategies

Research Gap: Current detection methods predominantly focus on visual artifacts in faces or frames, often neglecting the integration of other modalities such as audio, speech-text alignment, and contextual semantics. Existing multimodal approaches typically rely on late fusion or modality-specific feature concatenation, which may fail to capture subtle cross-modal inconsistencies.
Future Direction: Large-scale multimodal pretraining, leveraging contrastive and alignment objectives between modalities (e.g., audio–visual sync, phoneme–viseme correspondence), can enhance cross-domain generalization. Pretext tasks such as temporal coherence prediction or source–target identity disentanglement should be explored to improve manipulation awareness. Robust fusion mechanisms, including cross-attention and uncertainty-aware weighting, can further mitigate modality noise.

B. Federated and Privacy-Preserving Detection Models

Research Gap: Most deepfake detectors are trained on centralized datasets, which raises privacy concerns and limits collaborative development across platforms due to data-sharing restrictions. Moreover, few works address the robustness of detection models under non-IID data distributions or malicious client attacks
Future Direction: Federated learning frameworks incorporating secure aggregation, differential privacy, and Byzantine-resilient optimization could enable collaborative training without raw data exchange. Edge-level personalization layers can adapt global models to local content patterns while maintaining privacy guarantees. New evaluation protocols are needed to benchmark performance in privacy-preserving and adversarial environments.

C. Lightweight Architectures for Edge Deployment

Research Gap: State-of-the-art deepfake detection models are computationally expensive, hindering deployment in latency-sensitive environments such as live streaming or mobile messaging applications.
Future Direction: Model compression techniques, including neural architecture search (NAS) under FLOPs and energy constraints, knowledge distillation from high-capacity video transformers, and sparse temporal sampling strategies, can enable efficient edge deployment. Early-exit mechanisms with calibrated confidence scores could allow "anytime" detection, balancing latency with accuracy requirements.

D. Synthetic Data Generation for Training

Research Gap: Publicly available deepfake datasets are limited in scale, diversity, and modality coverage, often containing biases in ethnicity, gender, and environmental conditions. Synthetic datasets exist but frequently fail to capture the full diversity and realism of in-the-wild manipulations
Future Direction: Generating procedurally diverse synthetic datasets—varying lighting, pose, background, codec artifacts, and editing pipelines—can improve model generalization. Adversarial data generation, where generators and detectors co-evolve, offers a promising approach to closing the domain gap. Integration of weakly-labeled in-the-wild data, possibly through provenance-based cues, can further enhance robustness.

E. Explainable Deepfake Detection Systems

Research Gap: Most deepfake detectors operate as black-box models, providing little interpretability or actionable evidence for moderation, legal, or forensic purposes.
Future Direction: Future systems should integrate explainability techniques such as spatio-temporal artifact localization, counterfactual reasoning, and modality-specific anomaly attribution. Outputs should be accompanied by calibrated confidence scores and contextual evidence (e.g., audio–visual sync traces, metadata forensics, C2PA provenance alignment). Human–AI collaborative interfaces can enable investigators to validate, contest, or refine automated decisions.

F. Cross-Cutting Recommendations

In addition to the modality-specific gaps above, progress requires the development of:
OOD and Streaming Benchmarks: Standardized datasets for out-of-distribution generalization, multimodal manipulation detection, and real-time streaming evaluation.
Security-First Design: Treating detection as an adversarial security task, incorporating robustness certificates, adversarial training, and watermark/provenance fusion.
Policy-Aware Deployment: Aligning technical designs with ethical, legal, and governance frameworks, including transparency in uncertainty communication and clear pathways for appeal in case of false positives.
By addressing these research gaps through interdisciplinary collaboration, the deepfake detection community can move towards systems that are not only accurate and efficient but also ethical, privacy-preserving, and trustworthy.

VI. CONCLUSION

Hybrid approaches merge traditional computer vision techniques with deep learning architectures such as CNNs, RNNs, and Transformer-based models, capturing both fine-grained pixel artifacts and broader semantic inconsistencies. When coupled with multimodal fusion—aligning visual features with speech patterns, facial micro expressions, or background context—these systems significantly outperform single-technique or single-modality detectors, particularly under real-world compression and noise conditions. The benefits of this hybrid + multimodal fusion include enhanced accuracy, improved generalization to unseen manipulation types, and resilience against adversarial attacks. However, technical innovation alone is insufficient. Addressing the deepfake challenge demands collaborative, interdisciplinary progress involving AI researchers, media forensics experts, behavioural scientists, legal authorities, and policymakers. This collective effort can ensure the development of transparent, scalable, and ethically responsible detection systems capable of mitigating deepfake misuse while enabling positive applications of synthetic media.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Alanazi, S., Asif, S., & Moulitsas, I. (2024). *Examining the Societal Impact and Legislative Requirements of Deepfake Technology: A Comprehensive Study. International Journal of Social Science and Humanity*. Available from: https://doi.org/10.18178/ijssh.2024.14.2.1194

[2] Shahriyar, S. A., & Wright, M. (2022). *Evaluating Robustness of Sequence-based Deepfake Detector Models by Adversarial Perturbation.* Available from: https://doi.org/10.1145/3494109.3527194

[3] Shahzad, S. A., Hashmi, A., Peng, Y.-T., Tsao, Y., & Wang, H.-M. (2023). *AV-Lip-Sync+: Leveraging AV-HuBERT to Exploit Multimodal Inconsistency for Video Deepfake Detection. arXiv.Org*, abs/2311.02733. Available from: https://doi.org/10.48550/arxiv.2311.02733

[4] Altamimi, S., & Salameh, W. A. (2024). *Towards Analysis Detection of Deepfake Video via Deep Learning Models: A Review.* 87–92. Available from: https://doi.org/10.1109/ijcc64742.2024.10847278

[5] Gandhi, K., Kulkarni, P. A., Shah, T., Chaudhari, P., Narvekar, M., & Ghag, K. (2024). *A Multimodal Framework for Deepfake Detection.* Available from: https://doi.org/10.48550/arxiv.2410.03487

[6] Lin, C., Lee, J.-C., Wang, S.-J., Chiang, C.-S., & Chou, C. (2024). *Video Detection Method Based on Temporal and Spatial Foundations for Accurate Verification of Authenticity. Electronics*, 13(11), 2132. Available from: https://doi.org/10.3390/electronics13112132

[7] *Improving Deepfake Video Detection Using Data Augmentation Techniques.* (2022). Available from: https://doi.org/10.21203/rs.3.rs-1844392/v1

[8] Murugan, M. A., Mathu, T., & Priya, S. J. (2024). *Detecting Deepfake Videos using Face Recognition and Neural Networks.* 289–293. Available from: https://doi.org/10.1109/icc-robins60238.2024.10534025

[9] Bharadwaj, R., Ratnaparkhi, S., Rajpurohit, R., Rahate, K., Pandita, R., & Thosar, S. (2023). *Deepfake detection for preventing Audio and Video frauds using Advanced Deep Learning Techniques.* 1–7. Available from: https://doi.org/10.1109/iciics59993.2023.10421486

[10] Rathoure, N., Pateriya, R. K., Bharot, N., & Verma, P. (2024). *Combating deepfakes: a comprehensive multilayer deepfake video detection framework. Multimedia Tools and Applications.* Available from: https://doi.org/10.1007/s11042-024-20012-5

[11] Koçak, A., Alkan, M., & Arıkan, S. (2024). *Deepfake Video Detection Using Convolutional Neural Network Based Hybrid Approach. Politeknik Dergisi.* Available from: https://doi.org/10.2339/politeknik.1523983

[12] Singh, P., Chaudhary, K., Chaudhary, G., & Khari, M. (2022). *A Machine Learning Approach to Detecting Deepfake Videos: An Investigation of Feature Extraction Techniques. Journal of Cybersecurity and Information Management*, 9(2), 42–50. Available from: https://doi.org/10.54216/jcim.090204

[13] 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), 2398. Available from: https://doi.org/10.3390/electronics13122398

[14] Shirley, C. P., Jingle, B. J., Abisha, M. B., Venkatesan, R., Ram R V, Y., & Elango, E. (2024). *Deepfake Detection Using Multi-Modal Fusion Combined with Attention Mechanism.* 1194–1199. Available from: https://doi.org/10.1109/icses63445.2024.10763221

[15] *Multimodal Graph Learning for Deepfake Detection.* (2022). Available from: https://doi.org/10.48550/arxiv.2209.05419

[16] Arini, A., Setyaningrum, A. H., & Saputro, A. E. (2024). *Deepfake Video Classification Using Random Forest and Stochastic Gradient Descent with Triplet Loss Approach Algorithm.* 1–6. Available from: https://doi.org/10.1109/citsm64103.2024.10775699

[17] Jbara, W., & Soud, J. (2024). *DeepFake Detection Based VGG-16 Model*. 1–6. Available from: https://doi.org/10.1109/iccr61006.2024.10533024

[18] Cocchi, F., Baraldi, L., Poppi, S., Cornia, M., & Cucchiara, R. (n.d.). *Unveiling the Impact of Image Transformations on Deepfake Detection: An Experimental Analysis.* 345–356. Available from: https://doi.org/10.1007/978-3-031-43153-1_29

[19] Rathoure, N., Pateriya, R. K., Bharot, N., & Verma, P. (2024). *Combating deepfakes: a comprehensive multilayer deepfake video detection framework. Multimedia Tools and Applications.* Available from: https://doi.org/10.1007/s11042-024-20012-5

[20] Hu, J., Liao, X., Liang, J., Zhou, W., & Qin, Z. (2022). *FInfer: Frame Inference-Based Deepfake Detection for High-Visual-Quality Videos. Proceedings of the AAAI Conference on Artificial Intelligence*, 36(1), 951–959. Available from: https://doi.org/10.1609/aaai.v36i1.19978

[21] Sharma, V. K., Singh, S., & Caudron, Q. (2024). *Combating Deepfakes Using an Integrated Framework for Audio and Video Deepfake Detection.* Available from: https://doi.org/10.21203/rs.3.rs-4861782/v1

[22] Singh, V., Agone, B., More, A. W., Mengawade, A., Deshmukh, A., & Badgujar, A. (2024). *SAVANA- A Robust Framework for Deepfake Video Detection and Hybrid Double Paraphrasing with Probabilistic Analysis Approach for AI Text Detection. International Journal for Science Technology and Engineering*, 12(11), 2074–2083. Available from: https://doi.org/10.22214/ijraset.2024.65526

[23] Hernandez-Ortega, J., Tolosana, R., Fierrez, J., & Morales, A. (2020). *DeepFakesON-Phys: DeepFakes Detection based on Heart Rate Estimation. arXiv: Computer Vision and Pattern Recognition.* Available from: https://arxiv.org/abs/2010.00400

[24] *A New Approach to Detect Deepfake Video using Multi-Input Convolutional Neural Network.* (2022). Available from: https://doi.org/10.1109/sti56238.2022.10103229

[25] Mao, X., Sun, L., Zhang, H., & Zhang, S. (2023). *A DeepFake Compressed Video Detection Method Based on*

Dense Dynamic CNN. 12604, 1260409. Available from: https://doi.org/10.1117/12.2674838

[26] Lai, J., Huo, Y., Hou, R., & Wang, X. (2022). *A Universal Detection Method for Adversarial Examples and Fake Images*. *Sensors*, 22(9), 3445. Available from: https://doi.org/10.3390/s22093445

[27] Wang, C., Zhao, C., & Hu, G. (2022). *Multi-Definition Video Deepfake Detection via Semantics Reduction and Cross-Domain Training*. *IEEE International Conference on Multimedia and Expo*, 1–6. Available from: https://doi.org/10.1109/ICME52920.2022.9859736

[28] Lin, C., Lee, J.-C., Wang, S.-J., Chiang, C.-S., & Chou, C. (2024). *Video Detection Method Based on Temporal and Spatial Foundations for Accurate Verification of Authenticity*. *Electronics*, 13(11), 2132. Available from: https://doi.org/10.3390/electronics13112132

[29] Rana, P., & Bansal, S. (2024). *Exploring Deepfake Detection: Techniques, Datasets and Challenges*. *International Journal of Computing and Digital Systems*. Available from: https://doi.org/10.12785/ijcds/160156

[30] Altamimi, S., & Salameh, W. A. (2024). *Towards Analysis Detection of Deepfake Video via Deep Learning Models: A Review*. 87–92. Available from: https://doi.org/10.1109/ijcc64742.2024.10847278

[31] Helode, A., Yadav, A., Verma, V. P., & Srinivasa, K. G. (2024). *Fusion of Machine Learning and Deep Learning: A Hybrid Approach for Deepfake Detection*. 1–6. Available from: https://doi.org/10.1109/icccnt61001.2024.10724874

[32] Stankov, I., & Dulgerov, E. E. (2024). *Detection of Deepfake Images and Videos Using SVM, CNN, and Hybrid Approaches*. 1–5. Available from: https://doi.org/10.1109/et63133.2024.10721497

[33] Singh, D., Singh, P., & Bhandari, R. (2024). *Enhancing Deepfake Video Detection: A Hybrid CNN-LSTM Approach*. 130–135. Available from: https://doi.org/10.1109/tiacomp64125.2024.00031

[34] Dinçer, S., Ulutas, G., Ustubioglu, B., Tahaoglu, G., & Sklavos, N. (2024). *Golden ratio based deep fake video detection system with fusion of capsule networks*. *Computers & Electrical Engineering*. Available from: https://doi.org/10.1016/j.compeleceng.2024.109234