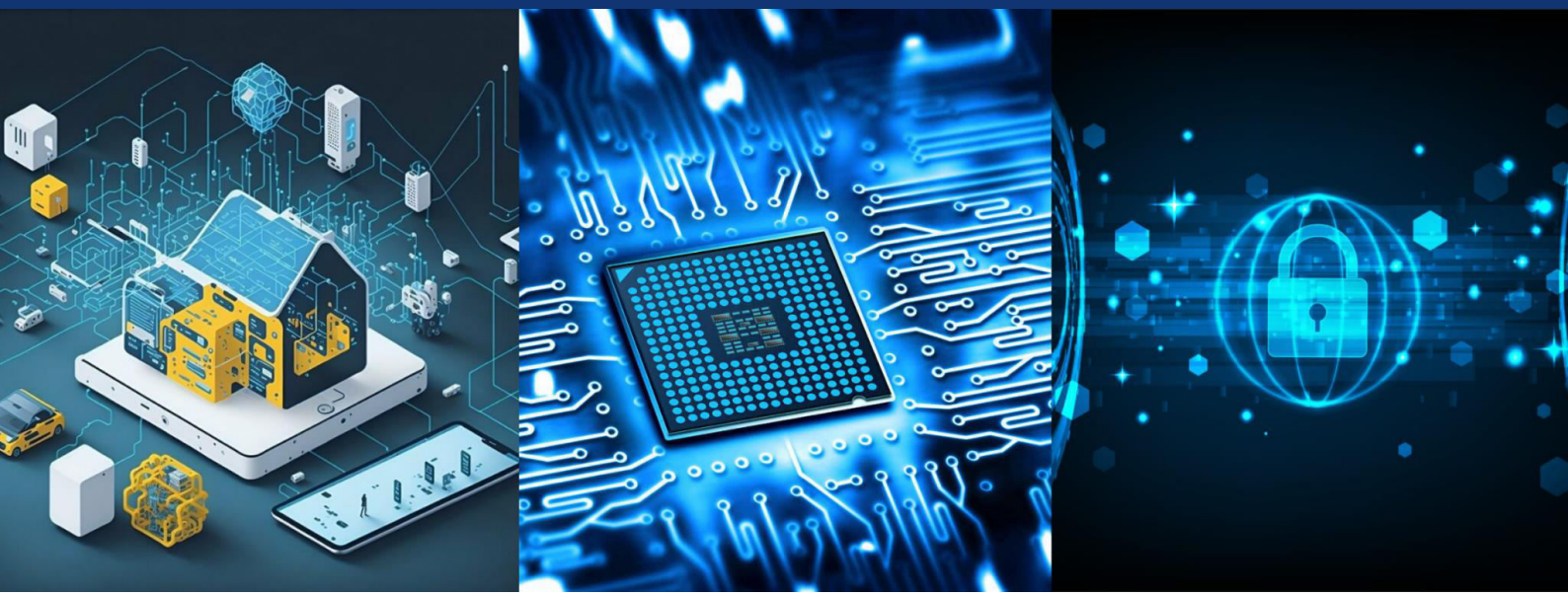
IJIRCCE

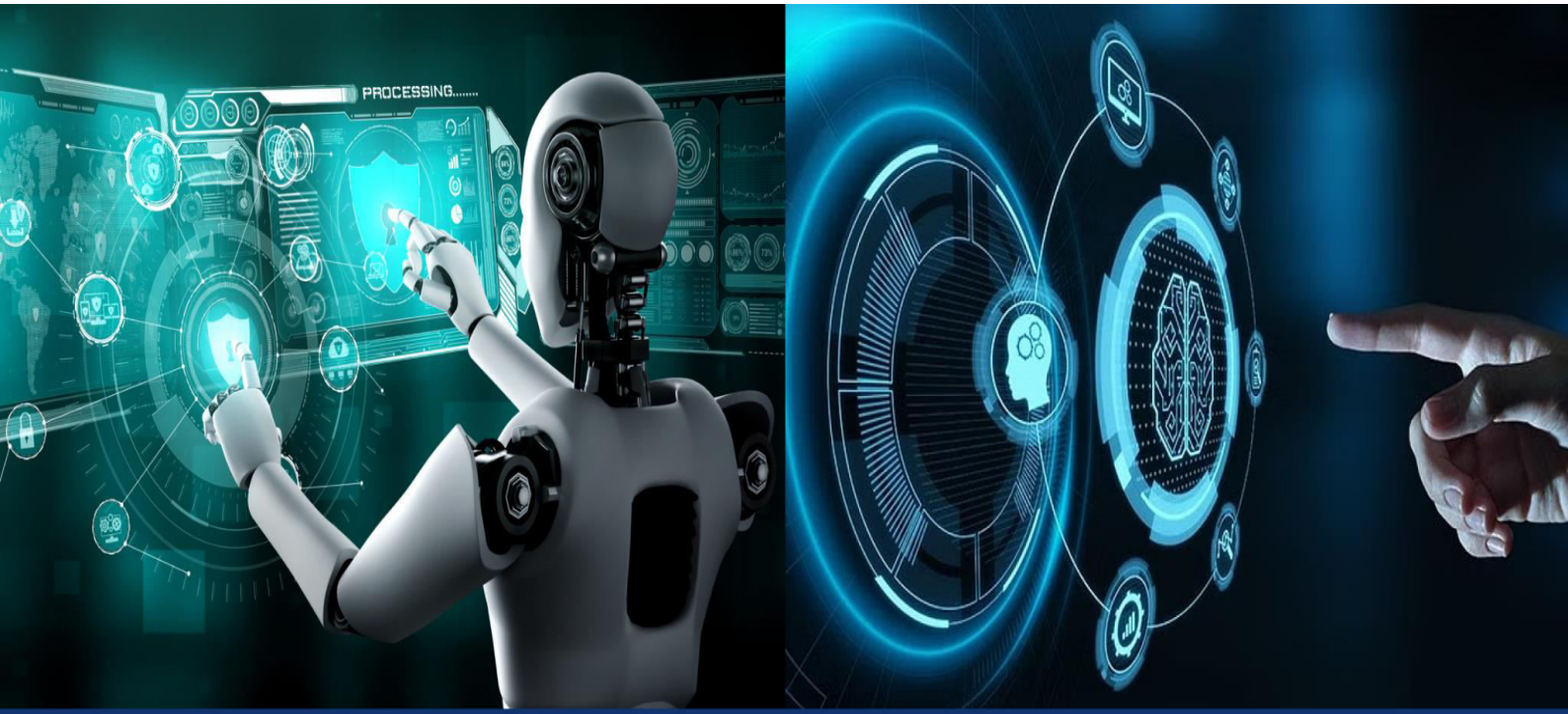
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

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



Impact Factor: 8.625

Volume 13, Issue 1, January 2025

© 2025 IJIRCCE | Volume 13, Issue 1, January 2025| DOI: 10.15680/IJIRCCE.2025.1301032

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

DeepDetect: A Multi-Modal Framework for Comprehensive Deepfake Detection

Nandini.N[1], Chidanandan V[2], Minal M[3], Nisarga K[3], Shameena[3], Kusuma.S[3]

Professor and HOD, Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bangalore, Karnataka, India[1]

Assistant Professor, Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bangalore, Karnataka, India[2]

Students, Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bangalore, Karnataka, India[3]

ABSTRACT: DeepDetect stands as a cutting-edge multi-modal detection system. It aims to fight the growing spread of deepfake content in audio, image, and video areas. As generative AI develops , deepfakes threaten digital security, media truth, and social trust. Current detection methods often focus on one type of media and have trouble adapting to different datasets and new fake techniques.To analyze audio, the system uses Mel Frequency Cepstral Coefficients (MFCC) with Support Vector Machines (SVM). This helps to examine spectral and time-based features. For images, it uses CNN structures like EfficientNet. These can spot small differences in texture, lighting, and facial details. Video detection combines CNN-RNN models to perform both space and time analysis. This catches odd things across video frames. These detection methods work together with a user-friendly interface. The interface uses Flask Streamlit, and React. This ensures users can access and interact with the system in real-time.Tests show the system is strong spotting fakes with over 90% accuracy in all types. DeepDetect's building-block setup makes it easy to grow, so it can work in busy places like fact-checking sites online security teams, and crime labs. Lots of trials prove it can handle real-life cases, including hard-to-spot changes and many different kinds of data.

KEYWORDS: Deepfake Detection, Multi-Modal Framework, Audio Forensics, Image Manipulation Detection, Video Anomaly Analysis, MFCC, Support Vector Machines (SVM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), EfficientNet, Temporal Analysis, Spatial Analysis, Media Integrity, Digital Security, Real-Time Detection.

I. INTRODUCTION

Generative AI has dramatically transformed the landscape of digital content creation, with unprecedented production of audio, images, and videos in unprecedented realism. On one hand, it has significantly changed entertainment, education, and creative industries, while, on the other hand, deepfake media pose serious threats-they are artificially created content meant to deceive. Deepfakes use highly advanced machine learning techniques to create realistic human-like voice, edit facial expressions and even manipulate videos. Most deepfakes have malicious intent; hence, the threat they pose is a risk to digital security, media integrity, and public trust.

The current techniques for deepfake identification are more based on a single media type, such as changed images or fake audio. However, the recent deepfake projects often include several types of formats, which demands a holistic approach capable of handling audio, images, and videos at the same time. Other problems like biased datasets, high fidelity changes, and real-time requirements further stress the limitations of such systems.

To overcome these challenges, the current study presents DeepDetect as a robust and extensible framework for deepfake detection, multi-modal analysis-specific. This system uses advanced machine learning and deep learning techniques. It includes Mel Frequency Cepstral Coefficients (MFCC) with Support Vector Machines (SVM) for audio detection, CNNs for identifying image manipulations, and hybrid CNN-RNN architectures for video content anomaly detection. The

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

modular architecture not only ensures flexibility for different media forms but also enables standalone implementation adapted to specific applications.

DeepDetect is further complemented by a user-friendly interface developed with Flask, Streamlit, and React, which encourages usability for users with technical expertise as well as those without. Extensive evaluations performed on a broad range of datasets demonstrate the robustness and outstanding accuracy of the system, achieving detection rates above 90% for various modalities.

II. RELATED WORK

Audio Deepfakes
Audio deepfakes can be tricky to spot because they sound so realistic. However, by focusing on subtle inconsistencies like unnatural changes in tone, pitch, or overall voice quality, they can be identified. One effective method involves extracting features from audio using MFCCs (Mel-Frequency Cepstral Coefficients) and analyzing them with advanced models like RNNs or CNNs. When paired with large datasets, these approaches can reliably separate real voices from fake ones.

Video Deepfakes
In video deepfakes, the devil is in the details. Look for oddities such as facial expressions that don't match the context, unnatural blinking patterns, or lip movements that are out of sync with speech. By studying individual frames with Convolutional Neural Networks (CNNs) and tracking motion over time using temporal models like LSTMs, inconsistencies across frames can be detected, exposing the fake.

Image Deepfakes
Images created using deepfake techniques often show subtle mistakes. These might include uneven lighting on the face, strange reflections in the eyes, or mismatched features like asymmetrical ears. Tools using CNNs are particularly effective here, as they analyze these images pixel by pixel. Other approaches focus on spotting unusual textures or irregularities in facial landmarks, which help reveal manipulations.

A Combined Approach for Better Detection
A single method may not always catch a sophisticated deepfake, but combining techniques can make detection much more reliable. This multimodal approach analyzes audio, video, and images together, allowing it to pick up on inconsistencies that may be missed when only one type of data is examined. Such strategies are becoming essential as deepfake technologies grow more advanced.

III. PROPOSED SYSTEM

The proposed system leverages a multimodal approach that combines audio, video, and image analysis to enhance accuracy in deepfake detection. Each type of input undergoes specialized processing to identify subtle inconsistencies often present in fake content. By integrating insights from all three modalities, the system offers a more reliable and robust detection framework, addressing the limitations of standalone methods.

Working Principle
The deepfake detection system works as follows:
Audio Processing: Audio clips are analyzed by extracting Mel-Frequency Cepstral Coefficients (MFCCs) to capture speech features. These are fed into machine learning models like CNNs or RNNs to classify the audio as 'real' or 'fake.'
Video Processing: The system detects and tracks faces in video frames, extracts spatial features using CNNs, and uses LSTMs to detect temporal inconsistencies, such as unnatural blinking or lip-syncing. The video is then classified as 'real' or 'fake.'
Image Processing: Still images are analyzed using CNNs to identify pixel-level artifacts, and facial landmarks are checked for misalignments. Based on these analyses, the image is classified as 'real' or 'fake.'

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

IV. SCOPE OF THE PROJECT

This project provides a holistic solution to the problem of deepfake detection by:

- Building a versatile system capable of detecting deepfakes across multiple media formats.
- Using advanced AI models to improve detection accuracy and adaptability.
- Offering seamless integration with user-friendly interfaces for practical deployment in various applications, such as media verification, cybersecurity, and forensic investigations.

This project aims to develop a robust deepfake detection system capable of identifying manipulated audio, video, and images using advanced AI models. The system leverages MFCC-based analysis for audio, CNNs and LSTMs for video temporal inconsistencies, and pixel-level anomaly detection with CNNs and facial landmark methods for images. Designed for real-time detection, it can be integrated into platforms like social media, news outlets, or video conferencing tools. With a focus on scalability and adaptability, the system is built to handle large data volumes and stay effective against evolving deepfake technologies, providing a versatile and practical solution for media verification and cybersecurity.

V. SYSTEM ARCHITECTURE

The system is designed to detect deepfake content across audio, video, and images using advanced AI models, integrating various stages from preprocessing to decision-making. It follows a modular and scalable architecture.

System Architecture Components:

Input Module: This is the entry point where multimedia content (audio, video, images) is received for analysis. It accepts input through different sources like file uploads, streaming services, or APIs.

Pre-Processing Module: Once the input data is received, it undergoes initial processing to prepare it for feature extraction. This involves steps such as noise removal in audio, resizing and detecting regions of interest in images, and extracting frames and adjusting resolution for video.

Feature Extraction Module: Key features from the pre-processed data are extracted. For audio, this includes speech features like MFCCs. For images, facial landmarks and texture inconsistencies are analyzed. Video analysis focuses on temporal inconsistencies and frame anomalies.

Classification Module: The extracted features are passed through classification models that determine whether the content is real or fake based on trained patterns. Machine learning frameworks like TensorFlow or PyTorch are used for this step.

Detection Module: After classification, the system makes a final decision. It uses a fusion algorithm to aggregate the results from different modalities (audio, video, image) and produces a confidence score to classify the content as either "Real" or "Fake."

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

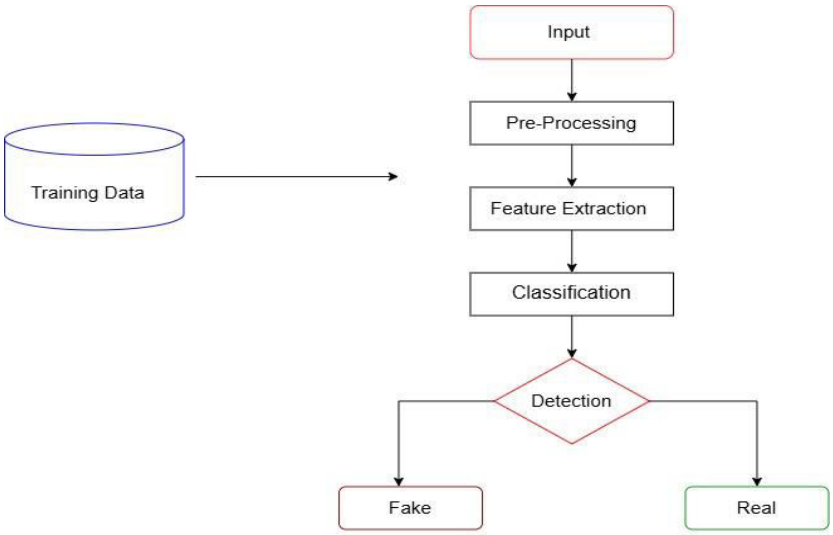


Fig:System Architecture

VI. METHODOLOGY

The research methodology outlines a structured approach to creating an effective deepfake detection system. It begins with a review of existing research, evaluating current generation and detection techniques, and identifying gaps. This review helps inform the choice of advanced methods like CNNs for image analysis, RNNs for video processing, and SVMs with MFCC features for audio classification.

The system adopts a multi-modal framework, with separate modules for audio, video, and image detection. Audio analysis involves extracting features such as MFCCs, pitch, and spectral contrast, while image processing is handled by CNN architectures like EfficientNet and ResNet to identify subtle visual anomalies. Video detection uses a hybrid CNN-RNN model to capture both spatial and temporal inconsistencies.

The system is trained and validated using diverse datasets like FaceForensics++, Celeb-DF, and DFDC, ensuring a wide range of scenarios are considered. Preprocessing steps, including normalization, resizing, and augmentation, improve the model's performance. The system is also designed for real-time detection, with optimized algorithms deployed on scalable cloud platforms like AWS or Azure.

Performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score, while robustness tests ensure the system can generalize well and resist adversarial attacks. A user-friendly interface, developed with frameworks like Flask or React, allows users to upload media files and view detailed reports, including confidence scores and detected anomalies.

Ethical concerns are addressed by prioritizing responsible AI practices and considering the societal implications of deepfake technology. The methodology also includes suggestions for future improvements, such as incorporating transformer models and enhancing real-time detection capabilities. This comprehensive plan ensures that the system is both accurate and scalable, making it suitable for a wide range of applications.

Ethical and Societal Topics

Deepfake technology can be harmful because it allows people to create fake content that looks and sounds real. This can hurt people's reputations, spread false information, and cause political problems. It's important for developers to make sure the technology is used responsibly. There have been real-life cases where deepfakes were used to cause harm, like in revenge porn and political manipulation, showing how important it is to have rules and transparency around the use of deepfakes.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

How Deepfakes Affect Trust in Media

Deepfakes make it hard to trust what we see and hear online. They can spread false information, especially in politics, where fake videos can change people's opinions. As deepfake technology improves, it's essential to protect the truth and prevent misinformation from spreading.

Case Study: Deepfake in Corporate Fraud – The "CEO Fraud" Incident

Deepfake technology was used in a fraud case in 2019. Scammers made a fake audio recording of a CEO's voice and tricked an employee into sending $243,000 to the wrong account. This shows how businesses can be vulnerable to such attacks and highlights the need for stronger security measures.

The Incident

The fraudsters used deepfake technology to make an audio recording of the CEO's voice sound real. The employee believed it and transferred the money. The fraud wasn't discovered until it was too late, resulting in financial and reputation damage for the company.

How It Happened

The scammers used advanced technology to make the CEO's voice sound real. The attack worked because there were no extra security checks, like voice recognition or multi-factor authentication, to verify the request.

What We Can Learn

From this incident, we can learn several important lessons:

1. **Use Multi-Factor Authentication (MFA):** Extra verification, like voice recognition, can stop fraudsters.
2. **Train Employees:** Employees should be taught about the dangers of deepfakes and how to verify requests.
3. **Use AI Tools for Detection:** AI tools can help detect deepfakes and prevent these attacks.
4. **Create Laws:** There should be laws to punish people who misuse deepfakes in business.

Conclusion

This "CEO fraud" case shows how dangerous deepfakes can be. Businesses need to improve security, train employees, and use AI tools to protect against fraud. By taking these steps, companies can reduce the risks of deepfake attacks and keep their reputation intact.

VII. CONCLUSION AND FUTURE ENHANCEMENT

The deepfake detection system marks a notable step forward in the battle against fake content in audio, images, and videos. Utilizing advanced machine learning techniques, it effectively identifies inconsistencies that suggest digital manipulation. Designed with user-friendliness in mind, this system caters to professionals from various backgrounds, making it accessible to a wide audience. While it already demonstrates impressive capabilities in detecting deepfakes, there are still many ways to enhance its performance. For example, improving real-time detection could increase its responsiveness, incorporating additional data sources would expand its analytical reach, and adapting the system for different languages and media formats would enhance its flexibility. These enhancements could significantly bolster its effectiveness in combating digital misinformation.

Future Improvement:

1. Real-Time Detection: Enhance the system's ability to analyze live video and audio streams with faster processing algorithms.
2. Integration of Modalities: Combine data from audio, video, and images to improve detection accuracy, particularly for more complex media types.
3. Adversarial Training: Implement methods to make the system more resilient against evolving deepfake generation techniques.
4. Cloud Integration: Utilize cloud platforms like AWS or Azure to expand the system's capacity for handling large datasets and accommodating high user volumes.
5. Advanced Architectures: Explore new machine learning models, such as transformers, to further enhance performance.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

6. Global Compatibility: Adapt the system to support a wider range of languages and datasets, making it useful on a global scale.
7. Interface Improvement: Refine the user interface to ensure that non-experts can easily navigate and use the system.
8. Ethical Collaboration: Work with policymakers to establish responsible guidelines for the use of AI in deepfake detection.
9. Forensic Adaptation: Modify the system for application in legal and cybersecurity investigations, broadening its scope to include digital forensics.

REFERENCES

1. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Use of Deep Learning to Detect Deepfakes. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
2. Explores the application of XceptionNet for detecting manipulated images in the FaceForensics++ dataset.
3. Wang, X., Zhang, Y., & Yu, H. (2020). A Hybrid CNN-LSTM Model for Deepfake Video Detection. Journal of Visual Communication and Image Representation.
4. Discusses combining CNNs and LSTMs to analyze spatial and temporal features in deepfake videos.
5. Zhao, Z., & Yan, L. (2022). Transformer-Based Methods for Deepfake Detection. Advances in Neural Information Processing Systems (NeurIPS).
6. Focuses on transformer models for detecting subtle temporal glitches in high-quality manipulated videos.
7. Smith, J., Liu, K., & Chen, X. (2020). Audio Deepfake Detection Using MFCC and Recurrent Neural Networks. IEEE Transactions on Audio, Speech, and Language Processing.
8. Examines the use of MFCC features and RNNs for detecting synthetic speech generated by voice cloning technologies.
9. He, X., Lin, Y., & Feng, S. (2023). Eye Gaze Consistency in Deepfake Image Detection. Pattern Recognition Letters.
10. Proposes a novel approach to detect deepfake manipulations by analyzing 3D gaze consistency.
11. Kim, J., & Park, T. (2020). Lighting and Shadow Mismatches in Image-Based Deepfakes. ACM Transactions on Multimedia Computing, Communications, and Applications.
12. Investigates physics-based methods for identifying manipulation in images through lighting and shadow inconsistencies.
13. Zhao, H., Xu, Z., & Wang, J. (2021). Domain Adaptation Techniques for Cross-Dataset Deepfake Detection. International Journal of Machine Learning and Applications.
14. Highlights the challenges of model generalization and proposes domain adaptation techniques for robust detection.
15. Nguyen, T. T., & Huynh, D. (2021). Real-Time Detection of Deepfake Content in Live Streams. Computers & Security.
16. Explores lightweight models for real-time detection of deepfake videos in live-streaming environments.
17. Kumar, A., & Singh, R. (2023). Multi-Modal Deepfake Detection: Challenges and Opportunities. Multimedia Tools and Applications.
18. Discusses integrating audio, image, and video modalities for improving robustness and accuracy in detection systems.
19. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems (NeurIPS).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details