IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.1412117

DEEPFAKE DETECTION: UNMASKING AI- GENERATED FORGERIES USING MACHINE LEARNING

Shiva Kumar D[1], A Saini[2], K Monica[3], Atiya Firdous[4]

Assistant Professor, CSE-Data Science, Rao Bahadur Y Mahabaleswarappa Engineering College, Ballari, VTU, Karnataka, India[1]

Students, CSE-Data Science, Rao Bahadur Y Mahabaleswarappa Engineering College, Ballari, VTU, Karnataka, India[2,3,4]

Abstract: Deepfake detection using machine learning is essential in safeguarding digital content authenticity. This project utilizes CNNs for image analysis, SVMs for audio verification, and Bayesian models for video scrutiny. By refining detection techniques, it ensures reliable identification of manipulated media and enhances digital security against evolving threats. system begins with data preprocessing, removing noise and extracting features essential for analysis. Machine learning models are trained on diverse datasets containing both genuine and synthetic content. Advanced classification algorithms then determine manipulation likelihood, continuously adapting to increasingly sophisticated deepfake generation methods for improved accuracy, By integrating multiple AI techniques, this project provides an automated solution for identifying manipulated content across various multimedia formats. Strengthening digital trust, it addresses growing concerns over misinformation while contributing to ethical AI applications that preserve content integrity, privacy, and authenticity in modern digital communication. The rapid growth of artificial intelligence has made it easier to create convincing fake media, posing serious risks in areas like politics, entertainment, and social media. As fake content becomes more widespread, effective detection methods are crucial. Current approaches struggle to keep up with evolving deepfake technologies, creating a need for reliable solutions. This paper proposes using convolutional neural networks to analyse facial features and motion inconsistencies in videos, aiming to improve detection accuracy. Additionally, audio analysis will be integrated to detect mismatches between sound and visuals, enhancing the model's effectiveness. The research emphasizes the importance of simple and effective methods to address the challenges of fake media.

Keywords: Deepfake detection, Machine learning, Digital content authenticity, Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), Image analysis, Video scrutiny, Data preprocessing, Feature extraction, Manipulation detection, Synthetic media, Classification algorithms, Adversarial threats, Multimedia forensics, Digital trust, Misinformation, Ethical AI, Content integrity, Privacy, Facial feature analysis, Motion inconsistencies, Audio-visual mismatch, Deepfake generation methods, Automated detection system, Reliable detection solutions.

I. INTRODUCTION

Deepfake detection using machine learning is essential in safeguarding digital content authenticity. This project utilizes CNNs for image analysis, SVMs for audio verification, and Bayesian models for video scrutiny. By refining detection techniques, it ensures reliable identification of manipulated media and enhances digital security against evolving threats. system begins with data preprocessing, removing noise and extracting features essential for analysis. Machine learning models are trained on diverse datasets containing both genuine and synthetic content. Advanced classification algorithms then determine manipulation likelihood, continuously adapting to increasingly sophisticated deepfake generation methods for improved accuracy By integrating multiple AI techniques, this project provides an automated solution for identifying manipulated content across various multimedia formats. Strengthening digital trust, it addresses growing concerns over misinformation while contributing to ethical AI applications that preserve content integrity, privacy, and authenticity in modern digital communication.

DeepGuard AI is a comprehensive deepfake detection application that combines cutting-edge machine learning with an intuitive web interface. Built with MobileNet transfer learning and TensorFlow 2.20.0, itprovides real-time detection for both images and videos with a modern, responsive user experience.

DOI: 10.17148/IJARCCE.2025.1412117

Deepfake Detection: Unmasking AI-Generated Forgeries focuses on identifying manipulated images and videos created using advanced AI techniques. Deepfakes pose serious threats by spreading misinformation, fraud, and identity misuse. As AI-generated content becomes more realistic, manual detection is no longer reliable. This project uses machine learning and deep learning methods to analyze facial and temporal inconsistencies. The goal is to accurately distinguish real media from AI-generated forgeries.

II. LITERATURE SURVEY

1. **Deepfake Detection Using Machine Learning Algorithms (2021) Authors**: Md. Shohel Rana Beddhu, Murali Andrew, H. Sung
This study focuses on detecting deepfakes using traditional machine learning techniques. The researchers extracted visual features from facial regions and applied feature selection and classification techniques. Various machine learning models were trained and tested on popular deepfake datasets such as FaceForensics++, DFDC, VFDD, and Celeb-DF.

Key Findings:

The results showed that traditional machine learning methods achieved accuracy of up to 99.84%. These methods were faster, more interpretable, and required fewer computational resources compared to deep learning models. However, their performance depended heavily on the quality of handcrafted features.

2. **Deepfake Detection Using Spatiotemporal Convolutional Networks (2021) Author**: Oscar de Lima
This research introduced spatiotemporal convolutional neural networks (CNNs) to capture both spatial and temporal information from videos. Instead of analyzing individual frames, the model studied motion inconsistencies and temporal artifacts across consecutive frames.

Key Findings:

The proposed approach outperformed existing frame-based detection methods. By analyzing video sequences instead of single images, the model achieved higher robustness and improved detection accuracy, especially for high-quality deepfake videos.

3. **Deepfakes Generation and Detection: State-of-the-Art, Challenges, Countermeasures, and Way Forward (2021)**

Author: Momina Masood

This paper provides a comprehensive survey of deepfake generation and detection techniques. It reviews various deep learning-based generation methods such as GANs and autoencoders and discusses multiple detection strategies including CNNs, frequency-domain analysis, and biological signal analysis.

Key Findings:

The study highlights major challenges such as dataset bias, generalization issues, and rapid improvements in deepfake generation. It also suggests future research directions, including multimodal detection and real-time detection systems.

III METHODOLOGY

1. **Data Collection & Preprocessing**

- Dataset Aggregation: Gather labeled datasets containing authentic and deepfake samples across different formats. Data Augmentation: Apply transformations (e.g., noise injection, compression artifacts) to enhance model generalization. Feature Extraction: Identify spatial (image), temporal (video), and spectral (audio) features.

2. **Multimodal Feature Fusion**

- Image & Video Analysis: Use CNNs and Vision Transformers to detect anomalies in pixel-level features and temporal inconsistencies. Audio Analysis: Apply recurrent networks (LSTMs/GRUs) and spectrogram-based classifiers to identify manipulated speech. Cross-Modal Fusion: Implement attention mechanisms or ensemble learning to integrate information across modalities.

DOI: 10.17148/IJARCCE.2025.1412117

3. **Model Training & Fine-Tuning**

- Adversarial Defenses: Train with GAN-generated deepfakes to improve robustness. Bayesian Calibration: Quantify uncertainty in predictions, enhancing classification confidence. Explainable AI Integration: Use interpretable methods (e.g., Grad-CAM, SHAP values) to improve trust in results.

4. **Detection & Classification Pipeline**

- Hierarchical Detection: First, classify content authenticity per modality; then, apply joint decision-making.
-
- Real-Time Adaptation: Implement a self-learning mechanism to evolve with new deepfake techniques. Deployment Considerations: Optimize for efficiency in edge devices and cloud environment.

Related Work in Deepfake Detection

1. Traditional Computer Vision Approaches

Early methods relied on inconsistencies in facial landmarks and temporal coherence. Limited effectiveness against sophisticated generation techniques. High computational overhead for real-time applications.

2. Deep Learning Based Methods

CNN architectures like ResNet, VGG for feature extraction. Transfer learning from ImageNet for improved performance. Ensemble methods combining multiple neural networks.

3. MobileNet Architecture

Howard et al. (2017) introduced depthwise separable convolutions. Significant reduction in parameters while maintaining accuracy. Optimal for mobile and edge deployment scenarios.

4. Current State-of-the-Art

Vision Transformers (ViTs) showing promising results. Attention mechanisms for focusing on manipulated regions. Multi-modal approaches combining audio and visual features.

Objectives

1. **Primary Objectives:**

- High-Accuracy Detection System Achieve >99% accuracy for image-based deepfake detection , Maintain >97% accuracy for video-based analysis, Implement confidence scoring for result reliability.
- Real-Time Processing Capability Process images in Real-Time Processing Capability, Process images in <2 seconds, Analyze video frames at 30fps, Optimize for CPU-based deployment.
- User-Friendly Web Interface, Develop responsive, mobile-compatible design, Implement drag-and-drop file upload functionality, Provide clear, interpretable results visualization.
- Production-Ready Deployment ,Ensure scalability for multiple concurrent users, Implement robust error handling and validation, Support various file formats and sizes.

2. **Secondary Objectives:**

- Educational Value, Demonstrate modern deep learning techniques, Showcase transfer learning implementation, Provide comprehensive documentation.
- Research Contribution, Benchmark performance against existing solutions, Document optimization techniques for mobile architectures, Validate effectiveness across diverse datasets.

Tools Used

Deep Learning Framework

- TensorFlow 2.20.0
- Keras 3.0 (integrated)
- MTCNN for face detection

DOI: 10.17148/IJARCCE.2025.1412117

Web Development

- Flask 3.0.3 (Backend framework)
- Bootstrap 5.3 (Frontend framework)
- HTML5, CSS3, JavaScript

Computer Vision

- OpenCV 4.10.0
- PIL (Python Imaging Library)
- NumPy for numerical operations

Development Tools

- Python 3.12
- VS Code IDE
- Git version control
- Virtual environment management

Software & Hardware Requirements

Software Requirements:

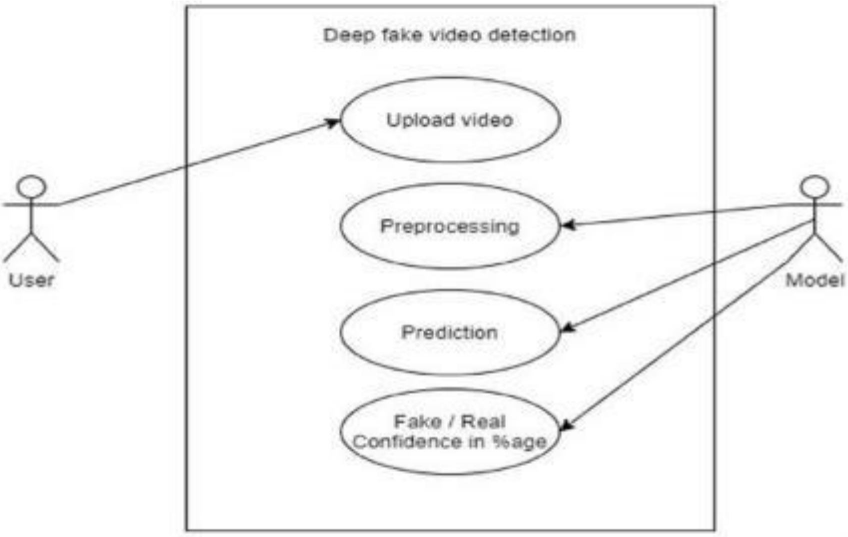
- Operating System: Windows 10/11, macOS 10.15+, Linux (Ubuntu 20.04+)
- Technology: Python 3.8+, TensorFlow 2.20.0
- Web Technologies: HTML5, CSS3, JavaScript, Bootstrap 5.3
- Web Server: Flask built-in server (development), Waitress (production)
- Database: File system based (no external database required)
- Python Version: 3.12 (recommended)

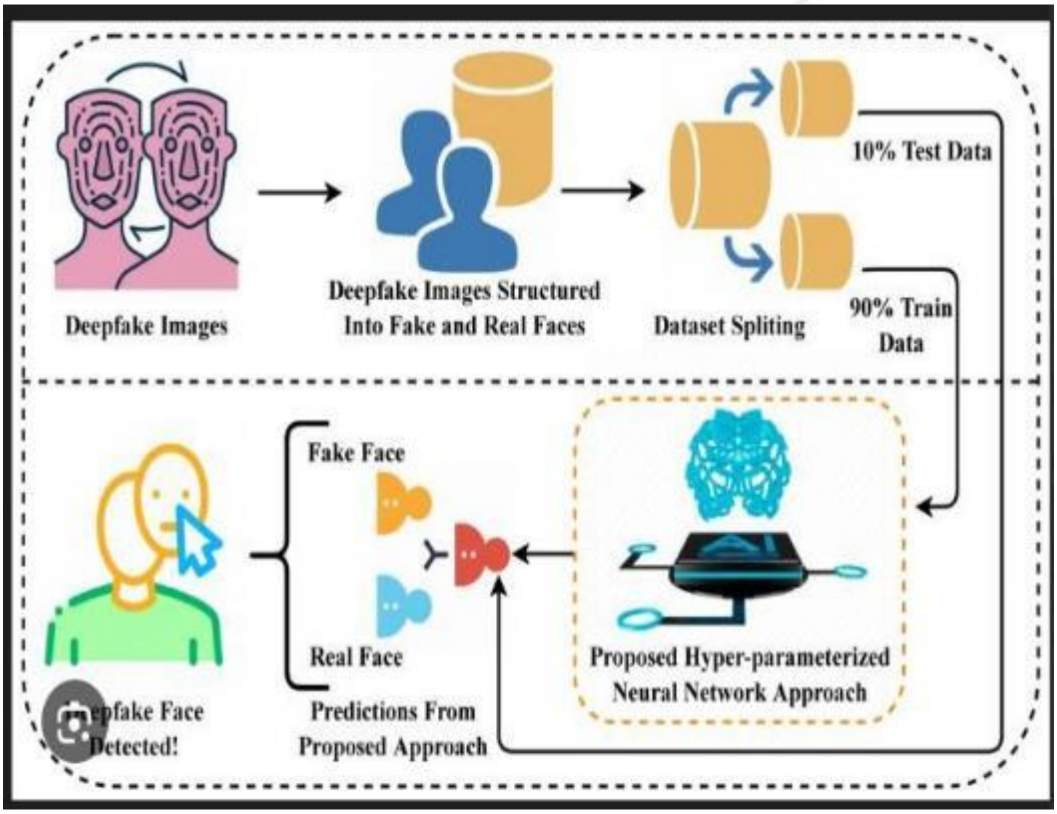
Hardware Requirements:

- Minimum: Intel i5/AMD Ryzen 5, 4GB RAM, 10GB storage
- Recommended: Intel i7/AMD Ryzen 7, 8GB+ RAM, SSD storage
- Optional: NVIDIA GPU (GTX 1060+) for acceleration

DOI: 10.17148/IJARCCE.2025.1412117

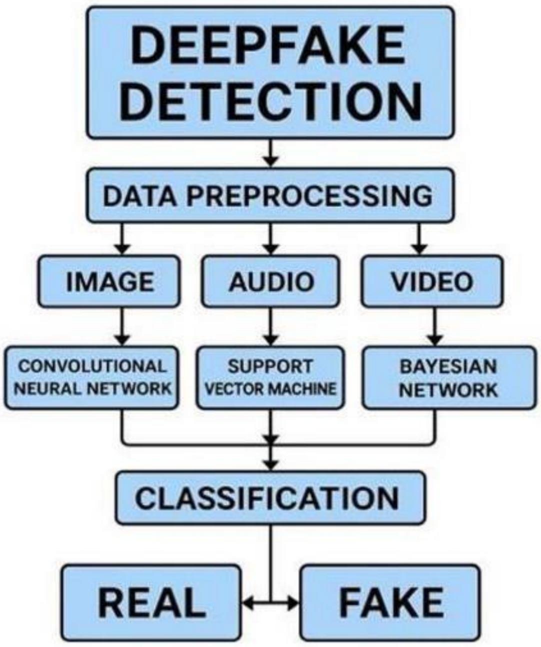
USE CASE DIAGRAM



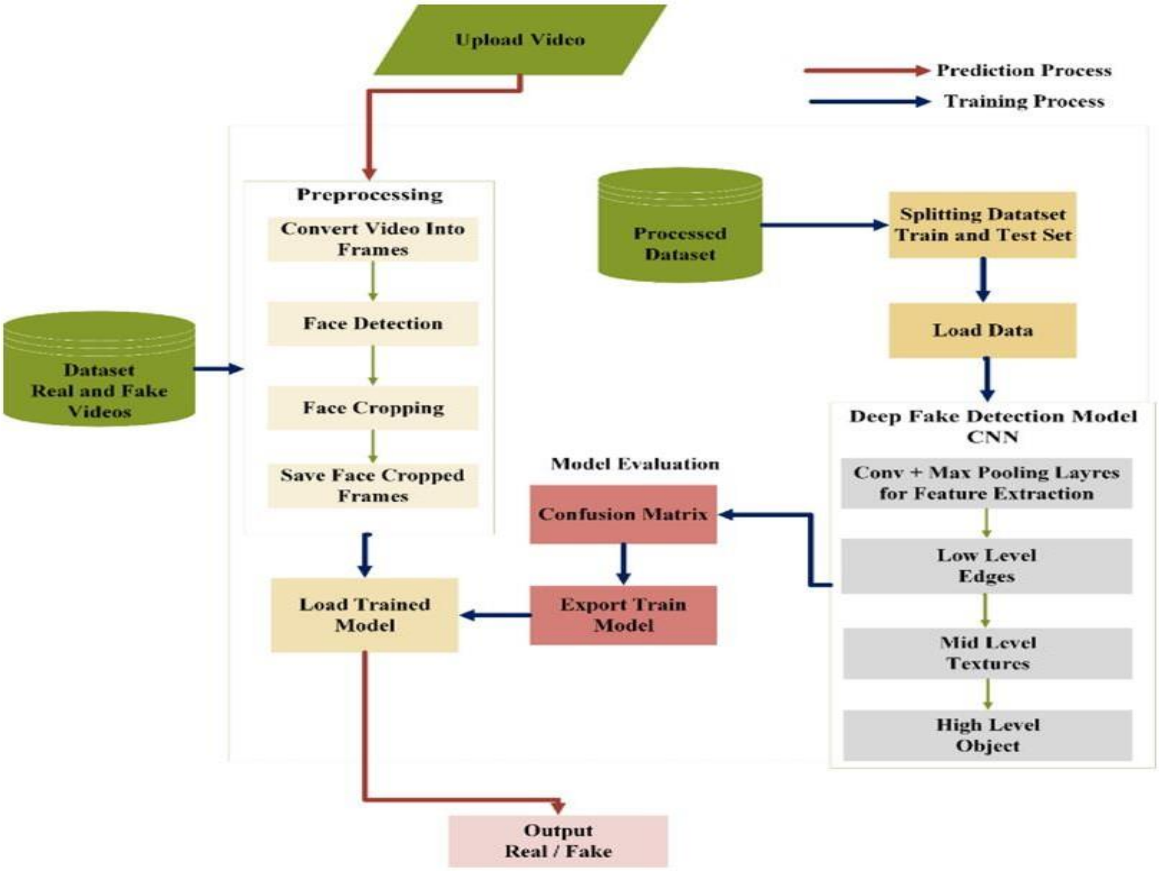


DOI: 10.17148/IJARCCE.2025.1412117

Deepfake Detection Workflow

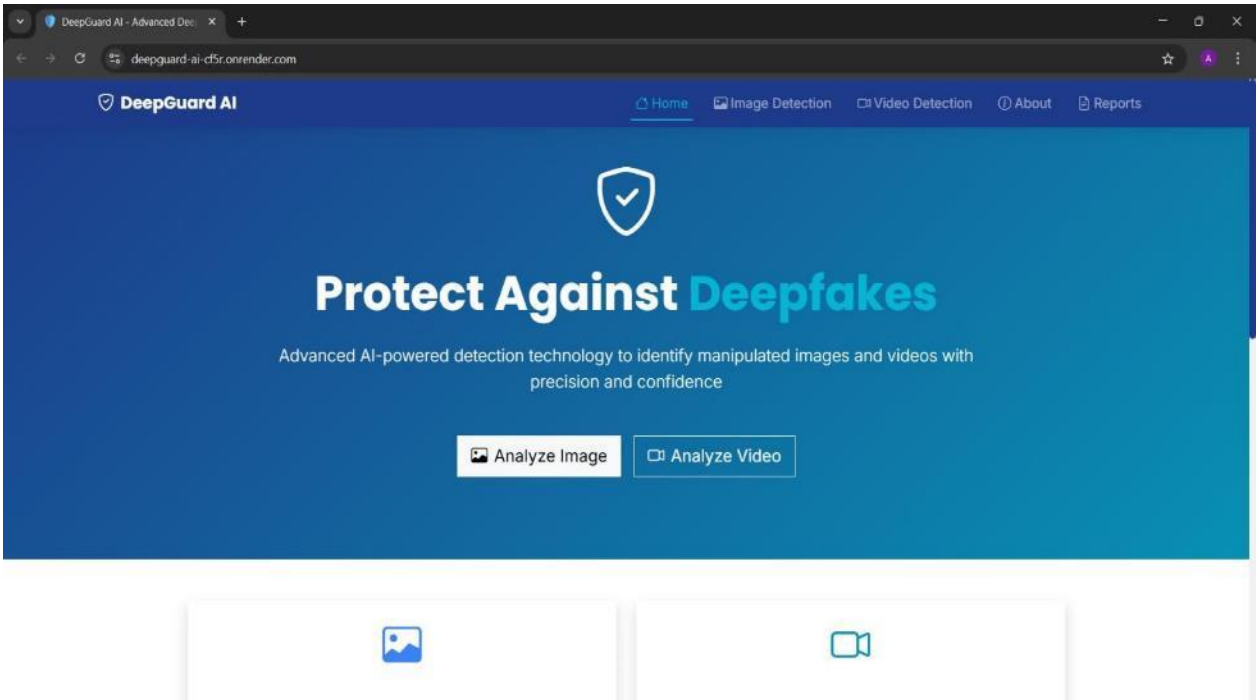


ER Diagram



DOI: 10.17148/IJARCCE.2025.1412117

HOME PAGE



Protect Against Deepfakes

Advanced AI-powered detection technology to identify manipulated images and videos with precision and confidence

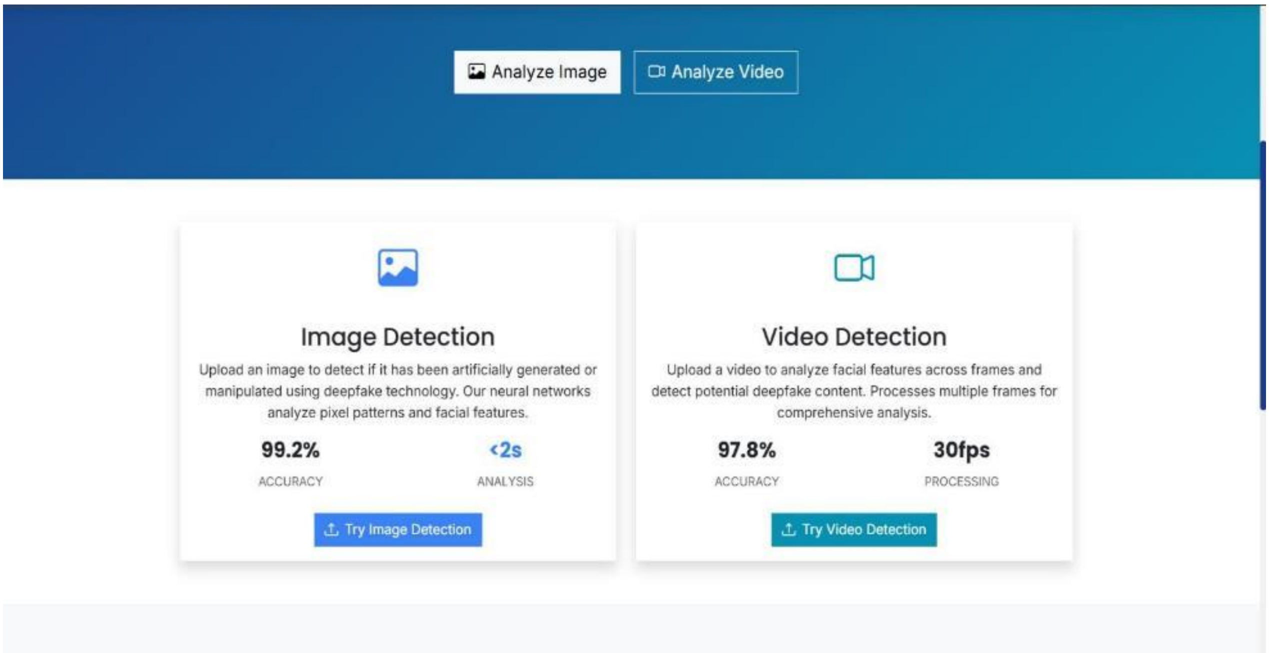


Image Detection

Upload an image to detect if it has been artificially generated or manipulated using deepfake technology. Our neural networks analyze pixel patterns and facial features.

99.2% ACCURACY

<2s ANALYSIS

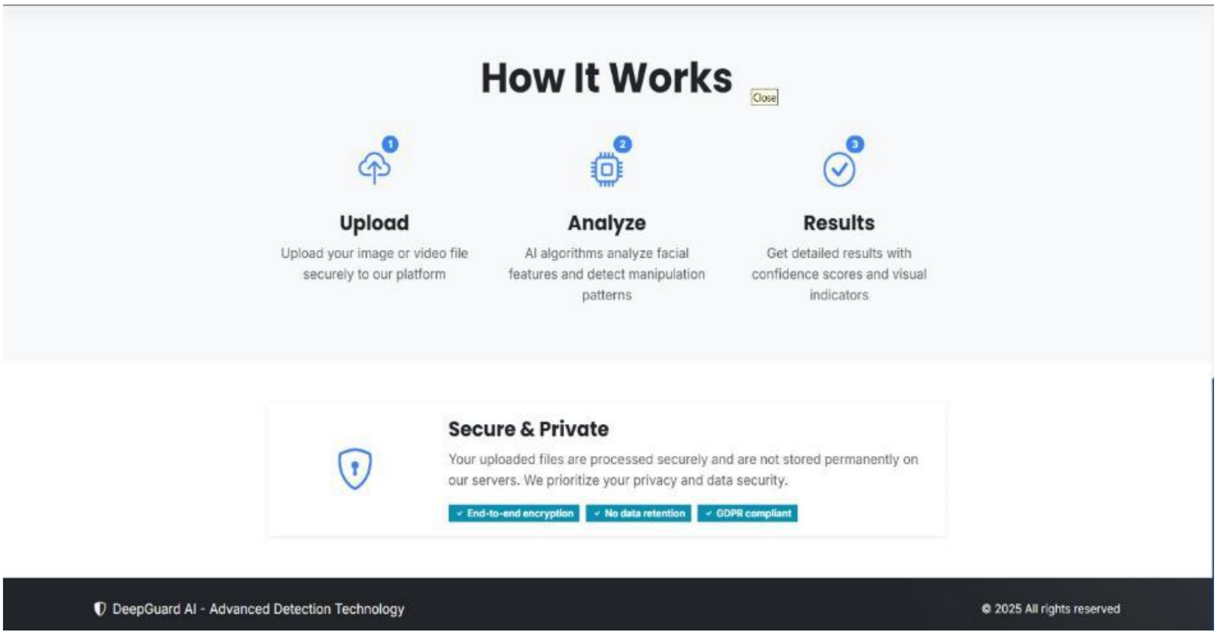
Video Detection

Upload a video to analyze facial features across frames and detect potential deepfake content. Processes multiple frames for comprehensive analysis.

97.8% ACCURACY

30fps PROCESSING

DOI: 10.17148/IJARCCE.2025.1412117



How It Works

1 Upload

Upload your image or video file securely to our platform

2 Analyze

AI algorithms analyze facial features and detect manipulation patterns

3 Results

Get detailed results with confidence scores and visual indicators

Secure & Private

Your uploaded files are processed securely and are not stored permanently on our servers. We prioritize your privacy and data security.

✓ End-to-end encryption
✓ No data retention
✓ GDPR compliant
DeepGuard AI - Advanced Detection Technology
© 2025 All rights reserved

Image & Video Detection

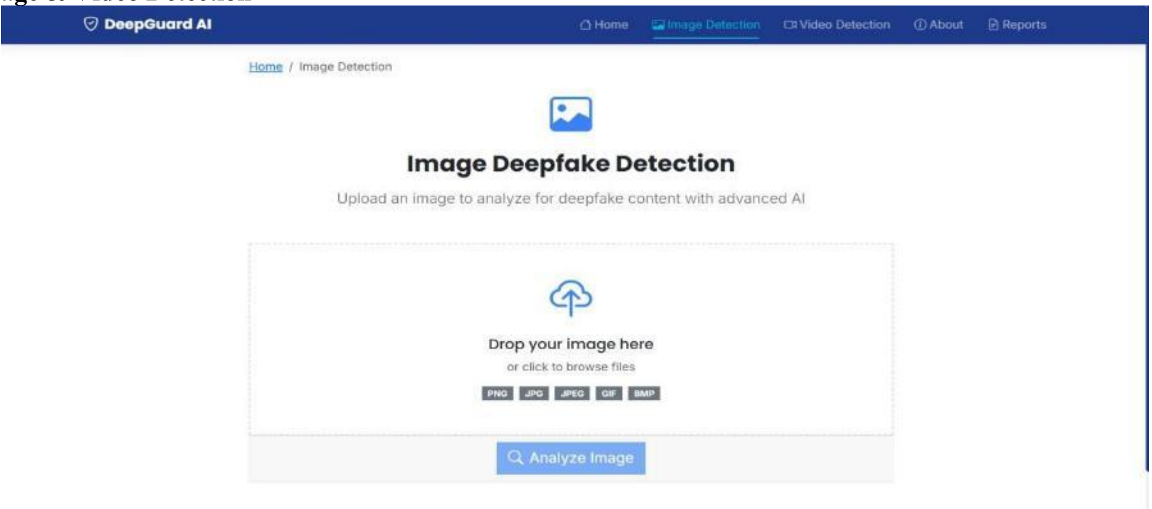
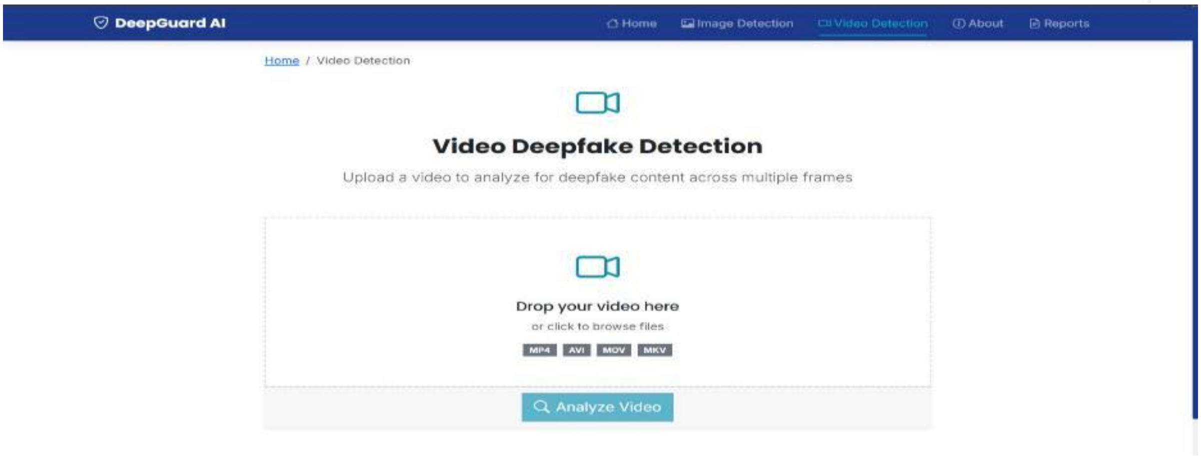


Image Deepfake Detection

Upload an image to analyze for deepfake content with advanced AI

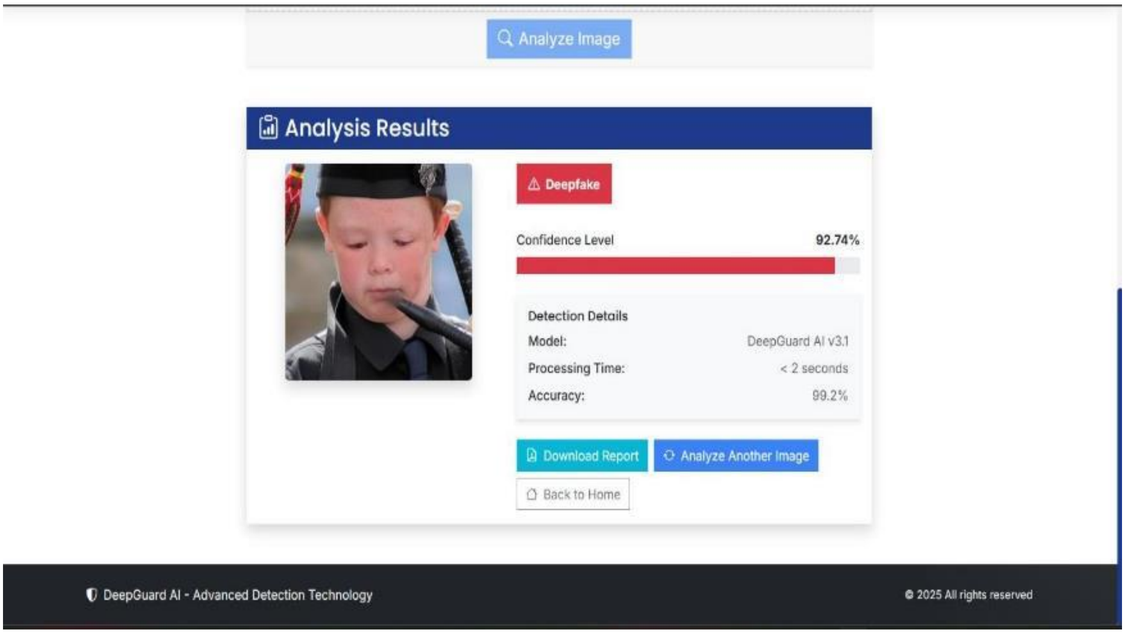


Video Deepfake Detection

Upload a video to analyze for deepfake content across multiple frames

DOI: 10.17148/IJARCCE.2025.1412117

Fake Image Detection Results



Confidence Level 92.74%

Detection Details

Model: DeepGuard AI v3.1

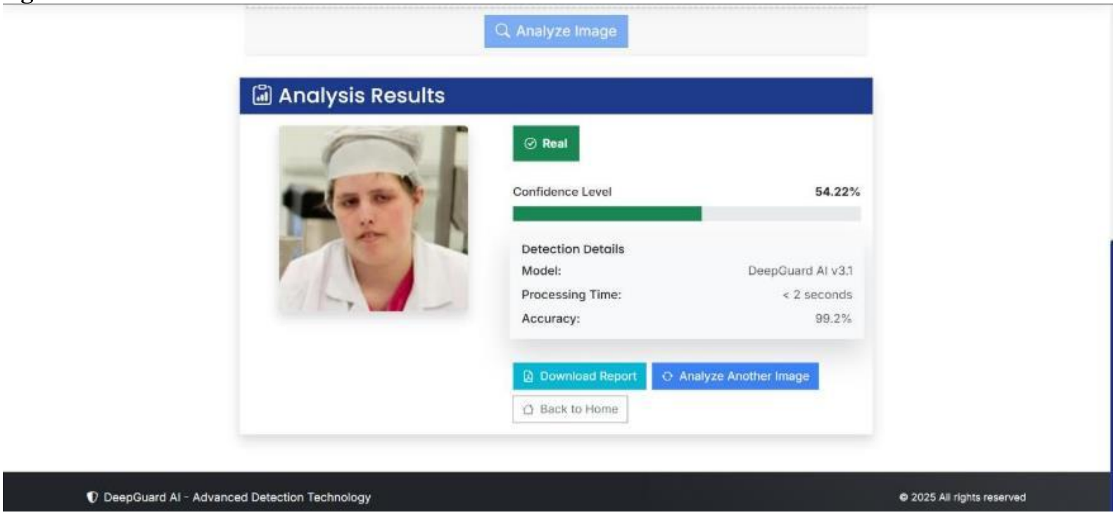
Processing Time: < 2 seconds

Accuracy: 99.2%

DeepGuard AI - Advanced Detection Technology

© 2025 All rights reserved

Real Image Detection Results



Confidence Level 54.22%

Detection Details

Model: DeepGuard AI v3.1

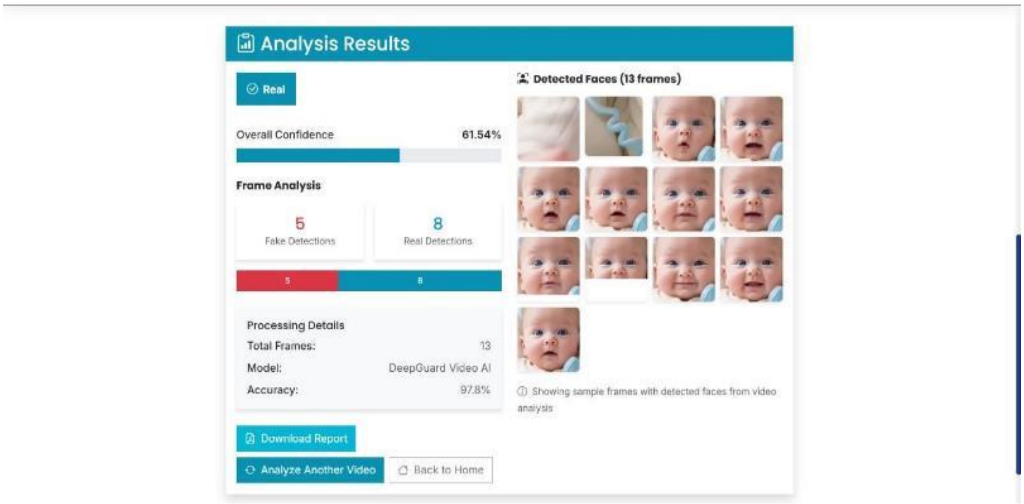
Processing Time: < 2 seconds

Accuracy: 99.2%

DeepGuard AI - Advanced Detection Technology

© 2025 All rights reserved

Video Detection Results



Overall Confidence 61.54%

Frame Analysis

5 Fake Detections

8 Real Detections

Processing Details

Total Frames: 13

Model: DeepGuard Video AI

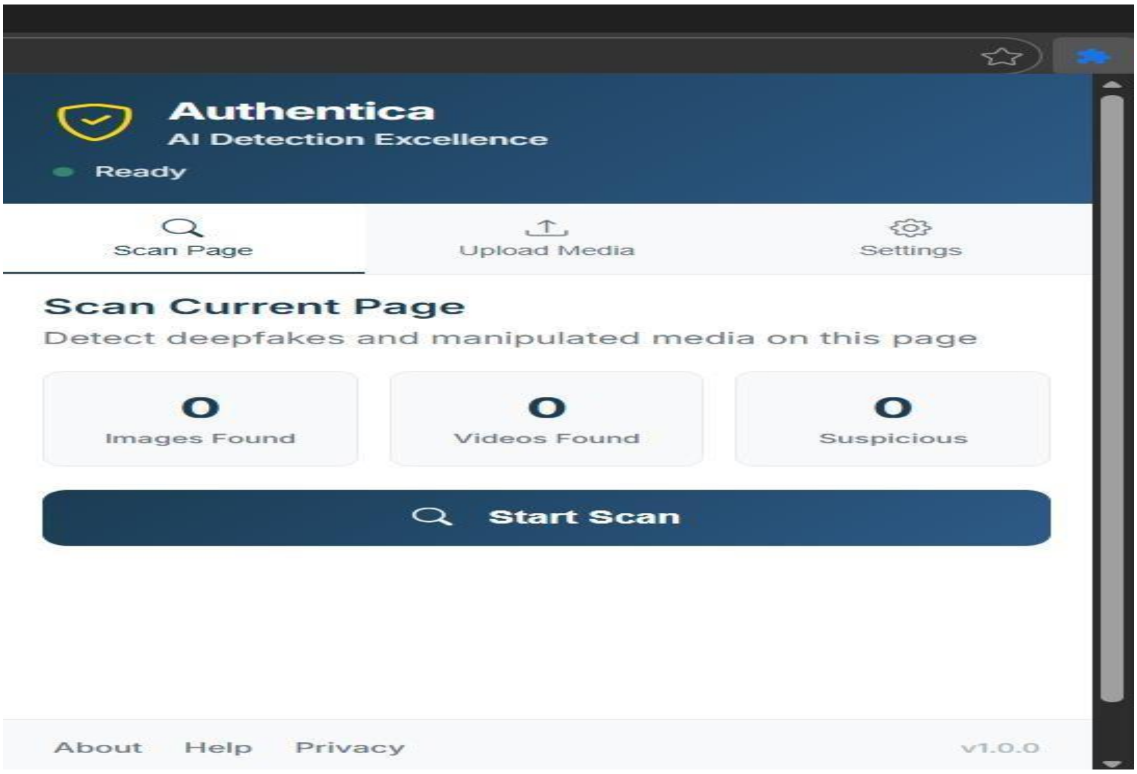
Accuracy: 97.8%

Detected Faces (13 frames)

ⓘ Showing sample frames with detected faces from video analysis

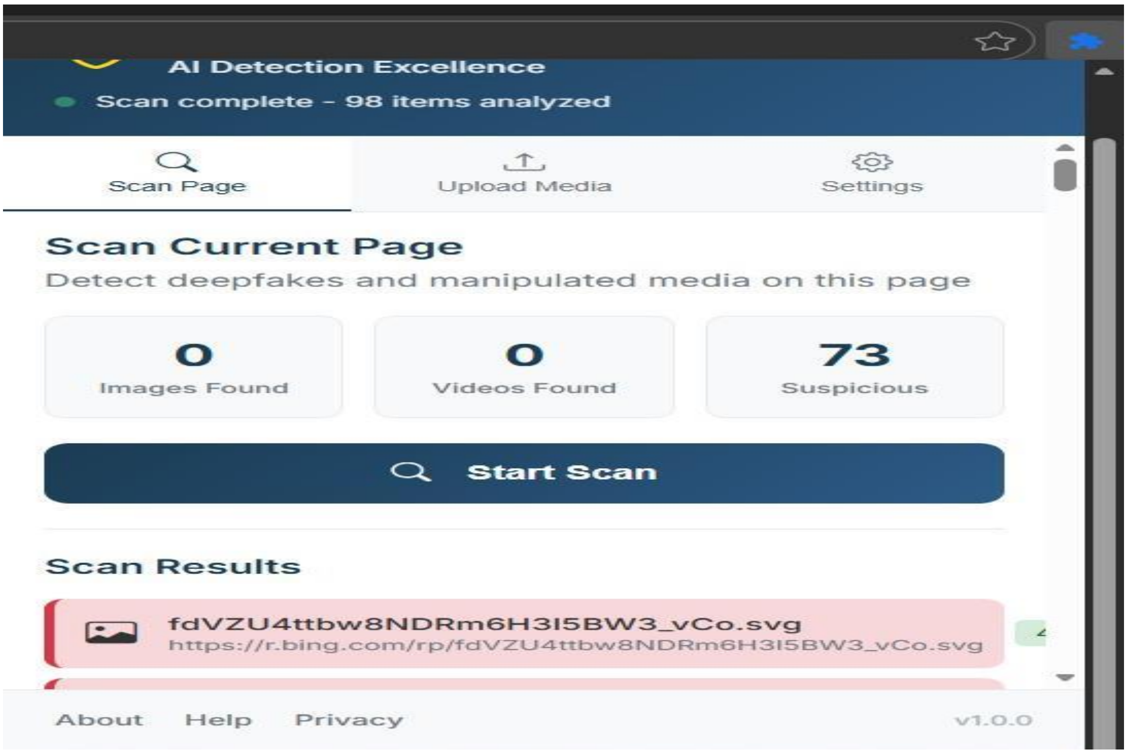
DOI: 10.17148/IJARCCE.2025.1412117

Extension Of Deepfake Detection through Online


Authentica
AI Detection Excellence
Ready

Scan Current Page
Detect deepfakes and manipulated media on this page

0 Images Found
0 Videos Found
0 Suspicious


AI Detection Excellence
Scan complete – 98 items analyzed

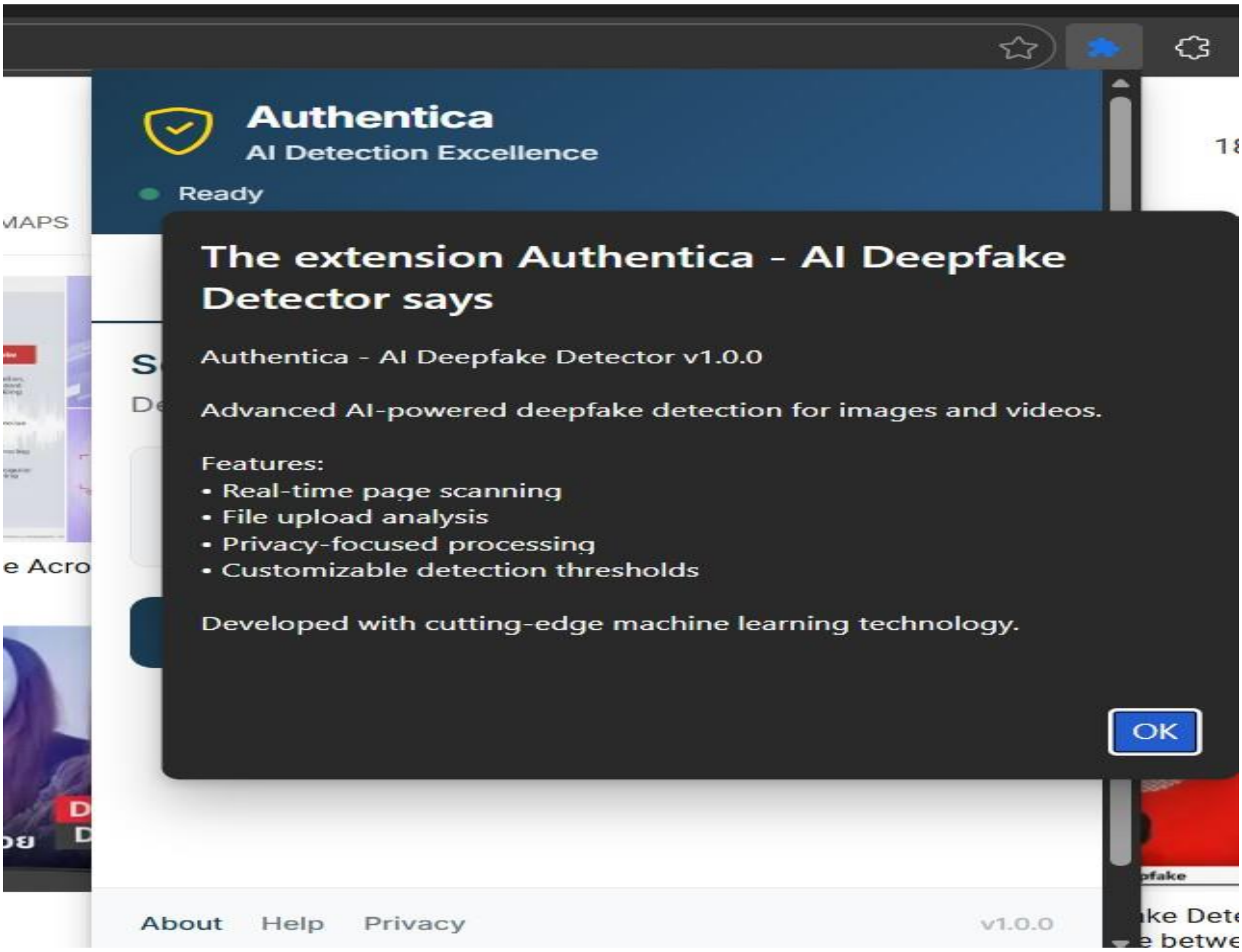
Scan Current Page
Detect deepfakes and manipulated media on this page

0 Images Found
0 Videos Found
73 Suspicious

Scan Results

fdVZU4ttbw8NDRm6H3I5BW3_vCo.svg
https://r.bing.com/rp/fdVZU4ttbw8NDRm6H3I5BW3_vCo.svg

DOI: 10.17148/IJARCCE.2025.1412117



The extension Authentica - AI Deepfake Detector says

Authentica - AI Deepfake Detector v1.0.0

Advanced AI-powered deepfake detection for images and videos.

Features:
- Real-time page scanning
- File upload analysis
- Privacy-focused processing
- Customizable detection thresholds

Developed with cutting-edge machine learning technology.

Link : https://deepguard-ai-cf5r.onrender.com/

The Deepfake Challenge:

Deepfake are synthetic media created using artificial intelligence, where a person appears to say or do things they never actually did. There AI-Generated videos and images can be incredibly convincing, making it difficult for the human eye to detect manipulation.

How DeepGuard AI Helps:

Our Solution analyzes digital media using advanced machine learning to detect artificial intelligence manipulation with high accuracy.

1. **Upload Media**: Simply upload an image or video file.
2. **AI Analysis**: our AI examines pixel patterns and inconsistences.
3. **Get Results**: Receive confidence scores and detailed analysis.

CONCLUSION

Deepfake detection in audio, video, and images is crucial for maintaining trust and authenticity in digital media. This project leverages machine learning models such as CNNs for image analysis, SVMs for audio verification, and Bayesian Networks for video detection. These models work together to identify manipulated content by analyzing subtle inconsistencies that may not be visible to the human eye. As deepfake technology continues to evolve, the system continuously refines its detection methods to stay ahead of emerging threats.

DOI: 10.17148/IJARCCE.2025.1412117

By integrating adaptive learning techniques, it enhances accuracy and reliability, ensuring privacy, security, and digital integrity. The project also supports real-time processing, allowing swift identification of deepfake content to prevent misinformation from spreading across online platforms. Deepfake detection through machine learning has become a critical area of research and development, driven by the rapid proliferation of AI-generated forgeries. By leveraging techniques like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and ensemble methods, detection systems can identify subtle artifacts and temporal inconsistencies in deepfakes.

Key advancements include:

1. **Robust Feature Extraction**: Analyzing spatial, temporal, and frequency-domain features to capture manipulation traces.
2. **Model Generalization**: Ensuring systems can detect deepfakes from unseen sources or generation methods.
3. **Real-Time Detection**: Optimizing models for low-latency applications in social media, news, and security. Despite progress, challenges remain, such as adversarial attacks, limited datasets, and the arms race between detection and generation techniques. Future work may focus on multimodal detection, explainability, and integrating detection systems with content provenance solutions like digital watermarking or blockchain.

Next Steps: - Explore multimodal detection (e.g., combining visual and audio cues). - Investigate adversarial training to improve model robustness. - Develop real-world datasets for training and e

Future Enhancements

Technical Improvements

- Advanced Architectures: Implementation of Vision Transformers (ViTs)
- Real-time Processing: WebRTC integration for live video analysis
- Multi-modal Detection: Audio-visual combined analysis
- Edge Deployment: Model optimization for mobile devices

Feature Additions

- API Development: RESTful API for third-party integration
- Batch Processing: Multiple file upload and analysis
- User Management: Authentication and user-specific history
- Advanced Reporting: Detailed analysis reports with visualizations

Scalability Enhancements

- Microservices Architecture: Containerized deployment with Docker/Kubernetes
- Cloud Integration: AWS/Azure deployment with auto-scaling
- Database Integration: PostgreSQL/MongoDB for result storage
- Caching System: Redis implementation for performance optimization

Research Directions

- Adversarial Robustness: Defense against adversarial attacks on detection models
- Explainable AI: Integration of LIME/SHAP for result interpretation
- Federated Learning: Privacy-preserving distributed model training

Challenges Overcome

Technical Challenges

- TensorFlow compatibility issues resolved with Keras 3.0 integration
- Memory optimization for CPU-based deployment
- Real-time face detection implementation with MTCNN
- Web application responsiveness optimization

Implementation Challenges

- Model size optimization while maintaining accuracy

DOI: 10.17148/IJARCCE.2025.1412117

- Cross-platform compatibility ensuring
- User interface design for technical and non-technical users
- Production deployment configuration

REFERENCES

[1]. Korshunov, P., & Marcel, S. (2018). Deepfakes: A New Threat to Face Recognition? Assessment and Detection. arXiv preprint arXiv:1812.08685. https://arxiv.org/abs/1812.08685

[2]. Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: a Compact Facial Video Forgery Detection Network. IEEE International Workshop on Information Forensics and Security (WIFS). https://arxiv.org/abs/1809.00888

[3]. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos. ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). https://ieeexplore.ieee.org/document/8683164

[4]. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection. Information Fusion, 64, 131–148. https://doi.org/10.1016/j.inffus.2020.07.007

[5]. Gandhi, K., Kulkarni, P., Shah, T., Chaudhari, P., Narvekar, M., & Ghag, K. (2024). A Multimodal Framework for Deepfake Detection. arXiv preprint arXiv:2410.03487. https://arxiv.org/abs/2410.03487

[6]. Wodajo, D., & Atnafu, S. (2021). Deepfake Video Detection Using Convolutional Vision Transformer. arXiv preprint arXiv:2102.11126. https://arxiv.org/abs/2102.11126

[7]. Saikia, P., Dholaria, D., Yadav, P., Patel, V., & Roy, M. (2022). A Hybrid CNN-LSTM Model for Video Deepfake Detection by Leveraging Optical Flow Features. arXiv preprint arXiv:2208.00788. https://arxiv.org/abs/2208.00788

[8]. Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2020). Emotions Don't Lie: An Audio-Visual Deepfake Detection Method Using Affective Cues. arXiv preprint arXiv:2003.06711. https://arxiv.org/abs/2003.06711

[9]. Howard, A. G., et al. (2017). "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv preprint arXiv:1704.04861.

[10]. Rossler, A., et al. (2019). "FaceForensics++: Learning to Detect Manipulated Facial Images." Proceedings of the IEEE/CVF International Conference on Computer Vision.