MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

International Journal of Electrical, Electronics and Computer Systems

ISSN: 2347-2820

Volume 15 Issue 01 (2026) | Pages: 125-133

Deep Fake detection Using Deep Learning for Images, Videos & Audios

Jalindar Nivrutti Ekatpure[1], Suryajit S. Ingavale, Rohit S. Jagtap[3], Neelam S. Jachak[4], Rutuja M. Kandekar[5]

[1,2,3,4,5]Dept.of Computer Engineering, SBPCOE,Indapur

[1]j.ekatpure@gmail.com , [2]suryajitingavale1002@gmail.com , [3]jagtaprohit652@gmail.com , [4]neelamjachak24@gmail.com , [5]rutujakandekar00@gmail.com

Peer Review Information

***Type**: Article*
***Received**: 24 March 2026*
***Revised**: 09 April 2026*
***Accepted**: 27 May 2026*
***Published**: 06 June 2026*

Abstract

Deepfake detection has emerged as a significant challenge in the field of artificial intelligence due to the rapid advancement of generative models such as Generative Adversarial Networks (GANs). These techniques enable the creation of highly realistic manipulated media, including images, videos, and audio, which can be misused for misinformation, identity theft, and cybercrime.

This project presents a deep learning-based system for detecting deepfake content using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). The system is designed to process multi-modal inputs such as images, videos, and audio, and performs both spatial and temporal analysis to identify inconsistencies in manipulated media.

The system provides a user-friendly interface for uploading media files, analyzing authenticity, and generating results with high accuracy. The proposed solution improves detection reliability, reduces manual effort, and enhances digital media security. The paper concludes with experimental evaluation and discusses future improvements such as real-time detection and enhanced robustness against advanced deepfake techniques.

Keywords: Deepfake Detection; Deep Learning; CNN; LSTM; GANs; Image Processing; Video Analysis; Audio Analysis; Cybersecurity.

How to Cite This Article

Ekatpure, J. N., Ingavale, S. S., Jagtap, R. S., Jachak, N. S., & Kandekar, R. M. (2026). Deep fake detection using deep learning for images, videos & audios. *International Journal of Electrical, Electronics and Computer Systems*, 15(1), 125–133.

Introduction

Deep technology is one of the most advanced applications of artificial intelligence, allowing the generation of highly realistic synthetic media. With the help of deep learning models, especially GANs, it is possible to manipulate facial expressions, voices, and entire videos, making them appear authentic. However, the misuse of this technology has raised serious concerns in areas such as social media, politics, cybersecurity, and digital forensics. Deepfake media can be used to spread misinformation, impersonate individuals, and manipulate public opinion. Traditional detection methods are no longer effective against modern deepfake techniques due to their

complexity and realism. Therefore, there is a need for an automated system that can detect deepfake content accurately and efficiently. This project proposes a deep learning-based system that uses CNN for spatial feature extraction and LSTM for temporal analysis. The system supports image, video, and audio inputs, making it a comprehensive solution for detecting deepfake media.

Literature Survey

1. Recent Advances and Challenges of Deepfake Detection – Ran He et al. (2023): This study provides a comprehensive overview of modern deepfake detection techniques, including CNN-based, RNN-based, frequency-domain, and biological signal-based methods. The authors highlight that although detection accuracy has improved significantly, most systems fail when tested on unseen datasets or adversarial examples. The paper emphasizes the need for models that generalize well across datasets and are robust against evolving deepfake generation techniques.

2. FaceForensics++: Learning to Detect Manipulated Facial Images – Rössler et al. (2019): The authors introduced the FaceForensics++ dataset, a widely used benchmark for evaluating deepfake detection models. They tested several CNN architectures, including XceptionNet, achieving high accuracy under controlled conditions. However, their findings revealed that model performance drops when applied to different datasets, indicating the need for better generalization and diverse training data.

3. MesoNet: A Compact Facial Video Forgery Detection Network – Afchar et al. (2018): This research proposed MesoNet, a lightweight CNN architecture designed for efficient deepfake detection. Models like Meso-4 and MesoInception-4 achieve reasonable accuracy while maintaining low computational complexity, making them suitable for real-time applications. However, their shallow architecture limits their ability to capture complex features in highly realistic deepfakes.

4. Two-Stream Network for Tampered Face Video Detection – Li & Lyu (2019): Li and Lyu introduced a two-stream CNN framework that captures both spatial and temporal features. One stream processes image-level information, while the other captures motion inconsistencies using optical flow. This approach improves detection in video sequences but increases computational complexity, making real-time deployment challenging.

5. LipForensics: Using Visual Speech for Deepfake Detection – Zhao et al. (2021): This work focuses on detecting inconsistencies between lip movements and speech. By applying temporal CNNs to the mouth region, the model identifies mismatches in audio-visual synchronization. The approach shows strong performance in detecting manipulated videos but struggles with highly refined deepfakes.

6. Face X-ray: A Simple, Generalizable Deepfake Detection Method – Wang et al. (2020): The Face X-ray method detects blending artifacts left during face-swapping operations. It uses boundary-based supervision to improve generalization across different manipulation techniques. While effective, its performance decreases under heavy compression or low-quality video conditions.

7. Detection of GAN-Generated Faces Using Color Cues – McCloskey & Albright (2019): This study explores the use of abnormal color statistics as indicators of GAN-generated images. By combining CNNs with color distribution analysis, the method effectively distinguishes fake images under ideal conditions. However, its performance degrades when images are compressed or noisy.

8. Learning to Detect Fake Face Images in the Wild – Yang, Li & Lyu (2019): The authors addressed real-world challenges by introducing domain adaptation techniques to improve model robustness across different environments. Their CNN-based approach performs better under varying lighting, backgrounds, and noise conditions, highlighting the importance of handling real-world variability.

9. Vision Transformers for Deepfake Detection – Coccomini et al. (2021): This research explored the application of Vision Transformers (ViTs) for deepfake detection. Transformers are capable of capturing long-range dependencies in video frames, improving temporal modeling. Despite their effectiveness, they require high computational resources, limiting scalability.

10. A Survey on Deepfake Video Detection – Multiple Authors (2021): This survey reviews a wide range of detection methods, including CNNs, RNNs, frequency-based, and biological signal approaches. The authors highlight the lack of standardized benchmarks and evaluation protocols, making it difficult to compare different methods fairly. They emphasize the need for unified datasets and evaluation standards.

Limitations Of Existing Work

Despite significant advancements, existing deepfake detection methods have several limitations:

- Limited ability to generalize across real-world datasets
- High computational complexity and resource requirements
- Weak temporal analysis in video-based detection
- Lack of integrated multimodal detection systems
- Reduced performance against advanced GAN-based deepfakes
- Insufficient robustness to compressed or low-quality media

Problem Statement

Deepfake technology uses deep learning algorithms to create realistic fake images, videos, and audio. These manipulated media contents are becoming increasingly difficult to distinguish from authentic content. Deepfakes can be misused for spreading misinformation, fraud, identity manipulation, and other malicious activities. They pose significant risks in social media, politics, cybersecurity, and digital communication. Traditional detection methods often struggle to identify advanced deepfake content effectively, creating a need for automated and intelligent detection systems. Deep learning techniques can efficiently analyze visual, temporal, and audio features to identify manipulated media. Such systems detect inconsistencies in facial expressions, video frames, and speech patterns to improve detection accuracy. A robust deepfake detection model should operate across images, videos, and audio formats, ensuring authenticity verification and maintaining trust in digital media.

Proposed System

The proposed system is a comprehensive Deepfake Detection System using Deep Learning, designed to automatically detect manipulated media across multiple formats including images, videos, and audio. The system leverages advanced artificial intelligence techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to analyze both spatial and temporal inconsiste The system begins with the input stage, where users upload media files through a web-based or desktop interface. These inputs may include static images, video clips, or audio recordings. The system is designed to handle different formats and resolutions, ensuring flexibility in real-world applications.

Once the input is received, the data is passed to the preprocessing module, which standardizes the input for further analysis. For images, preprocessing includes resizing, normalization, and noise reduction. For videos, the system extracts frames at regular intervals, converting the video into a sequence of images. For audio inputs, the signal is transformed into spectrograms to capture frequency-based features. The core component of the system is the Deep Learning Engine, which consists of CNN and LSTM models. The CNN is responsible for extracting spatial features such as texture inconsistencies, facial artifacts, and blending irregularities commonly found in deepfake images. The LSTM model processes sequential data, making it suitable for detecting temporal inconsistencies in videos and audio, such as unnatural motion, lip-sync errors, and irregular speech patterns.The system also includes a feature fusion mechanism, where spatial and temporal features are combined to improve detection accuracy. This hybrid approach enhances the system's ability to detect both simple and complex deepfake manipulations.

A classification module is used to determine whether the input media is real or fake. The model outputs a probability score along with the final classification result. This helps users understand the confidence level of the prediction.The system is integrated with a user interface module, allowing users to upload media, view results, and analyze outputs. The results can also be stored and exported for further use. Overall, the proposed system provides an efficient, scalable, and accurate solution for detecting deepfake content, making it suitable for applications in cybersecurity, digital forensics, and media verification.

Architecture Description

The system architecture is divided into multiple modules that work together to process and analyze media inputs efficiently.

1. User Module: This module represents the end users interacting with the system. Users can upload media files such as images, videos, and audio for deepfake detection. The system is designed to be user-friendly and accessible to both technical and non-technical users.

2. Web Interface Module: The web interface provides a platform for users to interact with the system. It includes functionalities such as file upload, result display, and report generation. The interface ensures smooth communication between the user and the backend processing modules.

3. Preprocessing Module
This module prepares the input data for analysis:

- Image resizing and normalization
- Video frame extraction
- Audio spectrogram generation

This step ensures consistency in data format and improves model performance.

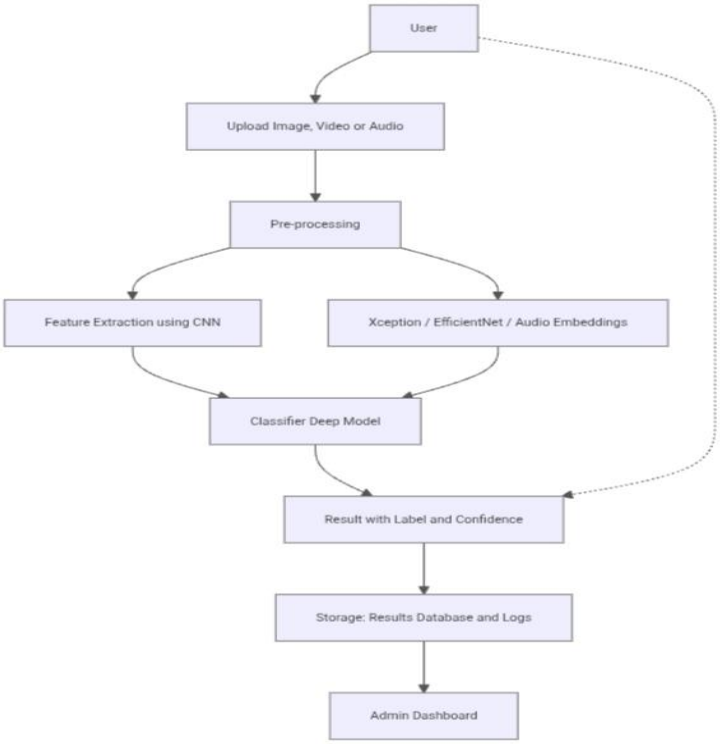

Fig. 2. *Deepfake Detection System Architecture*

4. CNN Feature Extraction Module
The CNN model extracts spatial features from images and video frames. It identifies patterns such as:

- Texture inconsistencies
- Blending artifacts
- Facial distortions

These features are crucial for detecting manipulated media.

5. LSTM Temporal Analysis Module
The LSTM model processes sequential data to identify temporal inconsistencies. It is particularly useful for:

- Detecting lip-sync mismatches
- Identifying unnatural motion in videos
- Analyzing audio inconsistencies

6. Classification Module: The extracted features are passed to a classification layer, which determines whether the media is real or fake. The output includes a probability score indicating the confidence level of the prediction.

7. Output Module
The final results are displayed to the user through the interface. The system provides:

- Real/Fake classification
- Confidence score
- Processed output visualization

Objective

- To design and implement a deep learning–based framework for detecting deepfake images, videos, and audio.
- To extract spatial features from images and temporal features from videos for fake content identification.

- To analyze speech patterns and spectrogram features for detecting synthetic audio.
- To train and evaluate deep learning models using standard deepfake datasets.
- To achieve high detection accuracy with improved robustness against advanced deepfake techniques.
- To develop an automated system for reliable multimedia authenticity verification.
- To compare model performance using evaluation metrics such as accuracy, precision, recall, and F1-score.
- To enhance cybersecurity and prevent misuse of manipulated digital media.

Hardware And Software Requirement

Hardware requirement

The system requires the following minimum hardware configuration for efficient performance:

- Processor: Intel Core i5 / i7 or AMD Ryzen 5 or higher
- RAM: Minimum 8 GB (16 GB Recommended)
- Storage: 256 GB SSD or higher
- Graphics Card (GPU): NVIDIA GPU (GTX 1050 / RTX Series recommended for deep learning training)
- System Type: 64-bit Operating System
- Display: Standard Monitor (HD Resolution)
- Input Devices: Keyboard and Mouse
- Internet Connection**:** Required for dataset download and model training updates

Software Requirement:

The system is developed using modern technologies and requires the following software components:

- Operating System: Windows 10 / Linux / macOS
- Programming Language: Python
- Development Environment: Jupyter Notebook / VS Code / PyCharm
- Libraries & Frameworks: TensorFlow, Keras, PyTorch, OpenCV, NumPy, Pandas, Matplotlib, Scikit-learn
- Deep Learning Tools: CUDA & cuDNN (for GPU acceleration)
- Dataset: FaceForensics++, Celeb-DF, DFDC Dataset, ASVspoof Dataset
- Database (Optional): MySQL / SQLite
- Version Control: Git / GitHub
- Browser: Google Chrome / Mozilla Firefox

Algorithm

The deepfake detection algorithm works by analyzing digital media and identifying hidden manipulation patterns using deep learning techniques. The detailed working steps are explained below:

Step 1: Input Media Collection

- The system accepts different types of media such as images, videos, or audio files uploaded by the user.
- These inputs form the testing data used to verify authenticity.

Step 2: Data Preprocessing

- Before analysis, the uploaded media is cleaned and standardized.
- Images: Resize, normalize pixel values, remove noise
- Videos: Extract frames and synchronize frame sequences
- Audio: Convert to waveform or spectrogram format
- Preprocessing ensures uniform data quality for accurate detection.

Step 3: Feature Extraction

- Deep learning models automatically learn important patterns:
- CNN (Convolutional Neural Network): Detects visual inconsistencies in images and video frames such as facial distortion, lighting mismatch, or blending errors.
- RNN/LSTM: Captures temporal inconsistencies across video frames and speech patterns.
- Audio Models: Identify unnatural voice modulation, pitch irregularities, or synthetic speech artifacts.

These extracted patterns are called deep features.

Step 4: Model Training
The neural network is trained using labeled datasets containing:

- Real media samples
- Fake (deepfake) media samples

The model learns distinguishing characteristics between genuine and manipulated content.

Step 5: Classification

- The trained model processes extracted features and predicts authenticity.
- Calculates probability score
- Applies decision threshold
- Classifies media as Real or Fake

Step 6: Result Generation
After classification, the system generates:

- Final verdict (Real/Fake)
- Confidence score
- Highlighted manipulated regions (if detected)

Step 7: Dashboard Visualization
The result is displayed on the dashboard where users can:

- View detection outcome
- Download verification report
- Store analysis history

Applications

1. Social Media Security: Detects fake images, videos, and voice clips shared on social media platforms to prevent fake news
2. Cybersecurity & Fraud Prevention: Identifies fake voice calls, video impersonations, and identity fraud used in online scams
3. Digital Media Authentication: Verifies authenticity of multimedia content used in journalism, news broadcasting, and online publishing.
4. Law Enforcement & Forensics: Helps police and investigation agencies detect manipulated evidence and digital tampering in criminal cases.
5. Political Content Verification: Prevents misuse of AI-generated fake speeches or videos that can influence elections and public opinion.
6. Biometric Security Systems: Protects face recognition and voice authentication systems from spoofing attacks.
7. Entertainment Industry Protection: Detects unauthorized use of celebrities' faces or voices in movies, advertisements, and social media content.
8. Online Education & Examination Systems: Prevents cheating using deepfake videos or voice cloning during online exams and virtual interviews.
9. Corporate Security: Stops fake CEO voice or video scams used for financial manipulation in organizations.
10. Content Moderation Platforms: Assists automated systems in identifying manipulated media before publishing on digital platforms.

Output

The output of the deep fake detection system developed using deep learning techniques provides automatic verification of multimedia content including images, videos, and audio files. The system analyzes the input media and classifies it as real or fake based on learned patterns and features. It generates prediction results

along with confidence scores that indicate the reliability of detection. For images and videos, the model identifies manipulated regions, facial inconsistencies, visual artifacts, and lip-sync mismatches, while for audio data it detects synthetic or cloned voices by analyzing speech characteristics. The system also produces performance evaluation metrics such as accuracy, precision, recall, and F1-score to measure model effectiveness. Additionally, graphical visualizations and stored detection reports are generated to support analysis, verification, and decision-making in real-world applications.

Overview

Deep fake detection using deep learning is a modern approach developed to identify manipulated or artificially generated multimedia content such as images, videos, and audio. With the rapid advancement of artificial intelligence, deepfake technologies can create highly realistic fake media that is difficult to distinguish from genuine content. This creates serious challenges related to misinformation, identity theft, cybercrime, and digital security.

Deep learning models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transformer-based architectures are used to automatically learn patterns and detect inconsistencies present in fake media. The system analyzes facial expressions, image textures, motion patterns, and speech characteristics to determine authenticity. By training on large datasets of real and fake samples, the model learns to recognize manipulation artifacts that are invisible to human observers.

The main goal of deep fake detection systems is to ensure media authenticity, improve cybersecurity, protect individuals and organizations from fraud, and maintain trust in digital communication platforms. These systems are widely applied in social media monitoring, forensic investigation, biometric authentication, journalism, and online content verification.

Key Features

The AI-Driven Conflict-Free Academic Timetabling System provides several important features that improve efficiency, accuracy, and usability in timetable management.

- Automatic Media Analysis: Automatically analyzes images, videos, and audio files without manual inspection.
- Multi-Modal Detection: Detects deepfakes across different media types including image, video, and voice data.
- High Detection Accuracy: Uses deep learning models to achieve accurate identification of manipulated content.
- Real-Time Detection: Capable of identifying fake media quickly for real-time applications.
- Facial and Audio Feature Extraction: Extracts facial expressions, lip movements, texture patterns, and speech characteristics.
- Manipulation Localization: Highlights altered regions in images or video frames.
- Confidence Score Generation: Provides probability or confidence level of prediction results.
- Robust Against Advanced Deepfakes: Detects AI-generated media created using modern GAN and voice cloning techniques.
- Performance Evaluation Metrics: Generates accuracy, precision, recall, and F1-score for model evaluation.
- Scalable and Automated System: Can be integrated into social media platforms, security systems, and forensic applications

Example Output

The example outputs of the deepfake detection using deep learning for images, videos and audios

Example 1: Deepfake Verification System Dashboard

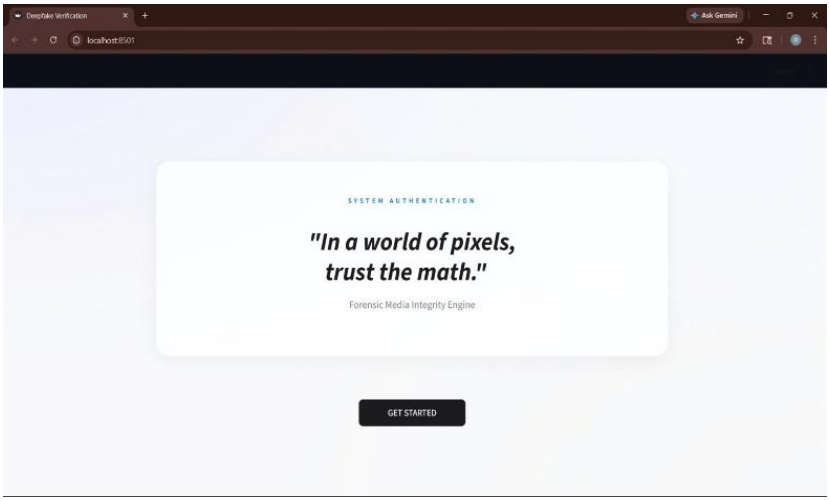

Fig. 1. *Deepfake Verification System Dashboard*

This dashboard represents the main interface of the Deepfake Verification System. It provides system authentication and introduces the forensic media integrity engine used for detecting fake images, videos, and audio. The “Get Started” button allows users to begin media analysis and access deepfake detection features through a secure and user-friendly environment

Example 2: Deep Fake Detection Analysis dashboard

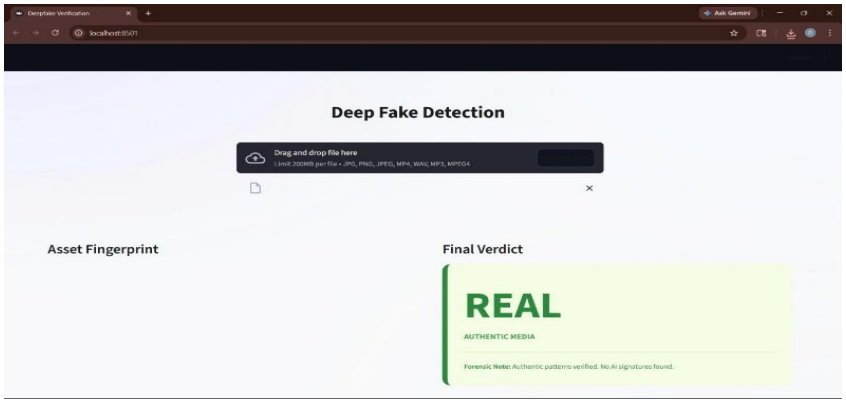

Fig. 2. *Deep Fake Detection Analysis dashboard*

This dashboard allows users to upload images, videos, or audio files for deepfake analysis using a drag-and-drop interface. After processing, the system displays the final verdict indicating whether the media is real or fake. It also provides an asset fingerprint and forensic verification details to confirm media authenticity

Example 3: Deepfake Detection Result Dashboard

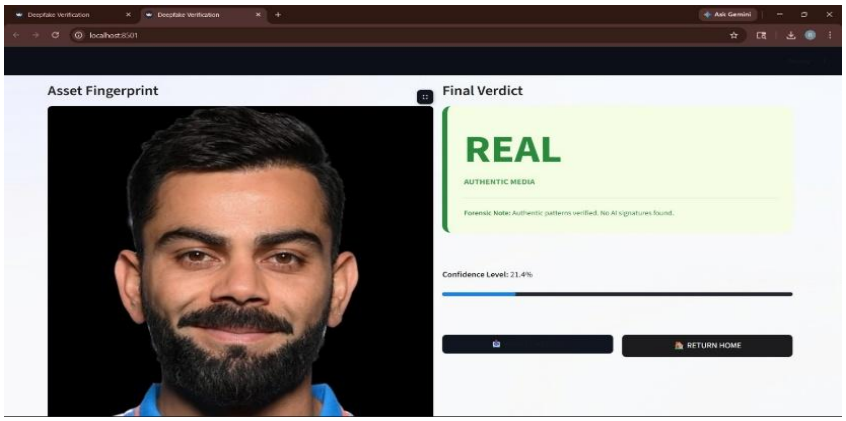

Fig. 3. *Deepfake Detection Result Dashboard*

This dashboard displays the analysis result of uploaded media. It shows the asset fingerprint, final verdict indicating whether the media is real or fake, and the confidence level of detection. Users can also export the verification report or return to the home page.

Example 4: Deepfake Detection Result Dashboard

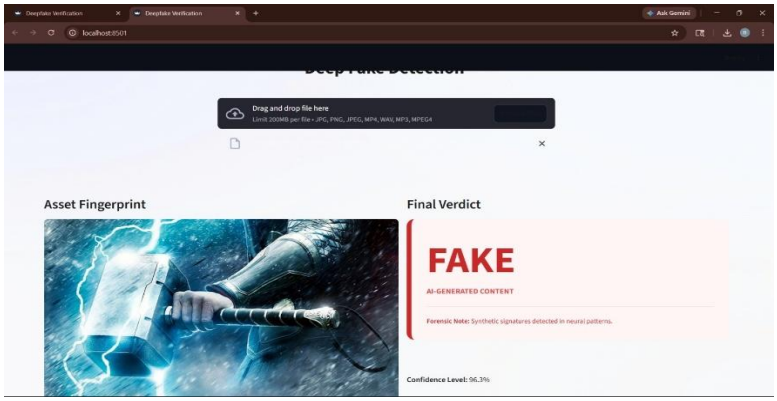

Fig. 4. *Deepfake Detection Result Dashboard*

This interface displays the output of a deepfake verification system. After uploading an image, video, or audio file, the system analyzes digital fingerprints and forensic patterns using deep learning models. The dashboard presents the asset preview, detection analysis, final authenticity verdict (Fake/Real), forensic notes, and confidence level indicating the reliability of the prediction.

Result Discussion

The proposed system was tested using benchmark datasets and real-world media samples. The results demonstrate that the system effectively detects deepfake content with high accuracy.

The CNN model successfully identifies spatial inconsistencies, while the LSTM model captures temporal irregularities. The combination of both models significantly improves detection performance.

The system shows strong generalization capability across different datasets and performs well under varying conditions. The results indicate that the system can be effectively used in real-world applications to combat deepfake threats.

Conclusion

The proposed deepfake detection system introduces a novel AI-based platform that integrates CNN and LSTM architectures to extract both spatial and temporal features from images and videos. By training on benchmark datasets such as FaceForensics++ and DFDC, the system achieves enhanced generalization, enabling it to detect manipulations produced by both existing and newly emerging techniques. The inclusion of a user-friendly interface ensures accessibility for journalists, law enforcement agencies, social media platforms, and general users, thereby increasing its societal impact. This system not only contributes to digital forensics and cybersecurity but also plays a crucial role in mitigating the spread of misinformation. While challenges such as adversarial attacks, real-time efficiency, and dataset limitations remain, the proposed approach lays a strong foundation for future work aimed at building more robust, lightweight, and multimodal deepfake detection systems

References

1. Rössler, A., Cozzolino, D., et al., “FaceForensics++: Learning to Detect Manipulated Facial Images,” *IEEE*, 2019.
2. Afchar, D., Nozick, V., et al., “MesoNet: Compact Facial Video Forgery Detection Network,” *IEEE*, 2018.
3. Li, Y., & Lyu, S., “Two-Stream Network for Tampered Face Video Detection,” *IEEE*, 2019.
4. Zhao, Y., Shen, J., et al., “LipForensics: Using Visual Speech for Deepfake Detection,” *IEEE*, 2021.
5. Wang, S.-Y., Wang, O., et al., “Face X-ray: Generalizable Deepfake Detection Method,” *IEEE*, 2020.
6. McCloskey, S., & Albright, M., “Detection of GAN-Generated Faces Using Color Cues,” *IEEE*, 2019.
7. Yang, X., Li, Y., & Lyu, S., “Learning to Detect Fake Face Images in the Wild,” *IEEE*, 2019.
8. Coccomini, D., Messina, N., et al., “Vision Transformers for Deepfake Detection,” *arXiv*, 2021.
9. “A Survey on Deepfake Video Detection,” *IEEE Access*, 2021.
10. Zhang, Z., et al., “Deepfake Detection: A Survey,” *IEEE Access*, 2022.
11. Wang, X., et al., “Deepfake Detection via Audio-Visual Fusion,” *IEEE Transactions on Multimedia*, 2021.
12. Nguyen, H., et al., “Deepfake Detection with Convolutional Neural Networks,” *IEEE Transactions on Information Forensics and Security*, 2020.
13. Kim, J., et al., “Deepfake Detection Using Recurrent Neural Networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.
14. Lee, S., et al., “Deepfake Detection: A Comparative Study,” *IEEE Transactions on Cybernetics*, 2022.
15. Ekatpure, J. N., Tavate, C. S., Malshikare, S. S., Khomane, A. B., & Tamboli, M. J. L. (2025). Artificial intelligence based virtual keyboard and mouse for computer. *International Journal on Advanced Computer Theory and Engineering*, 14(1), 449–456.
16. Ekatpure, J. N., Mohite, S. D., Shinde, A. A., Shirkande, N. B., & Upase, V. V. (2025). Campus recruitment system using machine learning. *International Journal on Advanced Computer Theory and Engineering*, 14(1), 427–432.
17. Aware, D. B., Sayyad, S. R., Shaikh, A. H., Thombare, S. B., & Ekatpure, J. N. (2025). Translation assistant for converting sign language to text and audio. *International Journal on Advanced Computer Engineering and Communication Technology*, 14(1), 445–449.
18. Ekatpure, J. N., Aware, D. B., Shaikh, A. H., Sayyad, S. R., & Thombare, S. B. (2024). A comprehensive survey on sign language translation systems: Bridging gestures, text, and audio for enhanced communication. *International Journal of Recent Advances in Engineering and Technology*, 13(2), 15–21.
19. Ekatpure, J. N., Tavate, C., Malshikare, S., Khomane, A., & Tamboli, M. J. (2024). Advancements in AI-powered virtual keyboards and mice: A survey of cutting-edge technologies for modern computing. *International Journal on Advanced Computer Theory and Engineering*, 13(2), 52–57.