IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

Deepfake Detection using AI

Prof. Priyanka. B. Palve[1], Miss. Shruti Tandale[2], Miss. Vishakha Mhaske[3], Miss. Vaishnavi Thombare[4], Miss. Vaishnavi Suse [5]

Prof. 1st Year Engineering Department, Adsul's Technical Campus, Ahilyanagar, India [1]
Students, AIDS Engineering Department, Adsul's Technical Campus, Ahilyanagar, India [2,3,4]
Student, Information Technology Engineering Department, Adsul's Technical Campus, Ahilyanagar, India [5]

Abstract: *The rapid advancement of deep learning technologies has led to the rise of deepfake media, posing serious threats to digital trust and security. This project presents an AI-based Deepfake Detection System that identifies manipulated images using advanced machine learning techniques. The system integrates a React frontend and a Node.js backend for seamless media upload and preprocessing. The processed content is analysis using deep learning models such as Pixel Error level analysis to detect visual and temporal inconsistencies. Based on the extracted features, the system classifies the media as real or deepfake and provides a confidence score along with an explanatory result. The proposed solution aims to enhance digital content verification and strengthen cybersecurity measures against synthetic media threats*

Keywords: Artificial Intelligence, Deepfake Detection, Pixel Error Level Analysis, Image Forensics, Machine Learning, Cybersecurity, Media Authentication

I. INTRODUCTION

In this research work presents a method that utilizes minimal training data and time to generate customized, photo-realistic talking head models. The technique employs few-shot learning, enabling the generation of satisfactory results from a single image, with improved fidelity using additional inputs. Unlike traditional warping-based approaches, the system synthesizes video frames directly using deep convolutional networks. The learning process is defined through adversarial training involving high-capacity generators and discriminators, which allows the system to quickly adapt to new identities through extensive meta-learning on large-scale video datasets. A person- specific parameter initialization further accelerates training and enhances performance. The proposed approach demonstrates the capability to produce lifelike talking head models of previously unseen individuals, including those depicted in portrait paintings. The paper discusses the ability of deepfake technology to produce artificial intelligence-generated digital content that looks real is examined closely. This research takes into account the wider societal ramifications as well as the complexities of AI algorithms in creating and detecting deepfakes. It highlights the critical requirement for advanced detection systems to stop exploitation and considers the continuous development of this powerful technology.

People share images widely on social media and digital platforms. With the help of artificial intelligence, it has become easy to create fake or manipulated media called deepfakes. These fake images look real and can be used to spread false information or harm individuals. It is difficult for humans to identify whether a media file is real or fake just by looking at it. This creates a need for a system that can automatically analyse and verify digital content. The proposed project develops an AI-based deepfake detection system that allows users to upload images and check their authenticity. The system is implemented as a web application using modern technologies such as React and Node.js, making it accessible and easy to use. The project aims to raise awareness about deepfakes and provide a technical solution for digital media verification.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

Recent years have seen notable advances in deep learning, a subset of machine learning techniques that combine representation learning with artificial neural networks. Deepfake technology, which automates the creation of synthetic video content, has raised concerns about election manipulation and cyberbullying. This study suggests an integrated system with a unique face forensics model and a convolutional approach for media manipulation detection in order to address these problems. In terms of construction, the system makes use of Convolutional Neural Networks (CNNs), which are often need to undergo substantial training on sizable datasets customized for each subject in order to produce convincing human head images. But in realworld applications, it becomes necessary to train customized talking head models using as little as one image as input. In order to address this difficulty, the system uses meta-learning methods on a sizable video dataset. It frames the learning procedure as adversarial training problems with high-capacity generators and discriminators, initializing parameters in a way that is unique to each individual to enable quick training despite the task's complexity. In the meantime, the detection side tackles the onslaught of altered media by utilizing deep learning and computer vision to create a hybrid face forensics framework that enhances detection performance by fusing traditional image forensic approaches with false face image forensic methods.

This framework, embodied in a convolutional neural network architecture with two distinct feature extractors, aims to simultaneously extract content and trace features from facial images, aiming to mitigate the adverse societal impacts of manipulated media. This study examines the relationship between deep learning and digital forensics, with a focus on Deepfake technology and the production and identification of altered material. The goal of the paper is to determine whether Convolutional Neural Networks (CNNs) with limited training data and time restrictions can produce individualized, photorealistic talking head models. In an effort to lessen the negative effects that these technologies have on society, it also explores the creation of a hybrid face forensics framework that uses convolutional methods to identify edited or altered media. The scope includes developing adversarial training strategies with high-capacity generators and discriminators, and investigating meta-learning techniques for few- and one-shot learning scenarios. In order to improve manipulation detection effectiveness, the research also looks into integrating traditional picture forensic methods with false face image forensic techniques. The study paper's overall goal is to advance knowledge and the creation of reliable techniques for manipulating media and identifying it, with ramifications for both digital integrity and the welfare of society.

Objective:

For the project focusing on the development of an integrated face forensics system to detect deepfake content effectively, the objectives are clearly outlined to address both the technical challenges and broader societal impacts. Here they are structured for clarity and focus:

• Develop an Advanced Detection Framework: Design and implement a hybrid face forensics model that integrates conventional image forensic techniques with cutting edge deep learning approaches, specifically convolutional neural networks (CNNs).

• Implement Few-Shot and One-Shot Learning: Incorporate meta-learning strategies to enable the system to effectively learn from a minimal number of images, including few-shot and one- shot learning capabilities. This is essential for adapting quickly to new and evolving deepfake methods with limited available data.

• Optimize Adversarial Training Techniques: Develop and refine adversarial training processes that involve high-capacity generators and discriminators. The objective is to enhance the model's ability to generalize from known to unknown manipulations, thereby improving its predictive accuracy against emerging deepfake technologies.

• Build and Validate with Diverse Datasets: Collect and compile a comprehensive and diverse set of data, including a custom DeepFake dataset alongside the public Face2Face dataset. Use these datasets to train and validate the effectiveness of the forensics framework under various conditions and across different types of deepfake manipulations.

• Ensure Robustness and Scalability: Ensure that the detection system is not only robust against a variety of deepfake attacks but also scalable and adaptable for widespread deployment across different platforms and technologies.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

II. LITERATURE SURVEY

In recent years, the proliferation of deepfake technology has raised concerns about its potential misuse, prompting researchers to develop various detection methods. This literature survey explores different approaches proposed by researchers to detect deepfake content across different media types, including images, videos, and audio. By analyzing the methodologies, limitations, and potential advancements outlined in these studies, we aim to provide insights into the evolving landscape of deepfake detection and mitigation strategies.

1. Liu and Du [1]: They delved into deepfake detection using Capsule Networks, which have shown promise in understanding hierarchical relationships in media. However, while their methodology offers insights into combating synthetic media manipulation, it lacks real time applicability and may not perform uniformly across various types of deepfakes.

2. Zang and Han [2]: Their use of Recurrent Neural Networks (RNNs) for deepfake detection focused on analyzing temporal dependencies within videos. While this approach addresses some challenges, such as the dynamic nature of deepfake manipulation, it may struggle with longer video sequences or subtle manipulations that could evade detection, posing limitations in real-world applications.

3. He et al. [3]: They prioritized high-level semantic features for deepfake detection, aiming to accurately identify manipulated content. However, by relying solely on highlevel features, their approach may overlook subtle manipulations or low-level artifacts, reducing its effectiveness in comprehensive detection across various manipulation techniques.

4. Guera and Abd-Almageed [4]: Their utilization of Capsule Networks for deepfake detection highlighted challenges regarding scalability and interpretability, particularly in complex scenarios. While Capsule Networks offer unique capabilities, their practical applicability and effectiveness in detecting diverse deepfake content require further refinement.

5. Zhang et al. [5]: By employing Convolutional Neural Networks (CNNs) for detecting manipulated facial content, they aimed to leverage the network's ability to learn complex patterns from images. However, their model's limited analysis of temporal dependencies in videos may hinder its ability to detect deepfakes relying on dynamic features rather than static facial characteristics.

6. Gül et al. [6]: Their introduction of an Attention- Based Detection Network effectively focused on specific facial regions, a critical aspect of deepfake detection. However, attention-based models may struggle to capture broader features necessary for comprehensive detection across various manipulation techniques and media types.

7. Alnaim et al. [7]: Their creation of a deepfake detection dataset focused on face masks addressed a pertinent need in the context of the infectious disease era. Nonetheless, attention-based models' limitations in capturing broader features could pose challenges in fully detecting manipulated content featuring individuals wearing face masks.

8. Hamza et al. [8]: Leveraging MelFrequency Cepstral Coefficients (MFCC) for deepfake audio detection, they emphasized the importance of high-quality training data and advanced audio processing techniques. However, further exploration of advanced audio processing methods could enhance their approach's efficacy in detecting and mitigating deepfake audio content.

9. Waseem and Abu Bakar [9]: Their comprehensive review of face and expression swap techniques laid a foundation for understanding these deepfake methods. To augment its utility, deeper exploration of specific detection methods alongside their respective strengths and weaknesses would provide valuable insights for effective countermeasures against sophisticated deepfakes.

III. SYSTEM ARCHITECTURE

The proposed Deepfake Detection System follows a modern full-stack, service-oriented architecture designed for scalability, usability, and efficient AI integration. The system consists of three primary layers: a React-based frontend for user interaction, a Node.js backend for request handling and orchestration, and an AI analysis service powered by Google Gemini AI. Users upload images through the web interface, which are securely transmitted to the backend

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

server. The backend performs validation, preprocessing, and forwards the media to the AI service for forensic analysis. The AI evaluates the content using advanced visual analysis techniques to identify manipulation artifacts and inconsistencies. The processed results, including the authenticity verdict, confidence score, and explanation, are returned to the frontend for display. This layered architecture ensures separation of concerns, improves maintainability, and supports future enhancements such as scaling, logging, and deployment to cloud environments.

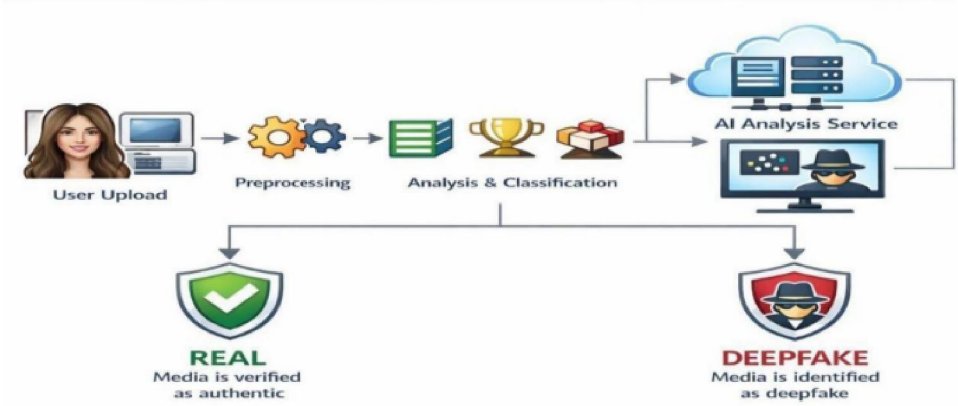

Figure1 Deepfake Detection System Architecture

A. Modules of the Proposed System

Module	Description
User Interface Module	Provides a web-based interface using React for users to upload images and view analysis results.
Backend Processing Module	Manages API requests, orchestrates communication between frontend and AI service, and handles preprocessing tasks.
AI Analysis Module	Integrates with Google Gemini AI to analyze media for manipulation artifacts and authenticity indicators
Logging & Monitoring Module	Maintains logs of requests and analysis outcomes for monitoring, debugging, and audit purposes.
Security & Access Control Module	Ensures secure handling of user data, validates requests, and protects APIs from misuse

Table 1: Module

B. Key Features

The proposed system offers an intuitive web-based interface for deepfake detection, enabling users to verify the authenticity of images with ease. It integrates AI-powered analysis to provide accurate detection along with confidence scores and explanations. The system is designed to be scalable, secure, and easily extendable for future enhancements.

Features	Benefit
Web-Based User Interface	access the system easily through any browser.
Confidence Score, Explanation	users understand the reliability and reasoning behind results.
Scalable Backend Design	Supports future growth and increased user traffic.
User Friendly Interface	Protects user data and ensures safe processing of uploaded files.

Table 2: key Features

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, May 2026
Impact Factor: 8.2

C. Workflow

The workflow of the proposed system starts with the user uploading an image through the web interface. The frontend securely transmits the media to the backend server, where validation and request handling are performed. The backend forwards the media to the AI analysis service for deepfake detection and forensic evaluation. The analysis results, including the verdict and confidence score, are processed and displayed to the user in a clear and user-friendly format

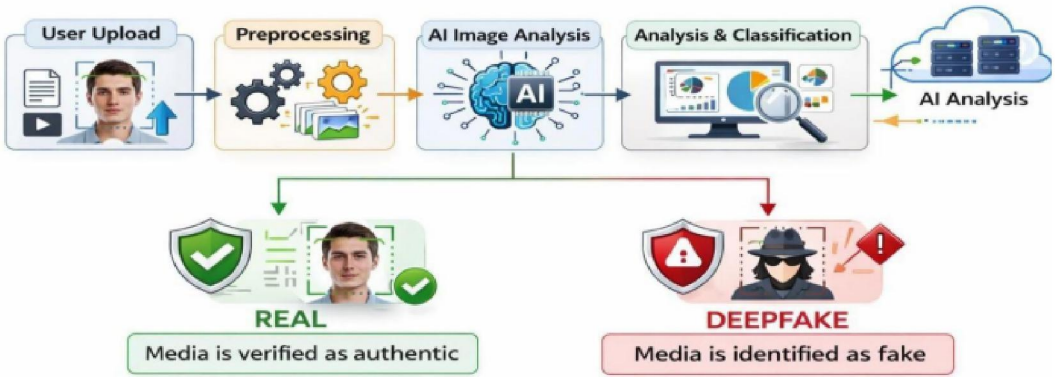

Fig. 2Workflow

Table 2: Workflow stage

Stage	Input	Process	Output
User Input	Image User	User uploads media through the web interface	Media file captured by frontend
Data Preprocessing	Media file	Frontend sends file to backend API	Validated upload request
Model Processing	Media file	Backend Performs validation and preprocessing	Prepared media for analysis
AI Analysis	Prepared media	Media sent to AI analysis service	Verdict, confidence score, explanation
Result Display	AI analysis results	Backend formats response and sends to frontend	Results displayed to the user

IV. DATAFLOW DIAGRAM

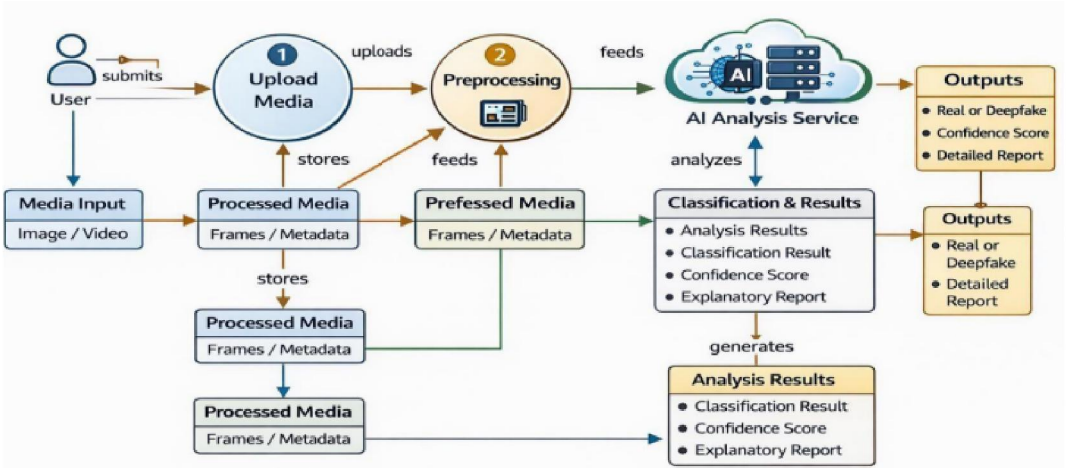

Fig.3: AI-Based Deepfake Detection System

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, May 2026 Impact Factor: 8.2

The data flow in the proposed Deepfake Detection System illustrates how information moves across different system components. The process begins when the user uploads an image through the web interface. The frontend captures the input and forwards it to the backend server for validation and preprocessing. The backend then transmits the media to the AI analysis service for deepfake detection and forensic evaluation. The AI service analyzes the content and generates detection results along with confidence metrics. These results are returned to the backend, processed into a user-friendly format, and finally displayed to the user. This structured data flow ensures secure handling, efficient processing, and clear presentation of analysis outcomes.

V. DISCUSSION & FUTURE SCOPE

Discussion:

The proposed Deepfake Detection System demonstrates the effective integration of AI services with modern web technologies to address the challenge of digital media authenticity. The system provides accessible and user-friendly verification of images with meaningful outputs such as verdicts and confidence scores. While the current implementation shows promising results, continuous improvements are required to keep pace with evolving deepfake techniques.

Future Work:

Future enhancements will focus on improving the accuracy and robustness of image-based deepfake detection. Advanced image forensics techniques such as high-frequency analysis, noise pattern detection, and metadata verification can be incorporated to identify subtle manipulation traces. The system can be extended to support high-resolution image analysis and batch image verification for large-scale content screening. Integration with browser extensions and mobile applications can improve accessibility and enable real-time image authenticity checks across digital platforms. Additionally, continuous model training using newly generated deepfake image datasets will help the system adapt to emerging image manipulation techniques and maintain long-term detection reliability.

VI. CONCLUSION

Deep Fake Creation is where users can make deep fakes from uploaded images or videos, and Deep Fake Detection is where they can check if uploaded video or audio content is real or fake (Figure 5). Image deepfake detection feature determines whether the uploaded image is real or fake. Audio deepfake detection feature, which determines whether the uploaded audio is real or fake. Video deepfake detection feature, which determines whether the uploaded Video is real or fake.

In conclusion, a series of convolutional layers converts input facial landmarks into output frames, with the generator network playing a critical part in this process. Through adaptive instance normalization, embedding vectors are modified throughout its operation. By running sets of frames from the same video through the embedder, averaging the resultant embedding, and utilizing the results, sets of frames from the same video can be processed via the generator to assist predictive modelling of the adaptive parameters during metal earning. In addition, the generator processes each frame landmark independently, and the resulting images are contrasted with the original data. This research demonstrates how well the generator network generates realistic facial pictures based on input landmarks, highlighting its potential uses in manipulation detection systems and personalized talking head models. These developments highlight deep learning's development and show how it might be used to solve problems in digital forensics and computer vision.

REFERENCES

[1].Y. Liu and S. Du, "DeepFake Detection Based on Capsule Network," in IEEE Access, 2021.
[2].S. Zang and W. Han, "Deepfake detection using recurrent neural network," in Multimedia Tools and Applications, 2021.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, May 2026 Impact Factor: 8.2

[3].Y. He et al., "Deepfake Detection Based on High-Level Information," in IEEE Transactions on Information Forensics and Security, 2021.
[4].D. Guera and W. Abd-Almageed, "Deepfake Detection Using Capsule Networks," in IEEE/CVF International Conference on Computer Vision Workshop, 2021.
[5].X. Zhang et al., "Deepfake Detection Using Convolutional Neural Networks," in IEEE Access, 2021.
[6].A. Gül et al., "Deepfake Detection via Attention Based Detection Network," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2021.
[7].N. M. Alnaim et al., "A Deepfake Face Mask Dataset for Infectious Disease Era with Deepfake Detection Algorithms," 2023.
[8].A. Hamza et al., "Deepfake Audio Detection via MFCC Features Using Machine Learning," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2023.
[9].S. Waseem and S. A. R. Abu Bakar, "DeepFake on Face and Expression Swap: A Review," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2023.
[10]. N. Waqas et al., "DEEPFAKE Image Synthesis for Data Augmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2022.
[11]. N. Waqas et al., "DEEPFAKE Image Synthesis for Data Augmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2022.
[12]. Y. Patel et al., "An Improved Dense CNN Architecture for Deepfake Image Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2023.
[13]. V.-N. Tran et al., "Generalization of Forgery Detection with Meta Deepfake Detection Model," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2022