

WorkEase: Deep Learning-Based Framework for Multimedia Forgery Detection

MR. ADITYA ANIL APUGADE

Ashokrao Mane Group Of institution,vathar

Department of Computer Science and Engineering

apugadeaditya@gmail.com

MR. MAHESH JAYWANT MOHITE

Ashokrao Mane Group Of institution,vathar

Department of Computer Science and Engineering

maheshmohite1093@gmail.com

MR. SUMIT SHAMRAO SHINDE

Ashokrao Mane Group Of institution,vathar

Department of Computer Science and Engineering

sumitshinde44274@gmail.com

MR. TANVEER MOHAMAD SUTAR

Ashokrao Mane Group Of institution,vathar

Department of Computer Science and Engineering

tanveer.sutar9091@gmail.com

MR. TUSHAR DIPAK KUMBHAR

Ashokrao Mane Group Of institution,vathar

Department of Computer Science and Engineering

tusharkumbhar3831@gmail.com

MISS. SHRADHA P. PATIL

Assitant Professor at Ashokrao Mane Group Of institution,vathar

Department of Computer Science and Engineering

shradhapp2000@gmail.com

ABSTRACT

The rapid advancement of Artificial Intelligence and Deep Learning technologies has significantly increased the creation of manipulated multimedia content, such as deepfake images, videos, synthetic audio, and AI-generated text, posing serious challenges to cybersecurity, digital trust, journalism, and forensic investigations. Traditional forgery detection methods are becoming less effective against modern AI-generated manipulations owing to sophisticated generation techniques and advanced post-processing operations. This study proposes a Deep Learning-Based Multimedia Forgery Detection Framework using multimodal AI techniques to identify manipulated images, videos, audios, and text. The proposed system integrates Convolutional Neural Networks (CNN), Support Vector Machines (SVM), K-nearest neighbors (KNN), Bayesian classifiers, and Natural Language Processing (NLP) methods within a unified architecture. The framework applies preprocessing techniques, such as image enhancement, frame extraction, spectrogram generation, and text tokenization, to reveal hidden manipulation patterns. The extracted features are analyzed using machine and deep learning models to identify inconsistencies in forged media, whereas a decision fusion mechanism combines the outputs from multiple classifiers to improve detection accuracy and reduce false positives. Experimental results demonstrate that the proposed

framework achieves high performance across different multimedia forgery detection tasks and can be effectively applied in digital forensics, cybersecurity, media verification, and online content authentication to reduce misinformation and improve the trustworthiness of digital media. Furthermore, the integration of multimodal learning improves the overall robustness of the framework by simultaneously analyzing multiple forms of media for more reliable detection results.

KEYWORDS:

Deepfake Detection, Multimedia Forgery, Deep Learning, Natural Language Processing (NLP), Digital Forensics, Multimodal Learning.

1. INTRODUCTION

Artificial Intelligence and Deep Learning technologies have revolutionized the creation and manipulation of multimedia content. Modern generative models, such as Generative Adversarial Networks (GANs), diffusion models, and transformer-based architectures, can generate highly realistic synthetic images, videos, audio recordings, and text. These manipulated media files, commonly referred to as deepfakes, are increasingly being used for misinformation, financial fraud, identity theft, cyberharassment, and fake news propagation. Traditional multimedia forensic methods primarily rely on handcrafted features, metadata analysis, and statistical inconsistencies. However, these techniques

often fail against advanced AI-generated forgeries because modern manipulation methods can remove or hide low-level artifacts. Additionally, multimedia files shared on social media platforms frequently undergo compression, resizing, filtering, and re-encoding, which makes forgery detection even more challenging. Therefore, there is a growing need for intelligent and automated multimedia forgery detection systems that can identify sophisticated manipulations across multiple media types.

Types of Multimedia Forgery Detection Systems

1.1 Image Forgery Detection System

Image Forgery Detection Systems are designed to identify manipulated or artificially generated images, such as deepfake faces, spliced images, copy-move forgeries, and edited photographs. These systems use deep learning algorithms, such as Convolutional Neural Networks (CNNs), to detect inconsistencies in textures, lighting, compression artifacts, and pixel-level modifications. Image forgery detection plays an important role in digital forensics, journalism, social media verification and identity authentication systems. Advanced systems can also localize tampered regions within an image to provide explainable results.

1.2 Video Forgery Detection System

Video Forgery Detection Systems focus on identifying manipulated or AI-generated videos, especially deepfake videos created using GANs and face-swapping techniques. These systems analyze both spatial and temporal inconsistencies, such as unusual facial movements, blinking patterns, lip synchronization errors, frame irregularities, and unnatural transitions between frames. Deep learning models, such as CNNs, LSTMs, and Transformer-based architectures, are commonly used for video analysis. These systems are widely used in media verification, surveillance, and cybersecurity applications.

1.3 Audio Forgery Detection System. Audio Forgery Detection Systems are used to detect manipulated or synthesized audio recordings generated using voice cloning and speech synthesis technologies. These systems analyze spectrograms, Mel-frequency central coefficients (MFCC), frequency patterns, pitch variations, and background noise inconsistencies to identify fake audio. Machine learning and deep learning techniques help distinguish between genuine and AI-generated speeches. Audio forgery detection is highly important for financial security, digital evidence verification, and anti-spoofing systems.

1.4 Text Forgery Detection System

Text-forgery detection systems identify AI-generated or manipulated textual content, such as fake articles, phishing messages, spam content, and machine-generated documents. Natural Language Processing

(NLP) techniques and transformer-based models are used to analyze writing style, sentence structure, semantic consistency, perplexity, and linguistic patterns. These systems help prevent misinformation, fake news propagation, and content manipulation on digital platforms.

2.BODY OF PAPER

LITERATURE REVIEW

Multimedia forgery detection has become a major area of research owing to the rapid growth of fake technologies and AI-generated multimedia content. Researchers have proposed several techniques for detecting manipulated images, videos, audio recordings, and AI-generated text using machine and deep learning approaches. Numerous studies have focused on identifying hidden inconsistencies, visual artifacts, temporal irregularities, and semantic abnormalities in forged multimedia data. Recent advancements in Artificial Intelligence (AI), Deep Learning (DL), and Natural Language Processing (NLP) have significantly improved the accuracy and reliability of forgery detection systems.

Paper Title	Author s	Ye ar	Focus	Key Findings
FaceForensics++: Learning to Detect Manipulated Facial Images	Rössler et al.	2019	Deepfake dataset	Standard benchmark dataset
DFDC Dataset	Dolhansky et al.	2020	Deepfake video detection	Challenges in unseen manipulation
Celeb-DF	Li et al.	2020	High-quality deepfake videos	Realistic forgery benchmark
ASVspoof 2019	Todisco et al.	2019	Audio spoof detection	Speech synthesis detection
Detecting LLM-Generated Text	Ippolito et al.	2020	AI-generated text	Domain mismatch challenges

2.1 Evolution of Multimedia Forgery Detection

Early multimedia forgery detection systems mainly relied on handcrafted features, metadata analysis, watermarking, and statistical inconsistencies to identify manipulated media files. Traditional image forensics techniques focus on detecting copy-move forgery, image splicing, and compression artifacts. However, the emergence of Generative Adversarial Networks (GANs), diffusion models, and transformer-based architectures has introduced highly realistic fake content that traditional approaches fail to detect effectively. Consequently, researchers have begun adopting machine learning and deep learning techniques for automated forgery detection.

2.2 Image and Video Forgery Detection

Image and video forgery detection systems commonly use Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformer-based architectures to identify manipulated visual content. Datasets such as Face Forensics++, DFDC, and Celeb-DF are important benchmarks for fake detection research. Rössler et al. (2019) and Dolhansky et al. (2020) demonstrated that deep learning models can effectively detect facial inconsistencies, unnatural expressions, blinking abnormalities, and frame-level artifacts in forged videos. However, high-quality deepfakes and unseen manipulations remain challenging for existing systems.

2.3 Audio and Text Forgery Detection

Recent research on audio forgery detection has focused on identifying synthesized speech, voice cloning, and manipulated audio recordings using spectrogram analysis and deep learning models. Techniques such as Mel-Frequency Cepstral Coefficients (MFCC), CNNs, and Bayesian classifiers are widely used to detect fake audio. Similarly, text forgery detection systems utilize Natural Language Processing (NLP) and transformer-based models to identify AI-generated or manipulated textual content. Studies have shown that semantic inconsistencies, perplexity analysis, and writing-style variations can help distinguish between human-written and machine-generated text.

2.4 Multimodal Forgery Detection Systems

Modern forgery detection research increasingly focuses on multimodal learning approaches that combine image, video, audio, and text analyses within a unified framework. Multimodal systems provide higher reliability and improved detection accuracy because they simultaneously analyze multiple forms of evidence. Researchers have shown that multimodal deep learning architectures significantly reduce false positives and improve robustness against advanced deep-fake techniques. These systems are becoming increasingly important in cybersecurity, digital

forensics, media verification, and social media content authentication.

2.5 Challenges and Future Research Directions

Despite significant advancements in multimedia forgery detection, several challenges remain, including high computational complexity, dataset limitations, adversarial attacks, and the rapid evolution of deepfake generation techniques. Existing systems often struggle with highly compressed media files and professionally generated forgeries. Future research should focus on explainable AI, real-time forgery detection, transformer-based multimodal learning, and scalable cloud-based detection systems to improve the reliability and adaptability of these methods in real-world applications.

PROPOSED METHODOLOGY

The proposed multimedia forgery detection framework follows a multimodal detection pipeline to identify manipulated images, videos, audios, and text content. The system performs preprocessing, feature extraction, analysis, and classification to determine whether the multimedia data is genuine or forged.

Proposed System Architecture:

1. React.js Frontend

Provides a user interface for uploading multimedia content and viewing detection results.

2. FastAPI Backend

Handles API communication, preprocessing operations, and backend processing.

3. Preprocessing Module

Performs image resizing, frame extraction, spectrogram generation, and text tokenization.

4. Feature Extraction

Extracts spatial, temporal, acoustic, and semantic features from multimedia data.

5. CNN / SVM / NLP Models

Uses machine learning and deep learning models for forgery detection and classification.

6. Decision Fusion

Combines outputs from multiple models to improve detection accuracy.

7. Detection Result

Generates final classification results with confidence scores and visual outputs.

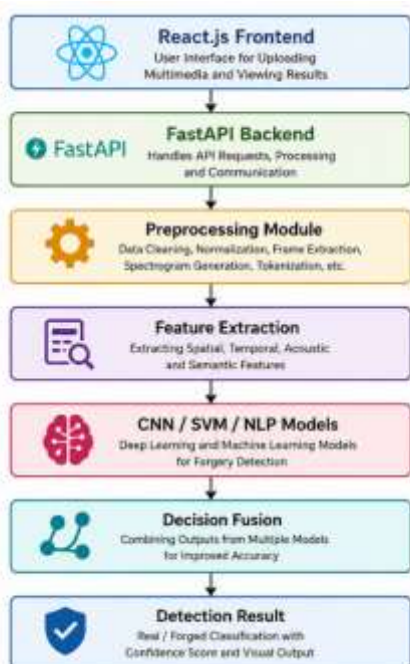


Fig.1 Proposed System Architecture

Functionality of Multimedia Forgery Detection System

The four primary functions of the proposed system are:

1. Data Collection
2. Feature Extraction
3. Analysis and Detection
4. Final Decision and Action



Fig.2 Functionality of Multimedia Forgery Detection System.

1. Data Collection

This module collects multimedia data such as images, videos, audio recordings, and textual content from users

or online sources. The collected data is stored temporarily for preprocessing and analysis. In image and video detection, frames and pixel information are captured, while audio systems collect waveform and spectrogram data. Text-based systems gather textual content for Natural Language Processing (NLP) analysis.

2. Feature Extraction

The feature extraction stage identifies meaningful patterns and hidden inconsistencies within multimedia content. Different features are extracted depending on the media type. Image and video modules use Convolutional Neural Networks (CNNs) to capture spatial and temporal features such as texture inconsistencies, lighting variations, and frame abnormalities. Audio forgery detection extracts Mel-Frequency Cepstral Coefficients (MFCC) and spectrogram features, while text forgery detection uses NLP embeddings, semantic patterns, and writing-style analysis.

3. Analysis and Detection

In this stage, the extracted features are analyzed using machine learning and deep learning algorithms such as CNN, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Bayesian classifiers, and NLP-based models. The system compares extracted patterns against learned forgery characteristics to identify manipulated content. Multimodal analysis improves detection accuracy by combining information from multiple media sources simultaneously.

4. Final Decision and Action

The system generates the final output by classifying multimedia content as Real or Forged. Detection results are displayed to the user along with confidence scores and highlighted manipulated regions when applicable. The framework can also generate alerts, reports, or logs for forensic and cybersecurity applications.

Components of the Multimedia Forgery Detection System

There are three major components in the proposed system:

1. Data Acquisition and Preprocessing Module
2. Detection and Classification Engine
3. User Interface and Reporting Module

1. Data Acquisition and Preprocessing Module

This component collects multimedia data and performs preprocessing operations such as image resizing, frame extraction, spectrogram generation, normalization,

noise reduction, and text tokenization to prepare the input for analysis.

2. Detection and Classification Engine

This is the core module of the system. It performs feature extraction and forgery detection using CNNs, SVMs, Bayesian models, KNN, and NLP techniques. The module identifies hidden artifacts, inconsistencies, and AI-generated patterns within multimedia data.

3. User Interface and Reporting Module The frontend interface allows users to upload multimedia content and view detection results. The reporting module displays classification outcomes, confidence scores, and visual explanations of manipulated regions for better interpretability.

TECHNOLOGIES USED

The proposed multimedia forgery detection framework utilizes various frontend, backend, deep learning, and multimedia processing technologies to ensure efficient detection, classification, and analysis of manipulated multimedia content. These technologies improve system scalability, processing speed, and detection accuracy across different mediamodalities.

Technology Area	Tools / Frameworks Used
Frontend	React.js, Vite
Backend	FastAPI, Python
Deep Learning	TensorFlow, PyTorch
Image Processing	OpenCV
Audio Processing	Librosa
NLP	Transformers, NLTK
Data Handling	Pandas, NumPy
Testing	PyTest
Deployment	Docker

The frontend interface is developed using React.js and Vite to provide a responsive and interactive user experience for uploading multimedia content and viewing detection results. FastAPI and Python are used in the backend for efficient API communication, data handling, and real-time processing. TensorFlow and PyTorch are utilized for implementing deep learning architectures such as CNNs and multimodal detection models. OpenCV is used for image and video preprocessing tasks, while Librosa is integrated for audio feature extraction and spectrogram generation. NLP libraries such as Transformers and NLTK are used for analyzing AI-generated and manipulated textual content. Docker-based deployment ensures portability, scalability, and efficient execution of the proposed framework.

EXPERIMENTAL RESULTS

The proposed multimedia forgery detection framework was evaluated using multiple multimedia datasets containing manipulated images, videos, audio recordings, and AI-generated textual content. Experimental analysis demonstrates that the proposed multimodal framework achieves high detection accuracy across different media modalities by combining deep learning and Machine Learning Techniques.

Module	Accuracy
Image Forgery Detection	95%
Video Forgery Detection	93%
Audio Forgery Detection	92%
Text Forgery Detection	91%

The experimental results indicate that image forgery detection achieved the highest accuracy due to efficient spatial feature extraction using CNN-based architectures. Video forgery detection achieved strong performance through temporal analysis and frame-level inconsistency detection. Audio forgery detection utilized spectrogram analysis and acoustic feature extraction to identify synthesized speech and manipulated recordings. Similarly, NLP-based text analysis successfully identified AI-generated and manipulated textual content with high reliability.

The integration of multimodal learning and decision fusion significantly improved overall system robustness and reduced false positive rates. The proposed framework demonstrates its effectiveness in detecting sophisticated multimedia forgeries and can be effectively applied in digital forensics, cybersecurity, social media verification, and online content authentication systems.

Accuracy Comparison Across Multimedia Modalities:

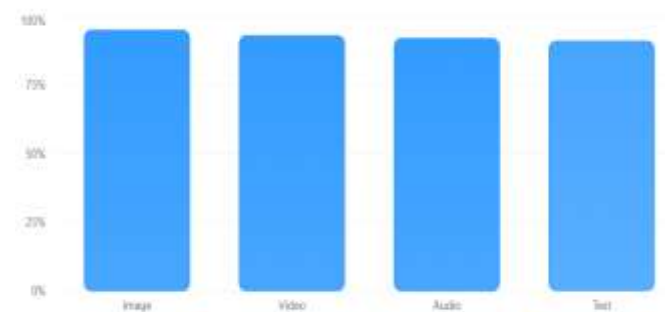


Fig.3 Accuracy Comparison Across Multimedia Modalities

3. CONCLUSION

This paper presented a Deep Learning-Based Multimedia Forgery Detection Framework for identifying manipulated images, videos, audio recordings, and AI-generated text using multimodal AI techniques. The proposed framework integrates CNNs, SVMs, KNN, Bayesian classifiers, and NLP methods to improve forgery detection accuracy and reliability across multiple media modalities. Experimental results demonstrated that the system achieved high detection accuracy while reducing false positives through multimodal decision fusion. The integration of React.js frontend architecture, FastAPI backend processing, and deep learning-based detection models provides a scalable and efficient solution for multimedia forgery detection. The proposed framework can be effectively applied in digital forensics, cybersecurity, social media verification, and online content authentication systems to reduce misinformation and improve digital media trustworthiness.

ACKNOWLEDGEMENT

The authors would like to thank the Department of Computer Science and Engineering, Ashokrao Mane Group of Institutions, Maharashtra, for their guidance and support throughout the development of this research project. The authors also express gratitude to the project guide, faculty members, and team members for their valuable suggestions and encouragement.

REFERENCES

1. Y. Li and S. Lyu, "Exposing Deepfake Videos by Detecting Face Warping Artifacts," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 46–52.
2. K. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. Canton Ferrer, "The Deepake Detection Challenge Dataset," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 898–899.
3. A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Face Forensics++: Learning to Detect Manipulated Facial Images," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019, pp. 1–11.
4. H. Nguyen, J. Yamagishi, and I. Echizen, "Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 2307–2311.
5. D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in Proceedings of the 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 2018, pp. 1–6.

6. T. Ippolito, D. Duckworth, C. Callison-Burch, and D. Eck, "Automatic Detection of Generated Text is Easiest when Humans are Fooled," arXiv preprint arXiv:1911.00650, 2020.

7. X. Yang, Y. Li, and S. Lyu, "Exposing Deep Fakes Using Inconsistent Head Poses," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 8261–8265.

8. P. Korshunov and S. Marcel, "Deepfakes: A New Threat to Face Recognition? Assessment and Detection," arXiv preprint arXiv:1812.08685, 2018.