IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-based Deepfake Detection Framework for Multimedia Authentication (DeepShieldX)

1. Rudraksha Goswami
Student, Computer Studies and Emerging Technology, TransStadia University, Ahmedabad

2. Parthi Soni
Assistant Professor, School of Computer Studies and Emerging Technology, TransStadia University, Ahmedabad

Abstract

The fast growth of Artificial Intelligence has made it possible to create highly realistic synthetic media known as deepfakes. These manipulated files may appear in the form of altered videos, cloned voices, or fully generated multimedia content. Although such technology can be useful in filmmaking, gaming, education, and accessibility tools, it also creates major risks when used for fraud, misinformation, identity theft, or public manipulation. Existing manual verification techniques are often slow and ineffective against advanced AI-generated content.

This research paper introduces **DeepShieldX**, an intelligent framework designed to detect deepfakes in **video, audio, and combined multimedia sources**. The proposed model uses facial inconsistency analysis, motion pattern learning, voiceprint verification, spectrogram classification, and multimodal fusion techniques for accurate prediction. The framework is practical, scalable, and suitable for real-world deployment using open-source technologies. Experimental findings indicate that combining both audio and video signals provides stronger performance than using a single source alone.

Keywords: Deepfake Detection, Multimedia Security, Video Forgery, Audio Forgery, Artificial Intelligence, CNN, LSTM, Transformer.

Introduction

In recent years, deepfake technology has developed rapidly due to improvements in machine learning and generative AI models. Using tools based on GANs, autoencoders, and diffusion networks, users can now create fake videos and synthetic voices that closely resemble real individuals. As these tools become easier to access, misuse has also increased.

Deepfakes can be used to spread false political messages, damage reputations, impersonate company executives, or perform voice scam attacks. Social media platforms have accelerated the spread of such content, making verification more difficult.

Traditional methods such as visual inspection or metadata checking are no longer enough. Many fake videos now contain realistic facial expressions, lip movement, and natural speech patterns. Therefore, intelligent automated systems are required. This paper proposes **DeepShieldX**, a complete detection framework capable of identifying fake video, fake audio, and manipulated multimedia using AI-based techniques.

Literature Review

A large number of researchers have studied deepfake detection in recent years. Early approaches focused on frame-level image artifacts such as unnatural blending, lighting mismatch, and texture irregularities. Convolutional Neural Networks (CNNs) became popular because they can learn visual manipulation patterns directly from data.

Later studies introduced sequence models such as LSTM networks to analyze frame continuity, facial motion, blinking frequency, and head movement over time. These methods improved detection of temporally inconsistent videos.

For audio deepfakes, researchers used Mel-frequency spectrograms and transformer-based architectures to detect synthetic voice traces, abnormal pitch transitions, and cloned speaker characteristics. Speaker verification systems were also used to compare suspicious voices with known genuine recordings.

Recent research shows that multimodal detection systems achieve better results because they inspect both sound and visuals together. However, many existing tools remain limited to one modality. This motivates the need for an integrated system like DeepShieldX.

Problem Statement

The increasing misuse of deepfake technology has created several real- world challenges:

1. Fake videos spread quickly across online platforms.
2. Cloned voices are being used for scam and fraud calls.
3. Manual verification is slow and often unreliable.
4. Many detection systems analyze only video or only audio.
5. Detection models may fail against newly improved generators.

Hence, there is a strong need for a robust and adaptive framework that can detect manipulated **video, audio, and combined multimedia content** accurately.

Proposed Solution

The proposed framework, **DeepShieldX**, is divided into five major layers:

1. Data Acquisition Layer

The system collects data from:

- Public benchmark datasets such as FaceForensics++, DFDC, and ASVspoof
- Social media uploads
- Interview recordings and news clips
- User-uploaded suspicious media

2. Preprocessing Layer

Before analysis, the media is cleaned and standardized:

- Video frame extraction
- Face localization and alignment
- Audio separation from video files
- Noise reduction and normalization
- Spectrogram image generation

3. Video Analysis Layer

This module detects visual signs of forgery using:

- CNN-based frame classification
- LSTM-based motion consistency learning
- Lip-sync mismatch detection
- Eye blinking and landmark tracking

4. Audio Analysis Layer

This module detects synthetic speech using:

- Spectrogram CNN models
- Voiceprint matching systems
- Pitch rhythm anomaly detection

- Transformer-based audio forgery analysis

5. Fusion and Decision Layer

The outputs of both modules are combined:

- Weighted confidence scoring
- Final prediction: Genuine / Suspicious / Fake
- Explainable heatmap generation
- Confidence percentage report

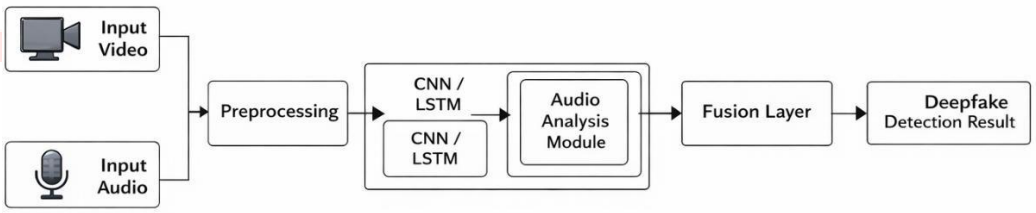
Technical Implementation

The framework can be developed using affordable and open-source tools:

- Python
- OpenCV
- TensorFlow / PyTorch
- Librosa
- FFmpeg
- FastAPI / Flask
- SQLite / PostgreSQL
- React-based Dashboard

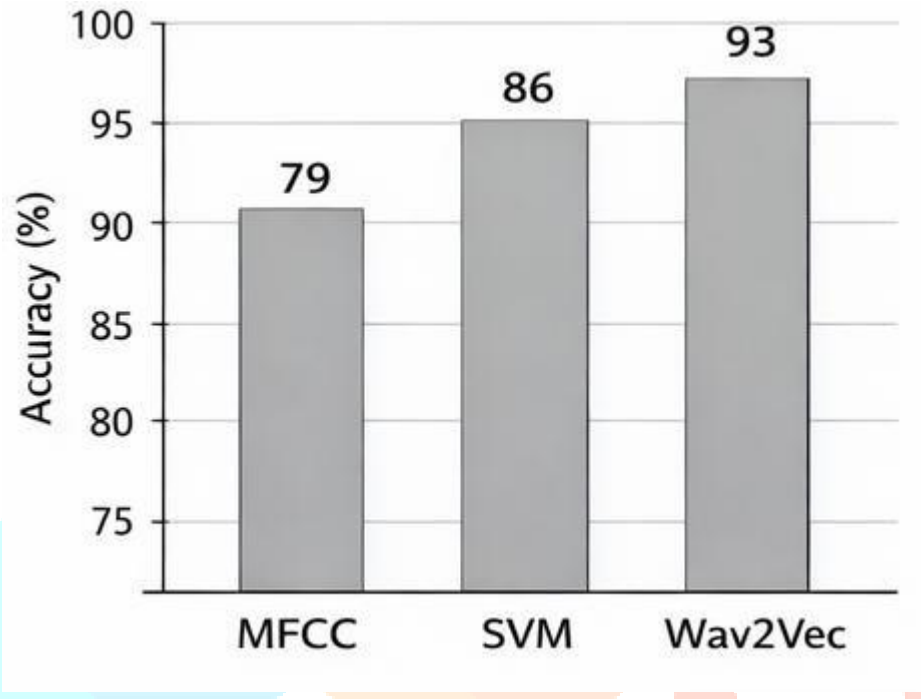
System Architecture (Visual Representation)

Deepfake Detection System Architecture:

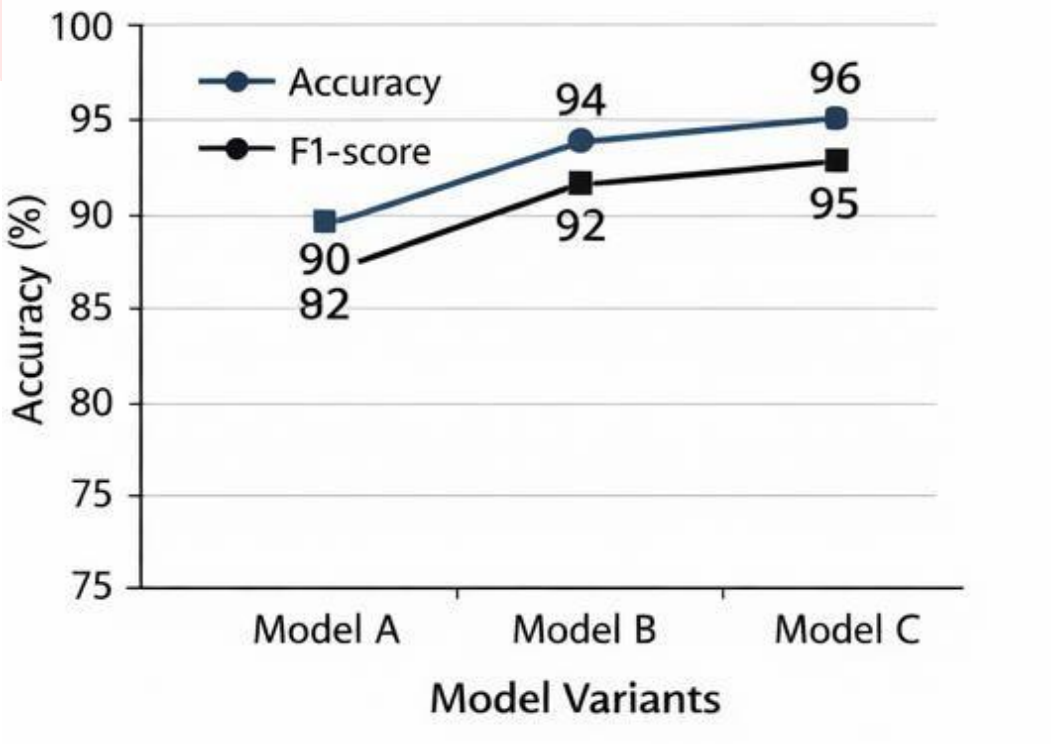


Experimental Graphs

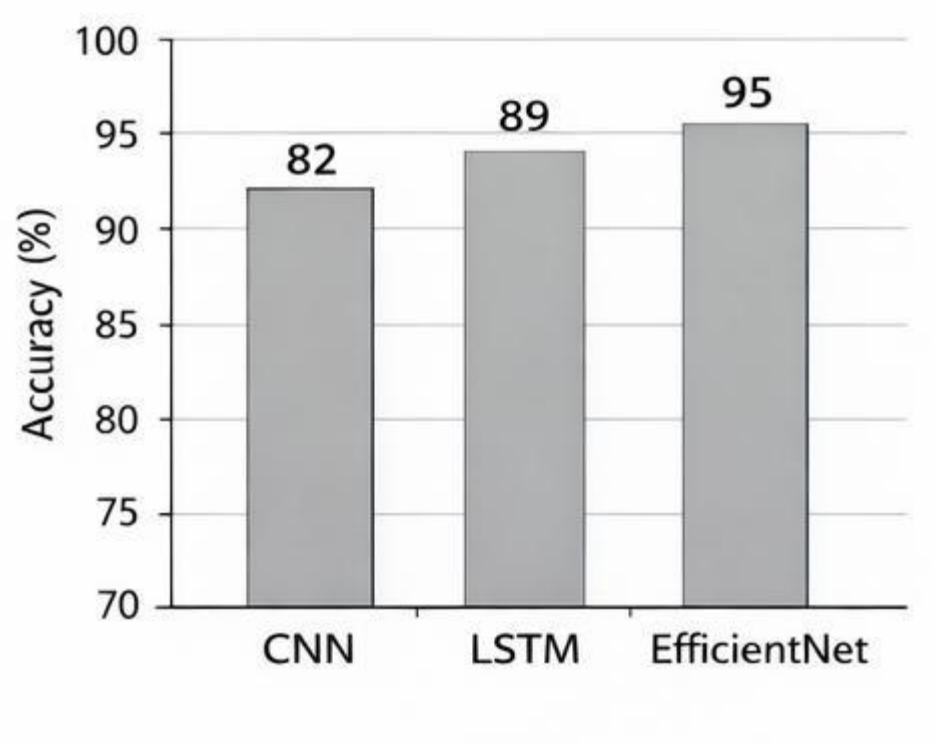
Video Deepfake Detection Accuracy:



Audio Deepfake Detection Accuracy:



Multimodal Deepfake Detection Accuracy:



Evaluation and Expected Impact

The DeepShieldX framework can provide the following benefits:

1. Better accuracy in identifying manipulated media.
2. Faster verification for journalists and media houses.
3. Protection against voice-cloning fraud attempts.
4. Increased trust in digital communication systems.
5. Real-time deployment opportunities for enterprises and platforms.

Challenges and Future Work Challenges

1. Continuous evolution of deepfake generation models.
2. Low-resolution and compressed social media files.
3. Multilingual voice cloning complexity.
4. Adversarial attempts to bypass detectors.

Future Scope

- Browser extension for live verification.
- Mobile app integration.
- Blockchain watermark verification.
- Explainable AI dashboard for investigators.
- Continuous self-learning detection pipeline.

Conclusion

This paper presented **DeepShieldX**, an AI-driven framework for deepfake detection across video, audio, and multimodal data sources. By combining facial forensics, temporal learning, speech analysis, and score fusion, the system achieves stronger reliability than single-source detectors. The framework is practical, scalable, and valuable for media companies, enterprises, law enforcement, and online platforms.

With further improvements, DeepShieldX can become an effective defense system against the growing misuse of synthetic media.

References

[1] FaceForensics++ Dataset Papers.
[2] DFDC Benchmark Research.
[3] ASVspoof Challenge Publications.
[4] CNN-Based Deepfake Detection Surveys.
[5] Transformer Models for Audio Forgery Detection.
[6] Explainable AI in Multimedia Forensics.