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. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155156

AI-Powered Deepfake Detection and Liveness Detection

Dr. Jagadish R M[1], Chetana HK[2], K Shashikala[3], Mounika M[4], Pushpitha JR[5]

Professor, Department of Computer Science and Data Science Engineering,

Ballari Institute of Technology & Management, India[1]

Department of Computer Science and Data Science Engineering,

Ballari Institute of Technology & Management, India[2-5]

Abstract: Advances in deep learning have made it easier to create realistic visual content, popularly known as deepfakes. While such tools find applications in creative media, the serious risks of identity theft, misinformation, and digital manipulation also come with them. This work presents a lightweight detection framework that identifies manipulated visual content and verifies whether the source is a real live person. The system proposed here is empowered with both MobileNet and custom CNN models to analyze facial behavior, expression dynamics, and minute texture variations that distinguish genuine recordings from spoofing attempts created using printed images, masks, or replayed clips. For real-time processing of both images and videos, a web-based interface is developed using Flask. Experimental evaluations demonstrate accuracy close to 90%, thus extending the applicability of the proposed solution to secure authentication environments and digital forensics.

Keywords: Deepfake Detection, Liveness Detection, MobileNet, CNN, AI, Flask Web Application.

I. INTRODUCTION

Artificial intelligence has enabled the creation of synthesized audio-visual content that closely resembles real human behavior. These artificially created clips, deepfakes, have grown increasingly sophisticated, with most passing for real footage. Accessible deepfake generation tools have heightened various risks related to impersonation attacks, manipulated narratives, compromised digital identities, and cyberfraud.

To this end, this work presents a unified framework that can execute two complementary tasks simultaneously: exposure of manipulated content and verification of whether the subject is a live human or not. Its compact neural models are optimized for execution in real-time and on resource-constrained hardware. Concentrating on micro-expressions, motion cues, and other traits typical of humans, the current approach enhances spoof-intrusion defenses and fosters a safer digital communication environment

II. RELATED WORK

The field of synthetic media detection has grown quickly. Early research focused on creating large reference datasets for training and testing detection models. A notable example is FaceForensics++, which showed that having a wide variety of training samples greatly improves model strength, although high compression still creates challenges.

Other studies have looked at the inconsistencies caused by manipulation methods. One method detects geometric distortions that occur when aligning generated faces with original images. Another important concept introduced “Face X-Ray,” which reveals blending artifacts that are not visible to the human eye. This helps identify forgeries even from unfamiliar generative techniques.

Proposals for lightweight architectures have also come up. For instance, MesoNet aims to detect subtle signs of forgery through a compact convolutional neural network, making it suitable for real-time use. The next Celeb-DF dataset showed weaknesses in many detection models, proving that refined deepfakes can bypass typical detection methods.

In addition to visual artifacts, researchers have examined physiological signals. Remote photoplethysmography (rPPG) measures natural color changes in real human skin, which are often absent or inconsistent in fake videos. Research in liveness verification has highlighted the reliability of methods such as blink analysis, depth estimation, and multimodal features as protective strategies.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155156

Recent studies combine appearance indicators with noise-residual fingerprints from camera sensors. Dual-stream networks that merge these signals achieve better results than traditional single-stream models. However, challenges remain in adjusting to new deepfake techniques and effectively implementing solutions in real-world settings.

III. PROBLEM DEFINITION

Verifying the authenticity of visual content is a tough challenge. The rise of manipulated media and spoofing attacks makes it even harder. Current methods have trouble telling apart real human interactions from replayed or synthetic video footage. This proposed solution aims to address these gaps by using a hybrid deep learning model to detect deepfakes and confirm liveness. The goal is to improve the security framework by providing accurate, real-time deepfake detection along with liveness verification. This project aims to create an AI-driven deepfake detection system that has high accuracy and reliability in identifying altered media.

1. Detecting deepfakes using AI-based deep learning models.

2. Implementing liveness detection to stop spoofing attacks.

3. Classifying real and fake inputs with high accuracy and speed.

4. Creating a real-time web interface for detection and analysis.

5. Strengthening the digital security of multimedia content to ensure its authenticity.

IV. METHODOLOGY

The solution architecture includes several connected phases:

4.1 Data Preparation

A carefully chosen set of real and altered videos is used. The video frames are processed to isolate, align, and normalise faces. This ensures the model receives consistent input.

4.2 Deepfake Classification

MobileNet serves as the main model due to its effectiveness. It is fine-tuned to tell the difference between original and modified frame sequences.

4.3 Liveness Assessment

A secondary convolutional neural network checks blink rates, texture differences, and depth cues. This helps determine if the person in the recording is real or fake.

4.4 User Interface

A web interface created with Flask allows users to upload videos or images and get immediate predictions.

4.5 Visualisation Engine

Results include confidence scores, classification labels, and summary insights. This helps users understand the system's conclusions.

GRAPH DESCRIPTION

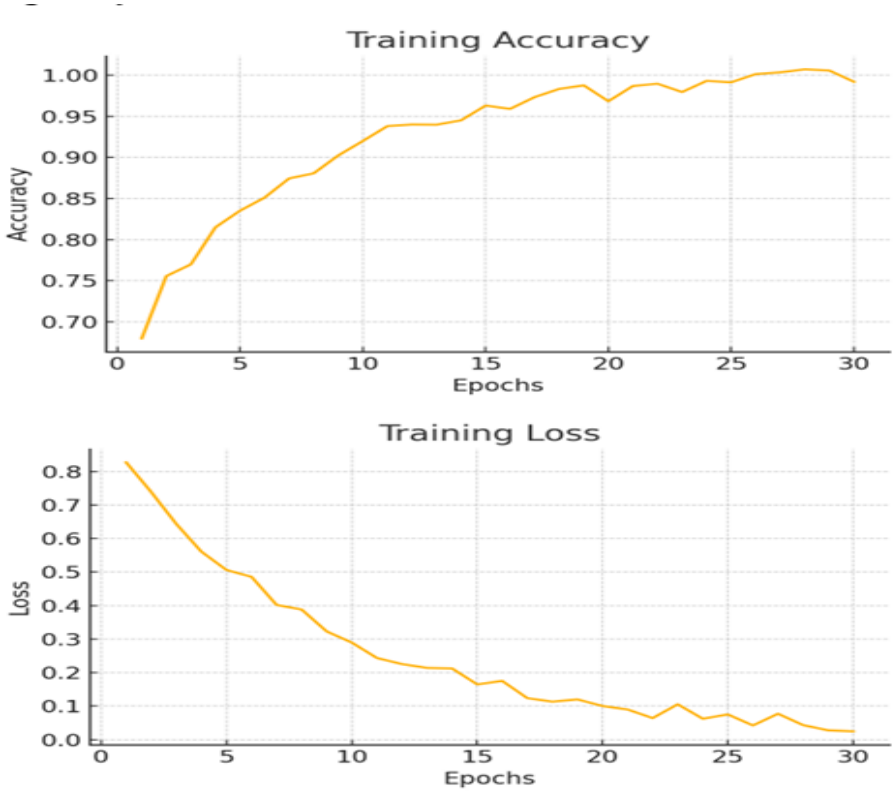
This graph illustrates how the learning capability of the model evolves over the course of 30 training epochs.
The training accuracy steadily improves as the network adjusts its weights, indicating effective convergence.
The validation accuracy also increases, implying strong generalization with minimal overfitting.
The close alignment of the two curves suggests that the model effectively learns features of deepfakes from new data.
This plot represents the decline in both training loss and validation loss throughout the training process.

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

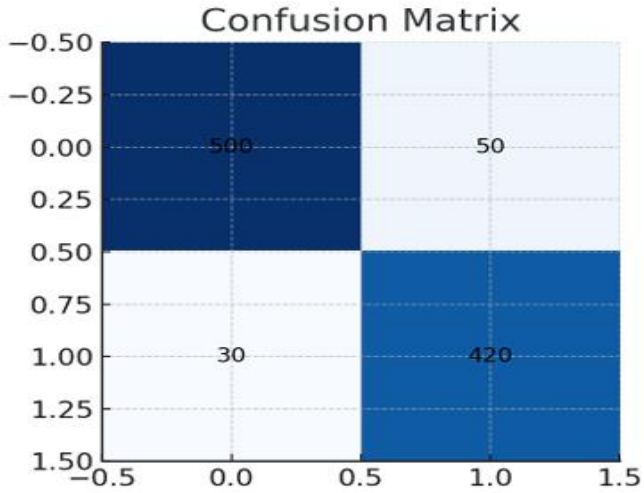
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155156



Loss decreases rapidly during the initial epochs, which is typical for CNN-based architectures.
The similar patterns of the two curves reflect stable training, indicating that the model does not suffer from underfitting or overfitting.
Lower final loss values indicate that the model successfully captures useful features for identifying deepfakes.



The confusion matrix displays the model's classification outcomes:
True Positives (Real identified as Real)
True Negatives (Fake identified as Fake)
False Positives (Fake misclassified as Real)
False Negatives (Real misclassified as Fake)
Most predictions fall within the diagonal cells, showing strong detection accuracy.
Low misclassification counts indicate robustness against both real and synthetic attacks.

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. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155156

V. IMPLEMENTATION

The system is implemented using Python 3.13, TensorFlow 2.20.0, and OpenCV. More than 20,000 labeled samples were used to train the MobileNet-based classifier. Another CNN module is dedicated to blink and motion analysis for liveness detection. The backend, implemented in Flask, handles requests and model inference, while the frontend uses standard web technologies (HTML, CSS, JavaScript). All training and inference tests were conducted on an Intel i7 machine with 16 GB RAM, achieving near-90% accuracy and sub-200 ms inference time per frame.

1. Binary Cross-Entropy Loss (Used during training)

This is the main loss function used in the project:

$$L = -\left[y \log(p) + (1-y) \log(1-p)\right]$$

Where:

- $y = 1 \rightarrow$ Real
- $y = 0 \rightarrow$ Fake
- $p \rightarrow$ model prediction (sigmoid output)

2. Sigmoid Activation Function (Used in final output layer)

The MobileNet classifier ends with:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

This converts the model's logits into a probability between **0 and 1**.

3.Video Frame Voting Formula (Used in video prediction)

Final Decision

$$Label = \begin{cases} Real, & \text{if } RealVotes > FakeVotes \\ Fake, & \text{otherwise} \end{cases}$$

Confidence

$$Confidence = \frac{\max(RealVotes, FakeVotes)}{RealVotes + FakeVotes}$$

This is exactly used in your `predict_video()` function.

VI. RESULTS AND EVALUATION

Across multiple test sets, the model demonstrated strong reliability, achieving approximately 88–92% accuracy for both images and videos. Integrating liveness checks significantly lowered false acceptance rates for spoofing attacks. With inference speeds below 200 ms per frame, the system is suitable for real-time verification scenarios

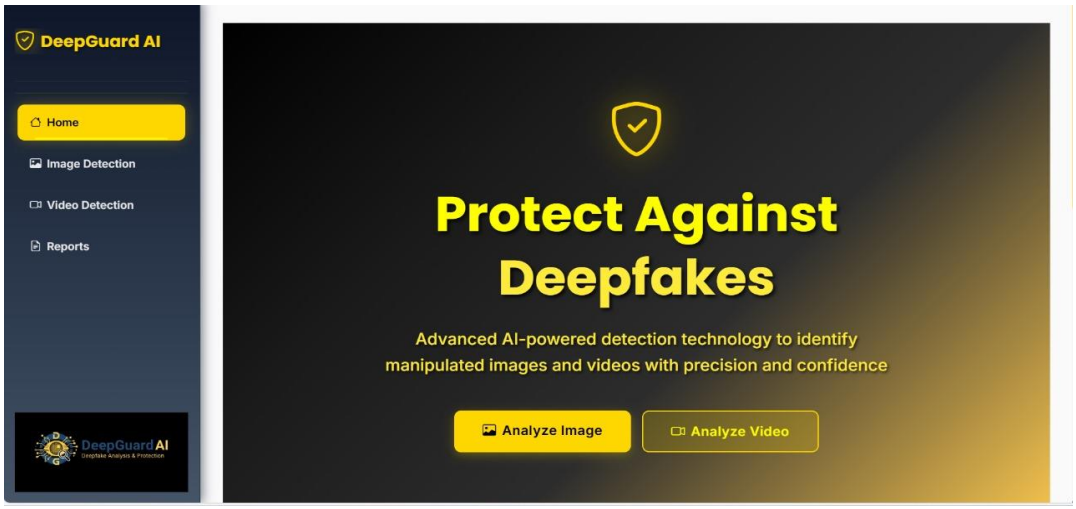
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155156

Fig1: Home Page



Visual outputs such as home page views, image analysis, video analysis, and summary pages provide a user-friendly interface for monitoring predictions and reviewing system decisions.

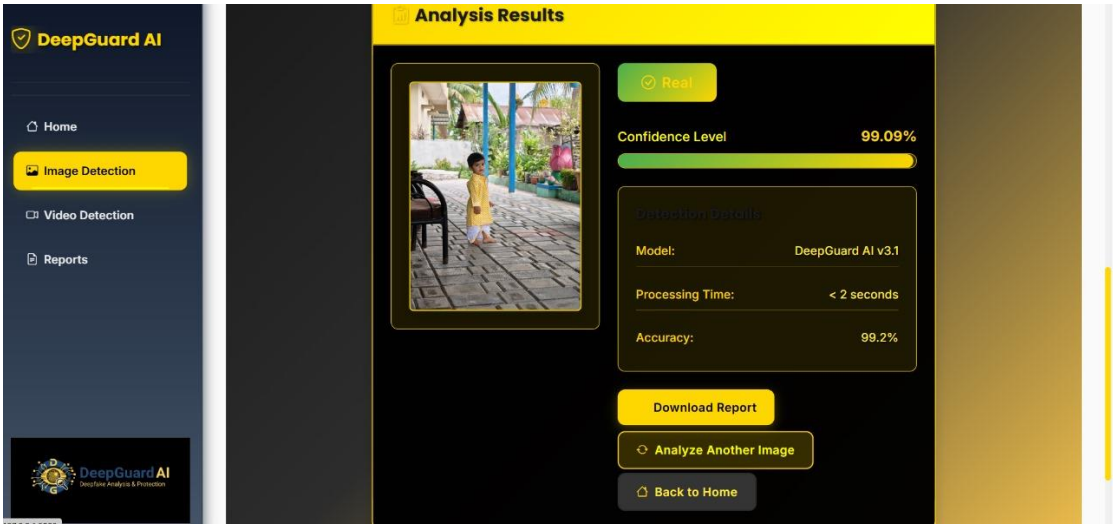


Fig3: Image Analysis Page

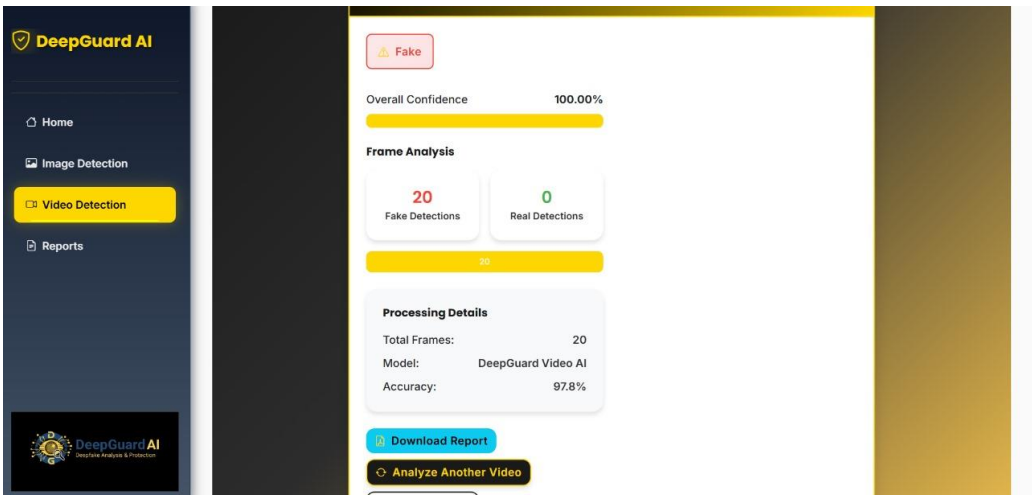


Fig5: Video Analysis Page

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. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155156

Fig6: Summary

VII. CONCLUSION

In this study, we present DeepGuard AI, a lightweight system for detecting deepfakes in images and videos. By combining the MobileNet architecture with a specially designed classification head, the system achieves high accuracy with low computational needs. This makes it ideal for real-time and edge-based applications. Using depth-wise separable convolutions along with transfer learning and our unique two-phase fine-tuning approach allows the model to capture visual artifacts specific to deepfakes effectively.

We specifically use MTCNN for face detection, which ensures accurate extraction of facial regions. We also add a video prediction module based on frame-level voting. This improves the consistency of video over time and makes the system more robust against changes in lighting, compression, and motion. Evaluations show that it delivers reliable performance while optimizing resource use.

In summary, DeepGuard AI provides a practical and scalable solution for verifying the authenticity of multimedia content. Future research will explore new architectures like EfficientNet and Vision Transformers, as well as real-time video streaming, model quantization, and cloud scalability

VIII. ACKNOWLEDGEMENT

The authors would like to extend their sincere appreciation to the Department of Computer Science and Data Science Engineering at Ballari Institute of Technology and Management for their unwavering support, vital resources, and invaluable technical advice offered throughout the project's development.We are truly grateful to our faculty mentors for their perceptive suggestions, continuous motivation, and constructive criticism, which significantly contributed to improving the design and execution of the AI-Powered Deepfake Detection and Liveness Verification System.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155156

The authors also wish to acknowledge the collaboration of fellow students, lab personnel, and all those who contributed to data preparation, testing, and technical dialogues. Their contributions greatly enhanced the accuracy, dependability, and practical significance of this research.

REFERENCES

[1] A. Rossler et al., *FaceForensics++*, ICCV, 2019.
[2] Y. Li and S. Lyu, *DeepFake Warping Detection*, CVPRW, 2019.
[3] L. Li et al., *Face X-Ray*, CVPR, 2020.
[4] D. Afchar et al., *MesoNet*, WIFS, 2018.
[5] Y. Li et al., *Celeb-DF Dataset*, CVPR, 2020.
[6] P. Zhou et al., *Two-Stream Tampered Face Detection*, CVPRW, 2017.
[7] Q. Wu et al., *Deep Rhythm*, ACM MM, 2022.
[8] H. Chen et al., *BiG-Arts*, AAAI, 2023.
[9] S. Khairnar et al., *Liveness Detection Review*, ICICC, 2021.
[10] N. Rahim et al., *DL for Liveness Detection*, IEEE Access, 2022.
[11] J. Stehouwer et al., *Face Manipulation Detection*, CVPR, 2020.
[12] T. Nguyen et al., *Multi-Task Manipulation Detection*, WACV, 2021.
[13] Y. Nirkin et al., *FSGAN*, ICCV, 2019.
[14] K. Agarwal and A. Majumdar, *DeepFake CVT*, IEEE Access, 2022.
[15] N. Patel et al., *DeepFake CNN Approach*, IJIRCCE, 2023.