Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Deepfake Detection Using CNNs

M Uday Kumar[1], A Rishikesh[1], G Ravi Kumar[2], K Sreeja[2]

[1]Assistant Professor, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India
[2]UG Students, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

Correspondence

M Uday Kumar

Assistant Professor, Department of CSE, Christu Jyothi Institute of Technology & Science, Jangaon, Telangana, India

- Received Date: 25 Jan 2026
- Accepted Date: 14 Mar 2026
- Publication Date: 20 Mar 2026

Keywords

Deepfake Detection, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Deep Learning, Generative Adversarial Networks (GANs), Django

Abstract

The rapid advancement of artificial intelligence and deep learning has led to the proliferation of deepfakes—synthetic media in which a person's likeness is convincingly replaced or manipulated using generative models such as GANs. While these forgeries pose severe ethical, social, and security threats, their increasing realism makes manual or traditional detection nearly impossible. This project proposes a Deepfake Detection System using Convolutional Neural Networks (CNNs) to automatically distinguish between authentic and manipulated facial videos or images. The system leverages spatial and temporal inconsistencies introduced during the synthesis process, learning deep feature representations that highlight subtle artifacts invisible to the human eye. The CNN model is trained on benchmark datasets of real and fake media, optimizing for high classification accuracy and robustness across varying compression levels and generation techniques. Experimental results demonstrate that the proposed CNN-based framework achieves effective deepfake identification, offering a scalable and efficient solution to combat digital misinformation and safeguard media integrity.

Introduction

The rapid advancement of artificial intelligence and deep learning technologies has significantly improved the ability to generate highly realistic synthetic media, commonly known as deepfakes. Deepfakes are digitally manipulated images or videos in which a person's face, expressions, or speech are artificially altered using deep learning techniques such as Generative Adversarial Networks (GANs) and autoencoder-based models. Although these technologies have useful applications, they also create serious risks when used maliciously to spread misinformation, create political manipulation, damage personal reputation, and generate misleading digital evidence. This project focuses on developing a Deepfake Detection System using Convolutional Neural Network (ResNeXt) and Long Short-Term Memory techniques to identify both spatial artifacts and temporal inconsistencies introduced during deepfake generation.

Background

The growth of digital media and online communication has transformed the way people consume and share information, making video content one of the most influential forms of communication. Earlier manipulation techniques relied on manual editing tools, which required significant technical skill and often left visible artifacts. However, with the emergence of GAN-based models and deep learning-driven face synthesis, fake videos can now be generated automatically with high realism. Because manual verification by experts is time-consuming and impractical for large-scale media analysis, automated deepfake detection methods using machine learning and computer vision are urgently needed.

Problem Statement

The rapid advancement of artificial intelligence and deep learning technologies has enabled the creation of highly realistic synthetic videos, commonly known as deepfakes. These manipulated videos are generated by replacing or altering facial expressions, lip movements, or identities in video content, making fabricated media appear authentic. Traditional methods of identifying manipulated videos rely mainly on manual observation or basic forensic analysis performed by experts. While obvious artifacts like unnatural blinking or lighting mismatches may sometimes reveal manipulation, these manual methods become highly unreliable when deepfakes are generated using advanced neural networks. Manual verification is also extremely time-consuming and impractical for handling the massive volume of digital video content uploaded online daily.

Literature Survey

Existing deepfake detection systems

Citation: Uday Kumar M, Rishikesh A, Ravi Kumar G, Sreeja K. Deepfake Detection Using CNNs. GJEIIR. 2026;6(3):0185.

primarily rely on conventional image-based analysis and basic deep learning methods for identifying manipulated facial content in videos. Most early systems focus on frame-level analysis, where individual video frames are extracted and processed independently using standard Convolutional Neural Network architectures or handcrafted feature extraction methods.

Several studies highlight specific approaches to exposing manipulated media. Li et al. (2018) proposed exposing deepfake videos by detecting face warping artifacts through a dedicated Convolutional Neural Network model. Another method by Li et al. proposed exposing AI-created fake videos by detecting eye blinking, which is a physiological signal often poorly represented in synthesized videos. Nguyen et al. utilized capsule networks to detect forged, manipulated images and videos in different scenarios, though this method incorporates random noise during the training phase which can cause failures on real-time data. Ciftci et al. proposed the detection of synthetic portrait videos using biological signals, extracting features like PPG maps from facial regions, though formulating a differentiable loss function to preserve biological signals remains highly complex. Because traditional models often process frames independently, they fail to capture temporal inconsistencies that naturally occur across video frames in high-quality deepfakes.

Proposed System

The proposed system introduces a hybrid deep learning architecture for deepfake detection by combining ResNeXt and Long Short-Term Memory networks. Unlike traditional frame-based approaches, the proposed system performs both spatial and temporal analysis to identify manipulated video content more effectively. Uploaded videos are first preprocessed using OpenCV, where facial frames are extracted, cropped, and standardized. The processed frames are passed to the ResNeXt network, which extracts deep spatial features such as facial texture irregularities, lighting inconsistencies, and edge distortions. These feature vectors are then sequentially fed into the LSTM layer, which analyzes frame-to-frame relationships and detects temporal inconsistencies commonly present in deepfake videos.

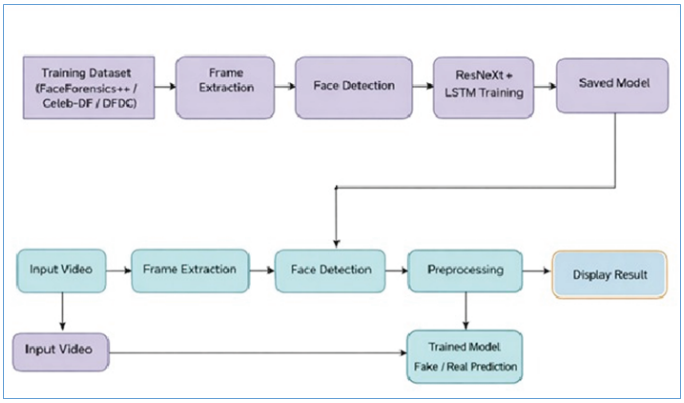
Modules used

- **User Interface Module:** This module provides a web-based interface developed using Django. It allows users to intuitively upload video files for verification and clearly displays the final prediction indicating whether the uploaded video is Real or Fake.
- **Video Processing Module:** This module extracts frames from uploaded videos and performs preprocessing operations such as resizing, normalization, and facial region extraction using OpenCV before passing data to the model.
- **Classification Module:** The classification module loads the trained hybrid deep learning model and performs inference on extracted video frames. The model analyzes spatial and temporal patterns within the video sequence to determine whether the uploaded video is fake or real.

Advantages of Proposed System

- **Hybrid Spatial and Temporal Analysis:** Combines CNN-based spatial feature extraction with LSTM-based temporal sequence learning.
- **Improved Prediction Accuracy:** Detects subtle visual and sequential inconsistencies in high-quality deepfake videos.
- **Better Generalization:** Training on multiple benchmark datasets improves performance on unseen manipulated content.
- **Reduced Dependence on Handcrafted Features:** The model automatically learns relevant deep features directly from video data.
- **Real-Time Prediction Capability:** Integration with a web interface allows immediate classification of uploaded videos.
- **Scalable Deployment:** The system can be extended for cloud deployment, browser integration, and live media verification

System architecture



The system architecture of the Deepfake Detection system is designed to process uploaded video data and classify it as Real or Fake using a hybrid deep learning model. In the training stage, benchmark deepfake datasets such as FaceForensics++, Celeb-DF, and Deepfake Detection Challenge Dataset are used. The input videos are processed using OpenCV to extract and normalize facial frames. The ResNeXt model extracts deep spatial features from each frame, capturing facial texture details, edge patterns, and visual inconsistencies. The extracted feature sequences are then provided to the Long Short-Term Memory model, which learns temporal dependencies and frame-to-frame inconsistencies. In the prediction stage, the trained model processes user-uploaded videos from the Django web interface frame by frame, ultimately generating a classification result that is displayed to the user.

Results

The output of the model classifies the uploaded video as "Real" or "Fake" along with the confidence of the model. The testing process demonstrated that the Deepfake Detection system performs reliably across different scenarios, successfully executing all test cases to produce expected outputs. The hybrid deep learning model accurately classified both real and manipulated videos, while the Django interface effectively handled uploaded inputs and displayed prediction results seamlessly. No major functional issues were observed during testing, indicating that the system satisfies its design objectives and provides stable performance for deepfake video classification.

Conclusion

This project successfully designed and implemented an automated Deepfake Detection system using a hybrid deep learning architecture that combines ResNeXt and Long Short-Term Memory models. The system analyzes uploaded video content and accurately classifies it as Real or Fake by leveraging

spatial feature extraction and temporal sequence analysis to capture facial inconsistencies and motion-related manipulation patterns. The integration of a Django-based web interface allows users to upload videos and receive predictions in an intuitive and user-friendly manner. Overall, this project demonstrates that hybrid deep learning models provide an effective and scalable solution for deepfake detection.

Future scope

- **RESTful API Deployment:** Deploying the trained hybrid deep learning model as an API to allow external applications, media verification platforms, and digital content systems to integrate real-time video verification.
- **Explainable Artificial Intelligence (XAI):** Integrating XAI techniques to highlight manipulated facial regions or temporal inconsistencies that influence the model's prediction, providing better transparency.
- **Real-Time Webcam Detection:** Extending the system to support real-time webcam-based deepfake detection by processing live video streams continuously.
- **Model Optimization:** Applying techniques such as quantization or pruning to reduce model size and improve inference speed, making it suitable for devices with limited computational resources.

References

1. Kaiming He, Xi Zhang, Shaoqing Ren, & Jian Sun (2016). Deep Residual Learning for Image Recognition. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.
2. Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, & Kaiming He (2017). Aggregated Residual Transformations for Deep Neural Networks. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.
3. Sepp Hochreiter & Jürgen Schmidhuber (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.
4. Andreas Rossler et al. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. Proceedings of IEEE International Conference on Computer Vision.
5. Yuezun Li, Siwei Lyu, "ExposingDF Videos By Detecting Face Warping Artifacts," in arXiv:1811.00656v3.
6. David Güera and Edward J Delp. Deepfake video detection using recurrent neural networks. In AVSS, 2018.