iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.83221

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Deepfake and Morphed Video Detection System Using CNN and LSTM

Sanjay S, R Marimuthu, Sankara Narayanan S T
Department of Cyber forensics Dr MGR educational Research and institute

Abstract— The rapid growth of deepfake technology has created serious challenges for digital trust and online security. Deepfakes are synthetic videos generated using advanced artificial intelligence techniques, especially Generative Adversarial Networks (GANs), which can realistically manipulate faces, expressions, and speech. Traditional detection systems mainly analyse individual video frames and often fail to identify inconsistencies that appear across time. This research proposes a hybrid deep learning model that combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to analyse both spatial and temporal features of videos. EfficientNetB0 is used to extract visual features from each frame, while stacked LSTM layers learn motion patterns and sequential inconsistencies. The model achieved 88.83% accuracy, 100% recall, and a 91.95% F1-score. A Flask-based web application was also developed for practical deployment. Results show that combining spatial and temporal learning significantly improves deepfake detection performance compared to CNN-only approaches.

Keywords— Deepfake Detection, CNN, LSTM, EfficientNetB0, Generative Adversarial Networks, Temporal Analysis, Video Classification, Deep Learning

I. INTRODUCTION

Advancements in artificial intelligence have enabled powerful media generation tools capable of producing highly realistic fake videos, commonly known as deepfakes. While these technologies offer benefits in entertainment and education, they also introduce risks such as misinformation, identity fraud, and reputational damage.

Most existing detection methods rely on analysing single images extracted from videos. Although these methods can identify visual artifacts, they ignore how facial movements evolve over time. Deepfake videos often contain subtle temporal irregularities such as unnatural blinking patterns, inconsistent lip synchronization, or sudden frame transitions that cannot be detected through frame-level analysis alone.

To overcome this limitation, this study introduces a hybrid CNN–LSTM framework that captures both visual details and temporal behaviour. The system also includes a web-based interface that allows users to upload videos and receive authenticity predictions, making the solution practical for real-world use.

II. SCOPE OF THE STUDY

This study focuses on developing a reliable deepfake video detection system using a hybrid CNN–LSTM deep learning model. The research aims to analyse both visual features within individual frames and motion patterns across video sequences to improve detection accuracy. The system is designed to identify manipulated facial videos created using modern deepfake techniques and provide authenticity predictions through a user-friendly web application. The proposed approach mainly concentrates on video-based analysis and practical deployment using accessible computing resources. While the study demonstrates effective detection performance, it is limited to visual data and short video sequences, leaving opportunities for future improvements such as audio analysis, real-time detection, and multimodal deepfake verification.

III. RELATED WORK

Early deepfake detection techniques used handcrafted features such as facial landmarks, head pose estimation, and eye-blink detection. While computationally efficient, these approaches became less effective as deepfake generation improved.

Deep learning models later introduced CNN-based detection, which automatically learned manipulation patterns from images. However, these models still analysed frames independently. Recent studies have explored sequence-based learning using LSTM and attention mechanisms to model temporal relationships between frames.

Despite improvements, challenges such as cross-dataset generalization, real-time detection, and robustness against advanced GAN models remain open research problems.

IV. PROPOSED SYSTEM ARCHITECTURE

A. Overall Workflow

The proposed system follows six main stages: (1) video upload through a web interface; (2) frame extraction using OpenCV; (3) image pre-processing and normalization; (4) spatial feature extraction using CNN; (5) temporal sequence modelling using LSTM; and (6) final classification into Real or Fake. The system outputs a prediction along with confidence scores and suspicious frame previews.

B. Pre-processing

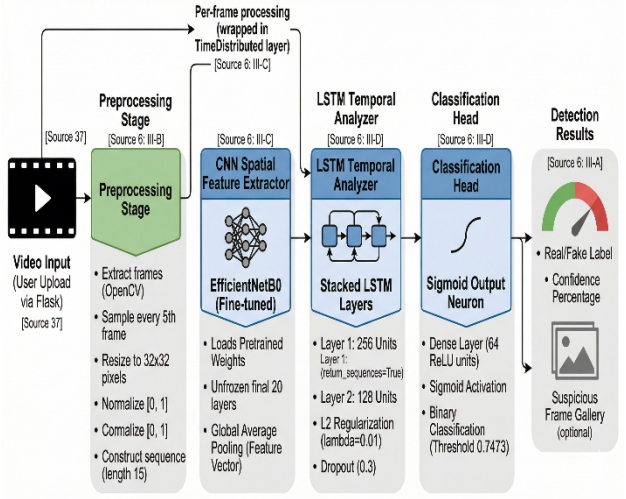
Videos are processed by extracting frames at regular intervals to reduce computation while preserving motion information. Each frame is resized to 32×32 pixels and normalized. Sequences of 15 frames are grouped together to provide sufficient temporal context for learning motion patterns.

C. CNN-Based Spatial Feature Extraction

EfficientNetB0, pretrained on ImageNet, is used as the backbone network. It learns visual characteristics such as texture inconsistencies, lighting mismatches, and facial blending artifacts. The Time Distributed layer enables frame-by-frame processing while maintaining sequence order.

D. LSTM-Based Temporal Modelling

The extracted features are passed to stacked LSTM layers that learn temporal dependencies across frames. These layers detect motion irregularities and evolving inconsistencies that typically appear in manipulated videos. The first LSTM layer comprises 256 units with return sequences=True, supplying the full temporal context to the second layer of 128 units. Both layers incorporate L2 weight regularization ($\lambda = 0.01$) and dropout (rate = 0.3) to prevent overfitting. A Dense layer with 64 RELU-activated units precedes the final sigmoid output neuron performing binary classification with an empirically determined threshold of 0.7473.



V. EXPERIMENTAL SETUP

A. Dataset

The dataset consists of real and fake videos collected from Kaggle and additional real-world sources. Fake samples include face-swapping and re-enactment manipulations. The data was split into 80% training and 20% validation sets.

B. Training Configuration

- Frame size: 32×32 pixels
- Sequence length: 15 frames

- Optimizer: Adam (learning rate: 0.00005)
- Loss function: Binary Cross-Entropy
- Epochs: 10
- Class imbalance handled using balanced class weights

VI. RESULTS AND DISCUSSION

A. Performance Evaluation

The proposed model achieved the results shown in Table I. The perfect recall indicates that all deepfake videos were successfully detected, which is particularly important for forensic and security applications where missing fake content can have serious consequences.

TABLE I PERFORMANCE METRICS OF THE PROPOSED CNN–LSTM MODEL

Metric	Score
Accuracy	88.83%
Precision	85.09%
Recall (Sensitivity)	100.00%
F1-Score	91.95%
ROC-AUC	90.09%
Optimal Threshold	0.7473

B. Comparison with CNN Baseline

A CNN-only model achieved only 72% accuracy, proving that temporal learning significantly enhances detection capability. The hybrid model improved both accuracy and reliability by capturing motion-based inconsistencies. Table II presents a direct comparison.

TABLE II COMPARISON OF CNN-ONLY VS. CNN+LSTM MODELS

Model	Architecture	Acc.	F1	AUC
CNN Only	EfficientNetB0+Dense	72%	0.77	0.60
CNN+LSTM	EfficientNetB0+LSTM	88.83%	0.92	0.90

C. Observations

The LSTM component successfully detected frame flickering patterns, irregular facial motion, lip-sync mismatches, and progressive visual anomalies. These temporal artifacts are invisible to frame-level classifiers but become apparent when sequential context is modelled.

D. Limitations

Some challenges remain, including reduced performance on high-quality GAN-generated videos, sensitivity to occlusions and shadows, and higher inference time for real-time applications. Future improvements may include attention mechanisms, model compression, and multimodal analysis.

VII. WEB APPLICATION IMPLEMENTATION

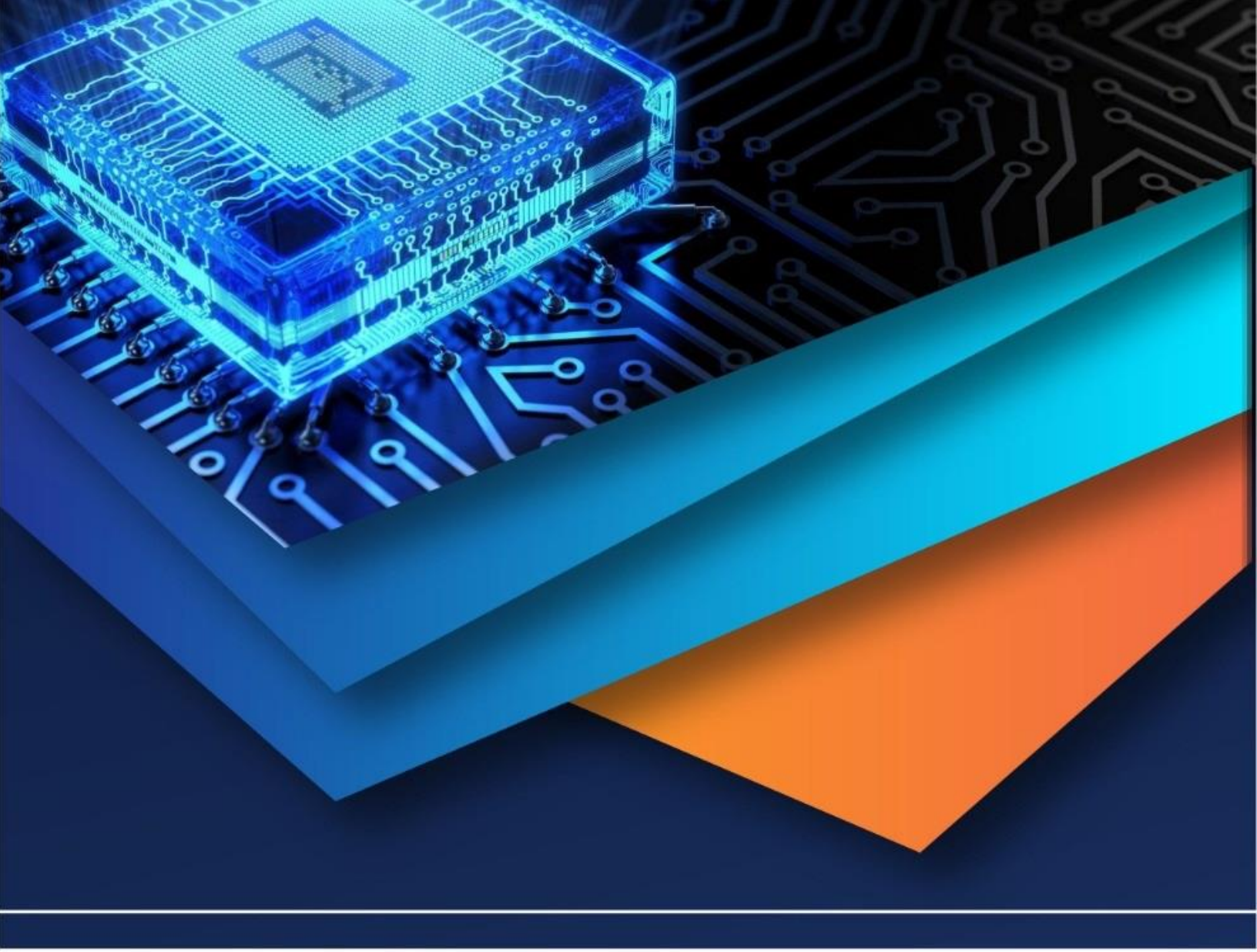
A Flask-based web platform was developed to demonstrate practical usage. Users can upload videos in common formats and receive authenticity predictions. The system uses TensorFlow, OpenCV, and PostgreSQL and runs efficiently even without GPU acceleration.

VIII. CONCLUSION AND FUTURE WORK

This study presents a hybrid CNN–LSTM model for deepfake detection that combines spatial and temporal learning. Experimental results demonstrate significant improvements over traditional frame-based approaches, achieving 88.83% accuracy and 100% recall. Future work will focus on: audio–visual multimodal detection, transformer-based temporal modelling, real-time optimization, adversarial robustness, and continuous learning from user feedback.

REFERENCES

[1] H. Cho Taliya et al., "Review: Deepfake Detection Techniques using DNN," ICAST 2023, pp. 480–484, Doi: 10.1109/ICAST59062.2023.10454938.
[2] T. Jung, S. Kim, K. Kim, "Deep Vision: Deepfakes Detection Using Human Eye Blinking Pattern," IEEE Access, vol. 8, pp. 83144–83154, 2020.
[3] Malik et al., "Deepfake Detection for Human Face Images and Videos: A Survey," IEEE Access, vol. 10, pp. 18757–18775, 2022.
[4] S. Waseem et al., "Deepfake on Face and Expression Swap: A Review," IEEE Access, vol. 11, pp. 117865–117906, 2023.
[5] M. S. Rana et al., "Deepfake Detection: A Systematic Literature Review," IEEE Access, vol. 10, pp. 25494–25513, 2022.
[6] Kaushal et al., "A Comparative Study on Deepfake Detection Algorithms," ICAC3N 2022, pp. 854–860, Doi: 10.1109/ICAC3N56670.2022.10074593.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY