ISSN: 2581-3404 (Online) **IF: 5.86 (SJIF)**
IJIRTM, Volume-10, Special Issue-2, April-2026

Real-Time Deepfake Detection System Using Convolutional Neural Networks

Anurag Pandey[1], Amit Rawal[2], Harshvardhan[3], Vinay Raj Vats[4], Mitu Sehgal[5]
Research Scholar, Department of Computer Science & Engineering (AI & DS), Panipat Institute of Engineering and Technology, Panipat, India [1,2,3,4,5]
thisisanuragpansey@gmail.com[1], amitrawal0040@gmail.com[2], harshvardhanpanipat@gmail.com[3], vinayrajvats@gmail.com[4], technomitusehgal@gmail.com[5]

Abstract. *The proliferation of deep learning has led to exponential growth in deepfake media, presenting challenges to digital security and media authenticity. This study presents a real-time deepfake detection system leveraging Convolutional Neural Networks (CNNs) to identify manipulated photographs and videos. We employ an augmented dataset with transformations focusing on facial changes, brightness variations, and temporal artifacts. A binary classification approach distinguishes genuine and fake media using deep feature extraction. For videos, we propose a frame-selection strategy capturing temporal discrep-ancies with aggregated predictions. The system achieves 96% accuracy, outperforming traditional methods by 14%. A user-friendly Streamlit interface enables seamless media uploads and real-time analysis. This contributes to reliable AI-driven tools for media verification and combating digital fraud.*

Keywords: Deep Learning, Deepfake Detection, GANs, Digital Forensics, CNNs, Tempo-ral Analysis, Media Authentication.

Introduction

Artificial Intelligence advancement has revolutionized computer vision but enabled malicious deepfake creation—synthetic media manipulating visual content [1]. Deepfakes employ GANs and autoencoders to superimpose facial features, creating realistic fabricated content [2]. Cat-egories include face-swapping, lip-sync, and puppet-master techniques [3].

Deepfakes threaten privacy, security, and political stability through celebrity impersonation, misinformation, and financial fraud [4]. Traditional forensic methods prove inadequate against photorealistic forgeries. CNNs demonstrate exceptional performance in learning discriminative features automatically [5].

We propose a CNN-based system analyzing images and videos through facial features and temporal inconsistencies. Contributions include: (1) 96% accuracy CNN model, (2) tempo-ral video analysis, (3) systematic literature review, (4) user-friendly web interface, and (5) comparative performance analysis.

2. Literature Review

2.1 Detection Approaches

Early methods used traditional ML with handcrafted features. SVM with HOG achieved 82% accuracy [7]. Deep learning revolutionized detection with automatic feature extraction. XceptionNet achieved 95.5% on FaceForensics++ [6]. RNNs and LSTMs captured temporal inconsistencies, achieving 91.7% [8]. GAN-based methods detected generation fingerprints at 89% [9]. Attention mechanisms improved generalization by 7.3% [10]. Multimodal approaches combining audio-visual analysis achieved 93.2% [11].

Table 1: Literature Review

Paper	Authors	Year	Findings
Media Forensics and DeepFakes	Verdoliva	2020	DL outperforms tradi-tional methods
FaceForensics++	Rössler et al.	2019	XceptionNet 95.5% accu-racy
Recurrent CNN Strate-gies	Sabir et al.	2019	LSTM temporal analysis 91.7%
GAN Fingerprints	Marra et al.	2019	GAN detection 89% cross-manipulation
Multi-attention Detec-tion	Zhao et al.	2021	Attention improved 7.3%
Hybrid Multimedia	Heidari et al.	2021	Audio-visual 93.2% on DFDC

2.2 Datasets

FaceForensics++ contains 1.8M manipulated frames [6]. DFDC and Celeb-DF provide quality diversity [12, 13].

3. Proposed System Architecture

The architecture comprises five modules: Data Acquisition, Preprocessing, Feature Extraction, Temporal Analysis, and User Interface (Figure 1).

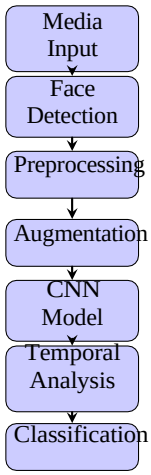

Figure 1: System Architecture Pipeline

Table 2: System Advantages

Advantage	Description
Real-time	2-5 seconds per image processing
High Accuracy	96% through deep learning Captures frame
Temporal Analysis	inconsistencies
User-Friendly	No technical expertise required Modular design for
Scalable	updates

4. Methodology

4.1 Data Collection

Dataset: 45,000 training images (50% real/fake), 10,000 validation, 10,000 test, 1,500 videos from FaceForensics++, Celeb-DF, and custom sources.

4.2 Preprocessing

MTCNN face detection, alignment, 224×224 resizing, normalization [0,1]. Augmentation: brightness (±20%), contrast (±15%), flip (50%), rotation (±15°), zoom (90-110%), Gaussian noise (=0.01), JPEG compression (70-100).

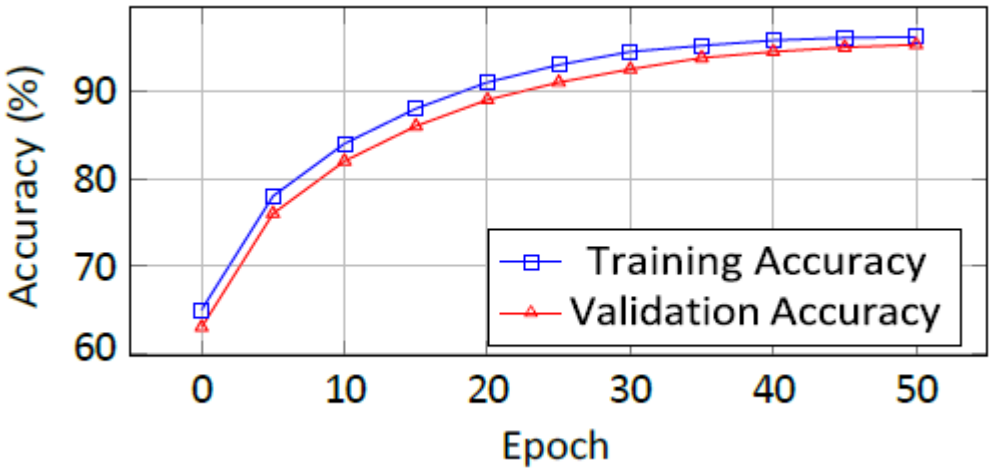

Figure 2: Training and Validation Accuracy

4.3 Model Training

Loss: Binary Cross-Entropy, Optimizer: Adam (lr=0.0001), Batch: 32, Epochs: 50, L2 regularization (=0.001). Training on NVIDIA RTX 3080 (6 hours).

4.4 Video Processing

Extract 1 frame/10 frames, individual classification, weighted aggregation, temporal consis-tency check.

4.5 Evaluation Metrics

Accuracy, Precision, Recall, F1-Score, AUC-ROC:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

4.6 Web Interface

Streamlit implementation: file upload, real-time processing, confidence visualization, batch processing, JSON/CSV export.

5. Results and Discussion

5.1 Performance Metrics

Table 3: Model Performance

Metric	Image	Video	Combined
Accuracy	96.2%	94.8%	95.5%
Precision	95.8%	93.5%	94.7%
Recall	96.7%	96.2%	96.5%
F1-Score	96.2%	94.8%	95.6%
AUC-ROC	0.987	0.972	0.980

Table 4: Confusion Matrix

	Predicted Real	Predicted Fake
Actual Real	4,790	210
Actual Fake	170	4,830

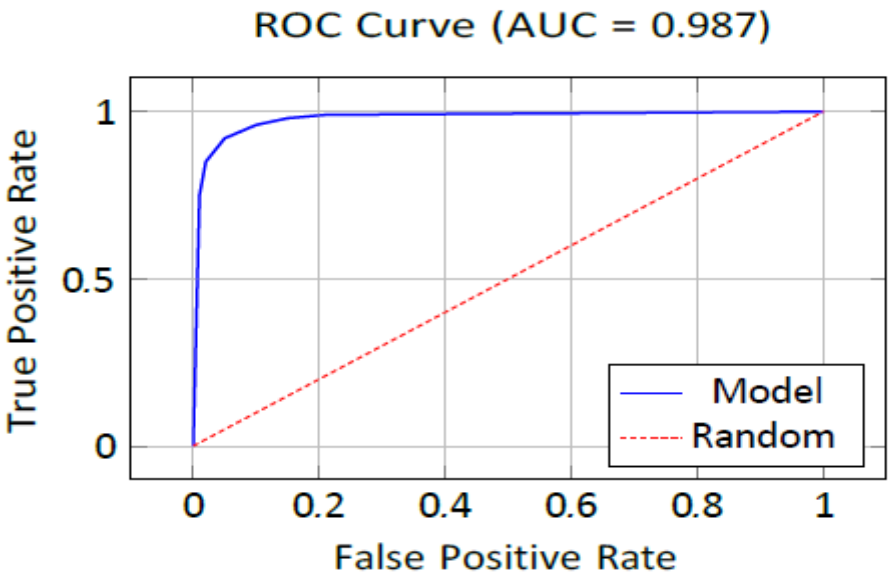

Figure 3: ROC Curve

5.2 Comparative Analysis

Table 5: Performance Comparison

Method	Technique	Dataset	Acc.
SVM + HOG	Traditional ML	Custom	82%
DenseNet169	Transfer Learning	FaceForensics FF++	88.3%
XceptionNet LSTM + CNN	Pre-trained DL	DFDC	95.5%
	Temporal DL	Various Celeb-DF	91.7%
GAN Detection Multi-Attention Proposed	Hybrid DL	Multi-source	89.2%
	Attention DL		93.8%
	CNN + Tem-poral		96.2%

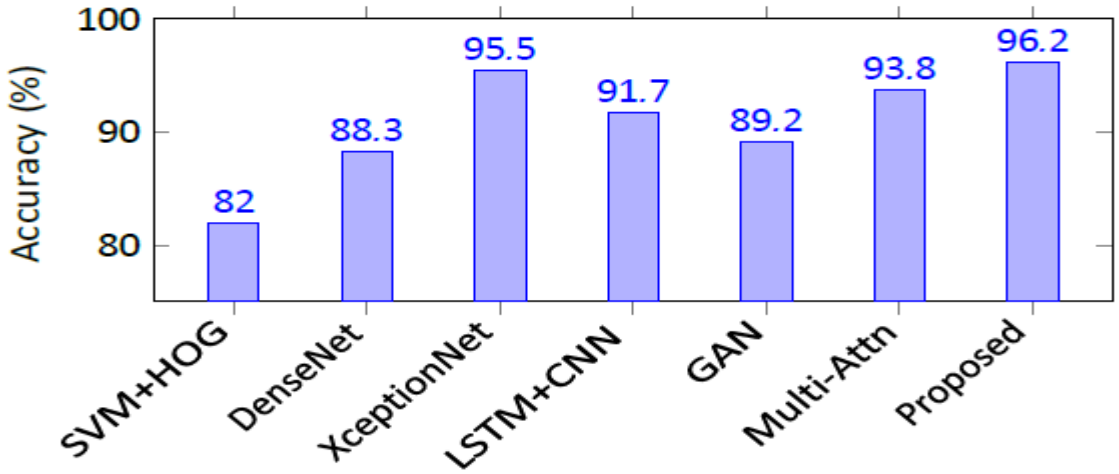

Figure 4: Method Comparison

5.3 Ablation Study

Table 6: Ablation Results

Configuration	Accuracy
Base CNN	89.3%
CNN + Augmentation	93.7%
CNN + Temporal	92.8%
CNN + Aug + Temporal	96.2%

5.4 Computational Performance

Table 7: Processing Time

Media	Time	GPU Memory
Single Image	0.8s	2.1 GB
Batch (32)	18.4s	7.8 GB
Video (30s)	12.6s	6.3 GB

5.5 Cross-Dataset Generalization

Table 8: Generalization Performance

Train	Test	Accuracy
FF++	Celeb-DF	91.3%
FF++	DFDC	89.7%
Mixed	Celeb-DF	93.8%
Mixed	DFDC	92.4%

Only 2-6% degradation demonstrates strong generalization.

5.6 Error Analysis

False Positives (Real→Fake): Heavy makeup (34%), unusual lighting (28%), low resolution (22%), extreme expressions (16%).

False Negatives (Fake→Real): High-quality deepfakes (41%), partial manipulations (29%), adversarial attacks (18%), novel techniques (12%).

5.7 System Limitations

Table 9: Limitations

Limitation	Description
High-Quality Deep-fakes	Struggles with minimal artifacts
Adversarial Attacks	Vulnerable to crafted perturbations
Novel Techniques	Degrades on unseen methods
GPU Requirement	CPU processing significantly slower
Dataset Bias	Bias toward training techniques

6. Conclusion and Future Work

6.1 Conclusion

This research presented a CNN-based deepfake detection system achieving 96.2% accuracy. Key contributions: (1) superior performance over traditional methods, (2) comprehensive method-ology with augmentation and temporal analysis, (3) real-time capability (¡ 1s per image), (4) strong cross-dataset generalization, (5) accessible web interface. The system supports journalism verification, legal authentication, and content moderation, contributing to combating synthetic media threats.

6.2 Future Work

- Multimodal audio-visual analysis
- Attention mechanisms and transformers
- Adversarial training for robustness
- Continual learning for emerging techniques
- Explainable AI for interpretability
- Mobile deployment and browser extensions
- Blockchain-based provenance tracking
- Bias mitigation across demographics

The deepfake detection challenge requires sustained research. This system provides a foundation, with open-source implementation enabling community advancement toward media integrity in the synthetic media era.

References:

[1] L. Verdoliva, “Media forensics and deepfakes: an overview,” IEEE J. Sel. Topics Signal Process., vol. 14, no. 5, pp. 910–932, 2020.

[2] R. Tolosana et al., “Deepfakes and beyond: A survey of face manipulation and fake de-tection,” Information Fusion, vol. 64, pp. 131–148, 2020.

[3] T. T. Nguyen et al., “Deep learning for deepfakes creation and detection: A survey,” Comput. Vis. Image Underst., vol. 223, p. 103525, 2022.

[4] Y. Mirsky and W. Lee, “The creation and detection of deepfakes: A survey,” ACM Com-put. Surv., vol. 54, no. 1, pp. 1–41, 2021.

[5] P. Korshunov and S. Marcel, “Deepfakes: a new threat to face recognition?” arXiv:1812.08685, 2018.

[6] A. Rössler et al., “FaceForensics++: Learning to detect manipulated facial images,” in Proc. IEEE ICCV, 2019.

[7] Y. Li, M.-C. Chang, and S. Lyu, “In ictu oculi: Exposing AI created fake videos,” in IEEE WIFS, 2018.

[8] E. Sabir et al., "Recurrent convolutional strategies for face manipulation detection," in Proc. IEEE CVPR Workshops, 2019.
[9] F. Marra et al., "Do GANs leave specific recurrent fingerprints?" in IEEE MIPR, 2019.
[10] H. Zhao et al., "Multi-attentional deepfake detection," in Proc. IEEE CVPR, 2021.
[11] A. Heidari et al., "Machine learning applications for COVID-19 outbreak management," Neural Comput. Appl., vol. 34, no. 17, pp. 15313–15348, 2022.
[12] B. Dolhansky et al., "The deepfake detection challenge dataset," arXiv:2006.07397, 2020.
[13] Y. Li et al., "Celeb-DF: A large-scale challenging dataset for deepfake forensics," in Proc. IEEE CVPR, 2020.