International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Deep Fake Detection: Using a web Based Convolutional Neural Network System

Saurabh Jain [1], Praveen Kumar Tiwari[2], Kalash Sharma[3] Kolla Charvi [4] ,Lalit Kumar Yadav [5] Praddyumn Raj Singh [6]

[1]*Guide Of Department of Computer Science Engineering, Babu Banarsi Das Institute of Technology and Management, Lucknow*
[2]*Bachelor of Technology in Computer Science Engineering, Babu Banarsi Das Institute of Technology and Management, Lucknow*

[3]*Bachelor of Technology in Computer Science Engineering, Babu Banarsi Das Institute of Technology and Management, Lucknow*
[4]*Bachelor of Technology in Computer Science Engineering, Babu Banarsi Das Institute of Technology and Management, Lucknow*
[5]*Bachelor of Technology in Computer Science Engineering, Babu Banarsi Das Institute of Technology and Management, Lucknow*
[6]*Bachelor of Technology in Computer Science Engineering, Babu Banarsi Das Institute of Technology and Management, Lucknow*

Abstract – The rapid proliferation of deepfake technology poses significant challenges to digital media authenticity, necessitating robust detection mechanisms. This paper presents a web-based deepfake detection system developed using Flask,
TensorFlow, and OpenCV, designed to classify uploaded videos as "REAL" or "FAKE" based on a pre-trained convolutional neural network (CNN) model. The system preprocesses video frames to a standardized (None, 128, 128, 3) input shape, leveraging a single-frame analysis approach for real-time classification. Key challenges, including model compatibility, input shape mismatches, and prediction biases, were addressed during development. Preliminary results indicate successful deployment on a local server, though limitations in model generalization were observed, with all test videos classified as "FAKE." This work highlights the feasibility of web-integrated deepfake detection and identifies areas for future enhancement

Deepfake detection has evolved alongside generative technologies. Early methods relied on visual artifacts, while modern approaches leverage deep learning. Li et al. proposed CNN-based detection using frame-level features, achieving high accuracy on datasets like Face Forensics++. Rossler et al. introduced the Face Forensics++ dataset, pairing real and manipulated videos, with Caption models expecting (128, 128, 3) inputs—similar to our system's model.

Key Words: Synthetic Media , Video Manipulation Detection Deep Fake , AI- generated Content , Deep Neural Network

1.INTRODUCTION :

This Deepfakes, synthetic media generated by artificial intelligence, have emerged as a dual-edged sword—offering creative potential while threatening misinformation and trust in visual content. With advancements in generative adversarial networks (GANs), detecting manipulated videos has become a critical research area. Traditional detection methods rely on manual analysis or desktop-based tools, limiting accessibility. This project introduces a web-based solution that integrates a pre-trained CNN model into a user-friendly interface, enabling non-experts to upload and analyze videos via a browser.

2. Methodology :

The development of the deepfake detection system followed a structured approach, combining frontend design, backend processing, and machine learning integration.

3.1 System Architecture :

Frontend: An HTML interface (index.html) with JavaScript for asynchronous file uploads via the Fetch API, styled with CSS (style.css).
Backend: Flask (app.py) serves the webpage, handles video uploads, and processes predictions, saving files to static/uploads/.
Model: A pre-trained CNN (deepfake_model.h5) expecting (None, 128, 128, 3) input, loaded with TensorFlow/Keras.

3.2 Video Preprocessing:

Frame Extraction: OpenCV's cv2.VideoCapture extracts a single frame from the uploaded video.

Resizing: Frames are resized to (128, 128) using cv2.resize() and normalized to [0, 1] by dividing by 255.0.
Input Formatting: A batch dimension is added with np. expand dims(), yielding (1, 128, 128, 3) for model input.

3.3 Classification:

The model outputs a scalar value (assumed sigmoid-activated), interpreted as a confidence score.
A threshold of 0.5 determines the label: "FAKE" if confidence > 0.5 else "REAL".

3.4 Deployment

 The system runs locally (http://localhost:5000) using Flask's development server, with debug mode enabled for real-time error tracking.

3.5 Development Challenges

Shape Mismatch: Initial attempts to process 10 frames ((1, 10, 128, 128, 3)) failed due to model incompatibility, resolved by switching to single-frame analysis.
File Format: Restricted to .mp4, .avi, and .mov via ALLOWED_EXTENSIONS, with codec support dependent on OpenCV's backend.

3. Results :

The system was tested on a local machine (Windows, Python 3.12) with sample .mp4 videos from various sources (phone recordings, stock clips). Key observations:
Deployment:: Successfully hosted at http://localhost:5000, with a responsive interface for video uploads.
Processing: Frames were consistently resized to (128, 128, 3), with model input shape verified as (1, 128, 128, 3).
Prediction: All tested videos were classified as "FAKE," with confidence scores ranging from 0.6 to 0.95 (e.g., Prediction: [[0.852]])

4. Conclusion

This project demonstrates a functional web-based deepfake detection system, integrating Flask, TensorFlow, and OpenCV to process video frames into a (None, 128, 128, 3) format for CNN-based classification. The system successfully uploads and analyzes .mp4, .avi, and .mov files, delivering predictions via a user-friendly interface. However, the uniform "FAKE" output indicates limitations in the pre-trained model's generalization, likely requiring retraining on a balanced dataset or adjustment of the classification threshold.

Future work includes:
Enhancing model accuracy with multi-frame analysis (e.g., (None, 10, 128, 128, 3) via LSTM).
Expanding supported formats with FFmpeg-backed OpenCV.
Deploying on a public server for broader accessibility.
This system lays a foundation for accessible deepfake detection, bridging the gap between AI research and practical application..

5. REFERENCES

1. A. Rossler et al., "FaceForensics++: Learning to Detect Manipulated Facial Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 12, pp. 2967-2981, 2020.

2. Y. Li et al., "Celeb-DF: A Large-scale Challenging Dataset for Deepfake Forensics," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 3207-3216

3. Goodfellow et al., "Generative Adversarial Networks," in Proceedings of Advances in Neural Information Processing Systems (NeurIPS), 2014, pp. 26722680.

4. T. Dolhansky et al., "DeepFake Detection Challenge Dataset," arXiv preprint arXiv:2006.07397, 2020.

5. P. Korshunov and S. Marcel, "DeepFakes: A New Threat to Face Recognition? Assessment and Detection," arXiv preprint arXiv:1812.08685, 2018.

6. . X. Yang et al., "Exposing Deep Fakes Using Inconsistent Head Poses," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019

7. . M. Verdoliva, "Media Forensics and DeepFakes: An Overview," IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910-932, 2020

8. . D. Guera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018.

.

9. Kundu, R., Balachandran, A., Roy-Chowdhury, A. K. (2025). TruthLens: Explainable DeepFake Detection. arxiv.org/abs/2503.15867

10. Chakraborty, R., Chakraborty, R., Rahimian, A. K., MacDougall, T. (2025). Training-Free DeepFake Detection. arxiv.org/abs/2503.15342

11. Hossain Shanto, M. D., et al. (2025). DFCon: Supervised Contrastive Learning for Deepfake Detection. arxiv.org/abs/2501.16704

12. Mehta, A., McArthur, B., Kolloju, N., Tu, Z. (2025). HFMF: Hierarchical Fusion for Deepfake Detection. arxiv.org/abs/2501.05631

13. Wyawahare, M., Tyagi, M., Rajasekaran, K. S. (2025). Comparative Analysis of Deepfake Detection Models. hrcak.srce.hr/en/326071

14. Deepfake-Eval-2024: A Multi-Modal In-the-Wild Benchmark of Deepfakes Circulated in 2024 Link: arXiv:2503.02857

15. SIDA: Social Media Image Deepfake Detection, Localization and Explanation with Large Multimodal Model Authors: Zhenglin Huang et al. Published: December 2024 Link: arXiv:2412.04292

16. A Improving Fairness in Deepfake DetectionAuthors: Yan Ju et al.Published: November 2023

17. Comprehensive Evaluation of Deepfake Detection Methods: Approaches, Challenges and Future ProspectsAuthor: Xixi Hu Published: February 2025Link: ITM Web of Conferences

18. SLIM: Style-Linguistics Mismatch Model for Generalized Audio Deepfake DetectionAuthors: Yi Zhu et al.Published: NeurIPS 2024Link: NeurIPS 2024

19. CrossDF:Link: arXiv:2310.00359 Authors: Shanmin Yang et al.Published: September 2023

20. Deepfake Detection via Joint Unsupervised Reconstruction and Supervised ClassificationAuthors: Bosheng Yan et al.Published: December 2022