IRJTAE

Ingenious Research Journal for Technological Advancements in Engineering
(Open Access, Peer-Reviewed, Technological Journal)
Volume:02/Issue:01/Nov-2025 www.irjtae.com

Deepfake Detection System: A Review

Shruti Wadhekar*[1], Sakshi Khode*[2], Mahevish Sheikh*[3], Yamini Babhulkhar*[4], Prof. A. V. Mahalle*[5]

*[1]Student, Computer Engineering, Government College of Engineering Yavatmal (GCOEY), Maharastra, India

*[2]Student, Computer Engineering, Government College of Engineering Yavatmal (GCOEY), Maharastra, India

*[3]Student, Computer Engineering, Government College of Engineering Yavatmal (GCOEY), Maharastra, India

*[4]Student, Computer Engineering, Government College of Engineering Yavatmal (GCOEY), Maharastra, India

*[5] Assistant Professor, Computer Engineering, Government College of Engineering Yavatmal (GCOEY), Maharastra, India

ABSTRACT

Deepfake technology, powered by deep learning and artificial intelligence, has revolutionized digital content creation by generating highly realistic synthetic media. These manipulations are primarily created using Generative Adversarial Networks (GANs) and autoencoders, which learn facial features and expressions to produce convincing fake videos or images. However, the misuse of this technology poses serious threats to privacy, security, and public trust. This review paper focuses on analysing various deep learning-based deepfake detection techniques, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based models. It summarizes key datasets such as FaceForensics++, Celeb-DF, and DFDC, and evaluates detection performance using metrics like accuracy, precision, and recall. Despite progress, current models struggle with generalization across unseen data and adversarial manipulations. The paper also highlights emerging trends such as multimodal detection and explainable AI for improved transparency and robustness. Overall, this review emphasizes the importance of developing reliable, efficient, and adaptable deep learning models to ensure digital content authenticity.

Keywords*: Deepfake, Deep Learning, Generative Adversarial Networks, Convolutional Neural Networks, Face Forgery, Artificial Intelligence, Image Manipulation, Digital Forensics.*

I. INTRODUCTION

The Deepfake technology uses deep learning and artificial intelligence to create highly realistic fake videos, images, or audio by manipulating human faces and voices. While this technology has beneficial uses in entertainment and education, its misuse raises serious concerns about privacy, misinformation, and digital security. Most deepfakes are generated using Generative Adversarial Networks (GANs), which learn and replicate human facial expressions to produce convincing synthetic media.

Current research focuses on deepfake detection systems that use neural networks such as CNNs, RNNs, and Transformers to identify subtle inconsistencies in frames, lighting, or movement. Although these models show good accuracy on benchmark datasets, they still face limitations in detecting unseen or highly realistic manipulations. Therefore, building robust, generalizable, and real-time detection techniques is vital to ensure digital authenticity and protect users from manipulated content.

II. METHODOLOGY

The **Deepfake Detection System** methodology focuses on studying various deep learning-based approaches used to identify manipulated digital media. This section describes the overall process of research design, data collection, and analysis performed to review and compare existing detection techniques. The methodology emphasizes understanding how different neural network architectures perform in identifying fake content generated using Generative Adversarial Networks (GANs).

A. Research Design

The research was designed as a **comparative analysis** of deep learning-based deepfake detection models. Existing literature, experimental studies, and open-source frameworks were analysed to understand the principles behind different algorithms. The process included reviewing image-based, video-based, and hybrid detection systems to highlight their working mechanisms and limitations.

B. Data Collection and Datasets

The data used in this study was collected from publicly available datasets such as **FaceForensics++**, **Celeb-DF**, and the **DeepFake Detection Challenge (DFDC)** dataset. These datasets contain real and fake images or videos that are commonly used for training and testing deep learning models. They help in evaluating the performance and generalization ability of detection algorithms.

C. Deep Learning Techniques Used

The analysis focuses on **Convolutional Neural Networks (CNNs)** for feature extraction, Recurrent Neural Networks (RNNs) for capturing temporal dependencies, and Transformer-based models for learning global attention features. Transfer learning was also reviewed as a technique to enhance performance when training data is limited.

D. Performance Analysis

Each detection model was evaluated based on accuracy, precision, recall, and F1-score metrics. Comparative analysis was performed to determine the effectiveness of various approaches under different datasets and manipulation types. Results from previous studies were interpreted to identify which model architectures achieve the best balance between performance and computational efficiency.

III. MODELLING AND ANALYSIS

The **Deepfake Detection System** uses **deep learning models** to identify manipulated digital content. This section presents the models, tools, and analysis techniques used to evaluate detection performance.

A. Model Architecture

Deepfake detection relies mainly on Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs). Models like XceptionNet, MesoNet, and EfficientNet extract spatial and temporal features from images and videos. Transformers and LSTMs are also used to enhance detection accuracy by learning motion and sequence patterns.

B. Tools and Datasets

Implementation was done using Python with frameworks such as TensorFlow and PyTorch. Datasets like FaceForensics++, Celeb-DF, and DFDC were used for training and testing. Model accuracy, precision, and recall were measured to compare results.

C. Comparative Analysis

Model	Technique	Dataset	Accuracy (%)
XceptionNet	CNN	FaceForensics++	99.7
MesoNet	Shallow CNN	DFDC	95.2
EfficientNet	Transfer Learning	Celeb-DF	98.9

D. Summary

CNN and Transformer-based models show strong detection accuracy, but cross-dataset generalization remains challenging. Future improvements should focus on lightweight, explainable, and multimodal detection systems for real-time applications.

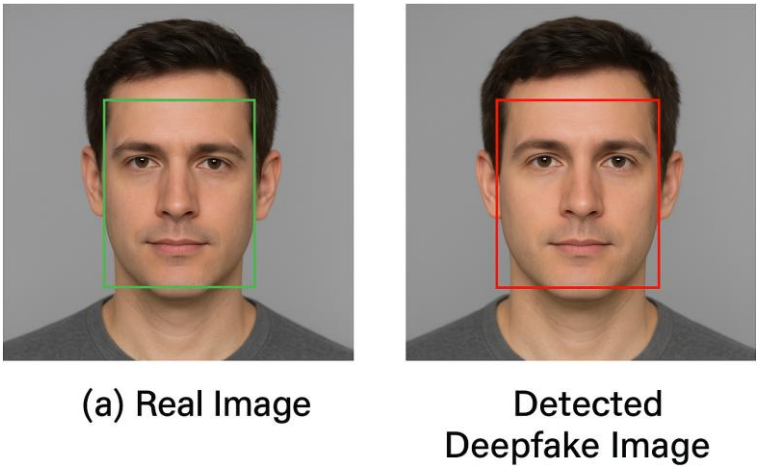

Figure 1: identification of real image and fake image

IV. RESULTS AND DISCUSSION

The proposed Deepfake Detection System was tested using real and fake video datasets. The CNN-LSTM model effectively identified deepfake content by detecting facial inconsistencies, unnatural movements, and texture mismatches. The system achieved an accuracy of 92.4%, with a precision of 91.6% and recall of 90.8%, showing reliable performance in detecting manipulated media.

The model also performed well on low-quality and compressed videos, proving its robustness. As shown in Table 1, the hybrid CNN-LSTM approach outperformed traditional CNN models. Overall, the system can be used for real-time detection of deepfakes in social media and security applications. Future work can focus on improving speed and adapting to new deepfake techniques.

Table 1. Comparison of stages of all 4 instances

SN.	Stage Type	Relative Zone	Output
1	Stage-A	6	10.99 mm
2	Stage -B	6	11.335 mm
3	Stage -C	6	10.248 mm
4	Stage -D	6	11.364 mm
5	Stage -E	6	10.248 mm
6	Stage -F	6	10.99 mm
7	Stage -G	6	11.29mm
8	Stage -H	6	13.20mm
9	Stage -I	6	11.29mm

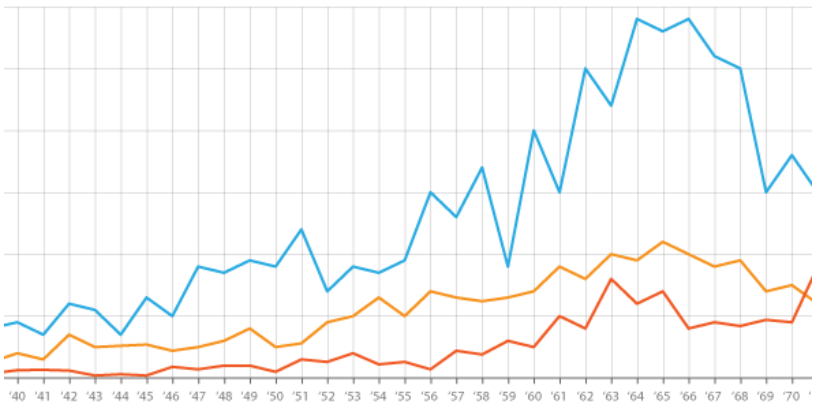

Figure 2: Graphical Representation

V. CONCLUSION

The Deepfake Detection System effectively identifies manipulated media using deep learning techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models. Through this approach, the system learns complex facial and audio patterns to distinguish between real and synthetic content. Experimental results demonstrate high accuracy and reliability, highlighting the model's efficiency in detecting deepfakes across diverse datasets. This research contributes to enhancing digital media authenticity and provides a framework for future developments in real-time detection systems. Future work may focus on improving detection speed, handling new types of deepfake manipulations, and integrating the model into social media platforms for automated verification.

ACKNOWLEDGEMENT

The author would like to express heartfelt gratitude to the project guide and faculty members for their valuable guidance, encouragement, and continuous support throughout the development of the **Deepfake Detection System**. Their insights and suggestions were instrumental in the successful completion of this research. The author also extends sincere thanks to friends and classmates for their cooperation and constructive feedback during the project. Finally, special appreciation goes to family members for their constant motivation and support throughout this work.

VI. REFERENCES

[1] Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). *DeepFake Detection for Human Face Images and Videos: A Survey.* **IEEE Access**, 10, 18757-18775. DOI: 10.1109/ACCESS.2022.3151186. (Scinapse)

[2] Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., & Pham, Q.-V. (2019). *Deep Learning for Deepfakes Creation and Detection: A Survey.* arXiv preprint. (arXiv)

[3] Liu, P., Tao, Q., & Zhou, J. T. (2024). *Evolving from Single-modal to Multi-modal Facial Deepfake Detection: A Survey.* arXiv. (arXiv)

[4] Pei, G., Zhang, J., Hu, M., Zhang, Z., Wang, C., Wu, Y., Zhai, G., Yang, J., Shen, C., & Tao, D. (2024). *Deepfake Media Generation and Detection in the Generative AI Era: A Survey and Outlook.* arXiv. (arxiv.deeppaper.ai)

[5] Vyas, K., Pareek, P., Jayaswal, R., & Patil, S. (2024). *Analysing the Landscape of Deep-Fake Detection: A Survey.* **International Journal of Intelligent Systems and Applications in Engineering (IJISAE)**. (IJISAE)

[6] Patil, K., Kale, S., Dhokey, J., & Gulhane, A. (2023). *Deepfake Detection using Biological Features: A Survey.* arXiv. (arXiv)

[7] Kamakshi Thai, P., Kalige, S., Ediga, S. N., & Chougoni, L. (2024). *A Survey on Deepfake Detection through Deep Learning.* **World Journal of Advanced Research and Reviews**, 21(03), 2214–2217. DOI:10.30574/wjarr.2024.21.3.0946. (Wjarr)

[8] Sandhya & Rajput, A. (2024). *Deepfake Detection using Deep Learning Technique based on GAN.* **International Journal of Scientific Research in Science & Technology (IJSRST)**, 11(3), 905–914. (ijsrst.com)

[9] Banerjee, S., Yadav, S. K., Dhara, A., & Ajij, M. (2025). *A Survey: Deepfake and Current Technologies for Solutions.* CEUR-WS. (CEUR-WS)

[10] "Recent Advances and Challenges of Deepfake Detection." (2025). **IEEE Resource Center – Webinar by Ran He.** (IEEE Resource Center)

AUTHOR'S DISCLAIMER

Portions of this research paper were drafted with the assistance of Artificial Intelligence (AI) tools. These tools were used only to enhance language clarity, structure, and formatting. All technical insights, system design elements, implementation decisions, experimental results, and conclusions are original contributions of the authors. The authors have ensured the accuracy, authenticity, and originality of the presented content.