IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

# Deepfake Detection

**Prof. Manoj Chittawar[1], Aakanksha Toutam[2], Sanika Gongale[3], Amey Zade[4], Kaushal Kamde[5], Vidish Worah[6]**

Guide, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2-6]
Rajiv Gandhi College of Engineering, Research and Technology, Chandrapur, India

**Abstract**: *The rapid advancement of deepfake technology has raised significant concerns regarding misinformation, security, and digital forensics. Various deepfake detection methods have been explored, leveraging both traditional machine learning (ML) techniques and advanced deep learning architectures. While early detection methods relied on handcrafted feature extraction, their effectiveness was often limited due to poor generalization and susceptibility to adversarial modifications. More recent approaches integrate deep learning frameworks such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to improve detection accuracy.*

*The project addresses the growing concern of deepfake detection by utilizing advanced techniques, including ResNext and LSTM models, along with deep learning methods. It features a Django web application designed to identify deepfake videos effectively. The process involves extracting frames from uploaded videos and splitting them into a specified number of frames. Subsequently, Python libraries for facial recognition and C++ visual tools are employed to detect faces in the video frames. The trained models, tailored to analyze various frame sequences, are then implemented to determine whether the video is authentic or artificially manipulated.*

**Keywords**: Deepfake detection, Machine Learning, Deep Learning, Convolutional neural network (CNN), ResNext, Long Short-Term Memory (LSTM)

## I. INTRODUCTION

With the rise in accessibility of deepfake technology, an increasing number of manipulated videos are circulating through social media. Deepfake refers to digitally altered media, such as images or videos, in which one person's likeness is substituted with another's. This phenomenon has become a significant concern in today's society. Deepfakes have been exploited for various harmful purposes, including swapping the faces of popular Hollywood figures with explicit content, as well as generating misleading political narratives. For instance, in 2018, a fabricated video of Barack Obama was created, showing him speaking words he never actually said. Similarly, during the 2020 US elections, manipulated videos of Joe Biden emerged, including one in which his tongue was unnaturally extended. The widespread use of deepfakes in this manner can lead to the rapid dissemination of false information, especially on social media platforms.

At the heart of deepfake technology are Generative Adversarial Networks (GANs), advanced deep learning models capable of creating realistic fake images and videos. These models are trained on extensive datasets to generate media that is increasingly difficult for humans to distinguish from real content. The more comprehensive the dataset, the more convincing and realistic the deepfakes become. The availability of a vast amount of videos, particularly of celebrities and politicians, makes it easier for individuals to generate highly convincing fake news or rumors, which can have a damaging effect on society.

Recent studies highlight that deepfake videos and images have spread rapidly across social media platforms, making detection increasingly critical. To address this challenge, several organizations, including DARPA (Defense Advanced Research Projects Agency), Facebook, and Google, have initiated research efforts aimed at detecting and preventing deepfakes. As a result, numerous deep learning techniques, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and hybrid approaches, have been proposed to identify fake media and combat the

DOI: 10.48175/IJARSCT-26342

spread of misinformation. Current research indicates that deep neural networks have made substantial progress in detecting and curbing fake news and rumors on social media platforms.

This paper provides a comprehensive survey of deepfake detection techniques using deep learning methods like RNNs, Convolutional Neural Networks (CNNs), and LSTMs. The primary goal of this review is to offer researchers a detailed overview of the current state of deepfake detection, including:

1) an in-depth summary of the existing research.
2) an exploration of the datasets commonly used in this domain.
3) an analysis of the limitations of current methods and suggestions for future improvements.

## II. PREVIOUS RESEARCH ON DEEPFAKE DETECTION USED IN THESE STUDIES

***Venkateswarlu Sunkari et al.***

**Paper**: *System Architecture for AI-Driven DeepFake Detection and Moderation on Social Media Platforms* (2023)

**Summary**: This paper proposes an AI-driven system architecture designed to detect and moderate deepfake content on social media platforms. The authors discuss integrating deep learning-based detection models with content moderation systems, offering a scalable solution for platforms to automatically identify and flag deepfakes. Their work emphasizes the importance of real-time detection systems to curb the spread of harmful manipulated media on social networks.

***Yash Doke el at.***

**Paper**: *Deep Fake Detection Through Deep Learning* (2021)

**Summary**: The authors propose a deep learning-based approach for detecting deepfakes in videos. They utilize convolutional neural networks (CNNs) and other advanced machine learning techniques to analyze subtle artifacts in deepfake videos that may not be visible to the human eye. The system demonstrated a high detection accuracy of 97.5%, showing significant potential for applying deep learning models in the identification of manipulated media.

***Raghava M S et al.***

**Paper**: *AI Deep Fake Detection Research Paper* (2023)

**Summary**: This paper provides a comprehensive overview of AI and deep learning algorithms used in the detection of deepfakes. The authors discuss various detection techniques, including facial recognition algorithms and neural networks. They highlight challenges such as the continuous evolution of deepfake technology and propose new avenues for future research to stay ahead of emerging techniques used to generate fake media.

***Darshan V Prasad et al.***

**Paper**: *Comparative Study on Deepfake Detection Methods* (2021)

**Summary**: This paper presents a comparative study of various deepfake detection methods. The authors analyze the strengths and weaknesses of several detection techniques, such as image-based, video-based, and audio-based methods. The research highlights the importance of combining techniques to improve detection accuracy and addresses challenges like the ever-improving quality of deepfake content.

## III. LITERATURE REVIEW

| Authors | Paper Name | Algorithms Used | Key Features |
|---|---|---|---|
| Rimsha Rafique el at., 2023 | Deep fake detection and classifcation using error-level analysis and deep learning | Error level analysis<br>Convolutional neural network (CNN)<br>K-nearest neighbors and support vector machine | The classification is then performed via SVM and KNN. The proposed method achieved highest accuracy of 89.5% via ResNext18 and KNN. |

| | | | |
|---|---|---|---|
| MD Shohel Rana el at., 2022 | Deepfake Detection | Machine learning based methods<br>Deep learning based methods<br>Statistical measurements based methods<br>Blockchain based methods | The experiment reveals the DeepfakeStack achieves 99.65% accuracy. |
| Arash Heidari el at., 2023 | Deepfake Detection using deep learning methods | Deep learning methods<br>CNN | CNN is the most frequently employed method in articles (61%) and is used in almost every category, particularly in image and video deepfake detection. |

## IV. PROPOSED WORK AND METHODOLOGY

***Algorithm used:***

Deep learning, also referred to as deep structured learning, is a branch of machine learning that focuses on using artificial neural networks with multiple layers of representation. This approach supports various learning paradigms, including supervised, semi-supervised, and unsupervised methods. Deep learning frameworks, such as deep neural networks (DNNs), deep belief networks (DBNs), deep reinforcement learning (DRL), recurrent neural networks (RNNs), and convolutional neural networks (CNNs), have been widely adopted across numerous fields.

Key applications include computer vision, speech recognition, natural language processing (NLP), machine translation, bioinformatics, drug discovery, medical image processing, material quality assessment, and game AI development. In many cases, deep learning models have demonstrated performance on par with or exceeding that of human experts.

***Data Preprocessing:***

- Collect and preprocess real and deepfake datasets, ensuring diversity in facial expressions, lighting conditions, and compression artifacts.
- Resize images to a standardized input size compatible with CNN architectures (e.g., 224x224 pixels).
- Normalize pixel values to enhance feature extraction accuracy.

***Feature Extraction using Deep Learning:***

- Use pre-trained CNN architectures (e.g., ResNext18, GoogLeNet, SqueezeNet) to extract low-level and high-level features from input images.
- CNNs capture spatial patterns, textures, and inconsistencies that indicate synthetic manipulations.
- The extracted feature vectors are stored for classification.

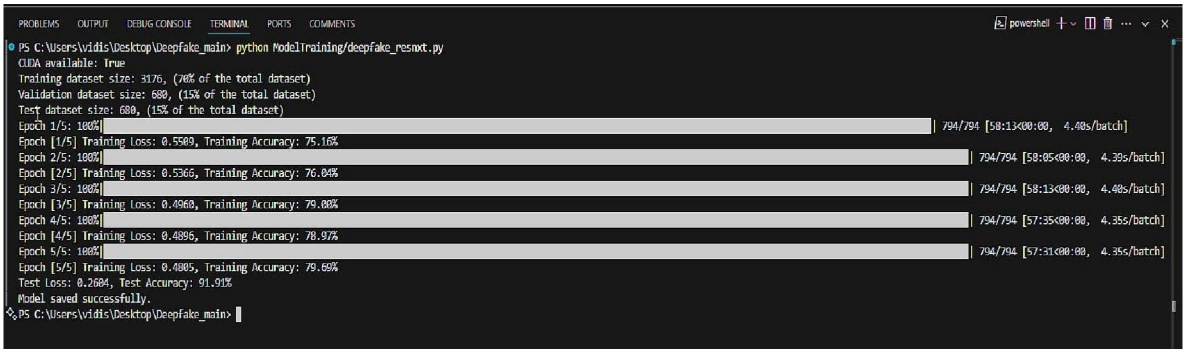
**LSTM:**

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture[1] used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition,[2] speech recognition[3][4] and anomaly detection in network traffic or IDSs (intrusion detection systems). A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs,

**Performance Evaluation:**

- Evaluate the proposed method using accuracy, precision, recall, and F1-score metrics.
- Used confusion matrices to analyze misclassification rates and improve model robustness.

## V. RESULT



```
PS C:\Users\vidis\Desktop\Deepfake_main> python ModelTraining/deepfake_resnxt.py
CUDA available: True
Training dataset size: 3176, (70% of the total dataset)
Validation dataset size: 680, (15% of the total dataset)
Test dataset size: 680, (15% of the total dataset)
Epoch 1/5: 100%|████████████████████████████████████████| 794/794 [58:13<00:00,  4.40s/batch]
Epoch [1/5] Training Loss: 0.5509, Training Accuracy: 75.16%
Epoch 2/5: 100%|█████████████████████████████████████████████| 794/794 [58:05<00:00,  4.39s/batch]
Epoch [2/5] Training Loss: 0.5366, Training Accuracy: 76.04%
Epoch 3/5: 100%|█████████████████████████████████████████████| 794/794 [58:13<00:00,  4.40s/batch]
Epoch [3/5] Training Loss: 0.4960, Training Accuracy: 79.00%
Epoch 4/5: 100%|█████████████████████████████████████████████| 794/794 [57:35<00:00,  4.35s/batch]
Epoch [4/5] Training Loss: 0.4896, Training Accuracy: 78.97%
Epoch 5/5: 100%|█████████████████████████████████████████████| 794/794 [57:31<00:00,  4.35s/batch]
Epoch [5/5] Training Loss: 0.4805, Training Accuracy: 79.69%
Test Loss: 0.2604, Test Accuracy: 91.91%
Model saved successfully.
PS C:\Users\vidis\Desktop\Deepfake_main>
```

The deep learning model for deepfake detection was trained using a dataset with 3176 training samples, 680 validation samples, and 680 test samples. Over the course of 5 epochs, the model demonstrated a steady improvement in performance. Initially, in the first epoch, the training loss was 0.5509 with a training accuracy of 75.16%. As training progressed, the model's accuracy improved, reaching 79.69% in the fifth epoch with a lower training loss of 0.4805. After training, the model was evaluated on the test dataset, achieving an impressive test accuracy of 91.91% with a test loss of 0.2604. The model was successfully saved at the end of the process, indicating readiness for deployment or further testing.

## VI. CONCLUSION

Deepfake detection has become a critical area of research as AI-generated synthetic media continues to advance, posing significant threats to security, misinformation, and digital forensics. While early detection techniques relied on traditional machine learning models, recent advancements leverage deep learning architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models to improve accuracy and robustness. Despite progress, challenges remain in real-world applicability, including generalization to unseen deepfake techniques, dataset limitations, and computational costs. Moving forward, future research should focus on developing scalable, real-time deepfake detection frameworks, integrating temporal analysis for video-based deepfake detection, and enhancing dataset diversity to improve model generalization. By continuing to refine detection methodologies, we can mitigate the risks associated with deepfakes and ensure greater trust and authenticity in digital media.

## REFERENCES

**[1].** Almars, A. (2021) Deepfakes Detection Techniques Using Deep Learning: A Survey. *Journal of Computer and Communications*, **9**, 20-35. doi: 10.4236/jcc.2021.95003.

**[2].** Multi-Attentional Deepfake Detection: Hanqing Zhao, Wenbo Zhou, Dongdong Chen, Tianyi Wei, Weiming Zhang, Nenghai Yu; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 2185-2194.

**[3].** J. Hui, How Deep Learning Fakes Videos (Deepfake) and How to Detect it, Jan. 2021, [online] Available:https://medium.com/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-it-c0b50fbf7cb9.

DOI: 10.48175/IJARSCT-26342

**[4].** J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt and M. Niessner, "Face2Face: Real-time face capture and reenactment of RGB videos", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2387-2395, Jun. 2016.

**[5].** L. M. Dang, S. I. Hassan, S. Im, and H. Moon, ‘‘Face image manipulation detection based on a convolutional neural network,’’ Expert Syst. Appl., vol. 129, pp. 156–168, Sep. 2019.

**[6].** U. Aybars Ciftci, I. Demir, and L. Yin, ‘‘How do the hearts of deep fakes beat? Deep fake source detection via interpreting residuals with biological signals,’’ 2020, arXiv:2008.11363.

**[7].** 7) [95] P. Charitidis, G. Kordopatis-Zilos, S. Papadopoulos, and I. Kompatsiaris, ‘‘Investigating the impact of pre-processing and prediction aggregation on the deepfake detection task,’’ 2020, arXiv:2006.07084.

**[8].** A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Nießner, "FaceForensics: A large-scale video dataset for forgery detection in human faces", arXiv:1803.09179, 2018.

**[9].** Afchar D, et al (2018) Mesonet: a compact facial video forgery detection network. In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE

**[10].** Zhang, T. Deepfake generation and detection, a survey. Multimed Tools Appl 81, 6259–6276 (2022). https://doi.org/10.1007/s11042-021-11733-y

**[11].** Chesney R, Citron DK (2018) Deep fakes: a looming challenge for privacy, democracy, and national security

DOI: 10.48175/IJARSCT-26342