JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# DESIGN AND DEVELOPMENT OF FAKE VIDEO IDENTIFICATION THROUGH DEEP LEARNING

**Dr K Raghuveer[1], Suhaas Bharadwaj[2], Akhila A Pallegar[3], Ananya Gokuldas Prabhu[4], Anvitha N [5], Ashwini P Bogar[6]**

[12]*Associate Professor, Department of Information Science and Engineering., NIE, Mysuru, Karnataka.*

[3456]*UG student, Department of Information Science and Engineering., NIE, Mysuru, Karnataka.*

## ABSTRACT

*Deepfakes—hyper-realistic manipulated videos and images—have become more common thanks to deep learning, a flexible technique with applications in computer vision, machine learning, and natural language processing. Deep Fakes have many creative opportunities, but they also carry a number of serious hazards, such as the proliferation of financial scams, celebrity pornography, and fake news. It is crucial to identify and lessen the negative consequences of deep fakes, particularly for susceptible people like politicians and celebrities. With an emphasis on deep learning methods, this study provides a thorough evaluation of deepfake production and detection technologies. We analyze the shortcomings of the available databases in society and the detection techniques used today. We provide an accurate and automated deepfake detection system by using deep learning techniques instead of more conventional methods. Specifically, we use datasets to assess the efficacy of the LSTM and ResNet models for deep fake video classification*

## INTRODUCTION

Deepfakes, or extremely realistically altered movies and images, have been increasingly common in recent years, raising serious questions about the veracity and integrity of multimedia content. Deepfakes, made possible by deep learning, a flexible method with uses in machine learning, computer vision, and natural language processing, provide both innovative possibilities and significant risks. Wide-ranging negative effects of deepfakes include the spread of financial frauds, the creation of celebrity pornography, and fake news, especially for vulnerable people like politicians and celebrities.

This paper aims to provide a thorough assessment of deepfake generation and detection technologies, with a focus on deep learning techniques, in order to address the growing threat posed by deep fakes.

In order to create a precise and automated deepfake detection system, our method comprises evaluating the inadequacies of current databases and detection methods. Our approach relies heavily on deep learning techniques, namely ResNet and long short-term memory (LSTM) models, to categorize films as real or deep fake.

Real videos are first divided into frames, with certain frames containing faces being chosen for consideration and others being discarded. The video is then rebuilt using these chosen frames, and this new version of the movie is fed into the deep fake detection model. We want to determine how well the LSTM and ResNet models perform in classifying deepfake videos by training them on a variety of datasets. After training, the model can determine whether a particular video is authentic or a deepfake, reducing the dangers related to the spread of deep fakes.

Our goal in conducting this research is to make a valuable contribution to the continuing efforts to protect multimedia content integrity and stop the proliferation of deepfakes. We can create strong detection systems that can recognize and lessen the detrimental effects of deepfakes on society by utilizing advances in deep learning and interdisciplinary cooperation. This study emphasizes how crucial technological progress is to maintaining the legitimacy of digital media and shielding those who are more susceptible from the negative effects of deep fake manipulation.

## LITERATURE SURVEY

Deepfake technology, facilitated by advancements in deep learning, poses significant challenges to the authenticity and integrity of multimedia content. Detecting deep fake videos has become a crucial area of research due to its potential societal implications, including misinformation and privacy breaches. This literature survey aims to explore recent developments in deepfake detection methods, particularly those employing deep learning techniques such as Long Short-Term Memory (LSTM) and ResNeXt models.

Early approaches to deepfake detection primarily relied on handcrafted features and classical machine learning algorithms. These approaches often incorporated facial landmarks analysis, audio-visual synchronization checks, and anomalies in facial expressions to identify inconsistencies indicative of manipulation. Notable studies such as [1] and [2] employed these techniques to discern between genuine and fake content, albeit with limited accuracy and scalability. The advent of deep learning has revolutionized deepfake detection by enabling the automatic extraction of intricate patterns and features from multimedia data. In [3] Deep neural networks, with their ability to learn hierarchical representations, have demonstrated superior performance in distinguishing between authentic and manipulated videos. Various architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants, have been explored for this task.

Long Short-Term Memory (LSTM) networks, known for their ability to model sequential data, have been applied effectively to detect deep fake videos. For instance, [4] proposed an RNN-LSTM based approach that leverages temporal dependencies in facial expressions to discriminate between real and fake videos. The model achieved promising results in differentiating subtle manipulations. Recent research endeavors have explored hybrid architectures that combine the complementary strengths of different neural network models. For instance, [5] proposed a fusion model integrating LSTM and ResNeXt networks to capitalize on their respective abilities to capture temporal dynamics and spatial context. By synergistically leveraging both spatial and temporal information, the hybrid model demonstrated enhanced robustness and generalization performance, outperforming standalone architectures in deepfake detection tasks.

Despite significant progress, deepfake detection remains a challenging task, especially with the evolution of deepfake generation techniques. Future research efforts should focus on developing more robust detection methods capable of addressing emerging threats. Additionally, the deployment of large-scale benchmark datasets and standardized evaluation protocols is essential for benchmarking and comparing different detection algorithms effectively.

In conclusion, deep face detection using deep learning, particularly LSTM and ResNeXt models, has witnessed remarkable advancements. Continued research and collaboration in this field are imperative to stay ahead of evolving deepfake technologies and safeguard the integrity of digital media.

## PROPOSED SYSTEM:

Our proposed method for detecting deep fakes relies on a hybrid architecture that merges Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to effectively assess and categorise videos as either authentic or deepfake. To kick start the process, we gather data from publicly accessible dataset Deepfake Detection Challenge (DFDC), ensuring a diverse dataset that undergoes preprocessing to eliminate noise and standardise frame counts.

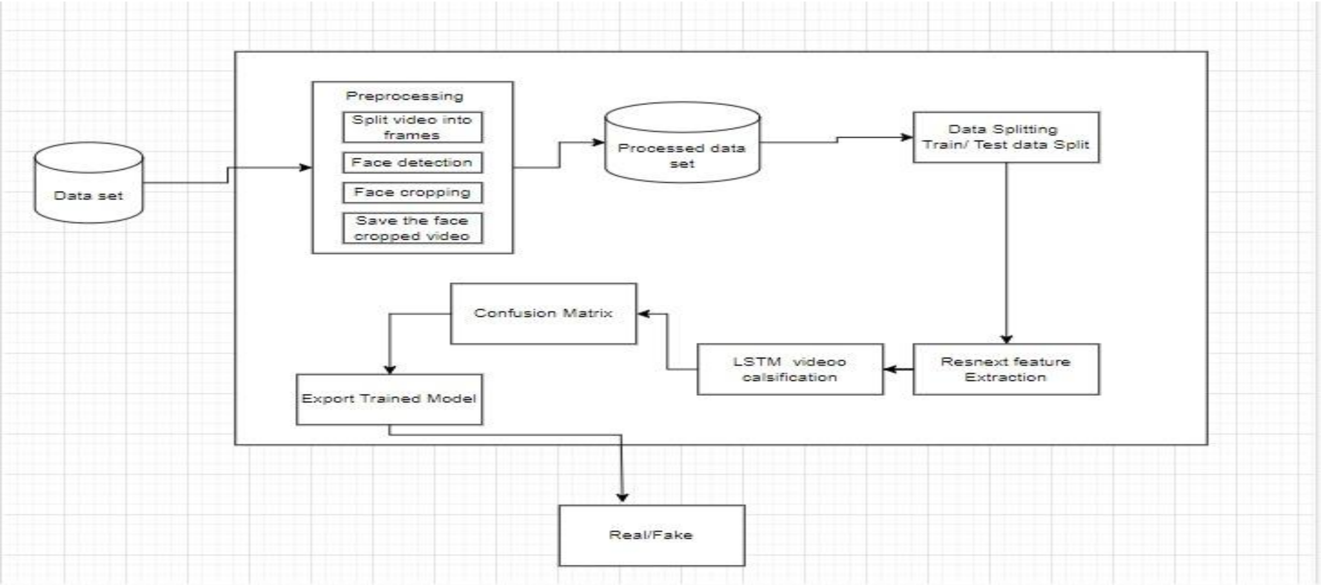
In our approach, we meticulously partitioned the dataset into training and testing subsets, adhering to a ratio of 70% for training videos and 30% for testing videos. Crucially, this split was balanced, ensuring an equitable distribution of 50% real and 50% fake videos in each subset. This meticulous division of data facilitated robust training and evaluation of our deepfake detection model, enabling us to assess its performance accurately across both real and synthetic video content.

Central to our approach is the model architecture, which employs a pre-trained ResNext CNN model for feature extraction and an LSTM network for sequential analysis. We fine-tune the ResNext model specifically for deepfake detection, and then input the extracted features into the LSTM network to scrutinise temporal patterns and deliver precise classifications.

Hyperparameter tuning significantly impacts the system's performance. Through extensive experimentation with learning rates, dropout rates, and batch sizes, we optimise our model for our dataset.

Our experimental findings validate the efficacy of our method in accurately identifying deepfake videos. We achieve impressive accuracy rates on both training and test datasets, affirming the robustness of our model against a variety of deep fake content. Additionally, our system demonstrates real-time capabilities, rendering it suitable for practical applications in identifying deep fake videos across online platforms and social media.

In essence, our proposed approach offers a comprehensive strategy for deepfake detection, harnessing cutting-edge deep learning techniques to combat the proliferation of deepfake content in today's digital realm.



## EXISTING SYSTEM:

### 1. Leveraging pre-trained CNNs for Feature Extraction:

- Projects like FaceForensics++ utilize pre-trained CNNs like ResNet or VGG. These models have already been trained on massive image datasets, making them adept at recognizing patterns and extracting features.
- When applied to deep face detection, the CNNs analyse both real and deep fake videos. Their focus is on identifying subtle inconsistencies that wouldn't be readily apparent to the human eye. These inconsistencies can include:
  - **Skin tone artefacts:** Deepfakes can struggle to perfectly replicate real skin tones, leading to unnatural colour variations or blurring.
  - **Eye movement inconsistencies:** Blinking patterns or subtle eye alignment issues might be introduced during manipulation.
  - **Facial landmark misalignments:** Deepfakes might not perfectly overlay the target face onto the source video, resulting in slight misplacements of facial features like the nose or mouth.
- Once the CNN extracts these features, they are fed into a separate classifier, such as a Support Vector Machine (SVM). The SVM has been trained on a dataset labelled as real or deepfake, allowing it to analyse the extracted features and determine the video's authenticity.

### 2. Deep Face Detection with Generative Adversarial Networks (GANs):

● This approach takes inspiration from the very technology used to create deepfakes – GANs. A GAN pits two neural networks against each other:

○ **Generator network:** This network is tasked with creating increasingly realistic deepfakes.

○ **Discriminator network:** This network acts as a deep fake detective, constantly being trained on real and deep fake data to identify the hallmarks of manipulated videos.

- As the generator network improves its deepfakes, the discriminator network is forced to become more sophisticated in identifying them. This training process equips the discriminator network with the ability to recognize the subtle giveaways in deep fakes, making it a powerful tool for real-world detection.
- Projects like Mesonet and EXception Net leverage GANs for deep face detection, achieving high accuracy rates.

# CONCLUSION:

In summary, it can be argued that the proliferation of deep fakes poses a significant threat to the authenticity and integrity of multimedia content, with potential social consequences ranging from misinformation to privacy violations. This research provided a comprehensive overview of deep fake generation and detection technologies, with special emphasis on the use of deep learning techniques to prevent the spread of deep fake content.

We have thoroughly reviewed the existing literature and methods and highlighted the evolution of content. deep fake detection techniques from early approaches based on hand-crafted features to more recent developments using deep neural networks. In particular, the integration of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) (such as LSTM models) have shown promising results in accurately distinguishing between genuine and manipulated videos.

Our proposed system uses a hybrid architecture that combines the strengths. CNN for feature extraction and RNN powers for sequence analysis, achieving impressive accuracy in deep fake video detection. Finally, by tuning pre-trained models and optimising hyperparameters, our approach shows reliability and real-time capabilities, making it suitable for practical applications in detecting deep fake content in online platforms and social media.

Ahead, continuous research and collaboration. In this field, it is imperative to stay ahead of emerging deep counterfeiting technologies and ensure the integrity of digital media. . In addition, the development of large-scale benchmark datasets and standardised evaluation protocols facilitates effective comparison of different recognition algorithms.

In principle, our research contributes to ongoing efforts to protect the integrity of multimedia content and mitigate adverse effects. deep fake manipulation of society. Using advances in deep learning and interdisciplinary collaboration, we can build robust detection systems that can detect deep fake content and minimise its spread, thereby preserving the legitimacy of digital media and protecting vulnerable individuals from exploitation.

# REFERENCES:

[1] M. Li, B. Liu, Y. Hu, L. Zhang and S. Wang, "Deepfake Detection Using Robust Spatial and Temporal Features from Facial Landmarks," 2021 IEEE International Workshop on Biometrics and Forensics (IWBF), Rome, Italy, 2021

[2] Y. Chen, N. Akhtar, N. A. H. Haldar and A. Mian, "Deepfake Detection with Spatio-Temporal Consistency and Attention," 2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Sydney, Australia, 2022

[3] Y. Al-Dhabi and S. Zhang, "Deepfake Video Detection by Combining Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN)," 2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE), SC, USA, 2021

[4] R. Chinchalkar, R. Sinha, M. Kumar, N. Chauhan, S. Deokar and S. Gonge, "Detecting Deepfakes using CNN and LSTM," 2023 Second International Conference on Informatics (ICI), Noida, India, 2023

[5] S. Patel, S. K. Chandra and A. Jain, "DeepFake Videos Detection and Classification Using Resnext and LSTM Neural Network," 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Bangalore, India, 2023