International Journal of Research Publication and Reviews
Journal homepage: www.ijrpr.com ISSN 2582-7421

# Deep Fake Video Detection Using Deep Learning

*Yash Doke[1], Prajwalita Dongare[2], Vaibhav Marathe[3], Mansi Gaikwad[4], Mayuri Gaikwad[5]*

[1]Yash Doke, Aspiring Computer Engineer, Pune, Maharashtra
[2]Prajwalita Dongare,Professor,Sinhgad Academy of Engineering Kondhwa(bk),Pune,India
[3]Vaibhav Marathe, Aspiring Computer Engineer, Pune, Maharashtra
[4]Mansi Gaikwad, Aspiring Computer Engineer, Pune, Maharashtra
[5]Mayuri Gaikwad, Aspiring Computer Engineer, Pune, Maharashtra

**ABSTRACT -**

In recent months, free software tools based on deep learning have made it easier to create believable face swaps in videos that leave few traces of manipulation, so-called " DeepFake Videos" (DF). Digital video manipulations have been demonstrated through good use of visual effects for several decades, recent advances in deep learning have led to a dramatic increase in the realism of fake content and the accessibility in which it can be created. These media are synthesized by AI (popularly known as DF). Artificially creating the DF with smart tools is an easy task. But when it comes to detecting these DFs, it's a big challenge. Because it is not easy to train the algorithm to detect the DF. We've taken a step forward in detecting DF using Convolutional Neural Networks and Recurrent Neural Networks. The system uses a convolutional neural network (CNN) to extract frame-level features. These functions are used to train a Recurrent neural network (RNN) that learns to classify when a video is manipulable or not and can detect the temporal inconsistencies between frames introduced by the DF creation tools. The expected result compared to a large number of fake videos collected from a standard data set. We show how our system can be competitive. The result of this task is the use of a simple architecture.

*Key Words*: Deepfake Video Detection, convolutional Neural network (CNN), recurrent neural network (RNN), Generative Adversarial Network(GAN), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM).

## 1. INTRODUCTION

The increasing sophistication of smartphone cameras and the availability of a good internet connection around the world has increased the reach of the growing use of social networking and media sharing Portals have made creating and streaming digital video easier than ever. Increasing computing power has made deep learning so powerful that it would have been impossible just a few years ago. Like any transformative technology and this has created new challenges. so-called "DeepFake" is generated by deep generative adversarial models that can manipulate video and audio clips. The spread of DeepFakes on social media platforms is widespread and leads to spam and false speculation information about the platform. These types of DeepFake are terrible and lead to threatening, and deceiving ordinary people.

To overcome this situation, Deep Fake detection is very important. Therefore, we describe a new method based on deep learning that can effectively distinguish fake AI-generated (DF) videos from real videos. It is incredibly important to develop technology that can detect counterfeits so that DeepFake can be identified and prevented from spreading across the internet. To recognize the DeepFake it is very important to understand how Generative Adversarial Network (GAN) creates the DF. GAN takes as input a Video and an image of a specific person ("target") and generates another video with the target person's face by replacing that of another person ('source'). The backbone of DF is deep adversarial neural networks trained on images of faces and videos of targets to automatically match faces and facial expressions from the source to the target. With the right post-processing, the resulting videos can achieve a high level of realism. The GAN splits the video into frames and replaces the input image in each frame. It also reconstructs the video. This process is usually accomplished through the use of automatic coders. We describe a new method based on deep learning that can effectively distinguish DF videos from real ones. Our method is based on the same process used to create the DF of GAN. The method is based on the properties of DF videos, due to the limitation of computing resources and production time, the DF algorithm can only synthesize face images of fixed sized which are subjected to an affine deformation that matches the face source settings. This warping leaves some distinguishable artifacts in the output deepfake video due to resolving inconsistencies between the warped facial area and the surrounding context. Our method detects such artifacts by comparing the generated facial areas and their surrounding regions by dividing the video into Frames Extracting features with a ResNet Convolutional Neural Network (CNN) and using Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) to capture the temporal Inconsistencies between introduced frames of GAN during DF reconstruction. To train the ResNext CNN model, we simplify the process by directly simulating the resolution inconsistency of Intune face wrappers.

## 2. LITERATURE SURVEY

Deep fake videos and their illicit usage are posing a danger to democracy, justice, and public trust. As a result, there is a greater demand for bogus video analysis, detection, and intervention. Some relevant words in deep fake detection are as follows:

A method to detect artifacts was employed in ExposingDF Videos by Detecting Face Warping Artifacts by comparing the generated face areas and their surrounding regions with a specific convolutional neural network model. There were two types of face artifacts in this work.

Their approach is based on the observation that the present DF technique can only produce images with a finite resolution, which then requires additional transformation to match the faces to be replaced in the source video.

Detecting Eye Blinking to Expose AI-Created False Videos describes a new method for detecting fake face videos made with deep neural network models. The method is based on detecting eye blinking in videos, which is a physiological signal that is not adequately represented in the fabricated videos. The method is tested on benchmark datasets of eye-blinking detection and exhibits promising results in detecting videos made with Deep Neural Network-based software DF.

Their detection system merely relies on the absence of blinking. However, certain other factors, such as teeth enchantment, wrinkles on faces, and so on, must be evaluated for the detection of a deep fake. All of these parameters are taken into account by our technique.

Detecting forged images and videos with capsule networks [3] employs an approach that employs a capsule network to detect forged, manipulated photos and videos in a variety of contexts, including replay attack detection and computer-generated video detection.

They used random noise in the training phase of their approach, which is not a suitable solution. Even if the model worked well on their dataset, it may fail on real-time data due to noise in training. It is proposed that our approach be trained on noiseless and real-time datasets.

The approach, Detection of Synthetic Portrait Videos Using Biological Signals [5], extracts biological signals from facial areas in authentic and fraudulent portrait video pairings. Use transformations to compute spatial coherence and temporal consistency, feature sets, and PPG maps to capture signal properties and train a probabilistic SVM and a CNN. The aggregate authenticity probabilities are then used to determine whether the video is fake or legitimate.

The Fake Catcher detects fake content with excellent accuracy, regardless of the generator, content, resolution, or video quality. Formulating a differentiable loss function that follows the recommended signal processing steps is a difficult procedure due to a lack of discriminator, which leads to a loss in their results to preserve biological signals.

### 3. PROPOSED SYSTEM

There are numerous tools available for making DF, however, there are few for detecting DF. Our approach to detecting the DF will make a significant contribution to preventing the DF from spreading over the internet. We will provide a web-based platform for users to post videos and determine if they are fake or real. This project can be expanded from creating a web-based platform to creating a browser plugin for automatic DF detection. Even large applications like WhatsApp and Facebook can integrate this project into their applications to easily identify DF before sending it to another user. One key goal is to assess its performance and acceptability in terms of security, usability, correctness, and reliability. Our technique is designed to detect all types of DF, including replacement, retrenchment, and interpersonal DF. Figure 1 depicts the proposed system's simple system architecture: -

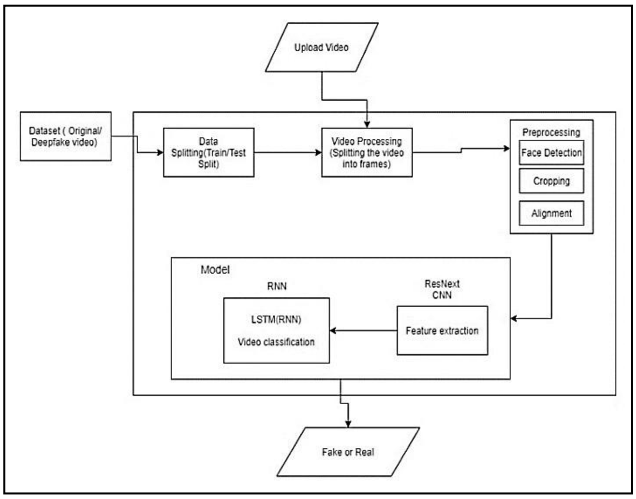

Fig. 1: System Architecture

**A. Dataset:**

We are employing a mixed dataset that has an equal number of videos from several dataset sources such as YouTube, FaceForensics++, and the Deep Fake Detection Challenge dataset[13]. Our newly created dataset includes 50% original footage and 50% altered deep fake video. The dataset is divided into two parts: 70% training and 30% testing.

**B. Preprocessing:**

The video is divided into frames as part of the dataset preprocessing procedure. Face detection and cropping the frame to include the found face come next. The mean of the dataset video is determined to maintain consistency in the number of frames, and a new processed face-cropped dataset is constructed using the frames that make up the mean. Preprocessing ignores the frames that don't contain any faces. It will take a lot of computing power to process the 300 frames in a 10-second video at 30 frames per second. Therefore, we are suggesting that for experimental purposes, the model be trained using only the first 100 frames.

**C. Model:**

The model comprises one LSTM layer followed by resnext50 32x4d. The preprocessed face-cropped videos are loaded by the data loader, who divides them into a train set and a test set. Additionally, the model receives the frames from the edited videos for training and testing in small batches.

**D. ResNext CNN for Feature Extraction:**

We suggest using the ResNext CNN classifier for properly recognizing the frame-level features rather than constructing the classifier from scratch to extract the features. The network will then be fine-tuned by adding any additional necessary layers

and choosing an appropriate learning rate to properly converge the gradient descent of the model. The sequential LSTM input is then taken from the 2048-dimensional feature vectors that remain after the last pooling layers.

**E. LSTM for Sequence Processing:**

Assume a 2-node neural network using a sequence of ResNext CNN feature vectors of input frames as input and the probabilities of the sequence being a deep fake video or an unaltered video. The design of a model to recursively process a sequence in a meaningful way is the main issue that needs to be addressed. We suggest using a 2048 LSTM unit with a 0.4 likelihood of dropout for this task to accomplish our goal. LSTM is utilized to process the frames sequentially so that the video's temporal analysis may be performed by contrasting the frame at second t with the frame at second t-n. where n is any frame number preceding t.

**F. Predict:**

The trained model receives a new video for prediction. Additionally, a fresh video is preprocessed to incorporate the trained model's format. The video is divided into frames, followed by face cropping, and the cropped frames are immediately sent to the trained model for detection rather than being stored locally.

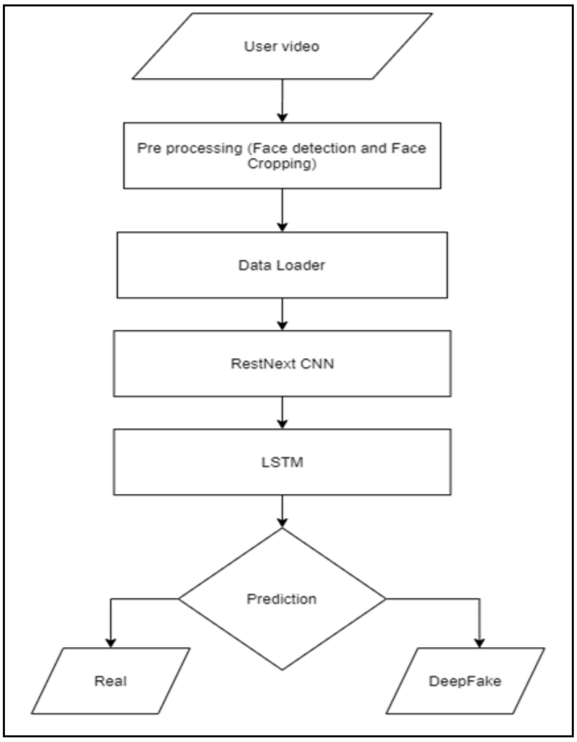

Fig. 2: Prediction flow

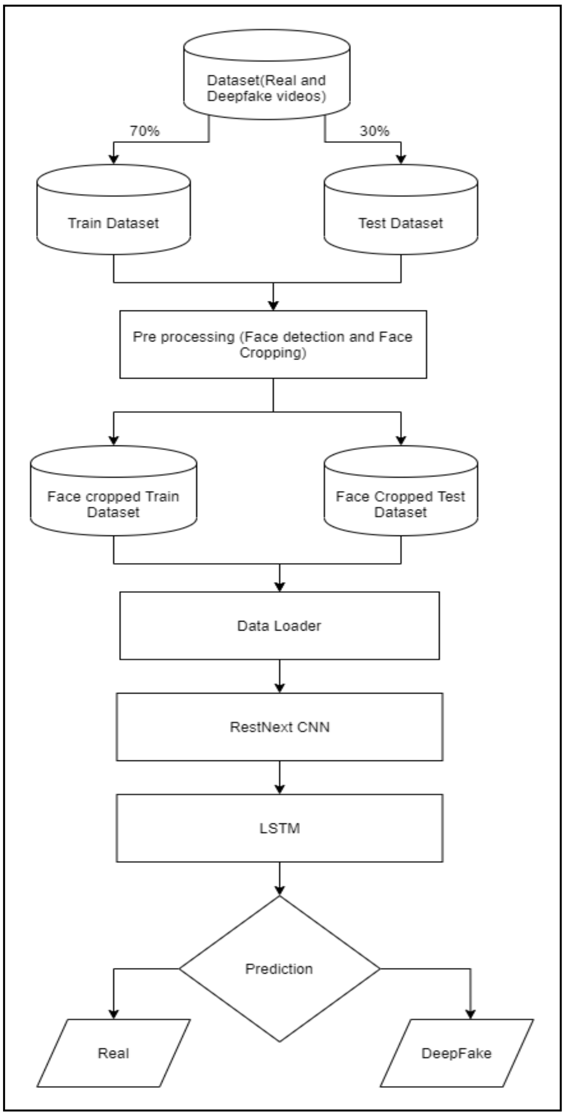
Dataset(Real and Deepfake videos)
70%
30%
Train Dataset
Test Dataset
Pre processing (Face detection and Face Cropping)
Face cropped Train Dataset
Face Cropped Test Dataset
Data Loader
RestNext CNN
LSTM
Prediction
Real
DeepFake

Fig. 3: Training Flow

## 4. RESULT

The model's output will include the model's confidence level and a determination of whether the video is authentic or a deep fake. In figure 3, one instance is displayed.

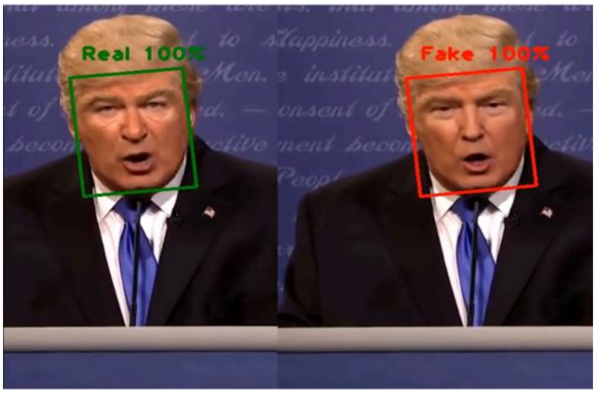

Fig. 3: Expected Result

## 5. CONCLUSION

We provided a neural network-based method for determining if a video is a deep fake or a real thing, along with the model's level of confidence. The deep fakes produced by GANs with the aid of Autoencoders serve as an inspiration for the suggested strategy. Our approach uses ResNext CNN for frame-level detection and RNN and LSTM for video classification. Based on the factors stated in the study, the suggested method is capable of determining if a video is a deep fake or real. We think it will deliver real-time data with extremely high accuracy.

## 6. LIMITATIONS

We did not account for the audio in our method. Due to this, the audio deep fake cannot be detected by our approach. But in the future, we consider achieving the identification of audio deep fakes also.

### REFERENCES

[1] Deng Pan, Lixian Sun, Rui Wang, Xingjian Zhang, Richard O. Sinnott, Deepfake Detection through Deep Learning, 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)

[2] Thanh Thi Nguyena, Quoc Viet Hung Nguyenb, Dung Tien Nguyena, Deep Learning for Deepfakes Creation and Detection: A Survey.

[3] SERGEY ZOTOV, ROMAN DREMLIUGA, ALEXEI BORSHEVNIKOV, DeepFake Detection Algorithms: A Meta-Analysis

[4] Exposing deepfakes using a deep multilayer perceptron–convolutional neural network model Santosh Kolagati, Thenuga Priyadharshini, V. Mary Anita Rajam∗

[5] Siwei Lyu, DEEPFAKE DETECTION: CURRENT CHALLENGES AND NEXT STEPS.

[6] Teng Zhang, Lirui Deng, Liang Zhang, Xianglei Dang, Deep Learning in Face Synthesis: A Survey on Deepfakes, 2020 IEEE 3rd International Conference on Computer and Communication Engineering Technology