

# Jaypee Institute of Information Technology

## **IsSwap?**

### **Deep Fake Detection using Deep Learning**

Supervised By:

**Dr. Rashmi Kushwah**

Assistant Professor (Senior Grade)

Dept. of Computer Science and Engineering

Aakriti Aggarwal	(9918103121)
Siddhant Wadhwa	(9918103198)
Pallav Gupta	(9918103134)
Nishit Anand	(9918103133)

## **ACKNOWLEDGEMENT**

We would like to place on record our deep sense of gratitude to Dr. Rashmi Kushwah, Assistant Professor (Senior Grade), Jaypee Institute of Information Technology, India for her generous guidance, help and useful suggestions.

We also wish to extend our thanks to other classmates for their insightful comments and constructive suggestions to improve the quality of this project work.

### **Name(s) of Students**

**SIDDHANT WADHWA (9918103198)**

**AAKRITI AGGARWAL (9918103121)**

**PALLAV GUPTA (9918103134)**

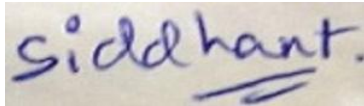
**NISHIT ANAND (9918103133)**

## DECLARATION

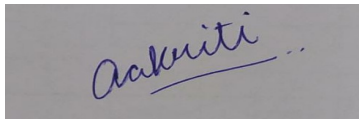
We hereby declare that this submission is our own work and that, to the best of our knowledge and beliefs, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma from a university or other institute of higher learning, except where due acknowledgment has been made in the text.

**Place: Jaypee Institute of Information Technology**

**Date:06 May 2021**

A handwritten signature in blue ink that reads "Siddhant" with a horizontal line underneath.

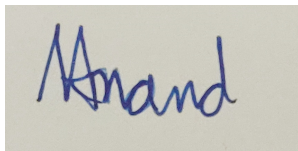
**Name:SIDDHANT WADHWA (9918103198)**

A handwritten signature in blue ink that reads "Aakriti" with a horizontal line underneath.

**Name:AAKRITI AGGARWAL (9918103121)**

A handwritten signature in blue ink that reads "Pallav" with a horizontal line underneath.

**Name:PALLAV GUPTA (9918103134)**

A handwritten signature in blue ink that reads "Anand" with a horizontal line underneath.

**Name:NISHIT ANAND (9918103133)**

## **CERTIFICATE**

This is to certify that the work titled “IsSwap?” submitted by Siddhant Wadhwa (9918103198), Aakriti Aggarwal(9918103121), Pallav Gupta(9918103134), Nishit Anand(9918103133) of B.Tech of Jaypee Institute of Information Technology, Noida has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of any other degree or diploma.

**Digital Signature of Supervisor**

**Name of Supervisor** Dr Rashmi Khushwah

**Designation** Assistant Professor (Senior Grade)

**Date** 6 May 2021

# TABLE OF CONTENTS

<b>ABSTRACT</b>	<b>3</b>
<b>INTRODUCTION</b>	<b>4</b>
<b>BACKGROUND STUDY</b>	<b>6</b>
<b>TOOLS AND TECHNOLOGIES USED</b>	<b>8</b>
<b>REQUIREMENT ANALYSIS</b>	<b>10</b>
<b>SYSTEM DESCRIPTION</b>	<b>11</b>
<b>ALGORITHM DESIGN</b>	<b>12</b>
1. Data Preprocessing	12
2. Model Architecture	13
3. Prediction WorkFlow	13
<b>RESULT</b>	<b>14</b>
Data Preprocessing Result:	14
Data Modelling Result:	17
<b>IMPLEMENTATION</b>	<b>19</b>
<b>LIMITATIONS</b>	<b>21</b>
<b>CONCLUSION</b>	<b>21</b>
<b>REFERENCES</b>	<b>22</b>
<b>TABLE OF CONTENTS</b>	<b>2</b>

# **ABSTRACT**

In this era of technology, the intake of information through digital media has grown exponentially and it has provided people with personal motives to spread falsified information among the masses to create biased opinions and a sense of unrest. The falsified information is provided to the peoples specially during elections to create political unrest among the masses or simply to spread a rumor. Since most of the information is consumed by people in the form of videos, it has become a great target spot for people with malicious intent. The explosive growth in deep fake video and its undetected use is a major threat to democracy, justice, and public trust. Due to this, there is an increased demand for fake video analysis, detection and intervention. The paper is aimed at overcoming the given challenge by providing a fast and reliable method to determine the authenticity of a given video.

# INTRODUCTION

In the current world scenario technology has taken over every aspect of human life and has become an essential part of the same. We use technology to complete routine tasks, this has made life simpler in more ways than we can imagine. The sheer amount of information we consume from digital media is enormous.

As our intake of information through digital media has increased, new problems have arisen. Over time with technological advancements people have found ways to use technology to falsify information and spread it to achieve personal vendetta. One of such practices is called [‘Deep-Fake’<sup>\[1\]</sup>](#). Deep Fake is a technique in which a video of a person’s face or body has been digitally altered so that they appear to be saying or doing something which they actually never have said or done. Deep Fakes are created by combining and superimposing existing images or videos using a deep learning technique known as GANs.

Information through digital media has grown exponentially and it has provided people with personal motives to spread falsified information specially during elections to create political unrest among the masses or simply to spread a rumor.

Users upload over [500 hours<sup>\[2\]</sup>](#) of fresh video content per minute which roughly translate to 7.2 lakhs of new content uploaded everyday and this is just on one platform, namely youtube.

Now the question arises, How do we know that the same technology we love and trust is not being used against us?

The explosive growth in deep fake video and its undetected use is a [major threat<sup>\[3\]</sup>](#) to democracy, justice, and public trust. Due to this there is an increased demand for fake video analysis, detection and intervention which is the main focus of our project along with spreading awareness about these deep-fakes and letting people know what serious implications these videos have in their daily life.

IsSwap? attempts to protect people from believing in false information which is being spread through these deep-fakes. It helps to identify such videos using different tools and technology. IsSwap is a web application which uses deep learning techniques to achieve the goal of fake video detection so that it becomes possible to determine the authenticity of a given video.

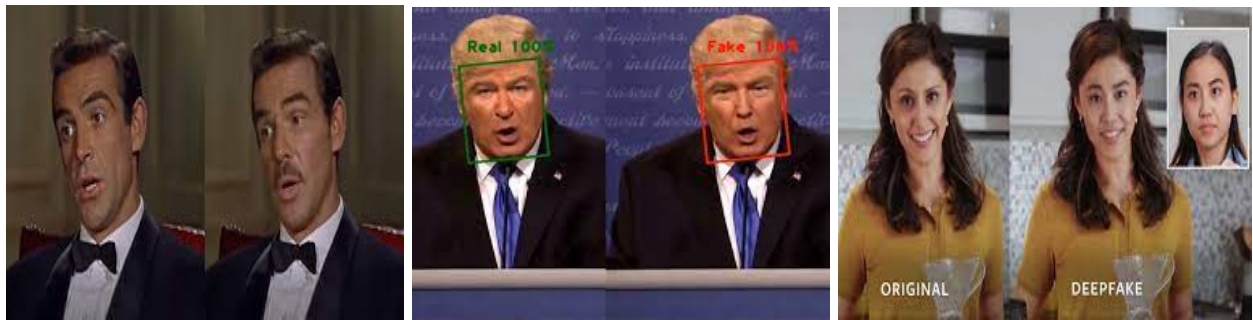


Fig1: Examples of Deep Fake



# BACKGROUND STUDY

In the proposed work of Deep Fake Detection, following libraries are required:

**1. Numpy:** Array-processing package. Provides a high performance multidimensional array object, and tools for working with these arrays.

**2. Pandas:** Fast, powerful, flexible and easy to use open source data analysis and manipulation tool.

**3. Keras:** There are two ways to build [Keras models: sequential and functional](#)<sup>[5]</sup>. The sequential API allows you to create models layer-by-layer for most problems. It is limited in that it does not allow you to create models that share layers or have multiple inputs or outputs.

[Sequential model](#)<sup>[6]</sup>: -It is a linear stack of layers.

A Sequential model is not appropriate when:

- Your model has multiple inputs or multiple outputs
- Any of your layers has multiple inputs or multiple outputs
- You need to do layer sharing
- You want non-linear topology (e.g. a residual connection, a multi-branch model)

**4. Pytorch:** It is a python library based on Torch library . it is mainly used in computer vision applications and natural language processing

**5. Neural Networks:** Neurons in the Neural Network are inspired from biological neurons. This Neural Network would be able to do various tasks like classifying

images, prediction, and so on. A [Perceptron<sup>\[7\]</sup>](#) is an algorithm used for supervised learning of binary classifiers. Binary classifiers decide whether an input, usually represented by a series of vectors, belongs to a specific class. In short, a perceptron is a single-layer neural network.

**6. Flask:** (Flux Advanced Security Kernel) It is a web framework. It is suitable for the development of web apps. We have chosen flask over django because it is lighter and much more explicit.

**7. [Tensorflow<sup>\[8\]</sup>](#):** It is a python library which has many uses . Its main uses include training and inference of deep neural networks

**8. [ResNeXt<sup>\[11\]</sup>](#)** is a simple, highly modularized network architecture for image classification. It is constructed by repeating a building block that collects a set of transformations with the same topology. It is a simple design which results in a homogeneous, multi-branch architecture that has only a few hyper-parameters to set. This strategy exposes a new dimension, which we call “cardinality” , as an essential factor in addition to the dimensions of depth and width.

**9. [XceptionNet](#)**

It is a deepfake detection method using convolutional neural network (CNN). Xception is a convolutional neural network pre trained on more than a million images from the ImageNet database.

The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images.

# TOOLS AND TECHNOLOGIES USED

## 1. Programming Language

- a. Python
- b. JavaScript
- c. HTML/CSS

## 2. Programming Framework

- a. PyTorch
- b. Flask

## 3. Neural Networks

- a. Convolutional  
Neural Network  
(CNN)
- b. Recurrent Neural  
Network (RNN)
- c. Xception Net

## 4. Libraries

- a. Numpy
- b. Pandas
- c. Matplotlib
- d. Seaborun
- e. Scikit Learn

## 5. IDE

- a. Google Colab
- b. Jupyter Notebook

## 6. Version Control

- a. Git

# REQUIREMENT ANALYSIS

For Deepfake Detection, we will be using Deep Learning, for that Neural Networks, Convolutional Neural Networks and Pytorch will be used. This requires a good amount of computational power and GPU power to do the required ML tasks quickly and efficiently.

We also require these things:-

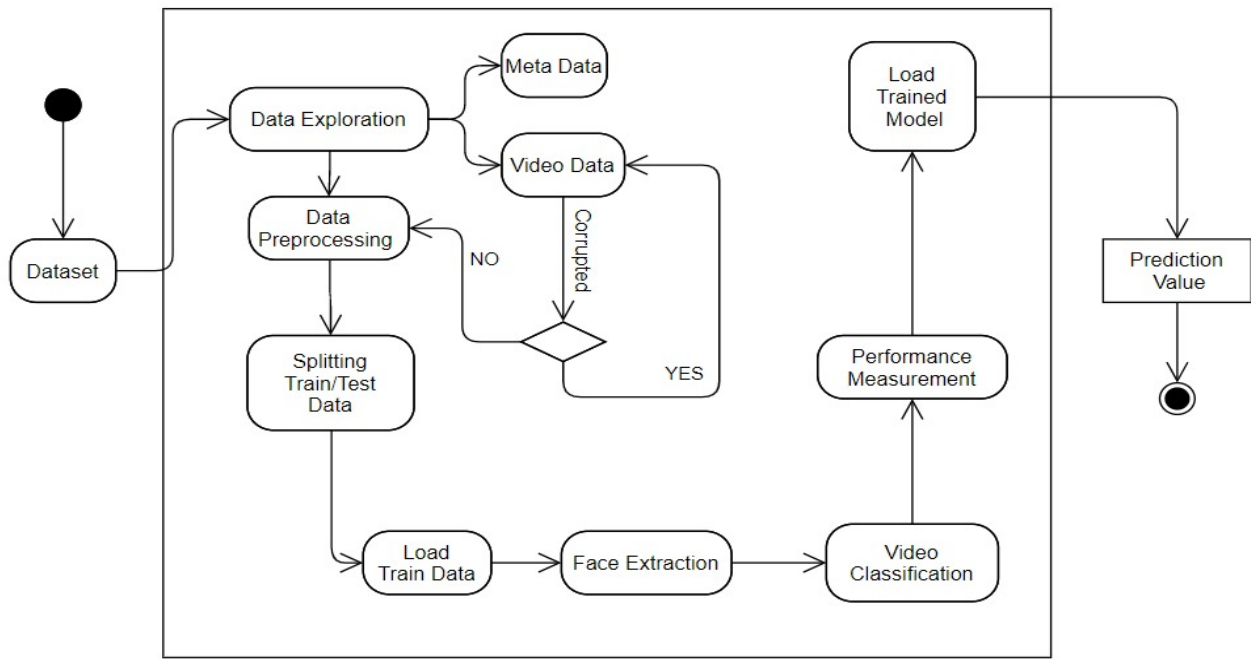
- For training our model we need training data. For this we need labelled data which tells which data is real and which is fake, so that the model can learn from it and then be able to identify deep fake videos on it's own.

The [data we are using<sup>\[12\]</sup>](#) is composed of mp4 files, split into compressed sets of ~10GB apiece. A metadata.json accompanies each set of mp4 files, and contains filename, label (REAL/FAKE), original and split columns, listed below under Columns

- For validation purposes, we also need testing data, so that we can check the performance of our model and see how well it can identify and differentiate deep fake videos from real ones.

test\_videos.zip - a zip file containing a small set of videos to be used as a public validation set.

# SYSTEM DESCRIPTION



There are two phases of this project. One from the user side i.e. user will upload a video to our model and secondly from the developer side i.e. developer will train the model based on the datasets.

Here we are using a dataset which consists of video and the corresponding labels in a .csv file. This dataset is fed into the model\_preprocess notebook. This notebook preprocesses the data. The preprocessed data will split into training and testing data for measuring the performance in 80:20 ratio.

From here we will train the model for 80% of the training data. We label the train videos with the corresponding label and then feed the data into the model.

# ALGORITHM DESIGN



## 1. Data Preprocessing

- Split the video into the frame.
- Validate the video to check if the video is corrupted or not. If it is then delete the video.
- A classifier is used to detect the faces in the frame.
- Remove the face from the frame. (Cropping)
- Resize the images so as to have fixed pixel size for better output.

## 2. Model Architecture

- ResNeXt-50 32\*4 dimension pre trained model for feature extraction. It consists of 50 layers with 32 nodes in each layer which is capable of learning

a large number of parameters.

- B. The output of ResNeXt is a pooling layer which gives us a feature vector which is then fed into a sequential layer.
- C. The output of Xception Net is the AdaptiveAvgPool2d layer.
- D. Sequential layer fed the input into the AdaptiveAvgPool2d layer. We have used 1 linear layer with the chance of Dropout of 0.5.
- E. The output of (d) is further processed by a linear layer.
- F. Finally the softmax layer is implemented which gives the output whether the video is Real or Fake.
- G. Train\_epoch trains the model based on the given number of epochs. Epochs is a hyperparameter that defines the number of times that the learning algorithm will work through the entire training dataset.
- H. Each epoch consists of a number of batches. In our function we are defining a range of parameters for each epoch. It will compute the best possible epoch value and batch size

### 3. Prediction WorkFlow

Prediction workflow starts as soon as the user uploads a video. Then the video enters the Video Preprocessing function where face extraction, feature extraction and splitting into frames is done. It is then loaded to the trained model.

Based on the model it calculates the label corresponding to the video.

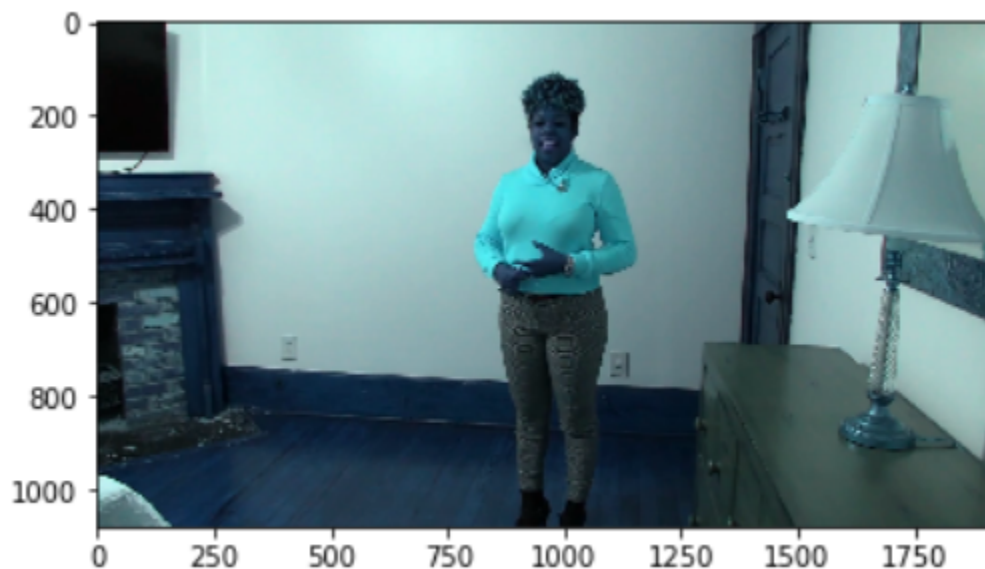
**If label  $\geq$  0.5; Fake Video      Else; Real Video**

# RESULT

## Data Preprocessing Result:

### 1. Frame Extraction from video

We know videos are a collection of images. Here we will need to extract frames from the input video. So to obtain image frames from the input video, we have used the OpenCV library. We take input as a video file and the output is the set of images which composed up the video given as input.

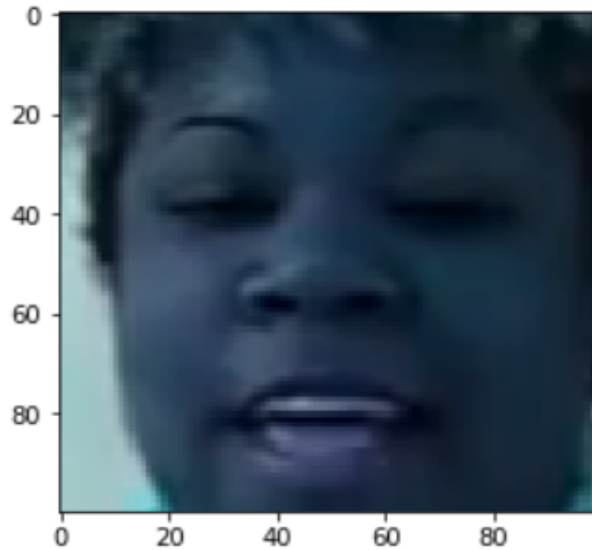


### 2. Face Extraction from frame

Now to extract faces from frames, we used HaarCascade Classifier. It recognizes the facial features and draws a bounding box around any face in an image. This tells us

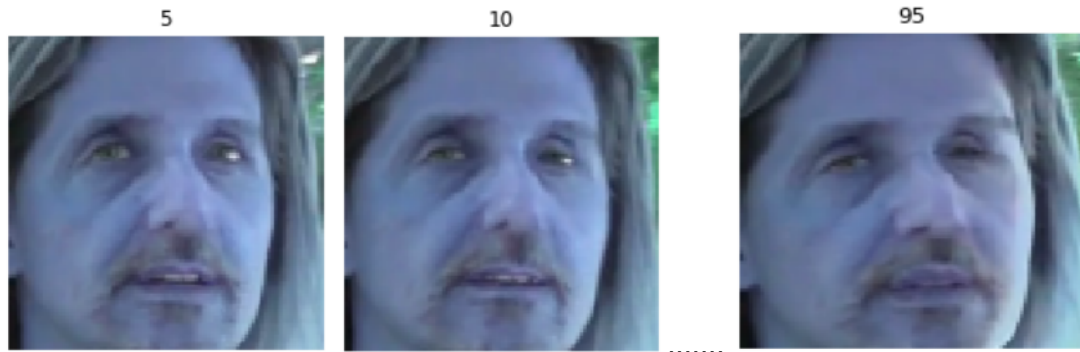


the location of a face in an image. It also tells us the coordinates of the bounding box drawn around the face. This information is used in the next step.



### 3. Cropping Faces in each frame

The bounding box coordinates we got from the last step are used here. Using the coordinates we will crop each image so that it only contains the face in each frame obtained from the video and nothing else. Thus, we cropped faces from the frame and extracted faces of every frame. These images are then resized. Frame size of each cropped image is (100,100,3). If there's no face captured, then we do not run the command `count+=1` and we go on to capture the next face and crop it.



## Data Modelling Result:

An important observation after training the final model is that as we are increasing the number of frames, i.e. as the sequence length is increasing, the accuracy also increases. We have checked the sequence length upto 100 frames. But by using Hyperparameter optimization we concluded that 64 frames gives the best accuracy.

We have trained the model using both the neural networks that are ResNet and Xception Net. The outcomes demonstrate two points:

1. ResNet independently performed less precisely as compared to the Xception Net network.
2. By applying the probability on each neural network's precision and performing the given below mathematical equation, produces a more desirable outcome to the model.

# IMPLEMENTATION



## ISSWAP?

Deepfake Detection system using deep learning

---

### UPLOAD VIDEO TO TEST

---

No file chosen

#### LOCATION

Jaypee Institute of Information  
Technology  
Noida, 201301

#### MADE BY

Aakriti Agarwal  
Nishit Anand  
Pallav Gupta  
Siddhant Wadhwa

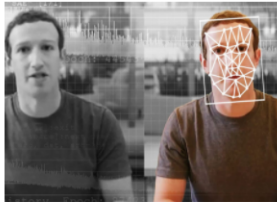
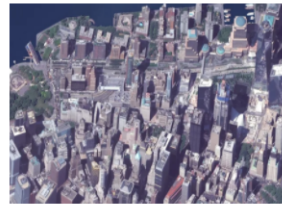
127.0.0.1:5000 says  
aassnauhq.mp4 is Fake

OK

---

## TOP STORIES

---



## LIMITATIONS

- This technology is not 100% accurate
- It requires access to the internet.
- The file has to be uploaded manually, insertion through URL not available.
- Haar cascades and blazeface don't seem to be very convenient tools. Setting up the parameters to match various images seems to be nearly manual

## CONCLUSION

IsSwap? is proposed to detect the fake face and general images generated by the state-of-the-art GANs.. Our results using a large collection of manipulated videos have shown that using a convolutional ResNET and Xception Net structure we can accurately predict if a video has been subject to manipulation or not within a few seconds. .The experimental results demonstrated that the proposed method outperformed other state-of-the-art methods in terms of precision and recall rate. Fake video detection is also an important issue, so in our future work, we will extend the proposed method to fake video detection, incorporating the object detection.

## REFERENCES

[1]

[https://www.researchgate.net/publication/337644519\\_The\\_Emergence\\_of\\_Deepfake\\_Technology\\_A\\_Review](https://www.researchgate.net/publication/337644519_The_Emergence_of_Deepfake_Technology_A_Review)

[2]

<https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/#:~:text=The%20platform%27s%20users%20upload%20more,of%20new%20content%20per%20day>

[3]

<https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=2ab780597494>

[4]

<https://matplotlib.org/stable/index.html>

[5]

<https://stackoverflow.com/questions/57751417/what-is-meant-by-sequential-model-in-keras>

[6]

[https://keras.io/guides/sequential\\_model/](https://keras.io/guides/sequential_model/)

[7]

<https://deeptai.org/machine-learning-glossary-and-terms/perceptron#:~:text=A%20Perceptron%20is%20an%20algorithm,a%20single%20layer%20neural%20network>

[8]

<https://www.tensorflow.org/tutorials>

[9]

<https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf>

[10]

<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

[11]

<https://github.com/facebookresearch/ResNeXt#:~:text=ResNeXt%20is%20a%20simple%20C%20highly,transformations%20with%20the%20same%20topology>

[12]

<https://www.kaggle.com/c/deepfake-detection-challenge/data>

\*\*\*\*\*