DEEPFAKE DETECTION

Submitted by: Bhawna Bisht
Mentor: Ms. Akshita Patwal Ma'am
University Roll No. :2021168

Introduction

A ground-breaking yet controversial application of Artificial Intelligence and Machine Learning. It leverages advanced technique like GAN to create hyper-realisitc synthetic media including audio/video.

spotlight

Deccan Chronicle

THE DEEPFAKE RESURRECTION

From black and white memories of great-grandparents to iconic figures like Bhagat Singh, here's how deep fake turns into a modern 'Ouija Board' breathing 'life' within the dead

Retd IPS officer's deepfake used to blackmail sr citizen

Avishek.Kumar
@timesgroup.com

Ghaziabad: Underlining the threat from AI generated deepfakes, criminals created a video featuring the face and voice of a retired IPS officer in UP Police to extort a senior citizen. Scared, the 76-year-old man made repeated payments to the scammers, believing them to be the IPS officer and faced with threats of police action over an image the blackmailers had that made it

GOT VIDEO CALL

- Arvind Sharma recently bought his first smartphone and opened Facebook A/C
- On Nov 4, he answered a Facebook video call. Sharma disconnected the call after seeing a nude woman
- An hour later, he got a video call on WhatsApp, and found a man in police uniform threatening him

cyber fraud attaining a lethal new dimension aided by

Deepfake: A type of manipulated media created using deep learning

Problem Statement:

Deepfakes can distort our perception of the truth and we need to develop a strategy to improve their detection. Deep Fakes are increasingly detrimental to privacy, social security, and democracy. We plan to achieve better accuracy in predicting real and fake videos.

Motivation:

RASHMIKA MANDANNA CONTROVERSY: HOW YOU CAN SPOT A DEEPFAKE

The original video features a British Indian woman (*left*). Mandanna's face is morphed onto it.

A DEEPFAKE video of actor Rashmika Mandanna posted online has triggered a controversy. The original video features a British Indian woman, and in the edited version, Mandanna's face is morphed onto it. With improvements in technology related to artificial intelligence (AI), deepfakes are becoming common on the internet. Here's how they can be spotted:

Unnatural eye movements

Deepfake videos often exhibit unnatural eye movements or gaze patterns. In genuine videos, eye movements are typically smooth and coordinated with the person's speech and actions.

Mismatches in colour, lighting

Deepfake creators may have difficulty replicating accurate colour tones and lighting conditions. Pay attention to any inconsistencies in the lighting on the subject's face and surroundings.

Audio quality

Deepfake videos often use AI-generated audio that may have subtle imperfections.

Strange body posture, movement

Deepfakes can sometimes result in unnatural body shapes or movements. For example, limbs may appear too long or short, or the body may move in an unusual or distorted manner. Deepfakes may also struggle to maintain a natural posture.

Artificial facial movements

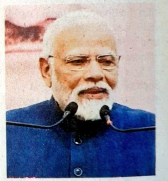
Deepfake software may not always accurately replicate genuine facial expressions. Look for facial movements that seem exaggerated, out of sync with speech, or unrelated to the context of the video.

You can also take a screenshot of the video and run a reverse image search to check the source and the original video. Go to images.google.com, click on the camera icon that says 'Search by image'. You can then upload the screenshot and Google will show you if visuals associated with it are taken from other videos.

ANKITA DESHKAR

Modi warns against deepfakes, asks media to be on guard

The Hindu Bureau
NEW DELHI



Narendra Modi

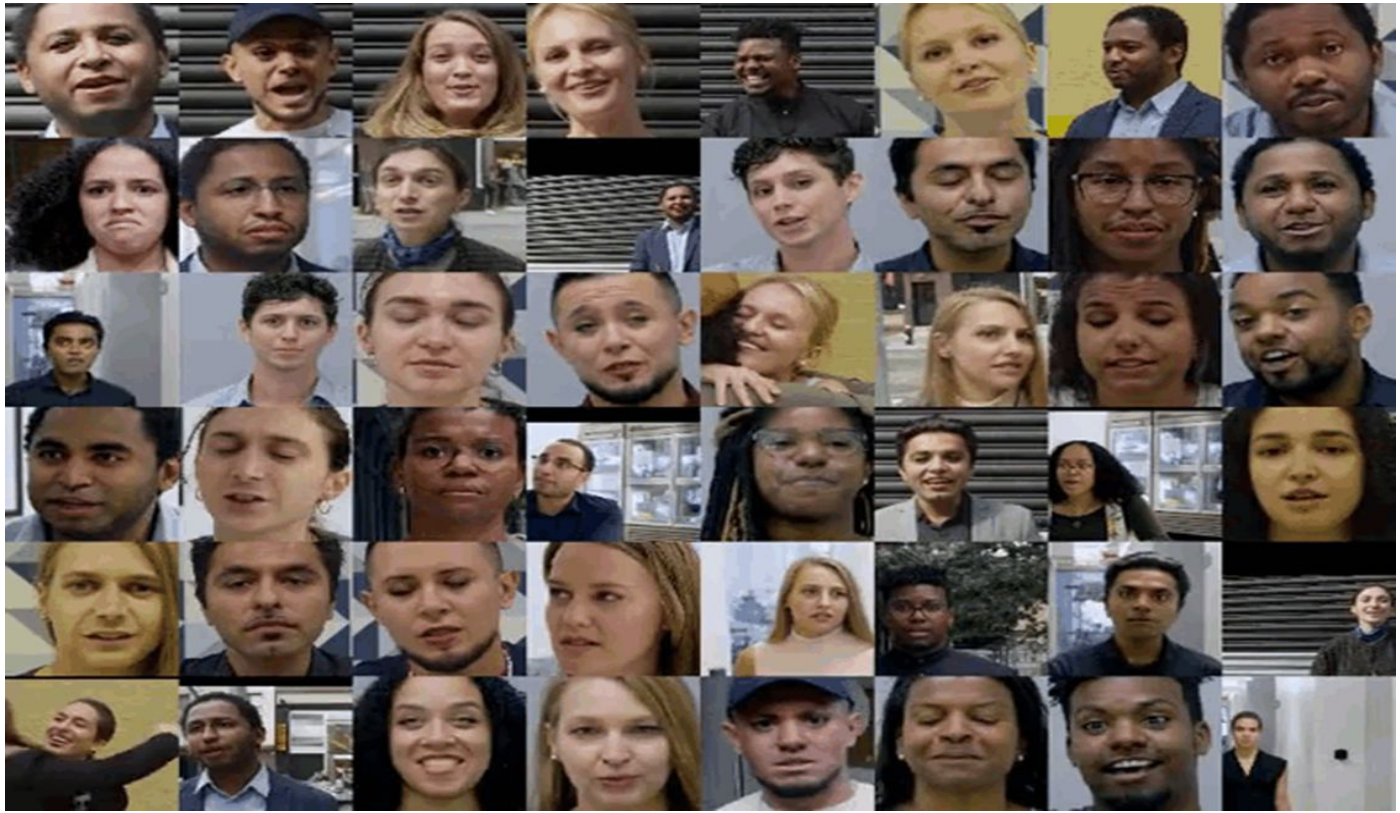
Prime Minister Narendra Modi said here on Friday that the misuse of artificial intelligence for creating deepfakes was problematic, and asked the media to educate people about such activities. He was speaking to presspersons at a Diwali *Milan* organised by the BJP at its national headquarters here.

He said many deepfakes generated with AI assistance appeared real, and the consequent disinformation could lead to much

with a laugh adding that while he did play Garba in school, he had not done so since.

Mr. Modi's remarks come against the backdrop

Examples:



Methodology:

DATASET used- No specific datasets used.

base_path='/content/drive/MyDrive/DD Model/train_sample_videos'

Step 1: Frame Extraction

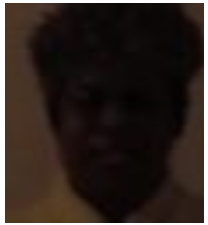
Step 2: Face Detection and Storing

Step 3: Dataset Splitting

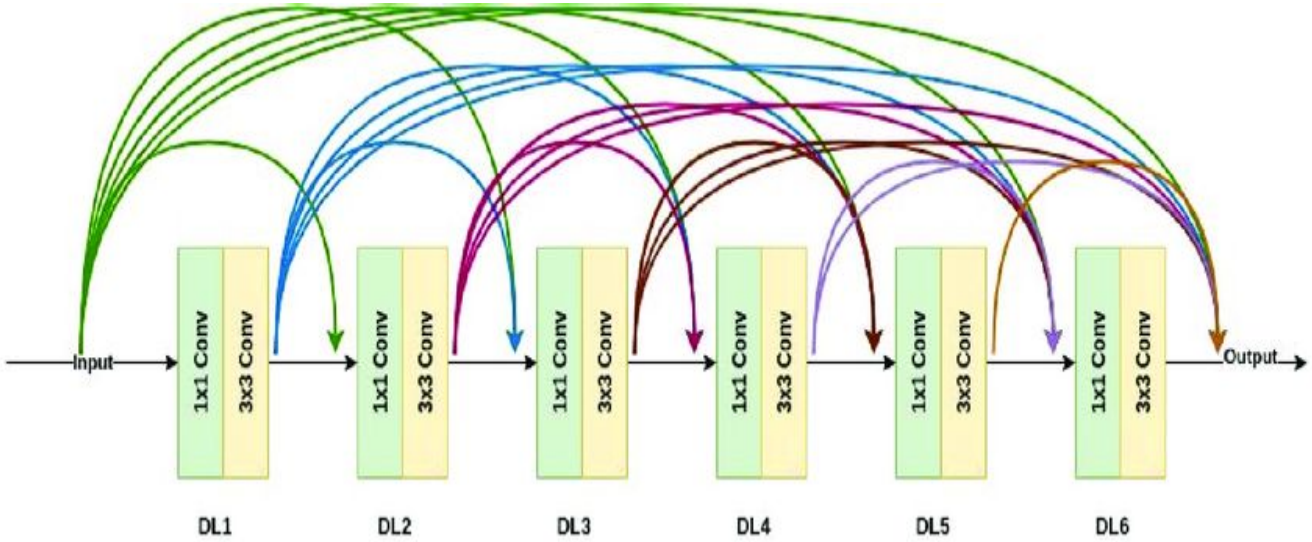
Train Dataset:

Test Dataset:

Validation Dataset:



Step 4: Model Training



Model Architecture:

```
16804768/16804768 ━━━━━━━━━━━━━━━━━━━━ 0s 0us/step
Model: "sequential"
```

Layer (type)	Output Shape	Param #
efficientnet-b0 (Functional)	(None, 1280)	4,049,564
dense (Dense)	(None, 512)	655,872
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 128)	65,664
dense_2 (Dense)	(None, 1)	129

```
Total params: 4,771,229 (18.20 MB)
Trainable params: 4,729,213 (18.04 MB)
Non-trainable params: 42,016 (164.12 KB)
```

Result and Discussion

```
print("Confusion Matrix:\n", cm)
print(f"Accuracy: {accuracy * 100:.2f}%")
print(f"Precision: {precision * 100:.2f}%")
print(f"Recall: {recall * 100:.2f}%")
print(f"F1 Score: {f1 * 100:.2f}%")
```

```
Confusion Matrix:
 [[1 1]
 [0 4]]
Accuracy: 83.33%
Precision: 80.00%
Recall: 100.00%
F1 Score: 88.89%
```

ACCURACY: %of correct predictions out of all predictions.

PREDICTION: % of true positive predictions out of all the predictions made by the model.

Recall: % of true positive predictions out of all actual positive predictions

F1 Score: Harmonic mean of precision and F1 score.

```
Choose Files df1.jpg
• df1.jpg(image/jpeg) - 18083 bytes, last modified: 1/10/2025 - 100% done
Saving df1.jpg to df1 (2).jpg
File is an image. Processing image...
1/1 ━━━━━━━━━━━━━━━━━━━━ 3s 3s/step
Prediction: Fake with confidence 0.42
```

```
Choose Files ex2.png
• ex2.png(image/png) - 117842 bytes, last modified: 1/10/2025
Saving ex2.png to ex2.png
File is an image. Processing image...
1/1 ━━━━━━━━━━━━━━━━━━━━ 2s 2s/step
Prediction: Real with confidence 0.88
```

```
Choose Files ex3.png
• ex3.png(image/png) - 118142 bytes, last modified: 1/10/2025 -
Saving ex3.png to ex3.png
File is an image. Processing image...
1/1 ━━━━━━━━━━━━━━━━━━━━ 3s 3s/step
Prediction: Real with confidence 0.69
```

```
Choose Files ex1.png
• ex1.png(image/png) - 113157 bytes, last modified: 1/10/2025 - 100% done
Saving ex1.png to ex1 (1).png
File is an image. Processing image...
1/1 ━━━━━━━━━━━━━━━━━━━━ 2s 2s/step
Prediction: Fake with confidence 0.64
```

Discussion

The model achieved desirable accuracy in detecting real vs. fake faces. However, performance can still be improved with more advanced architectures and fine-tuning.

The model is able to classify the images as deepfake or pristine.

Conclusion

This project successfully demonstrates a deep learning approach to detecting deepfake media based on face detection and CNNs. With additional data and model improvements, it has the potential to become an effective tool for detecting fake media in real-world applications.

Improvements needed in the model are like this:

- More dataset training
- Fine-tuning
- More regularization

Future Work

```
Choose Files download (3).jpeg
• download (3).jpeg(image/jpeg) - 10745 bytes, last modified: 1/10/2025 - 100% done
Saving download (3).jpeg to download (3) (1).jpeg
File is an image. Processing image...
1/1 ━━━━━━━━━━━━━━━━━━━━ 2s 2s/step
Prediction: Real with confidence 0.53
```

```
Choose Files ex1.png
• ex1.png(image/png) - 113157 bytes, last modified: 1/10/2025 - 100% done
Saving ex1.png to ex1 (1).png
File is an image. Processing image...
1/1 ━━━━━━━━━━━━━━━━━━━━ 2s 2s/step
Prediction: Fake with confidence 0.64
```

As in the above images due to lack of variety in dataset, the model is still struggling to verify the correct images and in case of video the confidence it has on the vide is very low.

Work to be done:

- Experiment with more advanced models like EfficientNetB5 or B7 for better performance.
- Use transfer learning with pre-trained models on similar datasets.
- Perform cross-validation to ensure the model generalizes well to new, unseen data.
- Increasing the number of training dataset to increase the understanding for the model and get clearer references.


THANK
YOU