Minor Project 2

Title: DeepFake Face Detection

Mentored By:
Panduranga Raviteja

Presented by:
R2142211374 Tannu Khetan
R2142210980 Kushagra
R2142211184 Archit Joshi
R2142211037 Ananya Singh

Content

- Introduction
- Literature Review
- Objectives
- Methodology
- Working Model
- System Requirements
- Datasets
- Result
- Conclusion
- Future Scope
- References

UPES
UNIVERSITY OF TOMORROW

Problem Statement

In an era dominated by digital advancements, the rise of deepfake technology poses a significant threat to the authenticity of visual content. Deepfake face manipulation has the potential to deceive individuals, manipulate narratives, and compromise the integrity of digital media. Detecting and mitigating the impact of deepfake faces is a pressing need to ensure the trustworthiness of online information.

Our project focuses on developing an efficient and robust deepfake face detection system. By addressing the challenges associated with the proliferation of manipulated facial images, we aim to enhance the reliability of visual content and contribute to the safeguarding of authentic digital experiences.

Motivation

In an era where visual content plays a pivotal role in shaping opinions, maintaining trust in the authenticity of images is more crucial than ever. The rapid evolution of deepfake technology poses a serious threat to this trust, allowing for the creation of convincing yet entirely fabricated facial images. Our motivation is to combat the erosion of trust caused by deepfake faces, safeguarding the credibility of digital media and ensuring a more secure and reliable online environment for all.

UPES
UNIVERSITY OF TOMORROW

Abstract

The goal of this project is to create a powerful deep fake face detection system that can recognise and analyse AI generated faces. The study intends to reliably identify small visual indicators suggestive of deep fake modifications by utilising machine learning methods, such as convolutional neural networks (CNNs) , long short tearm memory (LSTM) and facial recognition models, in conjunction with advanced picture analysis techniques. To distinguish between real and fake face photos, the approaches combine extensive feature extraction, pattern recognition, and classification. It is anticipated that the project's discoveries and understandings will greatly increase facial image manipulation detection methods, answering the escalating worries about deepfake technology abuse. The findings of this study have broad significance for a variety of fields where maintaining the authenticity and integrity of visual output is crucial, such as media, forensics, and cybersecurity.

UPES
UNIVERSITY OF TOMORROW

Literature Review

Title	Link	Author	Remark
Detecting Deepfake Images Using Deep Learning Techniques and Explainable AI Methods	(PDF) Detecting Deepfake Images Using Deep Learning Techniques and Explainable AI Methods (researchgate.net)	Wahidul Hasan Abir; Faria Rahman Khanam; Kazi Nabiul Alam; Myriam Hadjouni; Hela Elmannai, Sami Bourouis; Rajesh Dey; Mohammad Monirujjaman Khan	This paper suggests a CNN-based technique for explainable deepfake picture recognition that makes use of LIME. It accomplishes 99.87% accuracy, underscoring the problems with detection and the effects deepfakes have on society.
Deep Fake Face Detection using Convolutional Neural Networks	Deep Fake Face Detection using Convolutional Neural Networks | IEEE Conference Publication | IEEE Xplore	Mj Alben Richards; E Kaaviya Varshini; N Diviya; P Prakash; P Kasthuri; A Sasithradevi	This study suggests a deep learning method for identifying deepfake faces in photos and videos by employing CNNs. With a dataset of actual and synthetic faces, the authors used transfer learning to reach an accuracy of 99.37%. While pointing out CNNs' promise for deepfake detection, they also point out some of its drawbacks, such as dataset size and generalisation capacity.
Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network	[Retracted] Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network (hindawi.com)	Hasin Shahed Shad; Md. Mashfiq Rizvee; Nishat Tasnim Roza; S. M. Ahsanul Hoq; Mohammad Monirujjaman Khan; Arjun Singh; Atef Zaguia; Sami Bourouis	This article discusses how to identify deepfake photos. It talks about the risks associated with it and how hard it is to spot them. CNNs are one technique that researchers have created to identify deepfakes. In this study, a dataset of real and fake photographs was utilised to identify deepfake images using eight CNN models. At 99%, the VGGFace model had the highest accuracy.

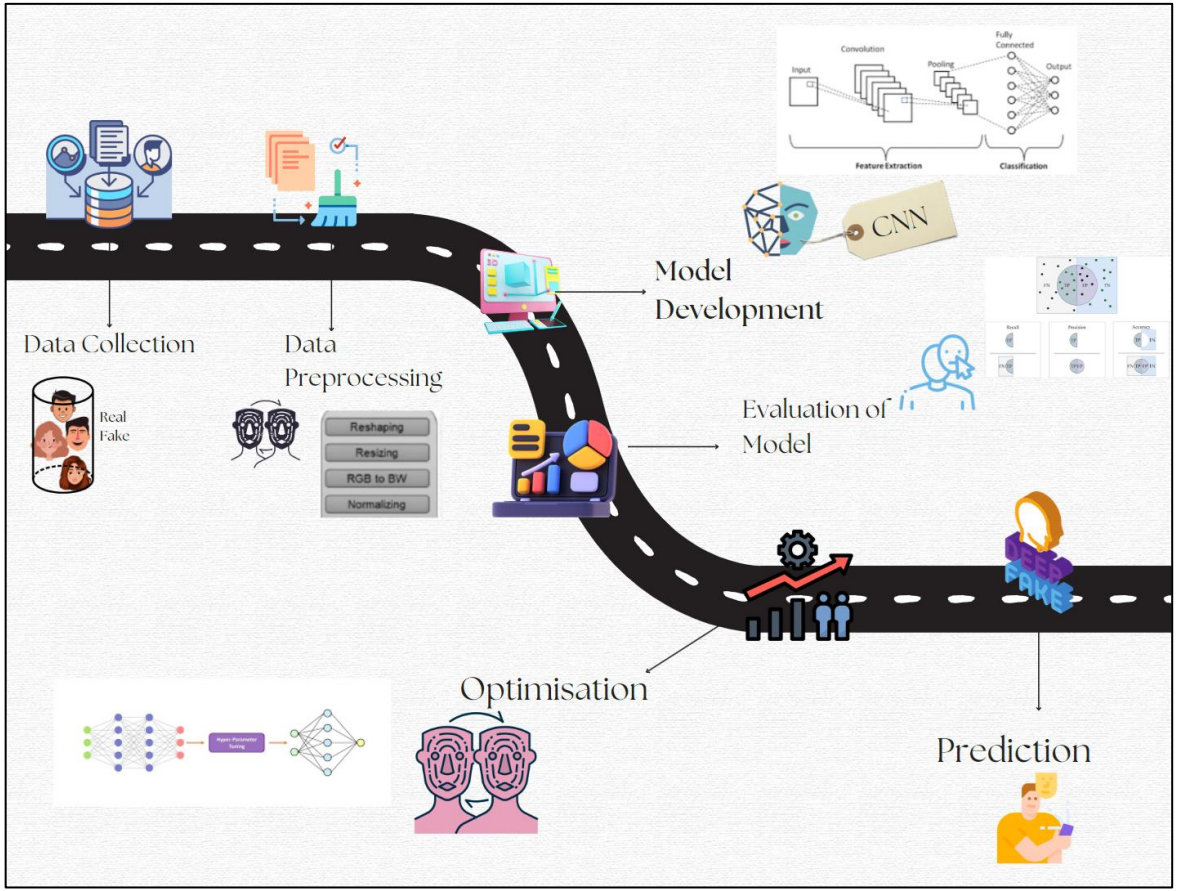
Literature Review

Title	Link	Author	Remark
Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review	(PDF) Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review (researchgate.net)	Remya Revi K. ; Vidya K R; M. Wilscy	This article discusses how to identify deepfake photos produced using GANs, or generative adversarial networks. It draws attention to how challenging detection is and examines a number of methods, such as custom-built CNNs and colour component analysis. The authors stress the need for additional study to create detection techniques that are more potent.
DeepFake Face Image Detection based on Improved VGG Convolutional Neural Network	DeepFake Face Image Detection based on Improved VGG Convolutional Neural Network | IEEE Conference Publication | IEEE Xplore	Xu Chang; Jian Wu; Tongfeng Yang; Guorui Feng	This research uses an upgraded VGG-based CNN to address deepfake detection. Using a dataset of actual and deepfake faces, the authors obtain 99.16% accuracy by suggesting changes to the VGG architecture. They stress how well their enhanced VGG model works for deepfake detection when compared to conventional CNNs, but they also admit that more research on a wider range of datasets is necessary.
Enhancing Deepfake Image Detection with Deep Convolutional Neural Networks	Enhancing Deepfake Image Detection with Deep Convolutional Neural Networks | IEEE Conference Publication | IEEE Xplore	Siddharth Bhamare; Shreeraj Bhamare	This study makes use of error level analysis (ELA), which may find inconsistent compression levels to identify photographs that have been altered. Although earlier studies laid the groundwork for the development of detection methods, this work presents an improved model that detects image alterations with over 95% accuracy.

Objective

- Our project aims at discovering the distorted truth of the deep fakes.
- Our project will reduce the Abuses' and misleading of the common people on the world wide web.
- Our project will distinguish and classify the photos as deepfake or pristine.
- Provide an easy-to-use system for used to upload the photo and distinguish whether the photo is real or fake.

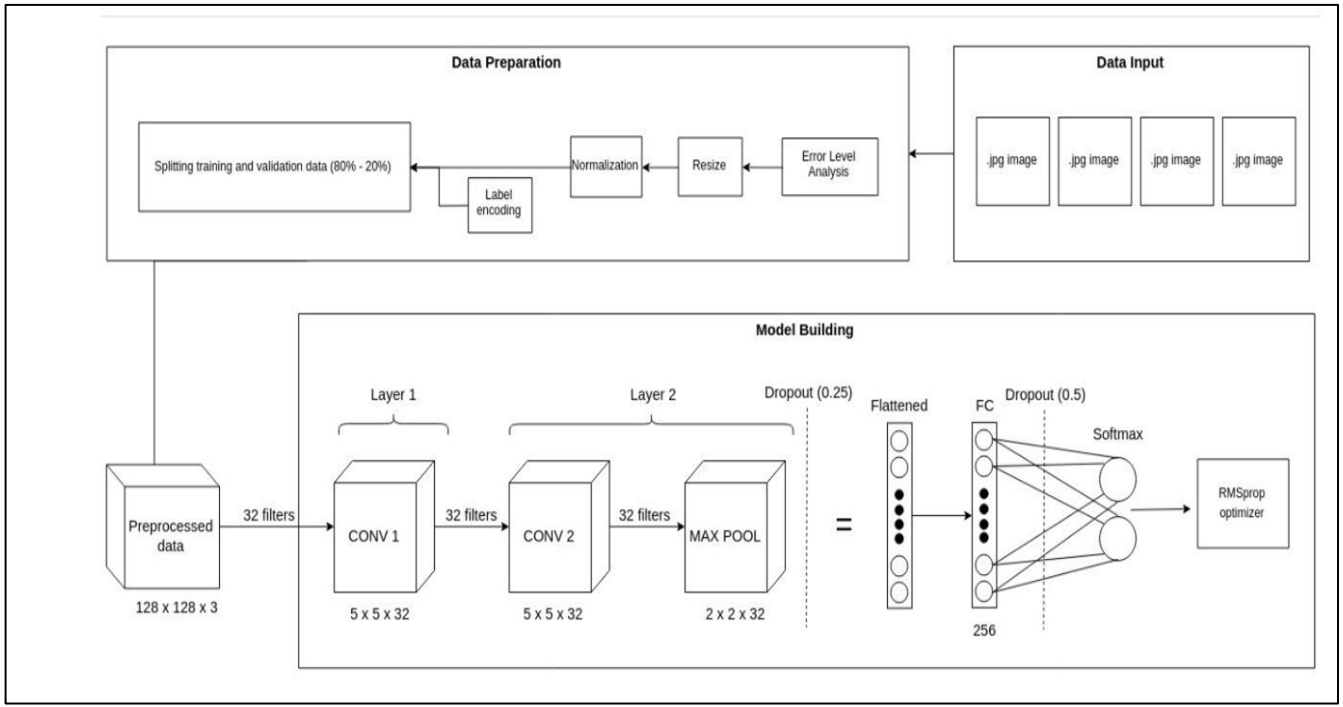
UPES
UNIVERSITY OF TOMORROW



Methodology

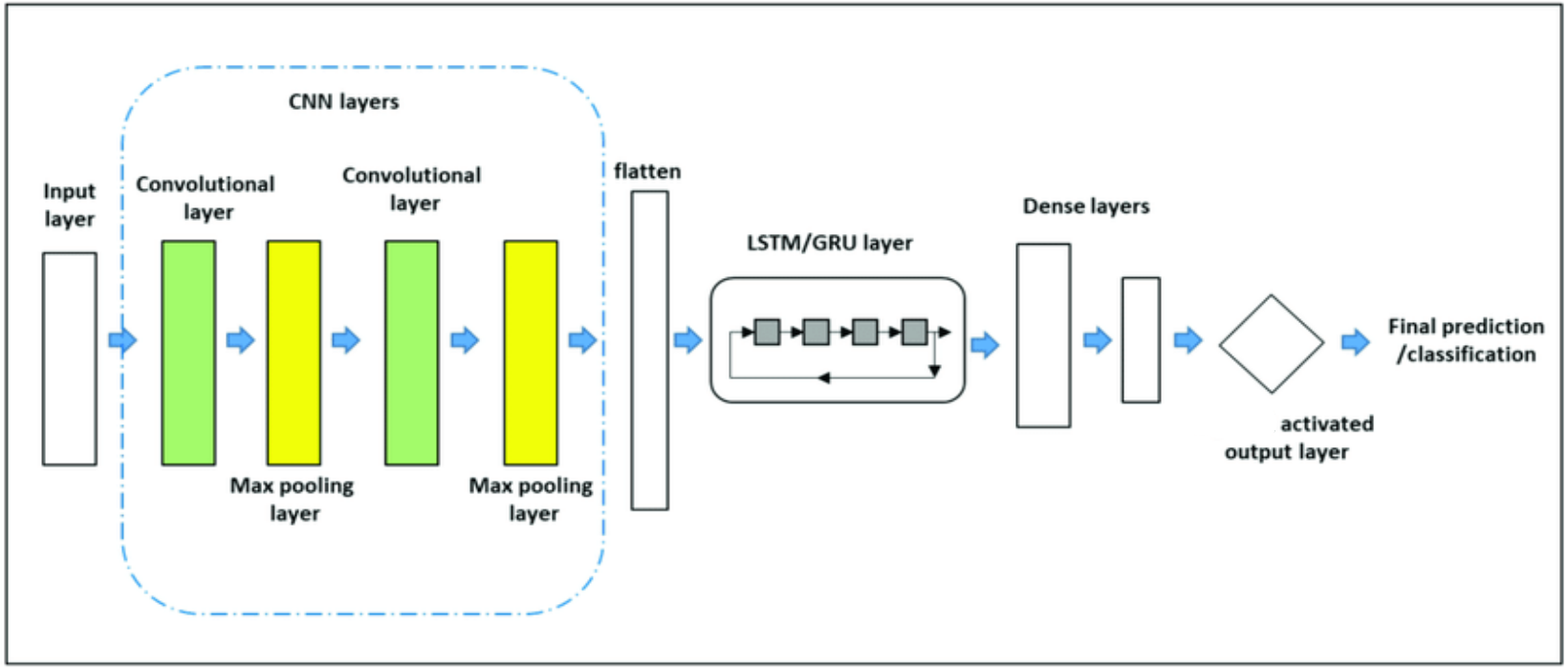
UPES
UNIVERSITY OF TOMORROW

Architecture



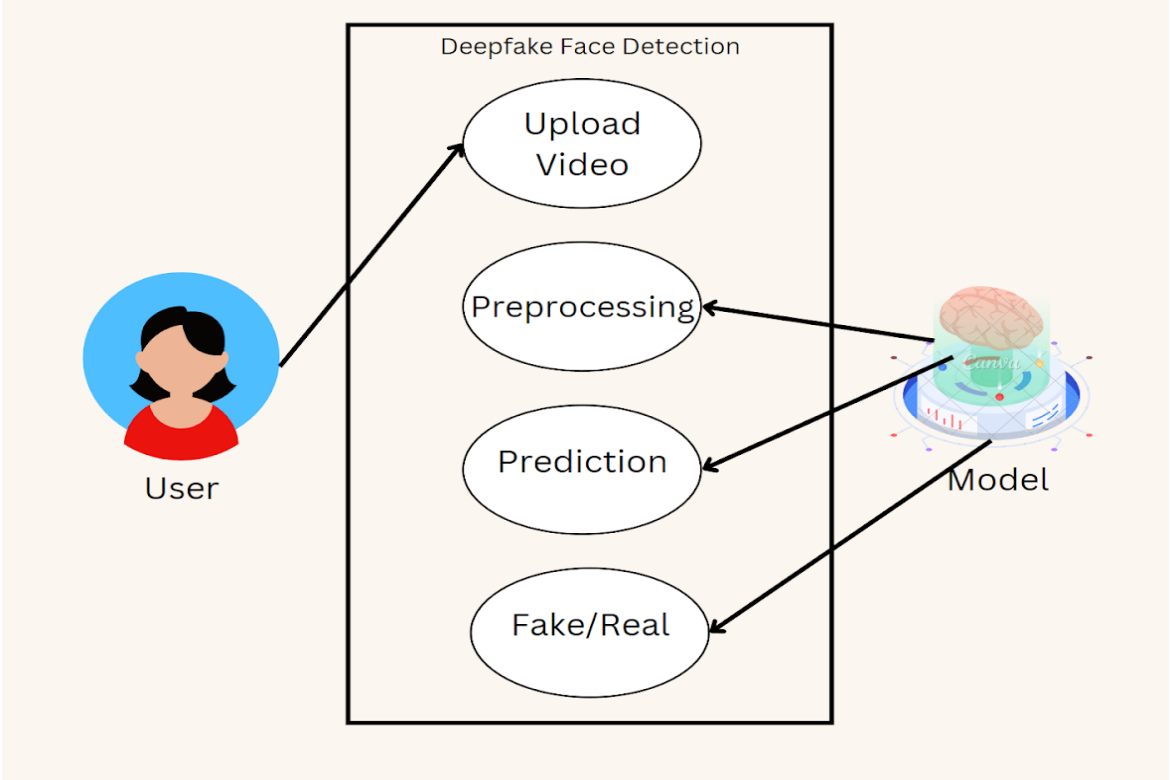
UPES
UNIVERSITY OF TOMORROW

Hybrid Model

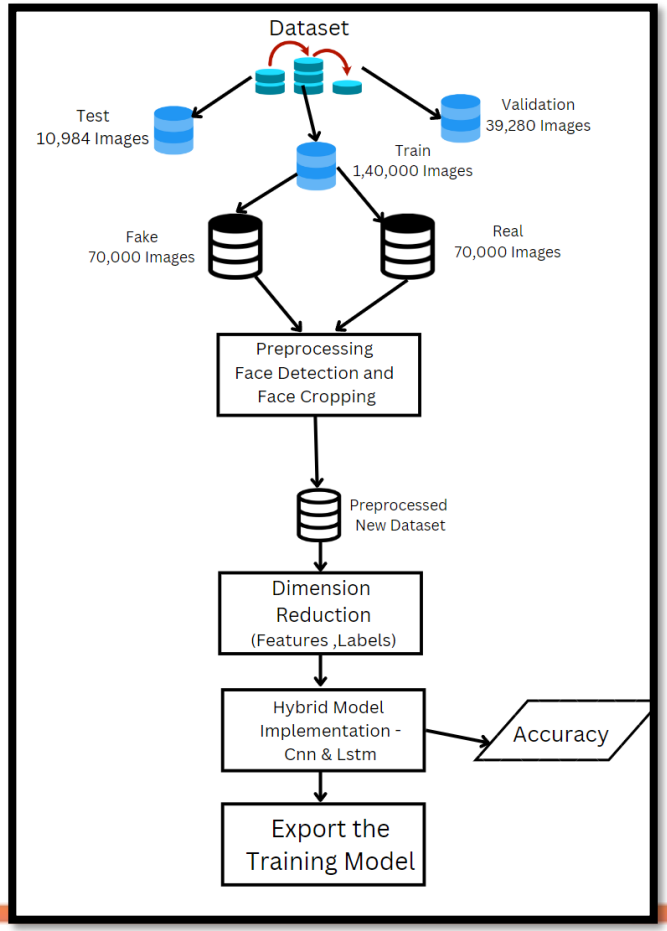


UPES
UNIVERSITY OF TOMORROW

Use Case View



UPES
UNIVERSITY OF TOMORROW



Training Workflow

UPES
UNIVERSITY OF TOMORROW

System Requirement :

Hardware:
CPU: Multi-core processor with 2.5 GHz or higher clock speed.
GPU: Dedicated GPU with CUDA support (for deep learning tasks).
Memory (RAM): Minimum 16 GB (more for larger models and datasets).
Storage: Adequate space for project files, datasets, and deep learning model checkpoints.

Software:
OS: Windows, macOS, or Linux (Ubuntu preferred for deep learning).
Programming language: Python (for deep learning frameworks compatibility).
IDE: Jupyter Notebooks, VS Code, or any Python-compatible IDE.
Deep Learning Framework: TensorFlow or PyTorch for implementing and training DeepFake detection models.
Image processing libraries: OpenCV for Python or PIL (Python Imaging Library).

Algorithm:
Deep Learning Model: Choose a suitable pre-trained or custom Convolutional Neural Network (CNN) architecture for face detection in DeepFake images. Common architectures include VGG, ResNet, or efficient variants like MobileNet.

Scalability:
Model Optimization: Implement model quantization or compression techniques to reduce the model size and memory footprint without sacrificing accuracy.
Batch Processing: Design the system to efficiently process batches of images to accommodate scalability requirements.

Datasets:

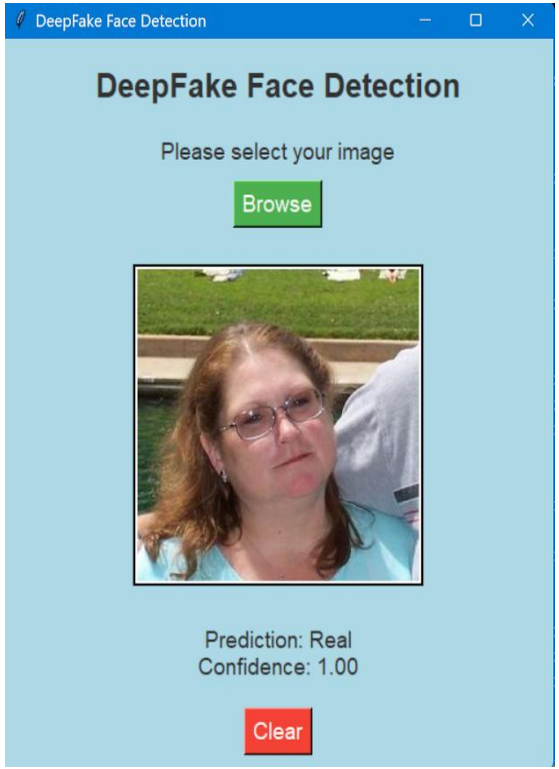
Images are divided into train, test and validation set. Each image is a 256 X 256 jpg image of human face either real or fake.

Folder	Description	Items
Train	This folder contains 2 directories of real or fake images.	70k - Fake 70k - Real
Test	This folder contains 2 directories of real or fake images.	5492 - Fake 5413 - Real
Validation	This folder contains 2 directories of real or fake images.	19.6k Fake 19.8k Real

https://www.kaggle.com/datasets/manjilkarki/deepfake-and-real-images

UPES
UNIVERSITY OF TOMORROW

Result



Accuracy –
87.99%

UPES
UNIVERSITY OF TOMORROW

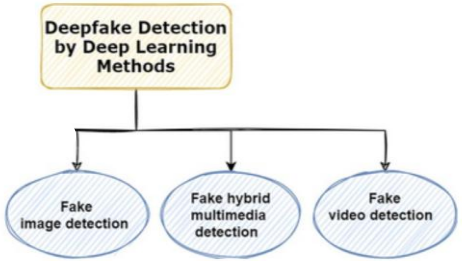
Conclusion

This project delves into the potential of CNNs and lstm for deepfake face detection using images, laying the groundwork for a powerful tool against online manipulation. While still under development, our initial exploration demonstrates promising results in differentiating real and manipulated faces. As we continue to refine our model and explore advanced techniques, we aim to achieve robust and accurate detection capabilities. This project holds immense potential to contribute to a safer digital landscape, empowering users to critically evaluate online content and combat misinformation. Stay tuned for further developments as we strive to bring this valuable tool to fruition.

UPES
UNIVERSITY OF TOMORROW

Future Scope

- Explore video and multimodal detection.
- Address data scarcity and generalizability.
- Partner with social media platforms and raise public awareness.
- Investigate new architectures like Transformers and Explainable AI.



UPES
UNIVERSITY OF TOMORROW

References:

1. Wahidul Hasan Abir; Faria Rahman Khanam; Kazi Nabiul Alam; Myriam Hadjouni; Hela Elmannai, Sami Bourouis; Rajesh Dey; Mohammad Monirujjaman Khan (July 2022) :- Detecting Deepfake Images Using Deep Learning Techniques and Explainable AI Methods
2. Mj Alben Richards; E Kaaviya Varshini; N Diviya; P Prakash; P Kasthuri; A Sasithradevi (September 2023) :- Deep Fake Face Detection using Convolutional Neural Networks
3. Hasin Shahed Shad; Md. Mashfiq Rizvee; Nishat Tasnim Roza; S. M. Ahsanul Hoq; Mohammad Monirujjaman Khan; Arjun Singh; Atef Zaguia; Sami Bourouis (December 2021) :- Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network
4. Remya Revi K. ; Vidya K R; M. Wilscy (February 2021) :- Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review
5. Xu Chang; Jian Wu; Tongfeng Yang; Guorui Feng (September 2020) :- DeepFake Face Image Detection based on Improved VGG Convolutional Neural Network
6. Siddharth Bhamare; Shreeraj Bhamare (February 2024) :- Enhancing Deepfake Image Detection with Deep Convolutional Neural Networks

UPES
UNIVERSITY OF TOMORROW

Thank You