Instructions to students

1. Find available datasets available for free download

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S NO | Dataset name | Download link | NO of Samples | Usage in papers last five years |
| 1 | DFDC  (Deepfake Detection Challenge) | https://www.kaggle.com/c/16880/datadownload/dfdc\_train\_all.zip | 124k videos | https://paperswithcode.com/dataset/dfdc |
| 2 | FF++ (FaceForensics++) | https://github.com/ondyari/FaceForensics | 1000 Original Videos  5000 Manipulated Videos | https://paperswithcode.com/dataset/faceforensics-1 |
| 3 | Celeb-DF | https://github.com/yuezunli/celeb-deepfakeforensics/tree/master/Celeb-DF-v1 | 590 Original Videos  5639 HQ DeepFake videos | https://paperswithcode.com/dataset/celeb-df |
| 4 | FFHQ  (Flickr-Faces-HQ) | https://github.com/NVlabs/ffhq-dataset | 70,000 HQ PNG Images (GAN Based) | https://paperswithcode.com/dataset/ffhq |
| 5 | DFFD  (Diverse Fake Face Dataset) | https://cvlab.cse.msu.edu/dffd-dataset.html | 1,00,000 Images (Pro-GAN)  2,00,000 Images (Style-GAN) |  |
| 6 | ForgeryNet | https://yinanhe.github.io/projects/forgerynet.html | 1,438,201(Real Images)  1,457,861 (Fake Images)  221,247 (Videos) | https://paperswithcode.com/dataset/forgerynet#:~:text=which%20are%20manipulated.-,ForgeryNet%20is%20by%20far%20the%20largest%20publicly%20available%20deep%20face,annotations%20(6.3%20million%20classification%20labels%2C |

2 .Read recent papers on deep fakes both supervised and unsupervised 10 papers each

concentrate only on 2024 papers (if papers are not available can use 2023 papers).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Title | Dataset Used | Preprocessing techniques used | Feature extraction techniques used | Deep Learning Modle /Classifier used | Accuracy achieved | Gaps identified |
| 1) DeepFake Detection Through Key Video Frame Extraction using GAN (2022) | -DFDC  -FF++ | -Splitting video into frames, face detection and cropping | ResNeXt 50 | Model- CNN  Classifier- GAN | 97.2 | Structure of deep neural networks is too vast to fit on edge devices due to memory constraints (reduce run-time memory and model size) |
| 2) Deepfake Face Detection Using Deep InceptionNet Learning Algorithm (2023) | -DFDC  -FF++ |  |  | InceptionNet(CNN Algorithm) | 93% | -Trained on DFDC tested on FF++, Accuracy=57.43  -Trained on FF++ tested on DFDC, Accuracy=60.51 |
| 3) Deepfake Video Detection System Using Deep Neural Networks (2023) | -Celeb-DF  -FF++ | -Frames are extracted from the videos, and the pixel values are normalized | ResNet-50 | LSTM (Video Classification) | -84.75 (20 epochs)  -87.48 (40 epochs) |  |
| 4) DFGNN: An Interpretable and Generalized graph neural network for deepfakes detection (2023) | -FF++  -WLRD  -Celeb-DF  -DFDC | -MTCNN | Pyramid ResNet | Model- GNN  Classifier- GraphNet, FFN | -FF++ (97.16%) | -Accuracy is less when performed cross corpora evaluation.  TrainSet(Celeb-DF)  TestSet(Other all)  (69.6%, 69.36%, 61.3%) |
| 5) DeepFakes detection across generations: Analysis of facial regions, fusion,  and performance evaluation | -UADFV  -FF++  -Celeb-DF  -DFDC | -Frames are extracted from the videos | -OpenFace2 (Facial Region Segmentation) | Fusion of CNN | 99% | -Only 90.6% accuracy for DFDC dataset |
| Deepfake Detection: A Systematic Literature Review | FaceForensics++ (FF++) Dataset | **Data augmentation** | Special artifacts-based features | Convolutional Neural Networks (CNN) | 99% | Public availability of datasets |
| Deepfake Detection: Analyzing Model Generalization Across Architectures, Datasets, and Pre-Training Paradigms | CelebDF-V2 | Face Cropping | DINO (self-supervised) | CNN Models | 92% | Dataset Challenges |
| Deepfake Detection on Social Media: Leveraging Deep Learning and FastText Embeddings for Identifying Machine-Generated Tweets | Deepfake Detection Challenge (DFDC) | Manual partitioning | **Term Frequency (TF)** and **TF-IDF** | ViT-Base models | 95% | Suboptimal generalization |
| DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era With Deepfake Detection Algorithms | DFFMD | Cropping | Convolutional layers | VGG19 | 77.80% | Lower Accuracy of Proposed CNN |
| 9 |  |  |  |  |  |  |
| 10 |  |  |  |  |  |  |
| 11 |  |  |  |  |  |  |
| 12 |  |  |  |  |  |  |
| 13 |  |  |  |  |  |  |
| 14 |  |  |  |  |  |  |
| 15 |  |  |  |  |  |  |
| 16 |  |  |  |  |  |  |
| 17 |  |  |  |  |  |  |
| 18 |  |  |  |  |  |  |
| 19 |  |  |  |  |  |  |
| 20 |  |  |  |  |  |  |

1. Title