Deepfake Dataset Pipeline Report

Dataset Summary

index	0
celeb_real	4
celeb_fake	4
ffpp_real	4
ffpp_fake	4

Cleaning Summary

- Low-res files removed.

Frame Extraction

535 frames were extracted in total from all available videos.

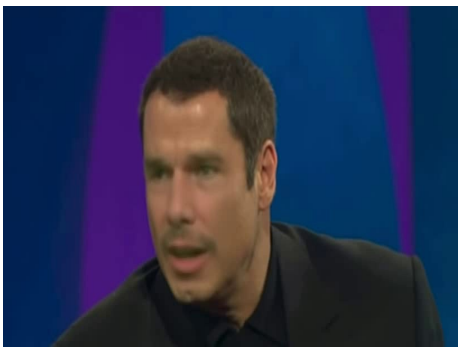
Frame Extraction

Real Frames Samples



Fake Frames Samples



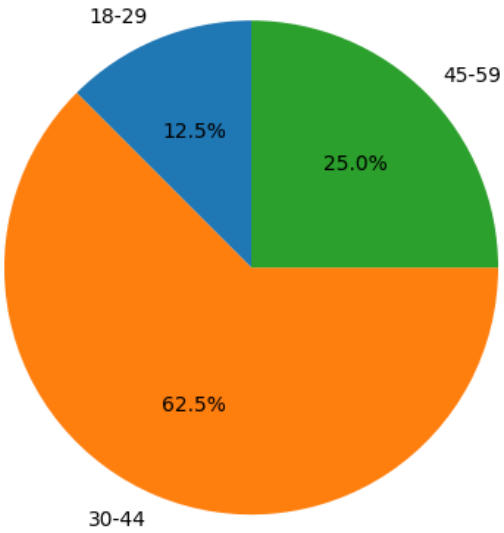


Age Annotation Summary

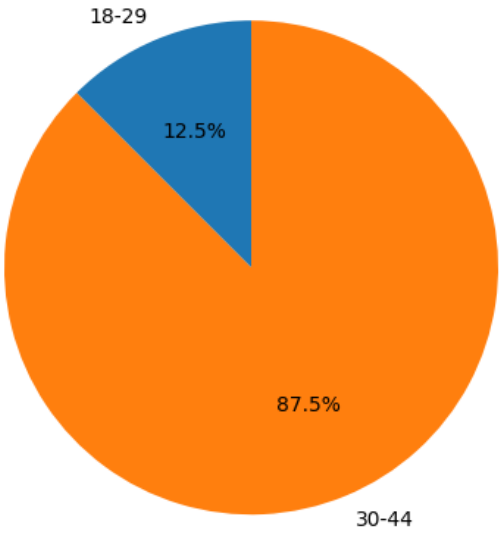
index	0
19-35	9
36-50	7

Deepfake Dataset Pipeline Report

Celeb Dataset



Faceforensics Dataset



Balanced Dataset Summary

index	0
19-35	8
36-50	6

Synthetic Data Summary

index	0
19-35	8
36-50	6

Train Test Split

Deepfake Dataset Pipeline Report

The balanced dataset was successfully split into training and testing sets.

70% of the data was used for training.

30% was used for evaluation.

The split was performed after age-group balancing to ensure fair representation of all age groups in both sets.

List of Model trained

XceptionNet, EfficientNet, LipForensics models were selected for training.

Evaluation on Balanced Dataset

The following table shows model evaluation results (AUC, pAUC, EER) on the balanced test dataset:

('Test Set', '')	('XceptionNet', 'AUC')	('XceptionNet', 'PAUC')	('XceptionNet', 'EER')	('EfficientNet', 'AUC')	('EfficientNet', 'PAUC')
Balanced	0.6045	0.6361	0.4451	0.5793	0.5437
Colab	0.4267	0.4821	0.5272	0.5433	0.5212
FaceForensics++	0.7528	0.7964	0.3631	0.4818	0.5368

('EfficientNet', 'EER')	('LipForensics', 'AUC')	('LipForensics', 'PAUC')	('LipForensics', 'EER')
0.466	0.4756	0.4328	0.5
0.4246	0.6338	0.5481	0.3972
0.5616	0.3178	0.4325	0.5955

Age-Specific Evaluation

The following table presents the model evaluation results grouped by age group and dataset source.

Metrics such as AUC, pAUC, and EER are reported per age group:

('overall', 'AUC')	('overall', 'PAUC')	('overall', 'EER')	('0-10', 'AUC')	('0-10', 'PAUC')	('0-10', 'EER')
0.7615	0.6773	0.2945	None	None	None
0.688	0.5852	0.3425	None	None	None
0.4299	0.4326	0.548	None	None	None

Deepfake Dataset Pipeline Report

0.5827	0.5816	0.5411	None	None	None
0.5004	0.4825	0.5049	None	None	None
0.3373	0.3811	0.6232	None	None	None
0.2139	0.353	0.6986	None	None	None
0.2609	0.3502	0.6575	None	None	None
0.4768	0.4765	0.5409	None	None	None
('10-18', 'AUC')	('10-18', 'PAUC')	('10-18', 'EER')	('19-35', 'AUC')	('19-35', 'PAUC')	('19-35', 'EER')
None	None	None	0.6732	0.6049	0.3679
None	None	None	0.5727	0.5105	0.4362
None	None	None	0.0	0.1945	1.0
None	None	None	0.7198	0.6975	0.3113
None	None	None	0.6718	0.6208	0.2872
None	None	None	0.3899	0.2716	0.6815
None	None	None	0.1616	0.3414	0.7736
None	None	None	0.2196	0.3571	0.7234
None	None	None	0.3311	0.3043	0.5361
('36-50', 'AUC')	('36-50', 'PAUC')	('36-50', 'EER')	('51+', 'AUC')	('51+', 'PAUC')	('51+', 'EER')
0.9375	0.9204	0.075	None	None	None
0.8606	0.8144	0.3381	None	None	None
0.7278	0.8134	0.3819	None	None	None
0.0	0.3192	1.0	None	None	None
0.0735	0.2819	0.9048	None	None	None
0.2731	0.445	0.6056	None	None	None
0.3775	0.4179	0.675	None	None	None
0.3103	0.3511	0.6147	None	None	None

Deepfake Dataset Pipeline Report

0.4472	0.5286	0.5528	None	None	None

Source-Based Evaluation

The following table shows model performance when trained on one dataset source and evaluated on another.

This cross-dataset evaluation helps assess how well models generalize across different data distributions.

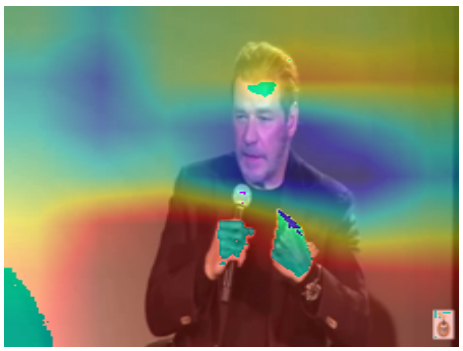
('Test Set', '')	('XceptionNet', 'AUC')	('XceptionNet', 'PAUC')	('XceptionNet', 'EER')	('EfficientNet', 'AUC')	('EfficientNet', 'PAUC')
Balanced	0.6109	0.6884	0.5	0.3472	0.5127
celeb	0.9609	0.9565	0.1238	0.8952	0.8997
FaceForensics++	0.4532	0.6164	0.5415	0.0407	0.3529

('EfficientNet', 'EER')	('LipForensics', 'AUC')	('LipForensics', 'PAUC')	('LipForensics', 'EER')
0.7261	0.6919	0.6646	0.4385
0.2331	0.8719	0.8591	0.2401
0.9246	0.5913	0.5621	0.5036

Grad-CAM Interpretability Visualizations

Grad-CAM visualizations highlight the regions of the input images that most strongly influenced the deepfake detection model's decisions. Below are examples of such heatmaps for real and fake frames:

Real Samples



Fake Samples



Deepfake Dataset Pipeline Report

Help & Instructions

This application builds and evaluates an age-diverse deepfake detection pipeline:

1. Import Videos (celeb-DF, FaceForensics++)

2. Preprocessing

3. Frame Extraction

4. Frame Preview

5. Age Annotation

6. Visualization

7. Dataset Balancing

8. Train-Test Split

9. Model Training

10. Evaluation

11. SimSwap Deepfake Generation

12. Grad-CAM

13. PDF Reporting

14. Dataset Export

Security Disclaimer

- Public datasets only: celeb-DF-v2, FaceForensics++

- Local processing only: no video files are uploaded or shared externally

- This tool is intended for educational and research purposes only

Data Availability Disclaimer

- Raw datasets are NOT downloadable via this tool

- Metadata such as paths, labels, and age group annotations can be exported

- Original datasets can be obtained from:

  UTKFace: https://susanqq.github.io/UTKFace/

Deepfake Dataset Pipeline Report

Celeb-DF-v2: https://github.com/DigitalTrustLab/celeb-DF

FaceForensics++: https://github.com/ondyari/FaceForensics

Deepfake Dataset Pipeline Report

Thank You

I appreciate your time exploring the Age-Diverse Deepfake Dataset Builder!

This tool was designed to support research and experimentation in fairness-aware deepfake detection.

Contact: unishajoshi@email.com

Institution: Department of Data Science, Grand Canyon University

Feel free to reach out for improvements, collaboration, or to share your findings!