DF-Platter: Multi-Face Heterogeneous Deepfake Dataset

Kartik Narayan*, Harsh Agarwal*, Kartik Thakral*, Surbhi Mittal*, Mayank Vatsa, and Richa Singh
IIT Jodhpur, India
{narayan.2, agarwal.10, thakral.1, mittal.5, mvatsa, richa}@iitj.ac.in

1. Video Generation Details

In this section, we provide an estimate of the generation time of videos through each generation algorithm, thereby accounting for the generation time of each set in the proposed DF-Platter dataset. The videos in different sets of the proposed dataset are generated using three techniques: FaceSwap [1], FaceShifter [2], and FSGAN [3]. For FaceSwap, each video is generated after 8 hours of training on 16 Nvidia A100 GPUs of 80GB memory each and 12 Nvidia V100 GPUs of 32GB memory each. Similarly, videos were generated using pre-trained weights of FaceShifter with default parameters on three Nvidia RTX 3090 GPUs of 24 GB memory each. Furthermore, the re-enactment generator of FSGAN is fine-tuned for each source video and inferencing is performed using 12 Nvidia V100 GPUs of 32GB memory each and an Nvidia DGX A40 GPU of 48 GB memory. The dataset generation is detailed in Table 1. The table reports the effective real-time for generating the proposed dataset after parallel processing.

For FSGAN [3] and Faceshifter [2], 65,948 videos were generated, encompassing all three sets. Videos generated using FSGAN [3] were trained on ten different GPUs in parallel. This reduced the effective time to generate one video to 0.8 minutes. Similarly, videos generated using Faceshifter were trained on four different GPUs in parallel. This helped reduce the effective video generation time to 1.25 minutes. Since Faceswap [1] requires more time for training, videos were generated using 16 different GPUs. This greatly reduced the effective generation time for one video to 30 minutes. Furthermore, some videos required training for an additional 16 hours as the resultant deepfake is poor. The generation is completed in approximately 116 days. For generating the compressed versions of the dataset, it took a total of 4 days on an AMD 24-core processor.

2. Results and Analysis

We evaluate various state-of-the-art deepfake detection algorithms on the proposed DF-Platter dataset. We perform thorough benchmarking experiments and report the results obtained. The models are trained and tested for different combinations of compression and resolution. In each table presented in this supplementary, we report the performance obtained by different deepfake detectors on the test set (using a single train and validation split). The models are trained on all the combinations (as described in the first column of the tables) other than those which are already covered in the main paper. Table 4 reports the results obtained on (c0, HR) videos of the DF-Platter dataset. From the table, we can observe that the models trained on (c23, HR) deepfake videos achieve the best performance, followed by the models trained on (c0, LR) videos. Table 5 reports the results obtained on (c23, HR) videos of the dataset when trained on different configurations of compression and resolution. For Set A, models trained on (c0, HR) and (c0, LR) videos of the dataset achieve competitive performance.

In Table 6, we report the performance obtained after testing the deepfake detection models on (c40, HR) videos of the dataset when trained on other variations of resolution and compression. The models trained on (c23, LR) and (c40, LR) videos of the dataset achieve competitive performance. In Table 7, it can be observed that the models trained on (c23, LR) performs the best after (c0, LR) videos of the dataset. This can be attributed to same observations made in the previous baseline experiments. The best performance is achieved through training on (c0, LR) videos as training and testing and performed on the same configuration of the dataset. Similarly, Table 8 reports the performance of different algorithms obtained on (LR, c23) videos of the dataset when trained on different combinations. For Set A, the models trained on (c0, LR) videos of the dataset perform the best while for Set B and Set C, the models trained on (c40, LR) perform the best after the models trained on the same setting of (c23, LR) videos. This can be attributed to the fact that models trained on higher compression and lower resolution are trained to distinguish between the artifacts of deepfake and compression. Due to this, the model is also able to generalize better on (c23, LR) videos of the dataset. The same inference can be made for the performance reported in Table 9, where the models trained on the

*Equal contribution by student authors.

Table 1. Generation time for videos in different sets of the proposed DF-Platter dataset using the FSGAN, FaceShifter and FaceSwap techniques. The time is reported in minutes unless explicitly specified otherwise.

Sets	**FSGAN**		**Faceshifter**		**FaceSwap**		**Time taken for each set (in mins)**
	Number of Videos	Time (in mins)	Number of Videos	Time (in mins)	Number of Videos	Time (in mins)	
Set A	65,348	52,279	65,348	81,685	-	-	133,964
Set B	300	240	300	375	300	11880	12,495
Set C	300	240	300	375	300	11880	12,495
Total time	52,759		82,435		18,000		**∼112 days**

Table 2. p-value and statistic value of McNemar test with DSP-FWA and CapsuleNet trained and tested on different compression factors and resolution.

Trained on	**Tested on**	**Set A**		**Set B**		**Set C**	
		p-value	**statistic**	**p-value**	**statistic**	**p-value**	**statistic**
c0, HR	c0, HR	0	4569.5887	6.1933	546.8192	2.17e-67	300.8265
c23, HR	c23, HR	0	9739.3265	1.92e-101	457.2272	4.37e-127	575.0954
c40, HR	c40, HR	0	282.6047	0.4174	0.6575	2.60e-08	30.9452
c0, HR	c0, LR	0	1626.338	7.44e-19	78.6415	1.47e-05	14.4084
c23, HR	c40, HR	0	7093.7296	5.16e-28	120.4029	2.54e-53	236.2825

Table 3. Performance Accuracy (%) of Capsule model when trained on (c0, LR) and (c40, HR) videos and tested on LR videos across different compressions of the DF-Platter dataset.

Trained on	**Uncompressed LR (c0, LR)**	**Compressed HR (c40, HR)**
Test LR, c0	93.10	84.42
Test LR, c23	91.46	83.98
Test LR, c40	84.70	82.39

(LR, c40) videos of the dataset perform the best.

We have also calculated the McNemar test results of the tables in the main paper to show the statistical significance of the accuracy values. Table 2 shows the McNemar results of the dataset when trained and tested on different compressions and resolutions. The results are calculated for DSP-FWA and Capsule. It can be inferred that the results are statistically significant except in two cases when (a) Trained on Set A (c0, HR) and tested on Set B (c0, HR) and (b) Trained on Set A (c40, HR) and tested on SetB (c40, HR).

Difference between uncompressed low-resolution (c0, LR) and compressed high-resolution (c40, HR): We trained the Capsule model on (c0, LR) and (c40, HR) videos from Set A and evaluate their performance on the test sets of Set A. From Table 3, it can be inferred that models, when trained on uncompressed LR deepfakes, perform better when detecting LR deepfakes across various compression levels. Additionally, the visual quality of Uncompressed LR deepfakes are better than Compressed HR deepfakes with BRISQUE scores of 33.03 and 46.24, respectively.

References

[1] Faceswap. https://github.com/MarekKowalski/FaceSwap/. [Accessed: 27-April-2022]. 1

[2] Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen. Faceshifter: Towards high fidelity and occlusion aware face swapping. *CoRR*, abs/1912.13457, 2019. 1

[3] Yuval Nirkin, Yosi Keller, and Tal Hassner. Fsgan: Subject agnostic face swapping and reenactment. In *IEEE/CVF ICCV*, pages 7184–7193, 2019. 1

Table 4. Performance obtained on the (c0, HR) test set of the proposed DF-Platter dataset for Sets A, B, and C. The models are trained for different combinations of resolution and compression as specified in the first column of the table. The performance is reported using Frame-level Accuracy (FLA) (%), AUC Scores, Face-wise Accuracy (FaceWA) (%), Face-wise AUC score (FaceAUC), and Video-level Accuracy (VLA) (%) as explained in the main paper.

Trained on	**Models**	**Set A**		**Set B**				**Set C**			
		Accuracy	**AUC**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**
c0, LR	MesoNet	80.95	0.7143	81.17	0.5527	60.76	64.97	83.02	0.7652	66.25	65.41
	Meso-Inception	68.92	0.7145	71.21	0.5996	45.58	44.79	73.26	0.7816	54.13	41.46
	FWA	62.53	0.7111	68.17	0.6355	40.00	37.92	66.62	0.7477	45.24	28.82
	Xception	78.86	0.6773	76.86	0.5450	55.24	58.98	78.59	0.7400	59.40	54.77
	DSP-FWA	79.50	0.5000	90.85	0.5000	90.00	99.78	86.96	0.5000	74.38	99.78
	Capsule	85.18	0.8602	79.08	0.6995	58.52	60.31	77.20	0.7979	60.25	53.44
c23, HR	MesoNet	82.39	0.5866	87.41	0.5494	77.82	85.81	86.21	0.6588	72.56	82.71
	Meso-Inception	81.91	0.6455	81.14	0.5589	63.30	67.18	82.56	0.7143	64.83	66.08
	FWA	84.89	0.6978	83.38	0.5880	68.46	74.28	83.16	0.7037	68.35	72.95
	Xception	81.02	0.5590	84.02	0.5293	69.18	77.38	82.62	0.5740	65.11	78.71
	DSP-FWA	91.24	0.8203	81.67	0.6060	61.04	63.64	83.99	0.8116	66.69	61.86
	Capsule	89.29	0.7501	87.12	0.5351	78.30	87.36	85.60	0.7299	71.34	75.83
c23, LR	MesoNet	81.74	0.6413	86.20	0.5587	74.74	82.26	84.99	0.6868	71.34	77.38
	Meso-Inception	80.00	0.7129	78.41	0.5656	60.60	63.41	79.93	0.7250	64.10	63.41
	FWA	75.47	0.7421	74.94	0.6088	51.50	51.44	73.88	0.7224	53.21	44.79
	Xception	80.49	0.6365	83.19	0.5531	67.06	74.50	83.03	0.6678	65.42	72.95
	DSP-FWA	81.77	0.7782	81.44	0.6154	63.48	66.08	81.21	0.7612	64.99	62.53
	Capsule	86.08	0.8448	82.77	0.6707	67.28	69.62	82.68	0.8265	66.74	60.98
c40, HR	MesoNet	81.35	0.5967	84.91	0.5555	71.48	78.71	82.42	0.6351	66.05	73.84
	Meso-Inception	80.70	0.6144	82.84	0.5580	67.96	74.94	80.24	0.6164	61.90	68.96
	FWA	83.22	0.6976	81.34	0.6038	60.66	64.75	80.29	0.7017	60.38	60.98
	Xception	80.63	0.5836	82.34	0.5453	66.80	74.28	81.22	0.6317	64.58	72.95
	DSP-FWA	90.88	0.8066	80.08	0.6058	58.44	62.08	80.93	0.7838	62.83	57.21
	Capsule	87.12	0.8014	79.38	0.6691	55.20	56.76	78.24	0.7569	57.69	52.11
c40, LR	MesoNet	73.56	0.5808	80.05	0.5444	60.66	64.97	79.39	0.6494	60.81	63.19
	Meso-Inception	74.41	0.6397	73.99	0.5442	49.08	50.55	74.55	0.6647	51.42	48.56
	FWA	71.44	0.6649	74.94	0.5586	51.40	53.22	76.15	0.6883	56.88	55.21
	Xception	76.22	0.5630	82.47	0.5396	65.12	72.06	82.43	0.6167	63.98	74.28
	DSP-FWA	74.76	0.6345	75.45	0.5664	52.44	56.98	77.18	0.7033	55.80	54.99
	Capsule	78.50	0.6299	86.92	0.5858	76.56	83.59	86.76	0.7638	73.72	78.27

Table 5. Performance obtained on the (c23, HR) test set of the proposed DF-Platter dataset for Sets A, B, and C. The models are trained for different combinations of resolution and compression as specified in the first column of the table. The performance is reported using Frame-level Accuracy (FLA) (%), AUC Scores, Face-wise Accuracy (FaceWA) (%), Face-wise AUC score (FaceAUC), and Video-level Accuracy (VLA) (%) as explained in the main paper.

Trained on	**Models**	**Set A**		**Set B**				**Set C**			
		Accuracy	**AUC**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**
c0, HR	MesoNet	82.98	0.5833	87.13	0.5467	77.26	84.26	87.23	0.6894	73.67	82.71
	Meso-Inception	83.21	0.6388	82.18	0.5508	64.06	71.62	82.54	0.6779	64.26	71.18
	FWA	83.43	0.6598	80.43	0.5764	61.90	67.18	79.04	0.6770	59.90	61.64
	Xception	80.26	0.5311	87.57	0.5328	79.44	89.80	85.33	0.5899	68.36	87.80
	DSP-FWA	89.58	0.7798	78.47	0.5631	57.10	60.75	82.97	0.7751	65.31	64.08
	Capsule	87.29	0.7032	86.17	0.5930	71.80	79.16	87.18	0.7587	72.55	75.83
c0, LR	MesoNet	80.51	0.6920	81.81	0.5487	62.08	66.52	83.56	0.7474	67.01	67.85
	Meso-Inception	68.34	0.6998	73.76	0.6041	50.52	51.88	75.52	0.7880	56.46	46.56
	FWA	61.06	0.6959	68.57	0.6097	40.24	38.36	67.88	0.7494	47.06	31.71
	Xception	77.60	0.6504	77.56	0.5442	56.26	60.31	79.65	0.7320	61.37	59.42
	DSP-FWA	81.31	0.5000	90.85	0.5000	90.00	99.78	86.98	0.5000	74.38	99.78
	Capsule	84.11	0.8453	78.34	0.6755	56.36	56.32	76.92	0.7787	58.69	51.66
c23, LR	MesoNet	81.29	0.6270	86.20	0.5551	74.80	81.60	85.53	0.6851	72.53	79.60
	Meso-Inception	78.58	0.6962	78.82	0.5718	61.12	64.75	80.50	0.7321	64.63	64.08
	FWA	74.22	0.7285	75.65	0.6052	52.50	54.10	75.01	0.7153	54.17	48.78
	Xception	79.68	0.6219	82.97	0.5541	66.42	74.50	83.25	0.6692	65.42	73.39
	DSP-FWA	80.09	0.7569	81.01	0.6048	62.82	65.19	81.25	0.7576	64.68	62.97
	Capsule	84.31	0.8229	82.32	0.6569	65.96	68.51	83.24	0.8239	67.90	63.19
c40, HR	MesoNet	82.00	0.5848	85.53	0.5550	73.12	80.27	82.99	0.6290	66.89	77.16
	Meso-Inception	81.26	0.6055	82.34	0.5453	66.92	73.84	80.35	0.6114	61.72	70.29
	FWA	83.22	0.6812	81.46	0.6013	61.18	64.75	81.06	0.7034	61.59	63.41
	Xception	81.16	0.5747	83.30	0.5588	68.46	75.17	81.50	0.6273	64.21	72.06
	DSP-FWA	89.21	0.7762	80.63	0.6024	60.16	65.19	81.22	0.7705	62.79	59.65
	Capsule	86.30	0.7763	80.28	0.6517	57.84	60.75	78.98	0.7381	58.78	55.88
c40, LR	MesoNet	74.23	0.5760	79.74	0.5423	60.02	63.86	79.19	0.6453	60.64	63.19
	Meso-Inception	74.64	0.6360	73.79	0.5437	49.06	52.77	74.40	0.6596	51.47	47.67
	FWA	71.75	0.6637	74.31	0.5420	51.18	53.88	76.43	0.6810	57.11	55.65
	Xception	76.59	0.5596	82.62	0.5372	65.46	73.17	82.35	0.6163	63.93	75.17
	DSP-FWA	74.92	0.6306	75.60	0.5633	52.68	57.65	77.52	0.6948	56.33	56.32
	Capsule	78.83	0.6156	86.76	0.5757	76.62	83.81	87.08	0.7486	73.92	80.27

Table 6. Performance obtained on the (c40, HR) test set of the proposed DF-Platter dataset for Sets A, B, and C. The models are trained for different combinations of resolution and compression as specified in the first column of the table. The performance is reported using Frame-level Accuracy (FLA) (%), AUC Scores, Face-wise Accuracy (FaceWA) (%), Face-wise AUC score (FaceAUC), and Video-level Accuracy (VLA) (%) as explained in the main paper.

Trained on	**Models**	**Set A**		**Set B**				**Set C**			
		Accuracy	**AUC**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**
c0, HR	MesoNet	82.02	0.5278	89.28	0.5307	83.44	91.57	87.50	0.6031	72.66	90.69
	Meso-Inception	82.98	0.5850	83.36	0.5436	66.48	73.61	82.84	0.6200	63.29	73.39
	FWA	83.09	0.5994	82.41	0.5627	65.70	70.95	78.28	0.5906	57.09	64.52
	Xception	81.46	0.5298	86.40	0.5163	76.44	87.14	83.67	0.5319	65.04	83.81
	DSP-FWA	86.11	0.6735	74.31	0.5105	49.98	54.32	76.88	0.6295	53.40	55.43
	Capsule	84.80	0.6112	86.45	0.5511	74.56	82.48	84.46	0.6238	66.61	78.71
c0, LR	MesoNet	82.46	0.6260	84.34	0.5293	69.46	75.83	85.86	0.7051	69.50	77.16
	Meso-Inception	75.74	0.6946	77.57	0.5754	55.46	59.42	79.65	0.7641	60.83	56.76
	FWA	67.83	0.6954	72.97	0.5875	46.16	44.79	73.76	0.7578	53.19	41.91
	Xception	80.69	0.5969	81.60	0.5295	64.00	71.18	82.21	0.7067	63.60	66.96
	DSP-FWA	81.31	0.5000	90.96	0.5000	90.00	99.78	87.68	0.5000	74.38	99.78
	Capsule	82.94	0.7880	78.20	0.6119	57.82	62.31	80.26	0.8048	64.08	57.65
c23, LR	MesoNet	82.05	0.5741	87.59	0.5370	78.38	87.14	86.42	0.6501	71.88	83.37
	Meso-Inception	80.96	0.6591	81.59	0.5527	64.54	69.18	82.38	0.6631	65.14	71.62
	FWA	78.43	0.7121	76.46	0.5816	52.04	54.32	77.57	0.7049	56.94	54.10
	Xception	81.10	0.5974	80.85	0.5367	61.28	68.29	81.34	0.6268	60.56	68.29
	DSP-FWA	80.40	0.6890	82.52	0.5746	65.46	70.95	82.01	0.6813	63.64	67.41
	Capsule	84.40	0.7491	83.13	0.6015	67.90	72.28	85.34	0.7641	70.76	74.06
c40, LR	MesoNet	75.61	0.5653	72.18	0.5139	47.46	50.11	78.84	0.6226	59.29	63.41
	Meso-Inception	75.88	0.6295	70.77	0.5329	40.74	42.79	71.72	0.6268	45.72	41.91
	FWA	72.45	0.6514	73.11	0.5363	47.08	48.34	75.29	0.6783	54.78	52.33
	Xception	77.25	0.5566	81.01	0.5405	61.98	66.96	81.70	0.6137	61.47	71.40
	DSP-FWA	75.16	0.6112	76.00	0.5598	53.18	56.54	77.61	0.6974	56.31	56.32
	Capsule	79.37	0.6413	83.46	0.5645	69.16	75.17	83.23	0.7251	67.55	69.84

Table 7. Performance obtained on the (c0, LR) test set of the proposed DF-Platter dataset for Sets A, B, and C. The models are trained for different combinations of resolution and compression as specified in the first column of the table. The performance is reported using Frame-level Accuracy (FLA) (%), AUC Scores, Face-wise Accuracy (FaceWA) (%), Face-wise AUC score (FaceAUC), and Video-level Accuracy (VLA) (%) as explained in the main paper.

Trained on	**Models**	**Set A**		**Set B**				**Set C**			
		Accuracy	**AUC**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**
c0, LR	MesoNet	83.59	0.5886	85.98	0.5230	75.62	84.48	87.78	0.6759	73.87	85.59
	Meso-Inception	86.97	0.7265	80.57	0.5618	61.34	66.30	83.26	0.7561	66.18	65.41
	FWA	87.10	0.7758	76.10	0.5781	52.12	53.88	78.27	0.7494	57.57	49.67
	Xception	83.44	0.5914	84.29	0.5217	71.12	77.61	84.80	0.6853	68.61	77.16
	DSP-FWA	80.29	0.5000	90.75	0.5000	90.00	99.78	88.20	0.5000	74.38	99.78
	Capsule	93.01	0.8276	79.18	0.6363	60.70	64.75	80.07	0.7956	64.69	59.20
c23, HR	MesoNet	80.69	0.5163	88.65	0.5297	82.24	91.80	85.98	0.5360	69.57	89.36
	Meso-Inception	81.01	0.5877	86.08	0.5648	74.94	82.48	85.42	0.6410	69.04	80.49
	FWA	82.08	0.5823	84.34	0.5384	72.40	80.49	83.85	0.5968	67.14	82.48
	Xception	80.89	0.5270	87.44	0.5317	79.10	87.80	83.91	0.5234	65.93	84.92
	DSP-FWA	85.18	0.6481	84.92	0.5350	71.96	77.16	87.19	0.7074	72.84	79.82
	Capsule	83.65	0.5946	86.91	0.6085	74.60	81.82	85.24	0.5586	68.35	85.81
c23, LR	MesoNet	82.14	0.5552	87.08	0.5374	78.30	85.81	87.12	0.6389	72.86	86.25
	Meso-Inception	85.07	0.6644	84.48	0.5560	70.64	78.05	85.28	0.6707	69.07	80.27
	FWA	86.92	0.7228	79.07	0.5590	57.34	60.75	81.82	0.7094	63.06	64.75
	Xception	82.88	0.5833	82.98	0.5313	66.96	72.95	84.64	0.6209	67.09	78.05
	DSP-FWA	86.93	0.6832	84.77	0.5455	74.18	82.26	85.53	0.6609	70.84	81.15
	Capsule	90.35	0.7627	86.08	0.6064	73.80	81.15	87.04	0.7407	73.16	78.05
c40, LR	MesoNet	77.93	0.5509	82.92	0.5386	68.26	74.50	81.98	0.6176	63.62	73.61
	Meso-Inception	78.84	0.6178	75.25	0.5332	53.44	56.98	77.90	0.6190	57.04	61.64
	FWA	80.07	0.6522	77.04	0.5369	56.78	61.42	76.87	0.6219	57.39	59.87
	Xception	78.68	0.5480	81.51	0.5221	65.28	72.95	81.39	0.5645	62.18	74.94
	DSP-FWA	78.79	0.5885	79.44	0.5392	59.94	64.75	79.64	0.6202	59.02	66.30
	Capsule	81.11	0.5753	85.32	0.5460	75.32	83.15	85.12	0.6353	70.03	83.15

Table 8. Performance obtained on the (c23, LR) test set of the proposed DF-Platter dataset for Sets A, B, and C. The models are trained for different combinations of resolution and compression as specified in the first column of the table. The performance is reported using Frame-level Accuracy (FLA) (%), AUC Scores, Face-wise Accuracy (FaceWA) (%), Face-wise AUC score (FaceAUC), and Video-level Accuracy (VLA) (%) as explained in the main paper.

Trained on	**Models**	**Set A**		**Set B**				**Set C**			
		Accuracy	**AUC**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**
c0, HR	MesoNet	81.53	0.5139	89.65	0.5176	86.00	96.23	87.77	0.5375	73.55	97.12
	Meso-Inception	82.32	0.5538	85.06	0.5292	72.42	80.93	84.82	0.5699	67.50	83.81
	FWA	82.13	0.5589	82.58	0.5379	66.34	71.40	80.95	0.5368	60.18	75.39
	Xception	81.29	0.5302	87.25	0.5132	79.26	89.14	85.12	0.5356	68.13	88.69
	DSP-FWA	84.60	0.6067	79.22	0.5303	60.02	64.75	79.48	0.5776	57.83	65.85
	Capsule	83.38	0.5659	82.40	0.5376	67.88	74.50	78.15	0.4975	55.93	69.62
c0, LR	MesoNet	83.37	0.5599	86.65	0.5245	77.40	85.81	87.57	0.6460	73.36	86.70
	Meso-Inception	86.07	0.6917	80.69	0.5698	61.76	65.41	83.12	0.7221	65.75	67.63
	FWA	86.25	0.7471	76.25	0.5836	52.44	53.88	77.44	0.7192	56.46	51.44
	Xception	83.29	0.5671	84.13	0.5168	70.58	77.83	84.37	65.66	67.47	76.50
	DSP-FWA	81.32	0.5000	90.74	0.5000	90.00	99.78	88.19	0.5000	74.38	99.78
	Capsule	91.46	0.7787	78.64	0.6249	59.38	64.97	78.66	0.7598	61.70	57.43
c23, HR	MesoNet	81.55	0.5130	88.50	0.5233	82.08	92.02	86.08	0.5257	69.80	92.02
	Meso-Inception	81.52	0.5818	85.74	0.5575	74.12	80.93	85.24	0.6387	68.76	80.27
	FWA	82.66	0.5792	84.24	0.5404	71.76	78.49	83.87	0.5957	66.84	81.37
	Xception	81.80	0.5241	87.24	0.5303	78.76	86.25	83.32	0.5070	65.02	82.71
	DSP-FWA	85.54	0.6427	83.98	0.5335	69.66	76.27	86.05	69.41	70.51	78.94
	Capsule	84.39	0.5962	85.02	0.5568	72.04	79.82	84.35	0.5586	66.46	83.37
c23, LR	MesoNet	82.77	0.5462	87.12	0.5351	78.30	87.36	86.83	0.6209	72.13	86.47
	Meso-Inception	85.77	0.6610	84.25	0.5441	70.50	78.27	85.44	0.6730	69.27	79.38
	FWA	86.97	0.7078	78.82	0.5566	58.00	63.86	80.44	0.6742	60.98	64.30
	Xception	83.58	0.5773	82.47	0.5292	65.72	72.28	84.32	0.6173	66.71	79.16
	DSP-FWA	87.14	0.6704	85.06	0.5581	74.02	81.60	85.42	0.6581	70.43	80.93
	Capsule	89.80	0.7345	85.89	0.6159	73.24	80.04	86.39	0.7205	71.95	78.71
c40, HR	MesoNet	81.18	0.5318	87.30	0.5233	79.10	88.69	86.34	0.5688	70.50	88.69
	Meso-Inception	81.25	0.5573	83.34	0.5303	68.42	75.61	80.95	0.5566	60.07	72.28
	FWA	82.70	0.5971	83.13	0.5610	67.44	74.72	81.91	0.6117	62.25	72.51
	Xception	81.26	0.5470	85.21	0.5301	73.46	81.82	84.82	0.5791	67.69	83.37
	DSP-FWA	85.38	0.6242	83.42	0.5465	69.54	75.83	83.62	0.6139	66.35	76.50
	Capsule	83.98	0.6714	80.88	0.5974	62.16	67.18	78.87	0.6165	57.83	63.64
c40, LR	MesoNet	78.71	0.5451	82.92	0.5288	68.58	75.61	82.04	0.6063	63.55	73.84
	Meso-Inception	79.83	0.6197	75.32	0.5362	53.82	58.54	77.94	0.6162	57.21	63.41
	FWA	80.53	0.6441	76.72	0.5311	56.74	60.98	77.68	0.6293	58.51	62.53
	Xception	79.41	0.5473	81.30	0.5260	65.04	71.40	81.43	0.5724	62.23	74.72
	DSP-FWA	79.51	0.5829	79.21	0.5368	59.34	64.75	79.69	0.6228	59.12	66.30
	Capsule	81.37	0.5587	84.64	0.5412	74.30	82.04	84.30	0.6175	68.58	80.27

Table 9. Performance obtained on the (c40, LR) test set of the proposed DF-Platter dataset for Sets A, B, and C. The models are trained for different combinations of resolution and compression as specified in the first column of the table. The performance is reported using Frame-level Accuracy (FLA) (%), AUC Scores, Face-wise Accuracy (FaceWA) (%), Face-wise AUC score (FaceAUC), and Video-level Accuracy (VLA) (%) as explained in the main paper.

Trained on	**Models**	**Set A**		**Set B**				**Set C**			
		Accuracy	**AUC**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**	**FaceWA**	**FaceAUC**	**FLA**	**VLA**
c0, HR	MesoNet	81.36	0.5060	88.91	0.5129	84.06	93.79	87.34	0.5264	72.61	97.12
	Meso-Inception	81.22	0.5205	85.52	0.5099	75.30	84.26	83.53	0.5267	65.22	83.15
	FWA	81.22	0.5290	83.13	0.5231	68.10	75.17	80.28	0.5009	58.84	75.39
	Xception	81.18	0.5207	85.59	0.5143	74.86	82.71	83.06	0.5021	63.82	84.26
	DSP-FWA	81.38	0.5607	71.54	0.4961	48.00	52.33	69.44	0.4472	44.61	54.99
	Capsule	81.81	0.5381	75.66	0.5067	59.88	66.74	69.42	0.4295	50.26	65.85
c0, LR	MesoNet	81.78	0.5171	87.01	0.5108	78.66	86.92	85.86	0.5490	70.07	88.25
	Meso-Inception	82.15	0.5793	80.78	0.5324	62.98	69.40	80.09	0.5696	58.58	69.18
	FWA	82.07	0.6272	79.26	0.5419	59.50	64.97	76.81	0.6086	54.99	58.09
	Xception	81.56	0.5202	84.66	0.5015	72.98	81.37	80.96	0.5426	61.92	76.72
	DSP-FWA	81.36	0.5000	90.74	0.5000	90.00	99.78	87.94	0.5000	74.38	99.78
	Capsule	84.70	0.6403	74.51	0.5947	52.36	55.21	74.10	0.5940	54.00	58.76
c23, HR	MesoNet	81.37	0.5049	88.41	0.5185	82.40	92.68	85.89	0.5114	69.34	92.68
	Meso-Inception	80.97	0.5466	84.35	0.5239	71.30	77.16	82.71	0.5848	62.86	76.27
	FWA	81.50	0.5390	84.18	0.5066	74.06	83.59	83.27	0.5255	66.38	85.81
	Xception	81.16	0.5069	86.48	0.5050	77.96	87.36	83.84	0.5141	65.74	85.59
	DSP-FWA	81.83	0.5786	81.51	0.5295	63.78	69.18	81.83	0.5839	61.69	73.17
	Capsule	81.93	0.5978	80.34	0.5194	62.02	70.51	79.26	0.5147	56.43	70.51
c23, LR	MesoNet	81.66	0.5142	86.91	0.5107	78.98	88.91	85.93	0.5548	69.45	90.24
	Meso-Inception	83.00	0.5784	85.37	0.5153	74.44	82.26	86.54	0.6093	71.24	87.36
	FWA	82.99	0.6122	75.83	0.5215	53.62	58.09	77.35	0.5907	55.40	60.98
	Xception	82.00	0.5441	76.75	0.4959	55.36	59.20	78.40	0.5478	56.94	66.74
	DSP-FWA	83.00	0.5603	85.04	0.5522	73.54	81.60	82.33	0.5653	65.34	79.60
	Capsule	83.46	0.5794	83.97	0.5255	70.34	76.72	83.30	0.5675	64.28	79.60
c40, HR	MesoNet	81.34	0.5168	87.25	0.5220	79.26	88.91	84.70	0.5231	67.60	88.47
	Meso-Inception	80.94	0.5298	83.53	0.5158	69.18	77.16	80.63	0.5255	58.43	72.51
	FWA	81.34	0.5499	82.43	0.5378	66.58	73.17	79.56	0.5286	57.26	71.62
	Xception	81.42	0.5291	84.81	0.5089	73.06	81.15	82.00	0.5259	63.27	79.82
	DSP-FWA	82.96	0.5630	81.35	0.5209	65.26	71.62	81.37	0.5150	60.98	77.61
	Capsule	82.39	0.6063	81.09	0.5670	64.06	70.95	78.60	0.5313	57.22	69.84
c40, LR	MesoNet	82.84	0.5571	78.49	0.5248	57.54	62.75	81.25	0.5847	61.54	72.95
	Meso-Inception	84.69	0.6358	72.18	0.5139	47.46	50.11	73.04	0.5690	48.53	52.33
	FWA	84.86	0.6494	74.24	0.5194	50.92	54.10	75.30	0.5780	53.70	58.54
	Xception	83.24	0.5644	79.67	0.5164	61.86	67.85	79.28	0.5352	58.66	71.84
	DSP-FWA	84.42	0.6017	74.68	0.5131	50.76	55.21	74.66	0.5833	50.51	53.88
	Capsule	84.82	0.6042	77.31	0.5549	59.90	65.19	75.09	0.5751	56.81	65.41