

Supplementary material for: 3DCAE-MRI: Overcoming Data Availability in Small Sample Size MRI Studies

Maria de Fátima Machado Dias, Tiago F.T. Cerqueira, João Valente Duarte, Miguel Castelo-Branco

Table S1: Demographics of the participants used to train and validate the autoencoder.

Repository	Site	Total of participants	Number males	Mean and standard deviation [years]	Min Age [years]	Max Age [years]
ABIDE I	California Institute of Technology	37	30	28.4 ± 10.7	17.0	56.2
ABIDE I	Carnegie Mellon University	26	20	26.8 ± 5.7	19.0	40.0
ABIDE I	Kennedy Krieger Institute	54	41	10.1 ± 1.3	8.1	12.8
ABIDE I	Ludwig Maximilians University Munich	56	49	25.7 ± 11.7	7.0	58.0
ABIDE I	NYU Langone Medical Center	182	145	15.3 ± 6.6	6.5	39.1
ABIDE I	Olin Institute of Livingat Hartford Hospital	16	14	16.9 ± 3.7	10.0	23.0
ABIDE I	Oregon Health and Science University	28	28	10.8 ± 1.9	8.0	15.2
ABIDE I	San Diego State University	36	29	14.4 ± 1.8	8.7	17.2
ABIDE I	Social Brain Lab	28	28	33.4 ± 6.8	20.0	49.0
ABIDE I	Stanford University	37	30	9.9 ± 1.6	7.5	12.9
ABIDE I	Trinity Centre for Health Sciences	49	49	17.2 ± 3.6	12.0	25.9
ABIDE I	University of California Los Angeles	97	86	13.0 ± 2.2	8.4	17.9
ABIDE I	University of Leuven	64	56	18.0 ± 5.0	12.1	32.0
ABIDE I	University of Michigan	143	116	14.0 ± 3.2	8.2	28.8
ABIDE I	University of Pittsburgh	55	48	18.9 ± 6.9	9.3	35.2
ABIDE I	School of Medicine	100	100	22.1 ± 7.7	8.8	50.2
ABIDE I	University of Utah School of Medicine	56	40	12.7 ± 2.9	7.0	17.8
ABIDE II	Yale Child Study Center	58	58	38.5 ± 15.5	18.0	64.0
ABIDE II	Barrow Neurological Institute	37	37	22.7 ± 4.4	13.8	30.7
ABIDE II	ETH Zurich	54	44	8.1 ± 1.1	6.2	10.7
ABIDE II	Erasmus University Medical Center Rotterdam	103	68	10.7 ± 1.6	8.1	13.9
ABIDE II	Georgetown University	26	20	24.8 ± 8.5	17.0	54.0
ABIDE II	Indiana University	55	25	20.1 ± 10.5	6.1	46.6
ABIDE II	Institut Pasteur and Robert Debré Hospital	28	28	23.6 ± 4.8	18.0	35.0
ABIDE II	Katholieke Universiteit Leuven	207	137	10.3 ± 1.3	8.0	13.0
ABIDE II	Kennedy Krieger Institute	74	67	9.9 ± 5.0	5.2	34.8
ABIDE II	NYU Langone Medical Center Sample 1	27	24	6.8 ± 1.1	5.1	8.8
ABIDE II	NYU Langone Medical Center Sample 2	93	57	10.9 ± 2.0	7.0	15.0
ABIDE II	Oregon Health and Science University	56	47	12.9 ± 3.1	7.4	18.0
ABIDE II	San Diego State University	41	37	11.1 ± 1.2	8.4	13.2
ABIDE II	Stanford University	42	42	15.2 ± 3.2	10.0	20.0
ABIDE II	Trinity Centre for Health Sciences	32	24	14.8 ± 1.8	12.0	17.8
ABIDE II	University of California Davis	31	25	10.8 ± 2.4	7.8	15.0
ABIDE II	University of California Los Angeles	37	35	13.5 ± 1.9	10.0	17.2
ABIDE II	Longitudinal Sample	26	20	9.8 ± 2.1	7.1	14.3
ABIDE II	University of Miami	34	26	14.9 ± 2.4	9.3	19.5
ABIDE II	University of Pittsburgh	32	27	20.9 ± 7.9	9.1	38.9
ABIDE II	University of Utah School of Medicine	18705	10233	75.3 ± 7.4	51.0	97.0
ADNI	-	1558	661	21.5 ± 2.9	19.0	35.0
GSP	-	1683	638	51.5 ± 25.3	18.0	96.0
OASIS1	-	1345	576	76.9 ± 7.6	60.0	98.0
OASIS2	-	2768	1185	70.7 ± 9.3	42.7	97.0
OASIS3	-	25	20	20.4 ± 7.7	13.0	40.0
FCP1000	AnnArbor a	36	17	348.0 ± 1732.7	19.0	9999.0
FCP1000	AnnArbor b	28	11	30.6 ± 9.2	23.0	54.0
FCP1000	Atlanta	23	8	29.3 ± 5.5	20.0	40.0
FCP1000	Baltimore	20	16	22.6 ± 4.6	19.0	38.0
FCP1000	Bangor	197	68	21.1 ± 1.8	18.0	26.0
FCP1000	Beijing Zang	26	12	29.9 ± 5.2	24.0	44.0
FCP1000	Berlin Margulies	198	68	20.9 ± 2.1	18.0	29.0
FCP1000	Cambridge Buckner	24	10	42.9 ± 20.4	20.0	71.0
FCP1000	Dallas	86	0	-	-	-
FCP1000	ICBM	12	9	23.6 ± 2.6	20.0	27.0
FCP1000	Leiden 2180	19	11	21.8 ± 2.7	18.0	28.0
FCP1000	Leiden 2200	37	13	25.8 ± 5.1	20.0	42.0
FCP1000	Leipzig	18	0	-	-	-
FCP1000	Milwaukee a	46	14	53.7 ± 5.9	44.0	65.0
FCP1000	Milwaukee b	16	9	68.3 ± 4.1	63.0	74.0
FCP1000	Munchen	12	0	-	-	-
FCP1000	NYU TRT session1b	18	10	31.6 ± 10.3	18.0	48.0
FCP1000	NewHaven a	15	7	27.6 ± 6.4	18.0	42.0
FCP1000	NewHaven b	84	40	24.3 ± 10.1	7.0	49.0
FCP1000	NewYork a	25	18	34.9 ± 9.6	20.0	50.0
FCP1000	NewYork a ADHD	20	1	40.0 ± nan	40.0	40.0
FCP1000	NewYork b	19	9	24.1 ± 3.9	21.0	39.0
FCP1000	Newark	9	0	-	-	-
FCP1000	Ontario	20	12	41.6 ± 11.2	20.0	55.0
FCP1000	Orangeburg	102	33	21.5 ± 0.6	20.0	23.0
FCP1000	Oulu	22	11	29.3 ± 3.3	21.0	35.0
FCP1000	Oxford	17	2	31.6 ± 7.6	22.0	46.0
FCP1000	PaloAlto	16	9	37.6 ± 8.7	25.0	54.0
FCP1000	Pittsburgh	19	10	25.9 ± 4.1	20.0	34.0
FCP1000	Queensland	31	13	25.3 ± 2.3	21.0	29.0
FCP1000	SaintLouis	14	0	-	-	-
FCP1000	Taipei a	8	0	-	-	-
FCP1000	Taipei b	8	0	-	-	-

Table S2: Demographics of the participants used to train, validate and test the problem-specific models.

Site	Total of participants	Number males	Mean and standard deviation [years]	Min Age [years]	Max Age [years]
Guys	312	139	50.73 $\pm$ 15.98	20.07	86.20
HH	179	85	47.63 $\pm$ 16.61	20.17	81.94
IOP	67	24	42.13 $\pm$ 16.60	19.98	86.32

Table S3: ANOVA results for the validation set comparing the different training strategies (off-the-shelf and fine-tuning learning using the 3DCAE-MRI, training the network from scratch and PCA) for the age prediction problem and for the different training dataset sizes. The significant p -values are in bold.

	ddof1	ddof2	F	p-value
Number of training samples				
25	3	87	55.16	4.54×10^{-20}
50	3	87	176.24	7.48×10^{-37}
100	3	87	433.09	3.64×10^{-52}
200	3	87	112.17	8.46×10^{-30}
300	3	87	173.41	1.37×10^{-36}

Table S4: Age prediction: for the statistically significant ANOVA comparing the MAE across different training strategies (off-the-shelf and fine-tuning learning using the 3DCAE-MRI, training the network from scratch and PCA) for different training dataset sizes. The significant p -values are shown in bold.

	A	B	T	dof	cohen	p-value
Number of training samples						
25	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	5.98	29.0	1.28	1.67×10^{-6}
25	3DCAE-MRI fine-tuning	PCA	-4.33	29.0	-0.94	0.00016
25	3DCAE-MRI fine-tuning	Training from scratch	-4.68	29.0	-1.06	6.23×10^{-5}
25	3DCAE-MRI off-the-shelf	PCA	-24.36	29.0	-3.91	7.38×10^{-21}
25	3DCAE-MRI off-the-shelf	Training from scratch	-11.26	29.0	-2.58	4.18×10^{-12}
25	PCA	Training from scratch	-1.96	29.0	-0.46	0.06
50	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	4.13	29.0	0.58	0.00028
50	3DCAE-MRI fine-tuning	PCA	-15.83	29.0	-3.84	8.33×10^{-16}
50	3DCAE-MRI fine-tuning	Training from scratch	-5.64	29.0	-1.22	4.24×10^{-6}
50	3DCAE-MRI off-the-shelf	PCA	-26.62	29.0	-4.84	6.23×10^{-22}
50	3DCAE-MRI off-the-shelf	Training from scratch	-9.56	29.0	-1.83	1.82×10^{-10}
50	PCA	Training from scratch	10.26	29.0	2.22	3.69×10^{-11}
100	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-1.13	29.0	-0.18	0.27
100	3DCAE-MRI fine-tuning	PCA	-34.15	29.0	-5.68	5.69×10^{-25}
100	3DCAE-MRI fine-tuning	Training from scratch	-6.93	29.0	-0.92	1.28×10^{-7}
100	3DCAE-MRI off-the-shelf	PCA	-28.12	29.0	-5.37	1.36×10^{-22}
100	3DCAE-MRI off-the-shelf	Training from scratch	-4.13	29.0	-0.73	0.00028
100	PCA	Training from scratch	23.48	29.0	4.33	2.03×10^{-20}
200	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-3.66	29.0	-0.55	0.001
200	3DCAE-MRI fine-tuning	PCA	-13.41	29.0	-3.02	5.83×10^{-14}
200	3DCAE-MRI fine-tuning	Training from scratch	-4.09	29.0	-0.67	0.00031
200	3DCAE-MRI off-the-shelf	PCA	-11.66	29.0	-2.50	1.8×10^{-12}
200	3DCAE-MRI off-the-shelf	Training from scratch	-1.17	29.0	-0.17	0.25
200	PCA	Training from scratch	10.84	29.0	2.21	1.02×10^{-11}
300	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-5.91	29.0	-0.70	2.03×10^{-6}
300	3DCAE-MRI fine-tuning	PCA	-18.20	29.0	-3.20	2.07×10^{-17}
300	3DCAE-MRI fine-tuning	Training from scratch	-6.28	29.0	-0.93	7.33×10^{-7}
300	3DCAE-MRI off-the-shelf	PCA	-17.07	29.0	-2.60	1.15×10^{-16}
300	3DCAE-MRI off-the-shelf	Training from scratch	-1.43	29.0	-0.23	0.16
300	PCA	Training from scratch	12.37	29.0	2.38	4.33×10^{-13}

Table S5: Age prediction: ANOVA comparing the means of MAE of the test set across different training strategies (off-the-shelf and fine-tuning learning using the 3DCAE-MRI, training the network from scratch and PCA) for the different training dataset sizes. The significant p-values are in bold.

	ddof1	ddof2	F	p-value
Number of training samples				
25	3	87	100.13	4.04×10^{-28}
50	3	87	86.56	4.97×10^{-26}
100	3	87	72.81	1.20×10^{-23}
200	3	87	60.25	3.59×10^{-21}
300	3	87	65.74	2.71×10^{-22}

Table S6: Age prediction: Post-hoc results for the statistically significant ANOVA comparing the mean MAE of the test set across different training strategies (off-the-shelf and fine-tuning learning using the 3DCAE-MRI, training the network from scratch and PCA) for different number of training instances. The significant p-values are shown in bold.

	A	B	T	dof	cohen	p-value
Number of training samples						
25	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	8.97	29.0	1.47	7.26×10^{-10}
25	3DCAE-MRI fine-tuning	PCA	-7.51	29.0	-1.83	2.81×10^{-8}
25	3DCAE-MRI fine-tuning	Training from scratch	-9.22	29.0	-1.65	3.99×10^{-10}
25	3DCAE-MRI off-the-shelf	PCA	-19.21	29.0	-4.09	4.88×10^{-18}
25	3DCAE-MRI off-the-shelf	Training from scratch	-13.46	29.0	-2.74	5.29×10^{-14}
25	PCA	Training from scratch	-2.51	29.0	-0.58	0.018
50	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	6.62	29.0	1.30	2.95×10^{-7}
50	3DCAE-MRI fine-tuning	PCA	-13.24	29.0	-2.61	8.04×10^{-14}
50	3DCAE-MRI fine-tuning	Training from scratch	-7.20	29.0	-1.48	6.26×10^{-8}
50	3DCAE-MRI off-the-shelf	PCA	-26.23	29.0	-5.15	9.45×10^{-22}
50	3DCAE-MRI off-the-shelf	Training from scratch	-9.73	29.0	-2.30	1.24×10^{-10}
50	PCA	Training from scratch	-0.16	29.0	-0.03	0.88
100	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	4.45	29.0	1.10	0.00012
100	3DCAE-MRI fine-tuning	PCA	-14.57	29.0	-3.04	7.14×10^{-15}
100	3DCAE-MRI fine-tuning	Training from scratch	-5.27	29.0	-0.85	1.19×10^{-5}
100	3DCAE-MRI off-the-shelf	PCA	-29.09	29.0	-6.35	5.2×10^{-23}
100	3DCAE-MRI off-the-shelf	Training from scratch	-6.26	29.0	-1.55	7.83×10^{-7}
100	PCA	Training from scratch	3.83	29.0	0.85	0.00063
200	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	3.51	29.0	0.91	0.0015
200	3DCAE-MRI fine-tuning	PCA	-7.90	29.0	-1.91	1.03×10^{-8}
200	3DCAE-MRI fine-tuning	Training from scratch	-8.76	29.0	-1.86	1.21×10^{-9}
200	3DCAE-MRI off-the-shelf	PCA	-16.00	29.0	-3.74	6.26×10^{-16}
200	3DCAE-MRI off-the-shelf	Training from scratch	-8.98	29.0	-2.44	7.2×10^{-10}
200	PCA	Training from scratch	-3.74	29.0	-0.97	0.00082
300	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	4.88	29.0	1.22	3.57×10^{-5}
300	3DCAE-MRI fine-tuning	PCA	-2.41	29.0	-0.49	0.023
300	3DCAE-MRI fine-tuning	Training from scratch	-7.64	29.0	-1.80	2.03×10^{-8}
300	3DCAE-MRI off-the-shelf	PCA	-12.40	29.0	-3.08	4.06×10^{-13}
300	3DCAE-MRI off-the-shelf	Training from scratch	-10.05	29.0	-2.63	5.84×10^{-11}
300	PCA	Training from scratch	-7.07	29.0	-1.71	8.98×10^{-8}

Table S7: Age prediction: ANOVA results comparing the stability metric across different training strategies (training the network from scratch, off-the-shelf and fine-tuning learning using the 3DCAE-MRI) for the different dataset sizes tested. The significant p-values are shown in bold.

Number of training samples	ddof1	ddof2	F	p-value
25	2	58	354.44	3.04×10^{-33}
50	2	58	92.00	1.02×10^{-18}
100	2	58	85.77	4.73×10^{-18}
200	2	58	70.86	2.67×10^{-16}
300	2	58	93.16	7.74×10^{-19}

Table S8: Age prediction: Post-hoc results for the statistically significant ANOVA comparing the stability metric across different training strategies (training the network from scratch, off-the-shelf and fine-tuning learning using the 3DCAE-MRI) for the different dataset sizes tested. The significant p-values are shown in bold.

Number of training samples	A	B	T	dof	cohen	p-value
25	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	27.22	29.0	5.22	3.36×10^{-22}
25	3DCAE-MRI fine-tuning	Training from scratch	-7.14	29.0	-1.49	7.49×10^{-8}
25	3DCAE-MRI off-the-shelf	Training from scratch	-20.72	29.0	-5.71	6.25×10^{-19}
50	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	13.34	29.0	3.29	6.68×10^{-14}
50	3DCAE-MRI fine-tuning	Training from scratch	-3.68	29.0	-0.72	9.40×10^{-4}
50	3DCAE-MRI off-the-shelf	Training from scratch	-11.02	29.0	-2.75	6.96×10^{-12}
100	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	11.50	29.0	2.57	2.52×10^{-12}
100	3DCAE-MRI fine-tuning	Training from scratch	-4.54	29.0	-0.79	9.15×10^{-5}
100	3DCAE-MRI off-the-shelf	Training from scratch	-10.77	29.0	-2.51	1.18×10^{-11}
200	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	10.42	29.0	2.06	2.6×10^{-11}
200	3DCAE-MRI fine-tuning	Training from scratch	-5.41	29.0	-1.03	8.02×10^{-6}
200	3DCAE-MRI off-the-shelf	Training from scratch	-9.64	29.0	-2.19	1.49×10^{-10}
300	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	10.64	29.0	2.36	1.6×10^{-11}
300	3DCAE-MRI fine-tuning	Training from scratch	-6.72	29.0	-1.56	2.26×10^{-7}
300	3DCAE-MRI off-the-shelf	Training from scratch	-11.81	29.0	-2.79	1.33×10^{-12}

Table S9: Sex prediction: ANOVA comparing the accuracy of the validation set across training strategies (off-the-shelf and fine-tuning learning using the 3DCAE-MRI, training the network from scratch and PCA) for different training dataset sizes. The significant p-values are in bold.

	ddof1	ddof2	F	p-value
Number of training samples				
25	3	87	14.11	1.42×10^{-7}
50	3	87	12.72	5.75×10^{-7}
100	3	87	14.12	1.40×10^{-7}
200	3	87	19.05	1.39×10^{-9}
300	3	87	14.49	9.76×10^{-8}

Table S10: Sex prediction: Post-hoc results for the statistically significant ANOVA comparing the accuracy of the validation set across training strategies (off-the-shelf and fine-tuning learning using the 3DCAE-MRI, training the network from scratch and PCA) for different training dataset sizes. The significant p-values are shown in bold.

	A	B	T	dof	cohen	p-value
Number of training samples						
25	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-5.69	29.0	-1.18	3.73×10^{-6}
25	3DCAE-MRI fine-tuning	PCA	-3.54	29.0	-0.80	0.0014
25	3DCAE-MRI fine-tuning	Training from scratch	-0.32	29.0	-0.08	0.75
25	3DCAE-MRI off-the-shelf	PCA	3.82	29.0	0.58	0.00065
25	3DCAE-MRI off-the-shelf	Training from scratch	5.05	29.0	1.13	2.2×10^{-5}
25	PCA	Training from scratch	3.25	29.0	0.73	0.0029
50	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-3.68	29.0	-0.80	0.00095
50	3DCAE-MRI fine-tuning	PCA	-1.07	29.0	-0.27	0.29
50	3DCAE-MRI fine-tuning	Training from scratch	2.35	29.0	0.54	0.026
50	3DCAE-MRI off-the-shelf	PCA	3.69	29.0	0.74	0.00092
50	3DCAE-MRI off-the-shelf	Training from scratch	5.42	29.0	1.48	8.0×10^{-6}
50	PCA	Training from scratch	3.61	29.0	0.94	0.0011
100	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-4.95	29.0	-0.94	2.95×10^{-5}
100	3DCAE-MRI fine-tuning	PCA	-3.36	29.0	-0.63	0.0022
100	3DCAE-MRI fine-tuning	Training from scratch	0.31	29.0	0.05	0.76
100	3DCAE-MRI off-the-shelf	PCA	1.56	29.0	0.29	0.13
100	3DCAE-MRI off-the-shelf	Training from scratch	5.17	29.0	1.14	1.59×10^{-5}
100	PCA	Training from scratch	3.81	29.0	0.77	0.00066
200	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-5.14	29.0	-0.92	1.73×10^{-5}
200	3DCAE-MRI fine-tuning	PCA	-4.58	29.0	-0.99	8.2×10^{-5}
200	3DCAE-MRI fine-tuning	Training from scratch	0.99	29.0	0.20	0.33
200	3DCAE-MRI off-the-shelf	PCA	-0.42	29.0	-0.08	0.68
200	3DCAE-MRI off-the-shelf	Training from scratch	5.43	29.0	1.10	7.63×10^{-6}
200	PCA	Training from scratch	5.82	29.0	1.17	2.65×10^{-6}
300	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-3.71	29.0	-0.80	0.00087
300	3DCAE-MRI fine-tuning	PCA	-5.09	29.0	-1.06	1.99×10^{-5}
300	3DCAE-MRI fine-tuning	Training from scratch	0.39	29.0	0.08	0.7
300	3DCAE-MRI off-the-shelf	PCA	-1.53	29.0	-0.24	0.14
300	3DCAE-MRI off-the-shelf	Training from scratch	3.94	29.0	0.95	0.00047
300	PCA	Training from scratch	5.02	29.0	1.24	2.38×10^{-5}

Table S11: Sex prediction: ANOVA comparing the accuracy of the test set across training strategies (off-the-shelf and fine-tuning learning using the 3DCAE-MRI, training the network from scratch and PCA) for different training dataset sizes. The significant p-values are in bold.

Number of training samples	ddof1	ddof2	F	p-value
25	3	87	26.74	2.38×10^{-12}
50	3	87	17.92	3.83×10^{-9}
100	3	87	11.83	1.42×10^{-6}
200	3	87	16.54	1.38×10^{-8}
300	3	87	17.10	8.12×10^{-9}

Table S12: Sex prediction: Post-hoc results for the statistically significant ANOVA comparing the accuracy of the test set across training strategies (off-the-shelf and fine-tuning learning using the 3DCAE-MRI, training the network from scratch and PCA) for different training dataset sizes. The significant p-values are shown in bold.

Number of training samples	A	B	T	dof	cohen	p-value
25	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-5.77	29.0	-1.37	2.99×10^{-6}
25	3DCAE-MRI fine-tuning	PCA	-6.75	29.0	-1.70	2.08×10^{-7}
25	3DCAE-MRI fine-tuning	Training from scratch	-0.39	29.0	-0.11	0.698
25	3DCAE-MRI off-the-shelf	PCA	-1.77	29.0	-0.33	0.0879
25	3DCAE-MRI off-the-shelf	Training from scratch	5.45	29.0	1.37	7.21×10^{-6}
25	PCA	Training from scratch	7.21	29.0	1.73	6.23×10^{-8}
50	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-2.98	29.0	-0.57	0.0058
50	3DCAE-MRI fine-tuning	PCA	-5.13	29.0	-1.13	1.78×10^{-5}
50	3DCAE-MRI fine-tuning	Training from scratch	2.23	29.0	0.54	0.0335
50	3DCAE-MRI off-the-shelf	PCA	-2.01	29.0	-0.46	0.054
50	3DCAE-MRI off-the-shelf	Training from scratch	4.56	29.0	1.04	8.58×10^{-5}
50	PCA	Training from scratch	6.28	29.0	1.57	7.41×10^{-7}
100	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-2.10	29.0	-0.47	0.044
100	3DCAE-MRI fine-tuning	PCA	-4.79	29.0	-0.96	4.53×10^{-5}
100	3DCAE-MRI fine-tuning	Training from scratch	1.41	29.0	0.31	0.169
100	3DCAE-MRI off-the-shelf	PCA	-4.19	29.0	-0.69	0.00024
100	3DCAE-MRI off-the-shelf	Training from scratch	2.95	29.0	0.79	0.0063
100	PCA	Training from scratch	4.90	29.0	1.22	3.34×10^{-5}
200	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-3.12	29.0	-0.68	0.0041
200	3DCAE-MRI fine-tuning	PCA	-5.37	29.0	-1.26	9.02×10^{-6}
200	3DCAE-MRI fine-tuning	Training from scratch	0.24	29.0	0.06	0.81
200	3DCAE-MRI off-the-shelf	PCA	-3.52	29.0	-0.72	0.0014
200	3DCAE-MRI off-the-shelf	Training from scratch	3.70	29.0	0.80	0.00091
200	PCA	Training from scratch	6.96	29.0	1.44	1.18×10^{-7}
300	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	-2.00	29.0	-0.45	0.055
300	3DCAE-MRI fine-tuning	PCA	-2.97	29.0	-0.59	0.0059
300	3DCAE-MRI fine-tuning	Training from scratch	3.66	29.0	0.74	0.0010
300	3DCAE-MRI off-the-shelf	PCA	-0.26	29.0	-0.06	0.80
300	3DCAE-MRI off-the-shelf	Training from scratch	5.05	29.0	1.11	2.21×10^{-5}
300	PCA	Training from scratch	5.70	29.0	1.28	3.66×10^{-6}

Table S13: Sex prediction: ANOVA results comparing the stability metric across different training strategies (training the network from scratch, off-the-shelf and fine-tuning learning using the 3DCAE-MRI) for the different dataset sizes tested. The significant p-values are shown in bold.

	ddof1	ddof2	F	p-value
Number of training samples				
25	2	58	132.62	2.31×10^{-22}
50	2	58	41.58	6.27×10^{-12}
100	2	58	30.11	1.08×10^{-9}
200	2	58	0.61	0.55
300	2	58	6.68	2.46×10^{-3}

Table S14: Sex prediction: Post-hoc results for the statistically significant ANOVA comparing the stability metric across different training strategies (training the network from scratch, off-the-shelf and fine-tuning learning using the 3DCAE-MRI) for the different dataset sizes tested. The significant p-values are shown in bold.

	A	B	T	dof	cohen	p-value
Number of training samples						
25	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	21.55	29.0	6.17	2.14×10^{-19}
25	3DCAE-MRI fine-tuning	Training from scratch	5.83	29.0	1.41	2.54×10^{-6}
25	3DCAE-MRI off-the-shelf	Training from scratch	-8.80	29.0	-2.24	1.1×10^{-9}
50	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	7.72	29.0	1.72	1.65×10^{-8}
50	3DCAE-MRI fine-tuning	Training from scratch	11.73	29.0	2.94	1.56×10^{-12}
50	3DCAE-MRI off-the-shelf	Training from scratch	-0.70	29.0	-0.17	0.49
100	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	7.26	29.0	1.41	5.4×10^{-8}
100	3DCAE-MRI fine-tuning	Training from scratch	6.05	29.0	1.39	1.39×10^{-6}
100	3DCAE-MRI off-the-shelf	Training from scratch	-2.10	29.0	-0.50	0.04
300	3DCAE-MRI fine-tuning	3DCAE-MRI off-the-shelf	3.25	29.0	0.81	2.91×10^{-4}
300	3DCAE-MRI fine-tuning	Training from scratch	3.54	29.0	0.84	1.39×10^{-4}
300	3DCAE-MRI off-the-shelf	Training from scratch	-0.35	29.0	-0.09	0.73