

Supplementary Information

2 0.0.1. Baseline age prediction performance

3 To establish a age prediction baseline performance, experiments both within individual sites and across different
4 sites were conducted. For within-site evaluation, we performed 10-fold train-test splits, using 20% of the data for
5 testing in each fold. A KRR model was again used to predict age, with the same hyperparameters were tuned in
6 the same fashion as before. The MAE and the coefficient of determination (R^2 score) on the test set were reported.

7 For cross-site evaluation, we trained a KRR model using all the available data on one site and evaluated its
8 performance on each of the remaining sites. The same procedure was used to determine the best KRR hyperpa-
9 rameters. It is timely to mention that no cross-validation was used in this experiment, as the entire datasets were
10 used as a train or test.

11 0.0.2. Baseline age prediction performance

12 The average MAE and R^2 score across all five sites, when the age prediction model was trained and evaluated
13 in the same site, were as follows: MAE=5.967 and R^2 =0.797, as detailed in table 1. Notably, sites IXI/HH and
14 IXI/IOP exhibited higher MAE and R^2 score values compared to the other three sites. This discrepancy can be
15 attributed to the smaller size of these two sites, resulting in reduced generalizability and weakening the predictive
16 power.

17 A notable observation from table 2 was the eNKI site’s superior capability to generalize across different sites,
18 likely due to its larger size compared to others. In contrast, the IXI/IOP site showed the least effective performance
19 in age prediction on other sites. This can be primarily attributed to its smaller size.

20 The average values for mean absolute error (MAE), mean squared error (MSE), and R^2 score across all 20 train-
21 test site combinations were recorded as 7.399, 87.579, and 0.7, respectively. These average numbers for cross-site
22 age prediction were elevated when compared to that of age prediction within a specific site, a difference attributed
23 to site-specific effects. This result suggests that a model trained on one site tends to learn the unique characteristics
24 or effects of that particular site. While this is helpful when performing predictions within the same site, these
25 learned site-effects are not generalizable to another distinct site, leading to diminished performance.

Supplementary Table 1: **Results of within-site age prediction.** The error values represent the error of KKR model on the test data averaged across 10 different test datasets within the same site. Abbreviations: MAE: mean absolute error, R^2 score: coefficient of determination.

Site	MAE	R^2 score
IXI/Guys	5.15 ± 0.47	0.827 ± 0.021
IXI/HH	5.856 ± 0.823	0.769 ± 0.058
IXI/IOP	6.536 ± 1.321	0.753 ± 0.088
eNKI	5.665 ± 0.357	0.829 ± 0.024
CamCAN	5.397 ± 0.322	0.867 ± 0.015

Supplementary Table 2: **Results of cross-site age prediction.** Train site is the site used as training data for kernel ridge regression model training and test site is the site used as test data for the model. Abbreviations: MAE: mean absolute error, MSE: mean squared error, R^2 score: coefficient of determination.

Train site	Test site	MAE	MSE	R^2 score
CamCAN	eNKI	6.927	75.901	0.758
CamCAN	IXI/Guys	6.94	74.93	0.702
CamCAN	IXI/HH	7.289	85.011	0.694
CamCAN	IXI/IOP	6.634	71.703	0.736
eNKI	IXI/Guys	5.736	50.353	0.79
eNKI	IXI/HH	6.213	62.599	0.775
eNKI	IXI/IOP	6.171	65.494	0.759
eNKI	CamCAN	5.85	54.734	0.841
IXI/Guys	IXI/HH	6.406	65.733	0.763
IXI/Guys	IXI/IOP	7.905	101.214	0.627
IXI/Guys	CamCAN	7.859	93.393	0.73
IXI/Guys	eNKI	7.464	88.221	0.72
IXI/HH	IXI/IOP	7.583	99.942	0.632
IXI/HH	CamCAN	8.746	118.023	0.658
IXI/HH	eNKI	8.537	112.112	0.643
IXI/HH	IXI/Guys	5.665	48.922	0.805
IXI/IOP	CamCAN	² 8.654	120.145	0.652
IXI/IOP	eNKI	8.668	116.954	0.628
IXI/IOP	IXI/Guys	7.301	82.689	0.671