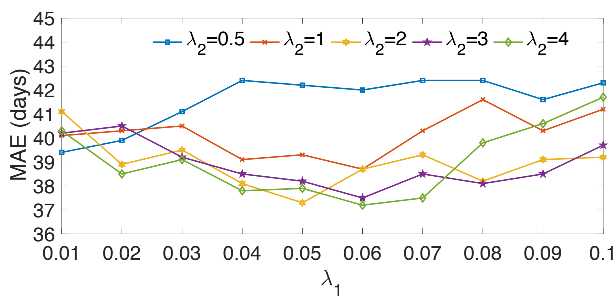
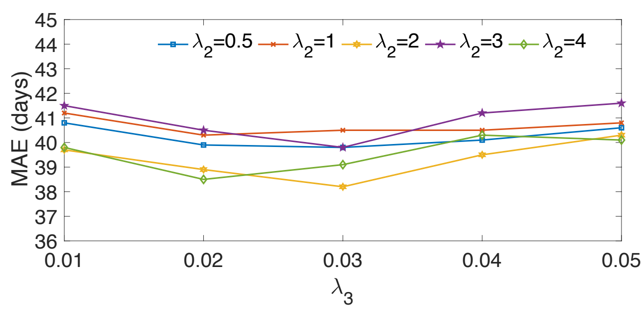
Supplementary Material

1. The impact of the trade-off parameters $\{\lambda_1, \lambda_2, \lambda_3\}$ in the loss function parameters on the performance of age prediction.

According to the different scales of the loss terms in the loss function of our model, the initial ranges of the trade-off parameters were set as $\lambda_1 \in \{0.01, 0.02, \cdots, 0.1\}$, $\lambda_2 \in \{0.5, 1, 2, 3, 4\}$, and $\lambda_3 \in \{0.01, 0.02, \cdots, 0.05\}$, respectively. Fig. S1 shows how the MAE of the age prediction varies with some specific combinations of $\{\lambda_1, \lambda_2, \lambda_3\}$. In each combination, five times of 10-fold cross validation were implemented. Based on the empirical results on different setting of the trade-off parameters in the loss function, grid search was implemented in the inner cross validation with the revised ranges of $\lambda_1 \in \{0.04, 0.05, 0.06\}$, $\lambda_2 \in \{2, 4\}$, and $\lambda_3 \in \{0.02, 0.03\}$.



(a) MAE of age prediction varies with different λ_1 and λ_2 ($\lambda_3 = 0.02$).



(b) MAE of age prediction varies with different λ_2 and λ_3 ($\lambda_1 = 0.02$).

Fig. S1. MAE of age prediction varies with different combination of $\{\lambda_1, \lambda_2, \lambda_3\}$.

2. The impact of the settings of dimension of latent space, common code, and specific code on the performance of age prediction.

In each combination, five times of 10-fold cross validation were implemented and the trade-off parameters in the loss function were fixed as $\lambda_1 = 0.02, \lambda_2 = 2,\ \lambda_3 = 0.02$.

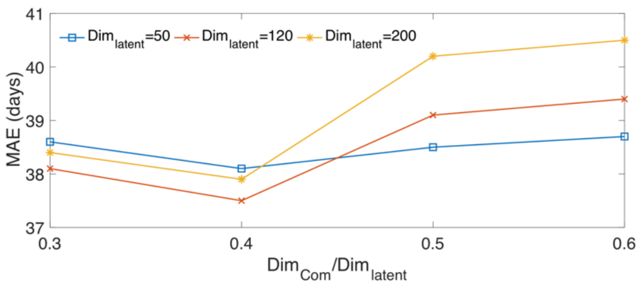


Fig. S2. The MAE of the age prediction varies along with the setting of $\mathrm{Dim}_{\mathrm{latent}}$ and $\mathrm{Dim}_{\mathrm{Com}}/\mathrm{Dim}_{\mathrm{latent}}$. $\mathrm{Dim}_{\mathrm{latent}}$ and $\mathrm{Dim}_{\mathrm{Com}}$ represent the dimension of the latent variable and the common code, respectively.

3. The comparison among DMM-AAE and some multi-modal regression methods on the synthetic dataset

In the 326 sMRI scans and 171 fMRI scans used in our experiments, each fMRI scan has a sMRI scan paired with it, which means we only handled the problem of missing fMRI scan. To simulate a scenario with missing data for both imaging modalities, we created a synthetic dataset by randomly deleting 5% sMRI from the original dataset in each fold and run the 10-fold cross validation 20 times. The results of the age prediction with the synthetic dataset are shown in Table SI. DMM- AAE maintains its advantage even when some sMRI data are missing. Compare with Table II, the MAE obtained by DMM- AAE increases 5.2 days when 5% sMRI scans are missing, while the increased MAEs obtained by other baseline methods are averaged to 8.0 days.

TABLE SI. THE COMPARISON AMONG DMM-AAE AND SOME MULTI-MODAL REGRESSION METHODS (5% STRUCTURAL MRI MISSING)

Fusion type	MAE	MRAE	r_L	r_R
Model-agnostic				
RF (Early).	62.6±0.7	0.23±0.003	0.920±0.023	0.949±0.011
SVR (Early)	65.6±0.9	0.28±0.004	0.921±0.019	0.948±0.010
GPR (Early)	78.4±0.4	0.27±0.002	0.885±0.033	0.921±0.023
PLSR (Early)	61.2±0.8	0.24±0.002	0.924±0.016	0.950±0.008
PLSR (Late 1)	58.9±0.9	0.25±0.003	0.926±0.012	0.951±0.006
PLSR (Late 2)	81.8±1.1	0.37±0.009	0.918±0.026	0.942±0.022
PLSR (Hybrid)	58.9±1.2	0.25±0.003	0.926±0.013	0.951±0.006
Model-based				
MKL	60.8±3.3	0.24±0.022	0.883±0.012	0.941±0.006
iMSF	64.8±1.2	0.24±0.016	0.887±0.008	0.938±0.002
AAE-based				
AAE (Early)	48.8±1.7	0.17±0.006	0.929±0.007	0.954±0.003
AAE (Late)	55.2±1.4	0.21±0.012	0.928±0.008	0.953±0.002
DMM-AAE	**42.8±1.2**	**0.15±0.004**	**0.935±0.0045**	**0.963±0.002**

4. Ablation study of DMM-AAE

An ablation study was implemented to analyze the effectiveness of $\mathcal{L}_{disen}$, $\mathcal{L}_{cross-recon}$, $\mathcal{L}_{adv}$, and embedded imputation strategy. The MAE, MRAE, r_L, and r_R were obtained from 5 times of 10-fold cross validation were reported in Table SII.

Table SII. ABLATION STUDY OF DMM-AAE

	MAE	MRAE	r_L	r_R
Without $\mathcal{L}_{disen}$	42.5$\pm$1.1	0.14$\pm$0.008	0.945 $\pm$ 0.005	0.967 $\pm$ 0.004
Without $\mathcal{L}_{cross-recon}$	39.1$\pm$1.4	0.13$\pm$0.006	0.950 $\pm$ 0.004	0.971 $\pm$ 0.003
Without $\mathcal{L}_{adv}$	40.9$\pm$1.3	0.13$\pm$0.007	0.946 $\pm$ 0.003	0.970 $\pm$ 0.003
Without imputation	40.2$\pm$1.2	0.12$\pm$0.004	0.951 $\pm$ 0.003	0.971 $\pm$ 0.002
DMM-AAE (Ours)	**37.6$\pm$1.3**	**0.11 $\pm$0.004**	**0.953 $\pm$ 0.003**	**0.975 $\pm$ 0.002**