

# Brain Age Gap Estimation from MRI Biomarkers: Benchmarking Machine Learning Regressors for an Interpretable Aging Marker



Sania Eskandari<sup>1</sup> – Luis Gonzalo Sanchez Giraldo<sup>1</sup> for the Alzheimer’s Disease Neuroimaging Initiative\*

<sup>1</sup>Department of Electrical and Computer Engineering, University of Kentucky, Lexington, KY, USA | ses235@uky.edu

## Rationale and Study Cohort

### Background and Objective

Brain-age modeling summarizes MRI-derived information as an estimate of biological brain aging. The **brain age gap (BAG)** quantifies deviation from chronological age and provides an interpretable marker for downstream neurodegenerative-risk analyses [1].

**Objective:** benchmark five regressors for age prediction from ADNI MRI summary biomarkers, consistent with prior comprehensive evaluations of machine-learning methods for brain-age prediction [2], and establish a transparent BAG baseline.

### Study Cohort

**1,556**  
Total participants  
49.7% women

**1245**  
Training set  
57.4% women

**311**  
Test set  
44.1% women

### MRI Biomarkers and Covariates

#### MRI-derived measures

- ▶ Hippocampal volume
- ▶ White matter hyperintensity (WMH) burden
- ▶ Global volumetric reference

#### Demographic covariates

- ▶ Sex and education

The feature set captures regional atrophy, lesion burden, overall brain size, and demographic context.

### Brain Age Gap and Clinical Relevance

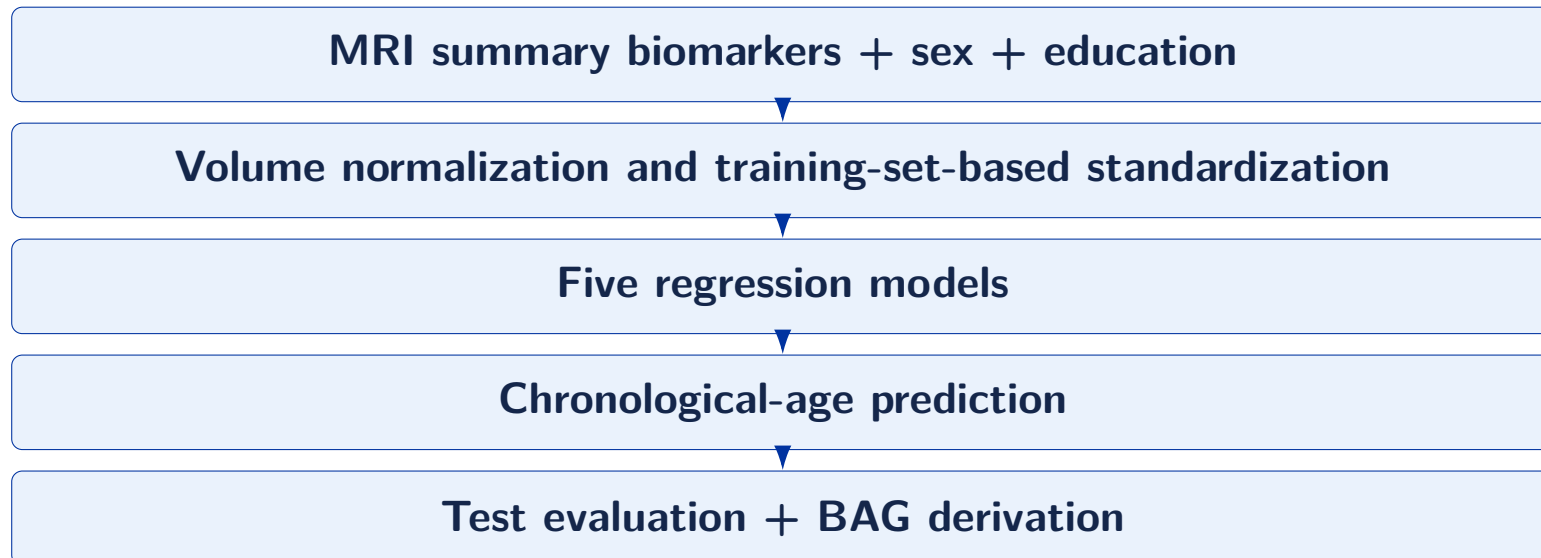
$$\text{BAG} = \text{Predicted Brain Age} - \text{Chronological Age}$$

**Positive BAG** indicates an older-appearing brain; **negative BAG** indicates a younger-appearing brain.

- ▶ Compact and clinically interpretable aging marker.
- ▶ Candidate feature for future disease-risk classification.

## Analytical Framework

### Modeling Workflow



### Models and Preprocessing

#### Regressors

- ▶ Linear Regression
- ▶ SVR
- ▶ KNN Regressor
- ▶ XGBoost Regressor
- ▶ TabPFNRegressor

#### Preprocessing

- ▶ Regional-volume normalization where applicable
- ▶ Continuous predictors were standardized using training-set statistics; the same transformation was applied to the test set for all models

### Evaluation Metrics

#### MAE

Mean Absolute Error

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Prediction error in years

Lower is better

$n$ : number of participants;  $y_i$ : chronological age;  $\hat{y}_i$ : predicted age;

#### RMSE

Root Mean Squared Error

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Prediction error in years

Lower is better

#### $R^2$

Coefficient of Determination

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Explained predictive variation

Higher is better

$\bar{y}$ : mean chronological age.

### Interpretability Results

**Method:** SHAP (SHapley Additive exPlanations) was applied to the tree-based model to quantify feature-level contributions to predicted brain age [3].

**Result:** Hippocampal volume and white matter hyperintensity (WMH) burden emerged as influential predictors of brain age.

- ▶ These findings indicate complementary contributions from regional atrophy and cerebral lesion burden.

## Results and Interpretation

### Key Finding

TabPFNRegressor achieved the strongest overall test performance.  
**MAE = 5.3666 y** | **RMSE = 6.6840 y** |  **$R^2 = 0.2502$**

### Performance Summary

| Model                  | MAE           | RMSE          | $R^2$         |
|------------------------|---------------|---------------|---------------|
| Linear Regression      | 5.5319        | 7.1031        | 0.1533        |
| SVR                    | 5.4939        | 6.8423        | 0.2143        |
| KNN Regressor          | 5.8431        | 7.4167        | 0.0768        |
| XGBoost Regressor      | 5.9214        | 7.5275        | 0.0516        |
| <b>TabPFNRegressor</b> | <b>5.3666</b> | <b>6.6840</b> | <b>0.2502</b> |

### Performance Interpretation

**Interpretability:** SHAP analysis identified hippocampal volume and white matter hyperintensity burden as influential contributors to predicted brain age.

- ▶ **TabPFNRegressor ranked first across all three evaluation metrics.**

- ▶ Compared with the second-best model, SVR, TabPFN reduced MAE by **0.1273 years** and RMSE by **0.1583 years**.

- ▶ Linear Regression remained competitive, whereas KNN and XGBoost showed higher prediction errors and lower  $R^2$  values.

- ▶ The best  $R^2$  of **0.2502** indicates that the evaluated summary biomarkers captured meaningful, but incomplete, age-related variation.

### Conclusions and Next Step

- ▶ TabPFNRegressor provided the strongest benchmark among the five evaluated regression models.
- ▶ MRI summary biomarkers supported compact and clinically interpretable brain-age estimation.
- ▶ The moderate predictive performance suggests that additional structural, clinical, or multimodal biomarkers may improve brain-age estimation.
- ▶ Future work will evaluate whether BAG provides incremental diagnostic value for supervised neurodegenerative-disease classification.

#### Selected References

- [1] Cole JH, Franke K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends Neurosci.* 2017;40(12):681–690.
- [2] Beheshti I, et al. Predicting brain age using machine learning algorithms: A comprehensive evaluation. *IEEE J Biomed Health Inform.* 2021;26(4):1432–1440.
- [3] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30.

#### ADNI Acknowledgment

\*Data used in preparation of this work were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI). ADNI investigators contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report.