Brain age prediction using machine learning techniques

J. Asher, K. T. Dang, P. Klopfenstein, M. Masters, J. Yeater

Supervised by Dr. A. Selvitella

Purdue University Fort Wayne

Brain age problem

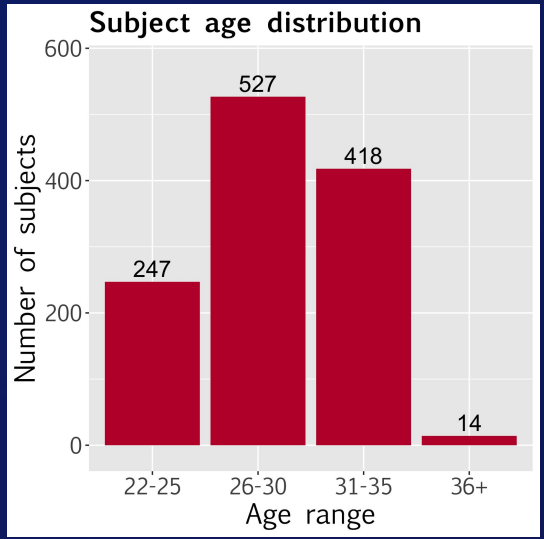
- Predict someone’s age from quantitative brain data
- Understand the aging population
  - Neurodegenerative diseases
- Has been researched before
  - Not using the HCP dataset
  - Not using our dimensionality reduction technique
- Explainable model

Human Connectome Project

- Government funded initiative
- Goal: map the brain’s neural pathways
- Research divided between two consortiums
  - Washington University in Saint Louis, Minnesota University, and Oxford University consortium
  - Harvard/Massachusetts General Hospital and UCLA consortium
- Released data for public use
  - Behavioural dataset
  - Restricted data

HCP Young Adults Behavioural Dataset

- Includes biometrics for the brain's regions (FreeSurfer data)
- 192 FreeSurfer variables:
  - 56 volume
  - 68 area
  - 68 thickness
- 1113 of the 1206 individuals have FreeSurfer data
- 4 age groups



Isoperimetric ratios

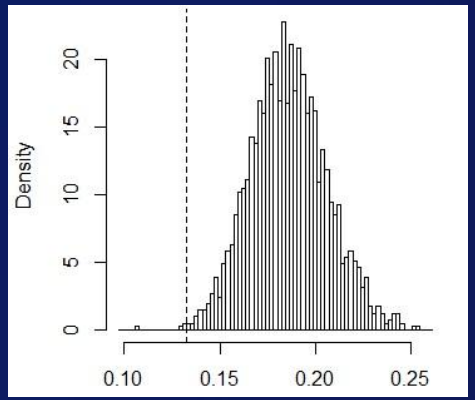
- Different ways to measure folding (gyrification)
  - Local gyrification index (LGI)
  - Isoperimetric quotients (IQs)
- Used isoperimetric ratios (IRs) instead of IQs
- The IR for a body of dimension n with boundary of area B (surface area) and volume V is B^n / V^{n-1}
  - 3d region IR: SA^3/V^2
  - 2d region IR: P^2/A

Isoperimetric ratios

- Could not calculate the 3d IR to measure folding
  - Volumes could not be paired with areas
- Used 2d IR instead, but with thickness replacing perimeter
  - T^2/A
  - Isoperimetric type ratio (ITR)

Multivariable regression model

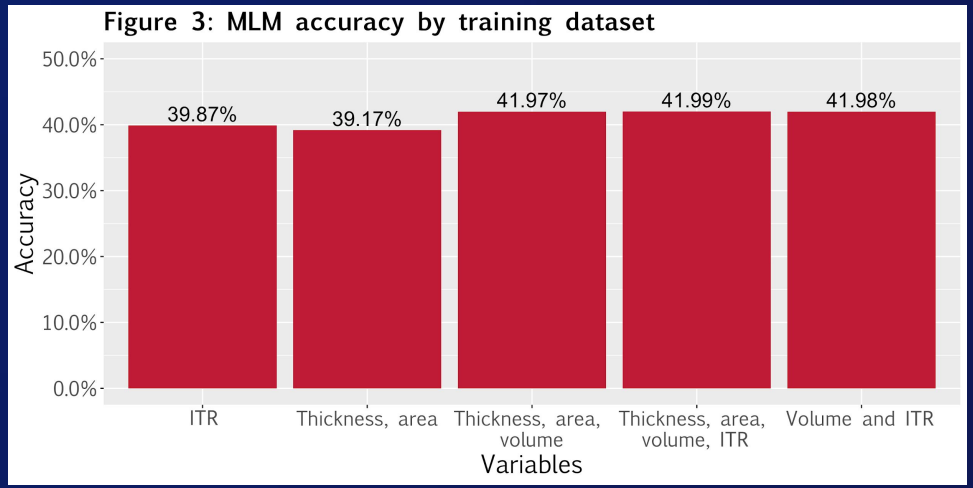
- Calculated a multivariable linear regression model with the FreeSurfer data and one with the IR data
- Used bootstrapping to construct confidence intervals for IR model
  - 2,000 bootstrap samples
  - Percentile confidence intervals
- IR linear model failed to preserve accuracy
  - Logistic model



Histogram of the coefficients of determination from bootstrap samples

Logistic model

- Calculated logistic models for the isoperimetric data and FreeSurfer data, along with other combinations
- Evaluated the models using a validation set approach
  - 10,000 iterations
- ITRs preserved prediction accuracy



Difficulty of the Problem

- Less metamorphic brain changes in young adults
- Missing full information about the brain regions
  - Could not pair volumes and surface areas
- Do not have the subjects' exact ages
  - Classification mistakes

Further analysis performed

- Group analysis
- Linear discriminant analysis
- Uncertainty quantification
- Visualization

	23.5	28	33
23.5	0.15151515	0.65800866	0.19047619
28	0.11316872	0.56790123	0.31893004
33	0.08616188	0.58224543	0.33159269

Matrix results from linear discriminant analysis with predicted on vertical axis and “actual” on horizontal axis

Conclusions

- Studied the brain age problem using the HCP Young Adult dataset
- Used biometrics of an individual to try to predict their age
- Estimated the variability of the coefficients of our prediction models via bootstrap
- Isoperimetric-type ratios provide an interpretable dimensionality reduction without sacrificing accuracy
- Future research
  - Regression problem with unrestricted data set
  - Brain imaging
  - Etc.

Thank you for your attention!