Predicting Age based on Bio-measurements

Peyton Ardoin, Khalil El-abbassi, Aditya Baisakh, Marilin Guerrero Laos, Tre' Henson, Matthew Lemoine, Austin Louque, Efren Mesino Espinosa, Aduragbemi Oyerinde, Shalini Shalini, Gowri Priya Sunkara

Department of Mathematics, Louisiana State University, Baton Rouge, Louisiana
Faculty Advisors: Dr. Nadejda Drenska, Dr. Peter R. Wolenski

History of the Project

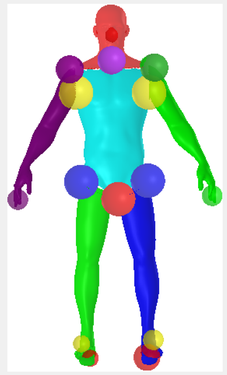

Figure 1. Example of an avatar from which body measurements are obtained.

- In previous projects, the Math Consultation Clinic predicted anthropometric and body-composition values from biomarkers.
- DXA scanning was the standard method for obtaining these values, but previous projects predicted them with high precision using machine-learning algorithms.
- A scan of a patient's body is taken, measurements are extracted, and those measurements are used to train and test machine-learning models.

Models used in predicting biological Age using Pennington Data

- Linear Regression
- Ridge Regression
- Bayesian Ridge Regression
- Random Forest
- XGBoost
- Neural Network

Models Studied and Final BSBA Biomarker Groups

Linear Regression models chronological age as a linear combination of the input biomarkers. Ridge Regression adds regularization to shrink coefficients, reducing overfitting when biomarkers are highly correlated. Bayesian Ridge Regression treats model coefficients probabilistically, accounting for uncertainty and providing more stable predictions.

Morphometric Biomarkers

- Waist, narrow waist
- Thigh, calf, and leg measurements
- Inseam and outside leg length
- Arm length and arm volume
- Leg surface area

Body Composition Biomarkers

- Appendicular Lean Mass (ALM)
- Bone Mineral Density (BMD)
- Total Body Volume
- Torso Volume

Linear Regression Results

Model	Biomarkers	Overall RMSE
Bayesian Ridge	46	13.278
Ridge	46	13.320
Bayesian Ridge	21	12.812
Ridge	21	12.867

- Reducing the model from **46 biomarkers** to **21 biomarkers** improved overall RMSE and produced a simpler, more interpretable BSBA model.

XGBoost

e**X**treme **G**radient **Boost**ing - also an ensemble algorithm but uses boosting vs. bagging (sequential vs. parallel)

RMSE 12.88

Root mean squared error, in years.

Observations

- Model performed best on the densely-sampled **5 - 12 age group** (RMSE $\approx$ 7.3)
- Error concentrates at the sparse **extremes**: teens and, especially, older adults (only 13 patients, RMSE $>$ 24)
- These thinly-sampled age bands inflate the overall RMSE disproportionately

Random Forest

- Random Forest is an ensemble of decision trees trained on bootstrapped samples.
- Each split selects the optimal feature and threshold by maximizing impurity reduction.
- Hyperparameters (number of trees, maximum depth, minimum samples, etc.) are optimized using randomized search.
- The optimized model is trained on 80% of the data and evaluated on the remaining 20%.
- Prediction performance is compared across biomarker subsets grouped by feature importance (high, medium, and low) and by biomarker interaction features.

Below are the results obtained from testing Random Forest.

Model	Error
Random Forest with 11 age brackets	11.6 years
Random Forest Tuned Hyper-parameters	10.7 years
Random Forest Bracketed markers	13.3 years
Random Forest Bracketed Markers with interactions	12.42 years

Table 1. Results from RF tests.

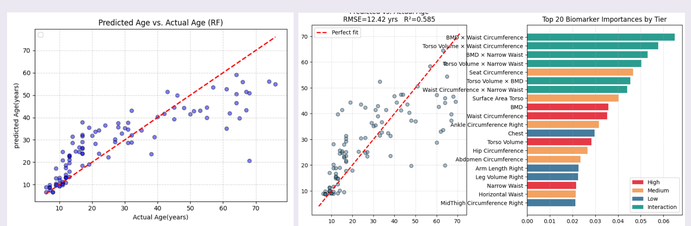

Figure 2. RF with tuned parameters and 11 age brackets, RF with tuned parameters and interacting tiered markers

Neural Network

- Model performance was evaluated using 10 repeated runs of 5-fold cross-validation, with 80% of the data used for training and 20% for testing in each fold.
- Network weights were updated after each mini-batch using backpropagation during training.
- The reported RMSE is the average across all folds and repeated runs for each epoch setting.

Group 1: The entire dataset

Group 2: Breaking the dataset into 16 age brackets

Group 3: Breaking the dataset into 3 using ALM and BMD

Epoch	Group 1	Group 2	Group 3
50	13.52 yrs	13.61 yrs	13.42 yrs
100	13.04 yrs	13.04 yrs	13.10 yrs
150	12.87 yrs	12.91 yrs	12.89 yrs
200	12.56 yrs	12.78 yrs	12.65 yrs
250	12.67 yrs	12.58 yrs	12.60 yrs
300	12.45 yrs	12.72 yrs	12.49 yrs
350	12.48 yrs	12.46 yrs	12.42 yrs
400	12.96 yrs	13.00 yrs	12.55 yrs

Table 2. Mean RMSE results from Neural Network

Acknowledgements

- We would like to thank Dr. Steven B. Heymsfield from Pennington Biomedical Research Center for giving us the datasets and supporting the evaluation of the models.
- We would like to thank Prof. Peter Wolenski and Dr. Nadejda Drenska for their help with this project and their guidance.

References

[1] G. Agrahari, K. Bist, M. Pandey, J. Kapita, Z. James, J. Knox, S. Heymsfield, S. Ramirez, P. Wolenski, and N. Drenska, "Predicting anthropometric body composition variables using 3d optical imaging and machine learning," *Frontiers in Bioinformatics*, vol. 6, p. 1722578, 2026.

[2] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[3] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, pp. 293–300, 1999.

[4] C. A. Fermín-Martínez, A. Márquez-Salinas, E. C. Guerra, L. Zavala-Romero, N. E. Antonio-Villa, L. Fernández-Chirino, E. Sandoval-Colin, D. A. Barquera-Guevara, A. Campos Muñoz, A. Vargas-Vázquez, C. D. Paz-Cabrera, D. Ramírez-García, L. M. Gutiérrez-Robledo, and O. Y. Bello-Chavolla, "Anthropoage, a novel approach to integrate body composition into the estimation of biological age," *Aging Cell*, vol. 22, no. 1, p. e13756, 2023.