Comparing Machine Learning Approaches to Predict Biological Age from Multi-Omics Data and Risk Factor Analysis in BioHEART-CT

Introduction & Aims

Chronological age is a well-known risk factor for many non-communicable diseases however measuring age based on the time elapsed is arbitrary. A more clinically relevant measurement, would capture the aggregate effect of cellular & biochemical processes in our body produced by genetic and environmental factors that translates to physiological impairment – termed biological age (Horvath, 2013).

The utility of biological age lies in its use of cross-sectional data. Traditionally, the study of risk factors on adverse events would require a survival analysis on longitudinal data – a non-trivial task to acquire. Biological age can be thought of as a surrogate – allowing the examination of contributions to risk of morbidity or life expectancy at any point in time. Moreover, patients with accelerated rates of aging can be identified for early intervention (Robinson et al., 2020).

The biomarkers used to predict biological age, initially used physical phenotypes such as declining cognition, muscular atrophy, expiratory volume and blood LDL. This later progressed to usage of cellular hallmarks of aging such as telomere attrition and cellular senescence (Rutledge et al., 2022). More recently the use of omics-based composite biomarkers has been popularised by the seminal DNAm clock which measures DNA methylation at influential CpG sites (Horvath, 2013).

Since then omics-based clocks such as proteomic age, mAA (metabolomics), iAge (immunomics) have been trained (Lehallier et al., 2019; Robinson et al., 2020; Sayed et al., 2021). These clocks exclusively use a single type of assay and interestingly some have shown low correlation with established methylation clocks indicating that they may be capturing a separate component of the aging process. Therefore, what remains unclear is whether a biological age derived from a panel of assays would perform better than any individual assay at capturing the aging effect. Yet, integrating multi-omics data appears to be promising as it holistically captures the complex biology that governs the flow from genotype to phenotype (Subramanian et al., 2020). Furthermore, it also remains to be seen which modelling approach is best suited to heterogenous data. Extensive work has shown elastic net regression to be a well-performing approach however it's possible that the interconnected nature of multi-omics may favour deep-learning approaches which have traditionally evaded success (Acharjee, 2012). Therefore, the first aim of the study is to;

(i) Find the best modelling technique for combining -omics assays to predict biological age and benchmark it against individual assays as this is currently unclear.

Historically, DNA methylation clocks have been well-linked to cancer-related mortality however struggle to correlate well with cardiovascular disease (CVD) outcomes and risk factors (Horvath, 2013). Interestingly, plasma proteomic clocks have been shown to include many CVD-associated proteins (Lehallier et al., 2019). However, what remains unclear is whether a biological age trained from such assays are also associated with late signs of disease that are precursors or comorbid to CVD. Yet, it is precisely that connection to observable signs of disease that would strengthen the evidence that the selected biomarkers indeed drive the spectrum of aging across the body. Therefore, the second aim of the study is to;

(ii) Determine the early risk factors which contribute to increased biological age calculated from multiple assays and its link to observable signs of disease.

Methods

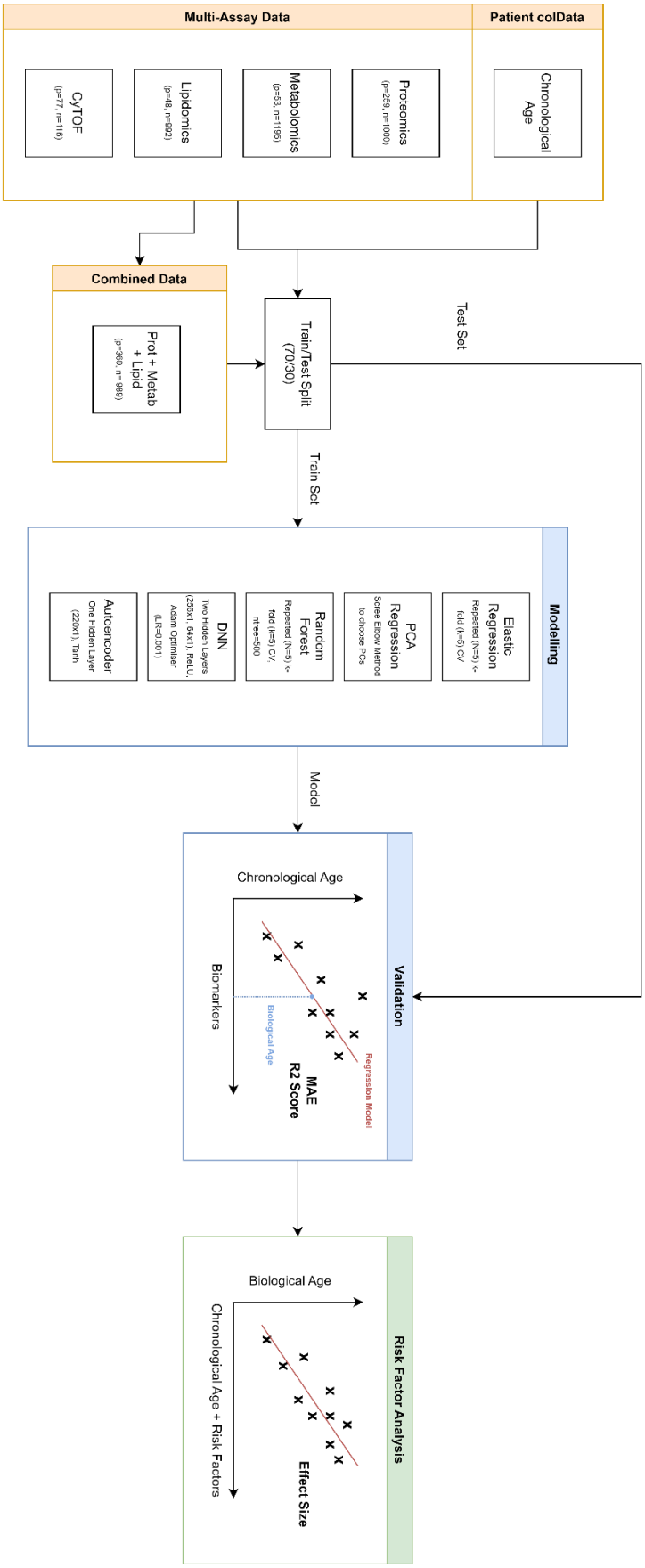

Figure 1 - Approach to modelling, validation and risk factor analysis.

Cohort information

The dataset used is derived from BioHEART-CT biobank – an Australian prospective, longitudinal cohort study of patients who had CT coronary angiograms with coronary artery calcium score (CACS) (Kott et al., 2019). The cohort contained 969 adult participants with 540 males (56%). The mean age was 60.8 (SD=12.2, range: 21-94 years). Most of participants (86.8%) were of European ethnicity. All patients were sourced from the Royal North Shore Hospital clinical site.

Multi-omics datasets were available as per *Figure 2*. The dataset contained metabolomics, proteomics, lipidomics and CyTOF data which was normalised by hierarchical remove-unwanted-variation (hRUV) method (Kim et al., 2021).

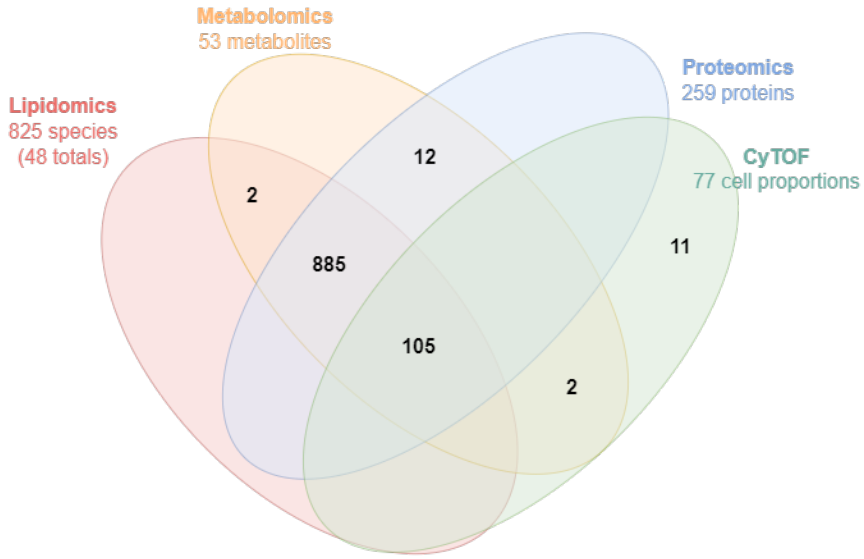

Figure 2 - Available patients for various combination of omics platforms.

Biological Age Modelling

As in *Figure 1*, All -omics datasets were split into independent training (70%) and test (30%) sets to prevent overfitting. All modelling was performed on the independent sets and out-of-sample performance was validated on the held-out test sets. All analysis took place in R ver. 4.1.0. Five common machine learning methods were chosen as candidates due to their previous use in omics analyses; elastic net, PCA regression, random forest, deep neural network (DNN), autoencoder regression (Acharjee, 2012).

In all modelling situations, the chronological age was regressed on either a subset of analytes or a dimensionally reduced representation of the assay data. In this sense, the model was 'tuned' on the chronological age as a calibration. The predictions made by the resulting model were then considered the biological age (bAge). All of the modelling was performed on metabolomics, proteomics, lipidomics, CyTOF separately. The modelling was then repeated on a concatenated set of lipidomics, proteomics and metabolomics which served as the 'combined assay'. CyTOF was omitted from this combined analysis because the intersection was insufficient as in *Figure 2.*

Elastic net regression was modelled using "glmnet" (Friedman et al., 2021). 5-fold repeated cross validation (n=5) using "caret" was performed to find optimal hyperparameters (Kuhn et al., 2022). This random search for *alpha* (describes the proportion of ridge and LASSO penalties) and *lambda* (severity of the penalty for large coefficients with small predictive value) hyperparameters optimised for minimum cross-validated RMSE.

PCA regression involved a multi-stage process. Firstly, single value decomposition of the assay matrix was performed to find principal components (PC). The 'elbow' of the scree plot indicated an abrupt drop in explained variance by the subsequent PCs and these PCs were dropped. Chronological age was then regressed on the PCs before the 'elbow', using multiple linear regression (MLR).

Random forest was modelled using Breiman's algorithm in "randomForest" (Cutler & Wiener, 2022). 5-fold repeated cross validation (n=5) using "caret" was used to tune hyperparameter *mtry* (number of randomly selected variables to split on at each branch) (Kuhn et al., 2022).

Neural network architecture consisted of two hidden fully connected layers (256, 64) using the ReLU activation function. The model training used an objective function of mean absolute error, an Adam optimiser (learning rate=0.001) and "keras" (Kalinowski et al., 2022).

Autoencoder regression was performed using 'h2o' (LeDell et al., 2022). The architecture consisted of a single hidden bottleneck layer (250), due to the comparatively low dimensionality of the data compared to typical autoencoder applications, and tanh activation function. The encodings stored within the deep features of the bottleneck layer were used to train a MLR with chronological age as the regressand.

Validation of bAge Models

Bootstrapped (n=200) confidence intervals for the R^2 and mean absolute error (MAE) were calculated for the fit on chronological age on the out-of-sample test data. This was performed separately for each model and assay.

Calculation of bAge

The best model was used to calculate the bAge – defined as the fitted value of chronological age regressed on the assay. By definition, this bAge is heavily correlated with chronological age. Therefore, in subsequent analyses, chronological age was included as a regressor to adjust for this dependency in a similar way to the original DNA methylation clock procedure (Horvath, 2013).

Risk factors of bAge

Risk factor analysis of bAge were adjusted for behavioural (BMI, smoking status, drinking status), demographic (chronological age, sex, ethnicity) factors. Firstly, relationships with early risk factors of cardiovascular disease were tested (total cholesterol, high-density lipoprotein (HDL), NT-proBNP, triglycerides (TG), C-reactive Protein (CRP), lipoprotein (a) i.e. Lp(a).) Secondly, relationships with late signs of disease were tested (hypertension, diabetes mellitus, osteoarthritis, osteoporosis, stroke, peripheral artery disease (PAD), deep vein thrombosis (DVT) and kidney disease).

The bAge was regressed on early risk factors and late signs of disease separately as the availability of lab data was diminished in comparison to that of disease status. Importantly, both analyses were adjusted for behavioural and demographic factors. The effect size or beta coefficients were then assessed for significance and magnitude with 95% confidence intervals.

Results

Benchmarking Multi-omics & Machine Learning Methods

Figure 3 shows that a clock trained on a combination of proteomics, metabolomics and lipidomics assays consistently performed better compared to lipidomics and metabolomics individually, however surprisingly performed similarly well with proteomics alone. Within most assays, elastic regression, autoencoder regression and random forest appear to be the most well-performing methods.

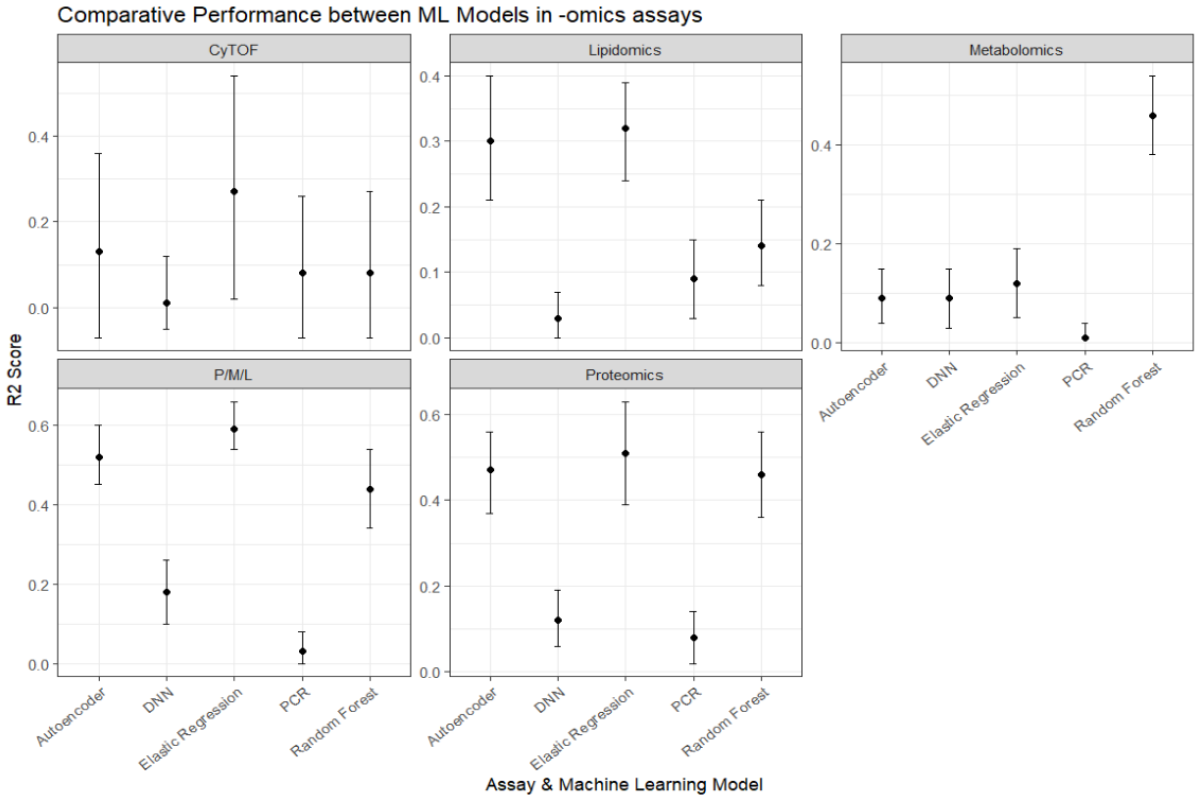

Figure 3 - Comparative performance between machine learning models (Autoencoder, DNN, Elastic Regression, PCR, Random Forest) in both individual assays (CyTOF, Lipidomics, Metabolomics, Proteomics) and also a combined assay P/M/L (Proteomics, Metabolomics, Lipidomics)

Table 1 shows the performance of the best model for chronological age of the combined assays. The current model performed worse from a R2 score 0.59 (CI: 0.54, 0.66) and MAE score 6.1 (CI: 5.59, 6.58) compared to all other models except iAge. It also had the smallest feature set.

	Current Model	DNAm (Hannum)	Proteomic (Lehallier)	Metabolomic (Robinson)	iAge (Sayed)
Model	Elastic Net	Elastic Net	Elastic Net	Elastic Net	Autoencoder
Training (N)	692	482	2817	2239	800
Test (N)	297	174	1446	2144	201
Feature Set (n)	360	485,577	2,925	28,941	50
Model Variables (n)	251	71	373	1311	50
R2	0.59	0.91	0.93	0.86	0.26*
MAE	6.1			3.71	15.2
RMSE	8.33		4.89		

*Table 1 - Comparison of the best biological age model in the current study compared to other notable biological clocks as well as their dataset, model and performance details. *termed average reconstruction errors in the original paper. Blocked cells mean un-reported values in original paper.*

Risk Factor & Observable Disease Analysis

Figure 4 shows non-smoking status (-2.08, CI: -3.48, -0.68), chronological age (0.64, CI: 0.61, 0.67) and Asian ethnicity (-1.96, CI: -3.41, -0.50) to be significant adjustment factors with respect to bAge ($p<0.05$). Diabetes mellitus (2.72, CI: 1.48-3.95), stroke (2.17, CI: 0.61-3.73), DVT (2.55, CI: 0.70-

4.40), and kidney disease (5.78, CI: 3.11-8.45), were late observable signs of disease that reflected in an increase in bAge ($p<0.05$).

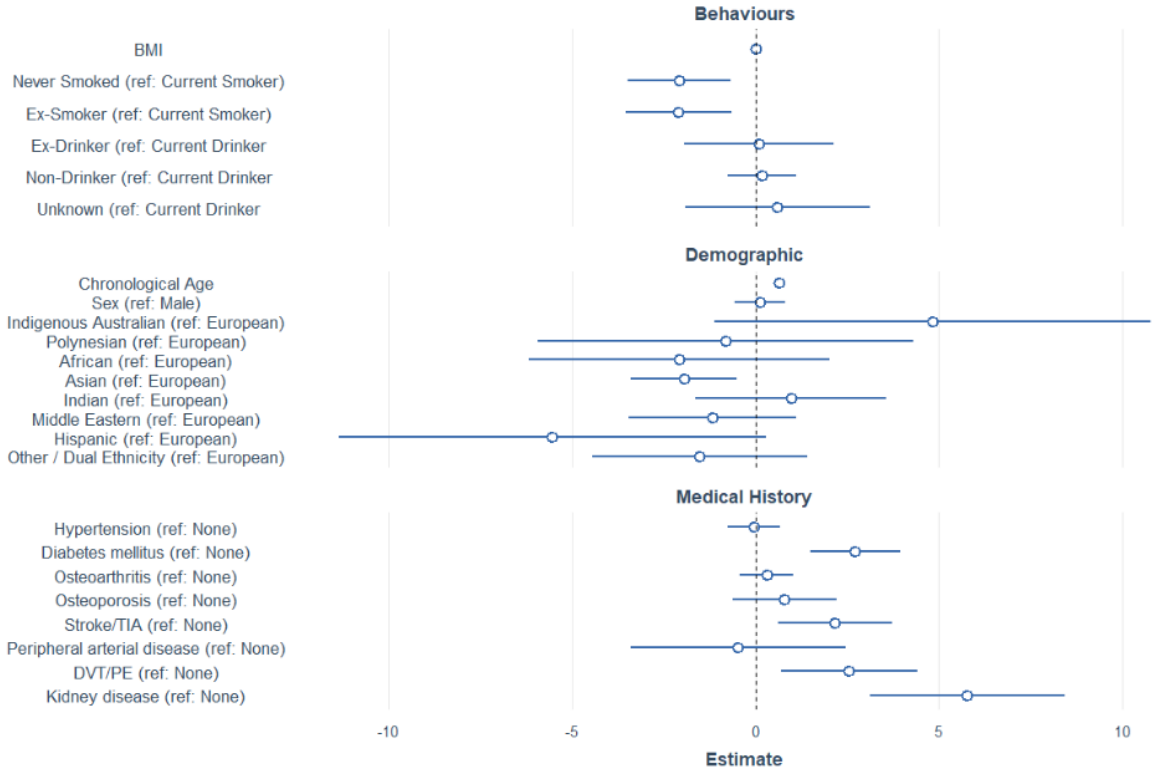

Figure 4 - Risk factors of increased bAge adjusted for behavioural and demographic factors. Estimate represents the effect size (95% C.I.)

High total cholesterol (-0.83, CI: -1.30, -0.35) was linked to decreased bAge. HDL (3.18, CI: 1.81, 4.55) and NT-proBNP (0.003, CI: 0.001, 0.004) were all linked to increased bAge. No significance effects were found for triglycerides, CRP and Lp(a).

Discussion

Single Assay versus Multi-assay

With reference to the first aim to compare multi-assay with single assays, *Figure 3* shows that a combination of proteomics, metabolomics and lipidomics fared better at predicting the chronological age compared to most other assays. Its superiority is unsurprising given the proven value of multi-omics integration over isolated platforms however what is of interest is the similar predictive performance of the proteome alone (Subramanian et al., 2020).

It's possible that proteins indeed contain the most informational content since they are the translational endpoint. This could explain why the proteome alone performs just as well in predicting chronological age as the combined assays. However, it's more likely that the unbalanced variable count of the proteomic assay containing 259 proteins, compared to the 53 metabolites, 48 lipid totals and 77 cell proportions, is 'washing out' the contributions of the other assays. Therefore, the combined assay model is functionally similar to the model built on proteome alone. Another possibility is that crucial metabolites and lipids which have explanatory power for the aging process, are simply missing from the small panel size available in this dataset. Indeed, many of the biomarkers

used in previous clocks are not available in this iteration of BioHEART-CT (Lehallier et al., 2019; Robinson et al., 2020).

Comparison of Machine Learning Methods

In line with the first aim to determine the best-performing machine learning method for combined assays, from *Figure 3*, it can be seen elastic regression consistently performs well. This is consistent with established benchmarks that have been conducted on single platform analysis (Acharjee, 2012). This is likely because the convex combination of both L1 and L2 regularization penalises excess covariates with minimal contribution to the response variable and encourages the final model towards parsimony. In comparison the worst performer was principal component regression which indicates that latent variables that give more weighting to predictors with greatest variance may not necessarily select predictors with the most explanatory relevance. This is particularly pertinent in multi-omics data which is typically characterised by uninformative between-sample and between-platform variance (Martorell-Marugán et al., 2019).

Contrary to initial belief, the 'vanilla' deep learning model did not show improved performance over regularised linear regression. It's known that DNNs need far more data compared to traditional ML modelling to avoid overfitting on noise (Zhang et al., 2018). Given that the characteristic noisiness of multi-omics dataset and the comparatively small dataset available as in *Table 1*, this could explain the poor performance. Interestingly, a different approach to deep learning was able to yield slightly better results – namely autoencoder regression. The deep-learning model in this instance was only used as a means of non-linear dimensional reduction via the bottleneck layer and was not involved in the inference step. The non-linear representation of the latent variables in autoencoder regression could explain the significant improvement over PCR which uses a linear combination of predictors in the principal components. However, the autoencoder's similar underlying reliance on identifying predictors of greatest variance likely explains why it still falls short of elastic regression.

Future Work on Multi-Omics Integration

Variable imbalance between platforms and model selection will always be a challenge in multi-omics integration. Future work could explore the use of established techniques such as sparse partial least squares (SPLS) as it works well on unbalanced samples sizes and datasets with high collinearity such as multi-omics. SPLS constructs linear combinations of the predictors i.e., latent variables, in a supervised manner with respect to the response variable whilst simultaneously applying regularisation penalties to reduce the erroneous inclusion of noisy but uninformative predictors (Chung & Keles, 2010). Importantly, this approach applies dimensional reduction but also considers the informativeness of the predictors it keeps, which may overcome the limitations of PCR and autoencoders. This could be implemented with 'mixomics' (Rohart et al., 2017).

Comparison to Existing Clocks

As in *Table 1*, the combined assay elastic regression model in this study performed worse than other omics models in literature in its ability to be tuned to the chronological age. This could be because the present dataset has far less samples diminishing the model's out-of-sample predictive power. It's also worth noting that the feature set available in the other datasets were much larger and closer to an un-targeted analysis (Lehallier et al., 2019; Robinson et al., 2020; Sayed et al., 2021). It's possible that features which would have been major contributors to the biological age prediction were not present in the current dataset.

Risk Factor & Observable Disease Analysis

For adjustment factors, *Figure 4* shows non-smoking status is associated with reduced bAge – a well-established risk factor for all-cause mortality (Chang et al., 2015). Unsurprisingly, chronological age is a significant adjustment factor where older individuals have increased bAge a priori. Interestingly, Asian ethnicity was the only race to be significantly associated with decreased bAge. It's possible that this is due to the lack of adjustment for lifestyle factors such as diet.

Assessing early observable risk factors as per the second aim, high total cholesterol (TC) was associated with decreased age although the interpretation is not all that useful given TC is an amalgamation of many categories of cholesterol of varying density and function (Birtcher & Ballantyne, 2004). What's ostensibly surprising is that increased HDL is associated with increased bAge. HDL has been traditionally perceived as 'good cholesterol' however more recent work has highlighted the deceptive nature of fixating on HDL serum concentration. Importantly, not all HDL particles are 'equal' and between particles, there are functional differences in the cholesterol molecules held within – becoming pro-atherogenic during onset of plaque development (Xu et al., 2013). Given plaque development is heavily tied to chronological age, HDL function or dysfunction evolves with age, a factor that confounds the interpretation here. NT-proBNP is associated with increased bAge which is expected given it is the gold standard for long-term independent prediction of mortality due to heart failure (Richards & Troughton, 2004).

Despite well-established links between triglycerides, CRP and Lp(a) to cardiovascular risk, the negative result likely reflects the nature of biological clock which is trained on chronological age rather than time-to-event for cardiovascular mortality; the component of aging measured here likely doesn't overlap perfectly with mortality outcome. It's been shown that clocks trained on longitudinal outcomes rather than age such as DunedinPoAm capture a different component of the aging process (Rutledge et al., 2022).

Assessing late observable signs of disease as per the second aim, diabetes mellitus, stroke, DVT and kidney disease were all associated with increased bAge. Whilst they are 'explainable' due to their association with increased all-cause mortality risk, it's worth noting the inherent selection bias present within BioHEART-CT where the inclusion criteria are patients under investigation of suspected coronary artery disease. Therefore, the bAge model here could be selecting biomarkers very specific to cardiovascular aging; this may explain why observable diseases linked to bAge in this study are also causally linked to CVD (Goldhaber & Bounameaux, 2012; Herzog et al., 2011). Importantly, this bAge may not reflect the overall aging process of the body which could limit the scope of its utility in assessing the aging status of other organ systems.

Future Work for bAge & Clinical Associations

Given the selection bias for suspected CAD within BioHEART-CT a worthwhile endeavour would be testing existing LipidClock, iAge and metabolomic clocks on patients in this study (Sayed et al., 2021; Unfried et al., 2022). This current study is limited to targeted data which is missing many of the predictors dictated by the aforementioned clocks however BioHEART-CT is an ongoing project and untargeted data will be made available. Correlation analysis between the biological ages generated by these other clocks and the current study's bAge on the same subjects will likely reveal a weak correlation due to this unique cohort. This would indicate that bAge is capturing a distinct element of the aging process and may be useful in identifying pertinent cardiovascular disease-specific phenotypes that are missed by existing clocks (Rutledge et al., 2022).

Conclusion

In sum the first aim was to determine what machine learning method was best suited to integrating multi-assay data to unlock its predictive advantage of biological age over single assays. Elastic regression was found to produce an effective, parsimonious model and deep-learning approaches were unsuccessful likely due to noisiness of omics data. The second aim was to determine the early risk factors which contribute to increased bAge and validate its link to observable signs of disease. Smoking, high NT-proBNP and interestingly, high HDL were associated with increased bAge. Increased bAge was also linked to history of diabetes mellitus, stroke, DVT and kidney disease. Future work should explore machine learning methods which simultaneously perform dimensional reduction and variable selection such as SPLS and test existing biological clocks on BioHEART-CT to assess whether bAge is capturing a distinct aspect of the aging process.

References

Acharjee, A. (2012). Comparison of Regularized Regression Methods for ~Omics Data. *Journal of Postgenomics Drug & Biomarker Development*, *03*(03). https://doi.org/10.4172/2153-0769.1000126

Birtcher, K. K., & Ballantyne, C. M. (2004). Measurement of Cholesterol. *Circulation*, *110*(11). https://doi.org/10.1161/01.cir.0000141564.89465.4e

Chang, C. M., Corey, C. G., Rostron, B. L., & Apelberg, B. J. (2015). Systematic review of cigar smoking and all cause and smoking related mortality. *BMC Public Health*, *15*(1). https://doi.org/10.1186/s12889-015-1617-5

Chung, D., & Keles, S. (2010). Sparse Partial Least Squares Classification for High Dimensional Data. *Statistical Applications in Genetics and Molecular Biology*, *9*(1). https://doi.org/10.2202/1544-6115.1492

Cutler, F. original by L. B. and A., & Wiener, R. port by A. L. and M. (2022, May 23). *randomForest: Breiman and Cutler's Random Forests for Classification and Regression*. R-Packages. https://cran.r-project.org/web/packages/randomForest/index.html

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., & Qian, J. (2021, June 24). *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models*. R-Packages. https://cran.r-project.org/web/packages/glmnet/index.html

Goldhaber, S. Z., & Bounameaux, H. (2012). Pulmonary embolism and deep vein thrombosis. *The Lancet*, *379*(9828), 1835–1846. https://doi.org/10.1016/s0140-6736(11)61904-1

Herzog, C. A., Asinger, R. W., Berger, A. K., Charytan, D. M., Díez, J., Hart, R. G., Eckardt, K.-U., Kasiske, B. L., McCullough, P. A., Passman, R. S., DeLoach, S. S., Pun, P. H., & Ritz, E. (2011). Cardiovascular disease in chronic kidney disease. A clinical update from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney International*, *80*(6), 572–586. https://doi.org/10.1038/ki.2011.223

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115

Kalinowski, T., Falbel, D., Allaire, J. J., Chollet, F., RStudio, Google, Tang [ctb, Y., cph, Bijl, W. V. D., Studer, M., & Keydana, S. (2022, May 23). *keras: R Interface to "Keras."* R-Packages. https://cran.r-project.org/web/packages/keras/index.html

Kim, T., Tang, O., Vernon, S. T., Kott, K. A., Koay, Y. C., Park, J., James, D. E., Grieve, S. M., Speed, T. P., Yang, P., Figtree, G. A., O'Sullivan, J. F., & Yang, J. Y. H. (2021). A hierarchical approach to removal of unwanted variation for large-scale metabolomics data. *Nature Communications*, *12*(1), 4992. https://doi.org/10.1038/s41467-021-25210-5

Kott, K. A., Vernon, S. T., Hansen, T., Yu, C., Bubb, K. J., Coffey, S., Sullivan, D., Yang, J., O'Sullivan, J., Chow, C., Patel, S., Chong, J., Celermajer, D. S., Kritharides, L., Grieve, S. M., & Figtree, G. A. (2019). Biobanking for discovery of novel cardiovascular biomarkers using imaging-quantified disease burden: protocol for the longitudinal, prospective, BioHEART-CT cohort study. *BMJ Open*, *9*(9), e028649. https://doi.org/10.1136/bmjopen-2018-028649

Kuhn, M., cre, Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2022, August 9). *caret: Classification and Regression Training*. R-Packages. https://cran.r-project.org/web/packages/caret/

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Aboyoun, P., Kurka, M., Malohlava, M., Rehak, L., Eckstrand, E., Hill, B., Vidrio, S., Jadhawani, S., Wang, A., Peck, R., Wong, W., & Gorecki, J. (2022, September 23). *h2o: R Interface for the "H2O" Scalable Machine Learning Platform*. R-Packages. https://cran.r-project.org/web/packages/h2o/

Lehallier, B., Gate, D., Schaum, N., Nanasi, T., Lee, S. E., Yousef, H., Losada, P. M., Berdnik, D., Keller, A., Verghese, J., Sathyan, S., Franceschi, C., Milman, S., Barzilai, N., & Wyss-Coray, T. (2019). Undulating changes in human plasma proteome profiles across the lifespan. *Nature Medicine*, *25*(12), 1843–1850. https://doi.org/10.1038/s41591-019-0673-2

Martorell-Marugán, J., Tabik, S., Benhammou, Y., del Val, C., Zwir, I., Herrera, F., & Carmona-Sáez, P. (2019). *Deep Learning in Omics Data Analysis and Precision Medicine* (H. Husi, Ed.). PubMed; Codon Publications. https://www.ncbi.nlm.nih.gov/books/NBK550335/

Richards, M., & Troughton, R. W. (2004). NT-proBNP in heart failure: therapy decisions and monitoring. *European Journal of Heart Failure*, *6*(3), 351–354. https://doi.org/10.1016/j.ejheart.2004.01.003

Robinson, O., Chadeau Hyam, M., Karaman, I., Climaco Pinto, R., Ala-Korpela, M., Handakas, E., Fiorito, G., Gao, H., Heard, A., Jarvelin, M., Lewis, M., Pazoki, R., Polidoro, S., Tzoulaki, I., Wielscher, M., Elliott, P., & Vineis, P. (2020). Determinants of accelerated metabolomic and epigenetic aging in a UK cohort. *Aging Cell*, *19*(6). https://doi.org/10.1111/acel.13149

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.-A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Rutledge, J., Oh, H., & Wyss-Coray, T. (2022). Measuring biological age using omics data. *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-022-00511-7

Sayed, N., Huang, Y., Nguyen, K., Krejciova-Rajaniemi, Z., Grawe, A. P., Gao, T., Tibshirani, R., Hastie, T., Alpert, A., Cui, L., Kuznetsova, T., Rosenberg-Hasson, Y., Ostan, R., Monti, D., Lehallier, B., Shen-Orr, S. S., Maecker, H. T., Dekker, C. L., Wyss-Coray, T., & Franceschi, C. (2021). An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. *Nature Aging*, *1*(7), 598–615. https://doi.org/10.1038/s43587-021-00082-y

Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights*, *14*. https://doi.org/10.1177/1177932219899051

Unfried, M., Ng, L. F., Cazenave-Gassiot, A., Batchu, K. C., Kennedy, B. K., Wenk, M. R., Tolwinski, N., & Gruber, J. (2022). LipidClock: A Lipid-Based Predictor of Biological Age. *Frontiers in Aging*, *3*, 828239. https://doi.org/10.3389/fragi.2022.828239

Xu, S., Liu, Z., & Liu, P. (2013). HDL cholesterol in cardiovascular diseases: The good, the bad, and the ugly? *International Journal of Cardiology*, *168*(4), 3157–3159. https://doi.org/10.1016/j.ijcard.2013.07.210

Zhang, Z., Zhao, Y., Liao, X., Shi, W., Li, K., Zou, Q., & Peng, S. (2018). Deep learning in omics: a survey and guideline. *Briefings in Functional Genomics*, *18*(1), 41–57. https://doi.org/10.1093/bfgp/ely030