

МАТЕМАТИЧЕСКАЯ БИОЛОГИЯ, БИОИНФОРМАТИКА/MATHEMATICAL BIOLOGY, BIOINFORMATICS

DOI: https://doi.org/10.60797/jbg.2026.33.2 EDN: LIZWIJ

METABOLOMIC MODELING OF BIOLOGICAL AGE AND AGE ACCELERATION USING REGULARIZED MACHINE LEARNING

Research article

Muhametshina R.D.[1, *], Muhametshina A.D.[2]
[1, 2] Lomonosov Moscow State University, Moscow, Russian Federation

* Corresponding author (1709mrd[at]gmail.com)

Suggested: 13.02.2026; Accepted: 07.07.2026; Published: 25.09.2026

Abstract

Biological aging is a complex and highly individual process that cannot be fully explained by chronological age alone. Plasma metabolomic profiles reflect the current physiological state of the body and may serve as informative biomarkers of biological aging. In this study, we developed and thoroughly evaluated a metabolomic age prediction model using regularized machine learning approaches. The analysis included 296 individuals with quantified plasma metabolites. After standard preprocessing and data transformation, chronological age was predicted using Elastic Net regression. Model performance was assessed with repeated nested cross-validation, which helped minimize overfitting and prevent information leakage between training and validation datasets. The primary Elastic Net model achieved a mean absolute error (MAE) of approximately 6 years and outperformed baseline models, including dummy regression and standard linear regression. To test the robustness of the detected metabolomic aging signal, we additionally compared the model with ridge regression and alternative feature selection approaches based on Random Forest feature importance and recursive feature elimination (RFE). Predictive performance remained generally stable across different regularized approaches, suggesting that aging-related metabolomic information is distributed across multiple correlated metabolic features rather than driven by only a few variables. Metabolic age acceleration was defined as the difference between predicted and chronological age and was further analyzed in relation to biological heterogeneity and menopausal status. In addition, conformal prediction was applied to estimate individualized prediction uncertainty and generate prediction intervals with controlled empirical coverage. Overall, the results show that plasma metabolomic profiles contain reproducible and biologically meaningful information associated with aging. By combining rigorous validation procedures, comparative regularized modeling, alternative feature selection methods, and uncertainty-aware prediction, this study provides a reliable framework for investigating metabolic aging and inter-individual variability.

Keywords: Metabolomics, Biological age, Aging, Machine Learning, Elastic Net, Ridge regression, Feature selection, Age acceleration, Metabolomic clock, Regularized regression, Plasma metabolites, Cross-validation.

МОДЕЛИРОВАНИЕ БИОЛОГИЧЕСКОГО ВОЗРАСТА НА ОСНОВЕ МЕТАБОЛОМНЫХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ РЕГУЛЯРИЗОВАННЫХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Научная статья

Мухаметшина Р.Д.[1, *], Мухаметшина А.Д.[2]
[1, 2] Московский государственный университет, Москва, Российская Федерация

* Корреспондирующий автор (1709mrd[at]gmail.com)

Предложена: 13.02.2026; Принята: 07.07.2026; Опубликована: 25.09.2026

Аннотация

Биологическое старение — это сложный процесс, который нельзя полностью описать только с помощью хронологического возраста. Метаболомные профили плазмы отражают текущее физиологическое состояние организма и могут использоваться как информативные маркеры биологического старения. В данном исследовании была разработана и оценена модель предсказания возраста на основе метаболомных данных с использованием методов регуляризованного машинного обучения. В анализ были включены данные 296 участников с количественно определёнными плазменными метаболитами. После стандартной предобработки и трансформации данных хронологический возраст прогнозировался с помощью регрессии Elastic Net. Для оценки качества модели использовалась повторяющаяся вложенная кросс-валидация, что позволило минимизировать риск переобучения и утечки информации. Основная модель Elastic Net показала среднюю абсолютную ошибку (MAE) около 6 лет и значительно превзошла базовые модели, включая dummy-регрессию и обычную линейную регрессию. Для проверки устойчивости выявленного метаболомного сигнала старения дополнительно проводилось сравнение с ridge-регрессией, а также с альтернативными методами отбора признаков на основе важности признаков Random Forest и рекурсивного исключения признаков (RFE). Результаты оказались достаточно стабильными для разных регуляризованных подходов, что указывает на то, что возрастные метаболические изменения распределены между множеством взаимосвязанных метаболитов. Метаболическое ускорение возраста определялось как разница между предсказанным и реальным возрастом и дополнительно анализировалось с учётом биологической неоднородности и менопаузального статуса. Для оценки неопределённости индивидуальных предсказаний использовался подход conformal prediction, позволяющий строить интервалы предсказания с контролируемым уровнем покрытия.

Полученные результаты показывают, что плазменные метаболомные профили содержат воспроизводимую и биологически значимую информацию, связанную со старением. Использование строгих процедур валидации, сравнительных методов регуляризованного моделирования, различных подходов к отбору признаков и оценки неопределённости делает предложенную модель надёжным инструментом для изучения метаболического старения и межиндивидуальных различий в процессах старения.

Ключевые слова: метаболомика, биологический возраст, старение, машинное обучение, Elastic Net, Ridge-регрессия, отбор признаков, ускорение старения, метаболомные часы, регуляризованная регрессия, плазменные метаболиты, кросс-валидация.

Introduction

Chronological age is a convenient but incomplete proxy for biological aging, as it does not fully capture substantial inter-individual differences in physiological state, functional reserve, and susceptibility to age-related disease [1], [17]. Individuals of identical chronological age may follow markedly different aging trajectories, motivating the development of biological age as a more integrative measure of organismal function and aging rate [2], [19]. In recent years, molecular aging clocks derived from epigenomic, transcriptomic, and metabolomic data have emerged as powerful tools for quantifying biological aging [3], [4], [5], [17]. Among these approaches, metabolomics occupies a distinct position because metabolites represent downstream products of cellular activity and reflect the combined influence of genetic background, environmental exposures, diet, medication use, and ongoing physiological processes [6], [20]. As a result, metabolomic profiles provide a dynamic representation of organismal state and may be particularly sensitive to functional aspects of aging and inter-individual heterogeneity. Large-scale population studies have consistently demonstrated age-associated alterations in circulating lipids, amino acids, sphingolipids, and related metabolites, many of which are linked to cardiometabolic risk, chronic inflammation, and age-related disease [7], [8], [9]. Because many metabolic alterations associated with aging are also observed in chronic diseases and metabolic disorders, distinguishing physiological aging from disease-related metabolic changes remains an important challenge for biological age modeling. Consequently, developing aging clocks using well-characterized healthy populations provides an essential reference for future applications in clinical cohorts. Several metabolomic age prediction models have reported moderate to high predictive accuracy [10], [11], [12], [18]. However, many previous studies have primarily focused on overall predictive performance, with less emphasis placed on robustness across modeling strategies, feature selection approaches, and uncertainty of individual predictions. In high-dimensional metabolomic datasets, where extensive multicollinearity between metabolites is common, regularization and feature selection become critical methodological considerations. Elastic Net regression combines L1 and L2 penalties, allowing simultaneous coefficient shrinkage and implicit feature selection while reducing overfitting.

Compared with epigenetic and transcriptomic clocks, metabolomic models offer complementary strengths. DNA methylation-based clocks often achieve high precision and long-term temporal stability [3], [4], whereas transcriptomic models capture dynamic gene expression patterns but may be sensitive to transient environmental effects [5]. Metabolomic profiles, in contrast, reflect integrated downstream physiology and may therefore provide functionally relevant information about current metabolic state and biological heterogeneity [6], [10], [11], [12]. Although this sensitivity may introduce greater short-term variability, it may also increase responsiveness to physiological transitions, environmental exposures, and lifestyle-related changes. In the present study, we developed a plasma metabolomics-based age prediction model using Elastic Net regression and evaluated its performance within a computationally intensive repeated nested cross-validation framework. To assess the robustness of the identified aging signal, we additionally compared alternative regularized and feature selection approaches, including ridge regression, Random Forest-based feature importance selection, and recursive feature elimination (RFE). Beyond predictive accuracy, we quantified metabolic age acceleration as a phenotype of aging heterogeneity and incorporated conformal prediction to estimate individualized prediction uncertainty. Together, this framework positions metabolomic clocks not only as age predictors, but also as quantitative tools for investigating metabolic remodeling and inter-individual variability during aging.

Research methods and principles

2.1. Study Cohort and Metabolomic Profiling

The study included 296 adult participants with available plasma metabolomic profiles and recorded chronological age. Data were obtained from the publicly available KarMeN (Karlsruhe Metabolomics and Nutrition) study, a comprehensive population-based metabolomics project designed to investigate the influence of sex, age, body composition, nutrition, and lifestyle factors on human metabolic profiles. The KarMeN cohort included healthy adult men and women aged 18 years and older and incorporated extensive phenotypic, anthropometric, clinical, and metabolomic characterization. Plasma metabolite concentrations were quantified using standardized high-throughput metabolomics platforms, including mass spectrometry–based analytical methods and nuclear magnetic resonance spectroscopy, ensuring reproducible and comparable measurements across samples. The dataset comprised several hundred metabolites, including lipids, lipoproteins, amino acids, phospholipids, sphingolipids, and other low-molecular-weight compounds reflecting systemic metabolic processes and physiological state [15]. The use of a healthy reference cohort allowed the model to characterize physiological metabolic aging while minimizing confounding effects introduced by chronic diseases and metabolic disorders.

2.2. Data Preprocessing

Metabolomic data preprocessing was performed within a unified machine learning pipeline implemented in Python using the scikit-learn framework. Prior to model development, metabolite concentrations were transformed using the Yeo-Johnson power transformation to reduce skewness and stabilize variance across features. Subsequently, robust scaling based on the interquartile range was applied to reduce sensitivity to outliers and non-normal feature distributions. Missing values were handled using median imputation performed independently within each training partition. To prevent information leakage, all

preprocessing operations, including imputation, transformation, and scaling, were recalculated separately within each cross-validation fold and applied only to the corresponding validation subset. No explicit outlier removal was performed, as regularized regression models are relatively robust to moderate deviations in feature distributions.

2.3. Age Prediction Model

Chronological age prediction was primarily performed using Elastic Net regression, which combines L1 (LASSO) and L2 (ridge) regularization. This approach enables simultaneous coefficient shrinkage and implicit feature selection, helping to control overfitting in high-dimensional metabolomic datasets characterized by extensive multicollinearity between metabolites. To evaluate robustness of the identified metabolomic aging signal, additional comparative models were implemented within the same analytical framework. These included ridge regression, standard linear regression, and two alternative feature selection strategies combined with Elastic Net regression. The first approach used Random Forest–based feature importance selection, in which top-ranked metabolites were selected according to feature importance scores derived from ensemble tree models. The second approach applied recursive feature elimination (RFE), which iteratively removed less informative variables based on model coefficients. A dummy regressor predicting mean chronological age was additionally included as a baseline reference model.

2.4. Cross-Validation and Model Evaluation

Predictive performance was assessed using repeated nested cross-validation. The inner loop was used for hyperparameter tuning, while the outer loop generated out-of-fold predictions for each participant. This procedure ensured that each predicted age was obtained from a model that had not been trained on that individual, providing an unbiased estimate of generalization performance. The outer cross-validation consisted of 5 folds repeated 10 times (total of 50 outer splits) to improve stability of performance estimates. Hyperparameter optimization for Elastic Net and ridge regression models was performed using randomized search with 50 parameter configurations sampled from predefined distributions. For Elastic Net, the regularization strength parameter was sampled from a log-uniform distribution ranging from 10^{-4} to 10, while the mixing parameter controlling the balance between L1 and L2 penalties was sampled from a uniform distribution between 0.05 and 0.95. Out-of-fold predictions were generated for all participants such that each predicted value was obtained from a model that had not been trained on the corresponding individual. This framework provides an approximately unbiased estimate of predictive performance and substantially reduces the risk of information leakage and overfitting. Model performance was evaluated using mean absolute error (MAE) as the primary metric and root mean squared error (RMSE) as a complementary measure. Variability across validation folds was additionally quantified to assess stability of model generalization.

2.5. Definition of Age Acceleration

Metabolic age acceleration was defined as the residual difference between predicted age and chronological age at the individual level:

$$\text{Age Acceleration}_{\,i} = \hat{Age}_i - Age_i$$

where $\hat{Age}_i$ denotes the predicted age for individual i, and $\text{Age}_{\,i}$ represents chronological age.

Positive values indicate accelerated metabolic aging, whereas negative values indicate a relatively younger metabolic profile compared to chronological age.

2.6. Prediction Uncertainty

Individual prediction uncertainty was estimated using a jackknife+ conformal prediction framework [13], [14]. Nonconformity scores were derived from absolute residuals obtained from out-of-fold predictions. Prediction intervals were constructed symmetrically around point estimates at a predefined significance level. Empirical coverage was calculated as the proportion of individuals for whom chronological age fell within the corresponding prediction interval, allowing assessment of interval calibration.

2.7. Feature-Level Statistical Analyses

To improve biological interpretability of the model, Spearman correlation coefficients were calculated between metabolite concentrations and chronological age. Multiple hypothesis testing was controlled using the Benjamini-Hochberg procedure with a false discovery rate (FDR) threshold of 5%. In addition, associations between metabolite concentrations and absolute prediction error were evaluated to identify metabolites linked not only to chronological aging, but also to prediction stability and model reliability. These analyses allowed differentiation between metabolites strongly associated with aging itself and those contributing to robust predictive performance within multivariate models.

Main results

3.1. Prediction of Chronological Age

The primary Elastic Net model trained on plasma metabolomic profiles demonstrated stable and reproducible performance in predicting chronological age under a computationally intensive repeated nested cross-validation framework. The validation procedure consisted of 5-fold outer cross-validation repeated 10 times, resulting in 50 independent outer validation splits. Out-of-fold predictions were generated for all individuals, ensuring that each predicted age was obtained from a model that had not been trained on the corresponding sample. The relationship between predicted and chronological age is presented in Figure 1. Overall, predicted values closely followed the line of identity, indicating strong agreement across the studied age range. The majority of observations clustered near the diagonal, suggesting that the model captured a substantial proportion of age-related metabolic variation. However, a gradual increase in dispersion was observed at higher chronological ages, likely reflecting greater inter-individual metabolic heterogeneity in older participants.

Across repeated validation runs, the Elastic Net model achieved a mean absolute error (MAE) of approximately 6.04 ± 0.52 years, while the root mean squared error (RMSE) was 7.28 years (Table 1). The relatively small standard deviation of MAE across validation splits indicates stable generalization performance and robustness to sampling variability. Mean prediction bias remained close to zero (−0.29 years), suggesting absence of systematic over- or underestimation. These

findings demonstrate that plasma metabolomic profiles contain sufficient information to estimate chronological age with moderate-to-high accuracy in a heterogeneous adult population.

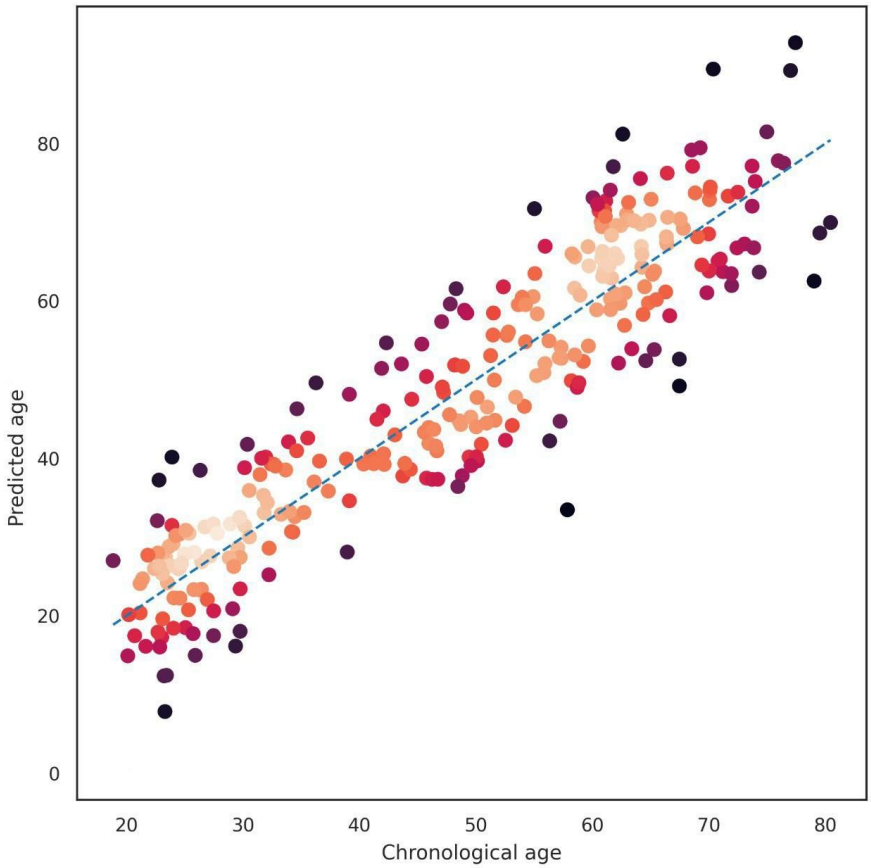

Figure 1 - Predicted and chronological age
DOI: https://doi.org/10.60797/jbg.2026.33.2.1

Note: relationship between predicted and chronological age for out-of-fold predictions; most points lie along the line of equality, indicating a good match between predicted and true age across the sample; however, the spread of points increases with chronological age, particularly in older age groups; this is consistent with the previously shown increase in prediction error in older age and reflects increased metabolic variability in older participants

Table 1 - Overall performance of the age prediction model

DOI: https://doi.org/10.60797/jbg.2026.33.2.2

Metric	Value (years)
Mean Absolute Error (MAE)	6.04 ± 0.52
Root Mean Squared Error (RMSE)	7.28
Mean prediction bias	−0.29

3.2. Prediction Error Across the Age Range

Prediction errors were further analyzed as a function of chronological age (Figure 2). Errors were symmetrically distributed around zero, indicating good overall calibration of the model. The near-zero mean error additionally supports absence of systematic prediction bias. Nevertheless, variability of prediction errors increased with age. Younger individuals exhibited relatively compact error distributions, whereas older age groups demonstrated broader dispersion. Stratified analysis by age quartiles confirmed this pattern, with the lowest MAE observed in middle-aged individuals and substantially larger prediction errors detected in the oldest participants (Table 2).

This age-dependent increase in prediction variability likely reflects increasing biological heterogeneity during aging. Older individuals may differ substantially in metabolic state owing to cumulative environmental exposures, lifestyle differences, physiological adaptation, and varying rates of biological aging. Although the present cohort consisted of healthy participants, inter-individual metabolic variability naturally increases with age. Consequently, chronological age becomes a progressively less precise proxy for metabolic state in later life.

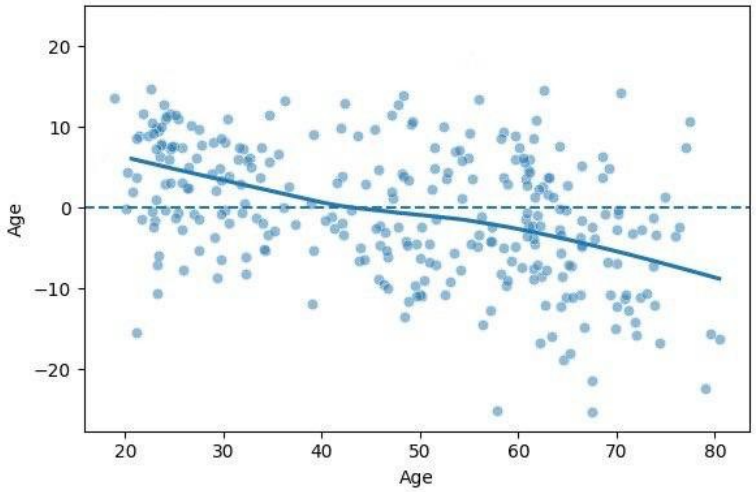

Figure 2 - Prediction error depending on age
DOI: https://doi.org/10.60797/jbg.2026.33.2.3

Note: dependence of prediction error on chronological age; the errors are distributed relative to the zero line without a pronounced systematic bias; the average error was −0.29 years, indicating no global bias in the model (Table 1)

Table 2 - Mean absolute error stratified by age quartiles

DOI: https://doi.org/10.60797/jbg.2026.33.2.4

Age quartile (years)	MAE (years)
18.9–30.6	4.98
30.6–49.6	4.66
49.6–62.0	6.47
62.0–80.4	8.19

3.3. Comparative Modeling and Alternative Feature Selection

To evaluate robustness of the identified metabolomic aging signal, additional regularized and feature selection approaches were assessed using the same preprocessing pipeline and a reduced validation framework (Table 3).

Ridge regression demonstrated predictive performance comparable to Elastic Net (MAE: 6.62), indicating that the aging-related metabolomic signal remained stable across different regularization strategies. Models incorporating Random Forest–based feature importance selection and recursive feature elimination (RFE) showed slightly reduced predictive accuracy (MAE: 7.00 and 7.47 years, respectively), although they continued to substantially outperform standard linear regression and the dummy baseline model (Table 3).

The reduced performance observed after explicit feature selection suggests that aging-related metabolomic information is distributed across multiple correlated metabolic variables rather than concentrated within a small subset of dominant metabolites. These findings support robustness of the proposed framework and indicate that the predictive signal captured by regularized models remains reproducible across alternative feature selection approaches.

Table 3 - Comparison with baseline models

DOI: https://doi.org/10.60797/jbg.2026.33.2.5

Model	MAE (years)
Dummy regressor	14.90
Linear regression	10.42
ElasticNet_RFE	7.47
ElasticNet_RFSelect	7.00
Ridge	6.62
Elastic Net	6.04

3.4. Age-Associated Metabolites

Spearman correlation analysis identified numerous metabolites significantly associated with chronological age after false discovery rate correction (Figure 3). The strongest positive correlations were observed for lipid- and lipoprotein-related measures, including LDL-associated fractions, sphingomyelins, and choline-containing phospholipids.

In contrast, several amino acids and lysophospholipids demonstrated negative correlations with age. This bidirectional pattern suggests that metabolic aging is characterized by coordinated remodeling across multiple biochemical pathways rather than uniform global increases or decreases in metabolite concentrations.

These observations are consistent with previous reports describing age-related alterations in lipid metabolism, lipoprotein composition, and amino acid homeostasis. Importantly, metabolites strongly associated with age were not necessarily those contributing most strongly to prediction stability, highlighting the distinction between biological association and predictive utility in multivariate models.

3.5. Metabolites Associated with Prediction Stability

To explore determinants of model robustness, we analyzed correlations between metabolite levels and absolute prediction error (Figure 4). Several metabolites, including carnitine, serine, and selected lipid fractions, were negatively correlated with absolute error, indicating that higher levels were associated with more accurate predictions.

This suggests that certain metabolic states are more “age-informative” and yield more stable age estimates. Conversely, other metabolic profiles may be influenced by factors unrelated to chronological aging, leading to increased prediction variability.

Notably, metabolites strongly correlated with age were not always those most strongly associated with prediction stability. This distinction highlights the difference between biological association and predictive utility in multivariate modeling.

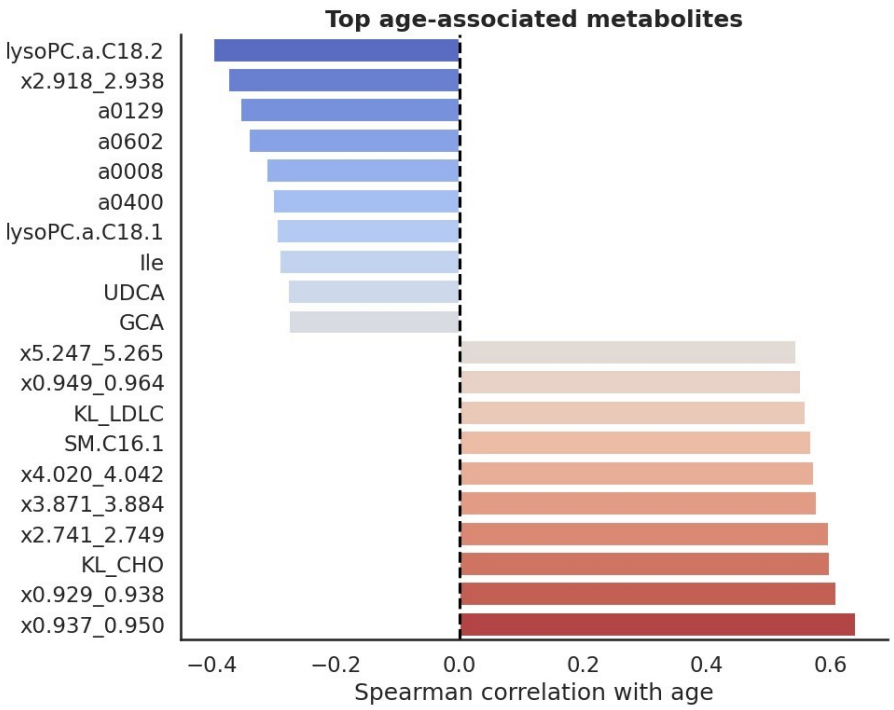

Figure 3 - Age-associated metabolites
DOI: https://doi.org/10.60797/jbg.2026.33.2.6

Note: spearman correlations between plasma metabolites and chronological age after FDR correction, highlighting lipid- and lipoprotein-related features as the strongest age-associated metabolites

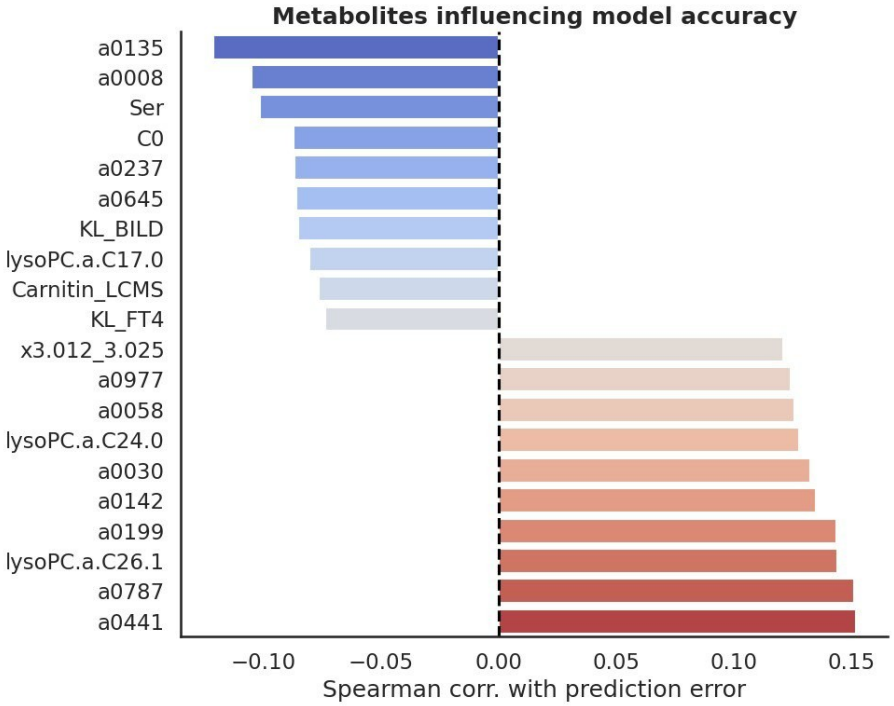

Figure 4 - Metabolites associated with prediction accuracy
DOI: https://doi.org/10.60797/jbg.2026.33.2.7

Note: spearman correlations between plasma metabolite levels and absolute age prediction error, identifying metabolites associated with increased robustness and reduced prediction uncertainty

3.6. Metabolic Age Acceleration

Metabolic age acceleration (predicted age minus chronological age) was centered around zero (mean −0.29 years; median −0.66 years; SD 8.37 years), as shown in Figure 5. Values ranged from −25.26 to +22.69 years, with 50% of participants falling between −5.99 and +6.12 years.

The distribution was approximately symmetric but exhibited substantial inter-individual variability. This variability increased with chronological age, mirroring the pattern observed in prediction error analysis and reinforcing the concept of increasing biological heterogeneity in later life.

Importantly, age acceleration should not be interpreted solely as model noise. While part of the variability reflects residual prediction error, systematic shifts may represent meaningful deviations from population-average metabolic aging trajectories.

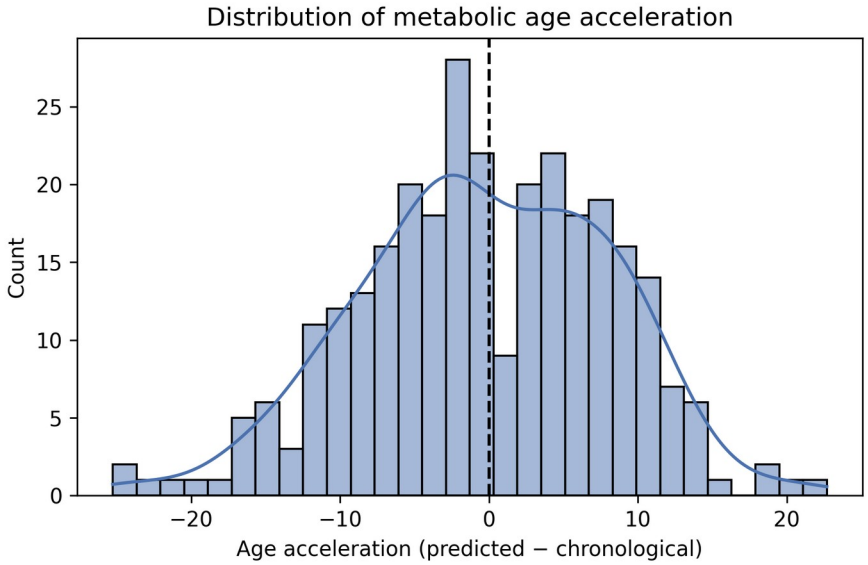

Figure 5 - Distribution of age acceleration
DOI: https://doi.org/10.60797/jbg.2026.33.2.8

Note: distribution of age acceleration values (predicted age minus chronological age) across the study population, showing a near-zero mean and substantial inter-individual variability

3.7. Sex Differences and Menopause

Age acceleration distributions were broadly similar between men and women (Figure 6), with no substantial difference in mean values. The overall overlap suggests that sex alone does not strongly drive metabolic age deviation in this cohort.

However, in a female-only subanalysis (n = 128), menopausal status was independently associated with higher metabolic age acceleration after adjustment for chronological age. Postmenopausal women exhibited a noticeable shift toward higher acceleration values (Figure 7).

This finding is consistent with known metabolic and hormonal transitions accompanying menopause and suggests that metabolomic age may capture biologically meaningful physiological shifts beyond chronological aging alone.

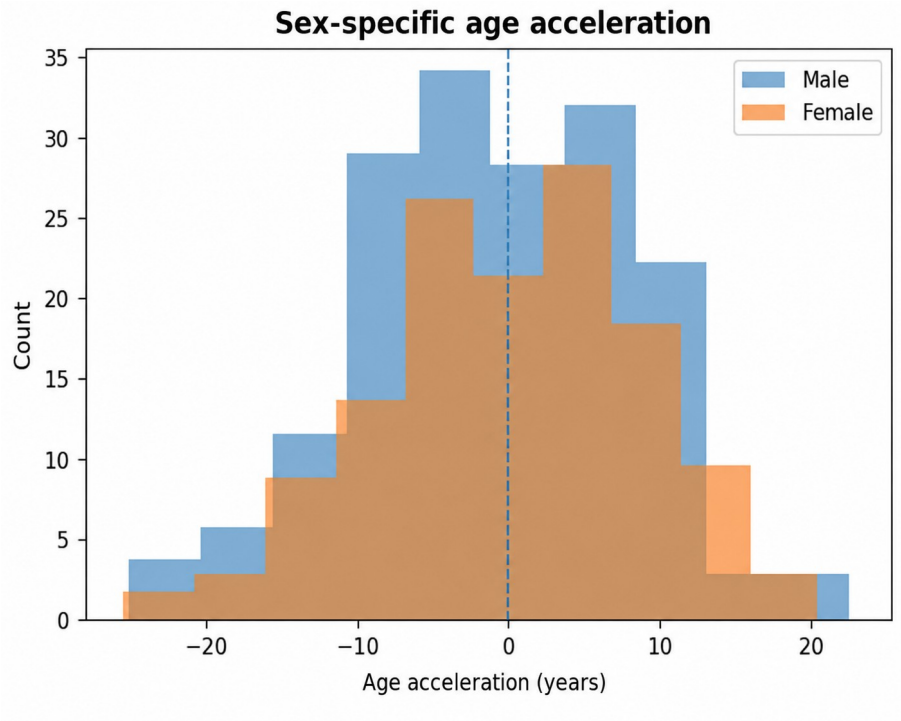

Figure 6 - Sex-specific distribution of age acceleration
DOI: https://doi.org/10.60797/jbg.2026.33.2.9

Note: distribution of age acceleration stratified by sex, showing broadly overlapping distributions with subtle differences in distribution shape

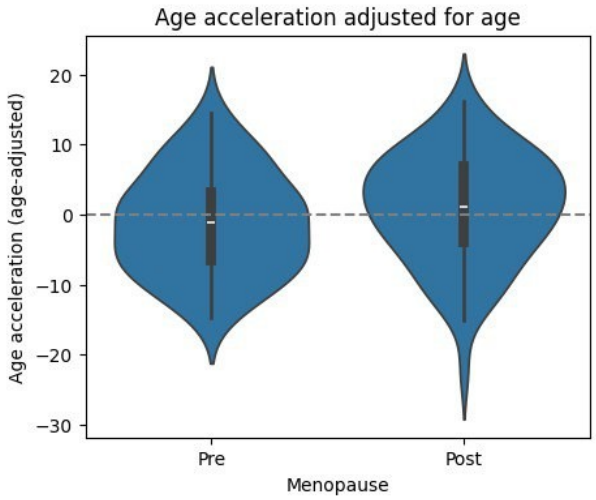

Figure 7 - Distribution of metabolic acceleration of age in premenopausal (Pre) and postmenopausal (Post) women
DOI: https://doi.org/10.60797/jbg.2026.33.2.10

Note: distribution of metabolic acceleration of age in premenopausal (Pre) and postmenopausal (Post) women after adjustment for chronological age; violin plots with median and interquartile range are shown; the horizontal dotted line corresponds to a zero value for age acceleration; in postmenopausal women, there is a shift in the distribution toward higher values of metabolic acceleration (p = 0.006)

Discussion

In the present study, we developed and rigorously evaluated a plasma metabolomics-based framework for predicting chronological age and quantifying metabolic age acceleration using regularized machine learning approaches. The primary Elastic Net model achieved a mean absolute error of approximately 6.04 years under a repeated nested cross-validation framework comprising 50 independent outer validation splits. This level of predictive accuracy is comparable with previously published metabolomic aging clocks and demonstrates that plasma metabolomic profiles contain reproducible information relevant to biological aging.

A major strength of the proposed framework is the integration of rigorous methodological safeguards designed to improve reproducibility and model reliability. Unlike many previous studies that primarily focused on predictive performance, our analytical pipeline combined strict leakage-free preprocessing, repeated nested cross-validation, comparative evaluation of multiple regularized models, alternative feature-selection strategies, and individualized uncertainty estimation using the jackknife+ conformal prediction framework. Together, these methodological components provide a robust assessment of model performance while minimizing optimistic bias arising from hyperparameter optimization and feature selection.

The comparative analyses further demonstrated that the metabolomic aging signal remained remarkably stable across different regularization strategies. Ridge regression achieved predictive performance comparable to Elastic Net, whereas explicit feature-selection methods based on Random Forest importance and recursive feature elimination resulted in only modest reductions in predictive accuracy. These findings suggest that age-related metabolomic information is distributed across numerous correlated metabolites rather than being driven by a limited number of highly informative biomarkers. Such observations agree with the complex and systemic nature of biological aging, which affects multiple interconnected metabolic pathways simultaneously rather than isolated molecular components.

The biological interpretation of the identified metabolic signatures is also consistent with current understanding of the aging process. Lipid- and lipoprotein-related metabolites, including LDL-associated fractions, sphingomyelins, and phospholipids, exhibited the strongest associations with chronological age. These metabolic alterations have previously been linked to age-related remodeling of lipid metabolism, chronic low-grade inflammation, altered membrane composition, and cardiometabolic dysfunction. Conversely, several amino acids and lysophospholipids demonstrated negative associations with age, supporting the concept that aging is characterized by coordinated metabolic remodeling across multiple biochemical pathways instead of uniform changes in metabolite concentrations.

Compared with epigenetic and transcriptomic aging clocks, metabolomic models provide complementary biological information. DNA methylation clocks generally achieve excellent predictive precision and long-term temporal stability, whereas transcriptomic clocks capture dynamic patterns of gene expression that may respond rapidly to environmental influences [22]. Metabolomic profiles represent downstream physiological processes and integrate the combined effects of genetic background, nutrition, lifestyle, environmental exposures, and ongoing metabolic regulation [20]. Although this integrated nature may introduce greater short-term variability, it also makes metabolomic clocks particularly suitable for investigating physiological transitions and functional aspects of biological aging.

The observed increase in prediction variability among older individuals likely reflects increasing biological heterogeneity during aging. As individuals grow older, cumulative environmental exposures, lifestyle differences, physiological adaptation, and natural variation in aging trajectories contribute to progressively greater diversity in metabolic profiles. Consequently, chronological age becomes a less precise surrogate for underlying biological state, resulting in wider prediction variability despite stable overall model performance. Rather than representing solely a limitation of the predictive model, this increasing dispersion may itself reflect an intrinsic biological property of the aging process.

An important consideration concerns the potential influence of chronic diseases and metabolic disorders on metabolomic age estimation. The present model was developed using the KarMeN cohort, which consists exclusively of healthy individuals and therefore provides a robust reference for characterizing physiological metabolic aging. However, metabolic disorders such as obesity, metabolic syndrome, type 2 diabetes mellitus, cardiovascular disease, chronic inflammatory conditions, and other age-related pathologies are known to substantially modify circulating metabolite profiles independently of chronological age. Consequently, these conditions may contribute to apparent metabolic age acceleration or alter the calibration of metabolomic aging models. Future studies should therefore evaluate the proposed framework in clinically diverse populations and investigate whether disease-specific recalibration or adaptation of metabolomic clocks is required to distinguish normal biological aging from pathological metabolic alterations [16].

The association observed between menopausal status and increased metabolic age acceleration further supports the biological relevance of the proposed framework. Menopause is accompanied by profound endocrine and metabolic changes, including alterations in lipid metabolism, body composition, insulin sensitivity, and inflammatory signaling [21]. The significant shift toward higher metabolomic age acceleration observed in postmenopausal women therefore appears biologically plausible and suggests that metabolomic aging captures meaningful physiological transitions extending beyond chronological age alone.

Several limitations should also be acknowledged. First, model development and validation were performed exclusively using a cohort of healthy adults. Although this design minimizes biological confounding and provides a reliable reference for physiological aging, the generalizability of the proposed metabolomic clock to individuals with chronic diseases or other clinically relevant conditions remains to be established. Second, the study is based on cross-sectional data, preventing direct assessment of longitudinal changes in biological aging within individuals. Future investigations using longitudinal cohorts, independent external validation datasets, and more diverse populations will be essential for evaluating the long-term stability, clinical applicability, and broader generalizability of metabolomic aging models.

Overall, the present findings demonstrate that plasma metabolomic profiles contain robust and biologically meaningful information related to aging. By integrating rigorous validation procedures, regularized machine learning, comparative modeling, feature-selection analyses, and individualized uncertainty quantification, the proposed framework provides a reproducible foundation for future research on biological aging. With further validation in longitudinal and clinically heterogeneous cohorts, metabolomic aging clocks may become valuable tools for investigating healthy aging, identifying early metabolic alterations associated with age-related diseases, and evaluating interventions aimed at promoting healthy longevity.

Conclusion

In this study, we developed and rigorously evaluated a plasma metabolomics-based model for predicting chronological age and quantifying metabolic age acceleration. The primary Elastic Net model demonstrated stable performance under a

computationally intensive repeated nested cross-validation framework and achieved accuracy comparable to previously reported molecular aging clocks. Beyond predictive performance, the identified metabolic patterns, particularly those related to lipid and lipoprotein metabolism, were biologically consistent with established mechanisms of age-related metabolic remodeling.

Importantly, comparative analyses using ridge regression and additional feature selection approaches demonstrated that the identified metabolomic aging signal remained robust across alternative modeling strategies. The relatively small differences in predictive performance between regularized models further suggest that aging-related metabolic information is distributed across multiple correlated metabolic pathways rather than driven by a small subset of individual biomarkers.

Our framework additionally integrates uncertainty quantification through conformal prediction, allowing age estimates to be interpreted alongside individualized measures of reliability. This feature enhances the practical value of metabolomic age prediction, particularly in research settings where inter-individual heterogeneity and model confidence are important considerations.

The proposed model was developed using a cohort of healthy individuals, providing a reliable reference for characterizing physiological metabolic aging. However, further validation in populations with chronic diseases, metabolic disorders, and independent external cohorts will be necessary to establish its broader generalizability and potential clinical applicability. In addition, longitudinal studies will be important for evaluating the ability of metabolomic aging clocks to monitor biological aging over time and to distinguish physiological aging from disease-associated metabolic alterations.

Taken together, our findings support the utility of plasma metabolomic profiles as informative indicators of biological aging. By combining interpretable regularized modeling, robust validation, alternative feature selection strategies, and uncertainty-aware estimation, this study provides a reproducible foundation for future investigations of metabolic aging. With further longitudinal validation and integration across independent cohorts, metabolomic clocks may contribute to monitoring aging dynamics, evaluating lifestyle or therapeutic interventions, and improving understanding of the metabolic mechanisms underlying age-related decline.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы на английском языке / References in English

1. Ferrucci L. Measuring biological aging in humans: A quest / L. Ferrucci, M. Gonzalez-Freire, E. Fabbri et al. // Aging Cell. — 2020. — 19(2). — P. e13080. — DOI: 10.1111/acel.13080

2. López-Otín C. The hallmarks of aging / C. López-Otín, M.A. Blasco, L. Partridge et al. // Cell. — 2013. — 153(6). — P. 1194–217. — DOI: 10.1016/j.cell.2013.05.039

3. Horvath S. DNA methylation age of human tissues and cell types / S. Horvath // Genome Biol. — 2013. — 14(10). — P. R115. — DOI: 10.1186/gb-2013-14-10-r115

4. Lu A.T. DNA methylation GrimAge strongly predicts lifespan and healthspan / A.T. Lu, A. Quach, J.G. Wilson et al. // Aging (Albany NY). — 2019. — 11(2). — P. 303–327. — DOI: 10.18632/aging.101684

5. Peters M.J. The transcriptional landscape of age in human peripheral blood / M.J. Peters, R. Joehanes, L.C. Pilling et al. // Nat Commun. — 2015. — P. 6:8570. — DOI: 10.1038/ncomms9570

6. Johnson C.H. Metabolomics: beyond biomarkers and towards mechanisms / C.H. Johnson, J. Ivanisevic, G. Siuzdak // Nat Rev Mol Cell Biol. — 2016. — 17(7). — P. 451–9. — DOI: 10.1038/nrm.2016.25

7. Lehallier B. Undulating changes in human plasma proteome profiles across the lifespan / B. Lehallier, D. Gate, N. Schaum et al. // Nature Medicine. — 2019. — 25(12). — P. 1843–1850. — DOI: 10.1038/s41591-019-0673-2

8. Auro K. A metabolic view on menopause and ageing. / K. Auro, A. Joensuu, K. Fischer et al. // Nat Commun.. — 2014. — P. 5:4708. — DOI: 10.1038/ncomms5708

9. Jylhävä J. Biological Age Predictors / J. Jylhävä, N.L. Pedersen, S. Hägg // EBioMedicine. — 2017. — 21. — P. 29–36. — DOI: 10.1016/j.ebiom.2017.03.046

10. Robinson O. Determinants of accelerated metabolomic and epigenetic aging in a UK cohort / O. Robinson, M. Chadeau Hyam, I. Karaman et al. // Aging Cell. — 2020. — 19(6). — P. e13149. — DOI: 10.1111/acel.13149

11. Hertel J. Measuring Biological Age via Metabonomics: The Metabolic Age Score / J. Hertel, N. Friedrich, K. Wittfeld et al. // J Proteome Res. — 2016. — 15(2). — P. 400–10. — DOI: 10.1021/acs.jproteome.5b00561

12. Cheng S. Distinct metabolomic signatures are associated with longevity in humans / S. Cheng, M.G. Larson, E.L. McCabe et al. // Nat Commun. — 2015. — P. 6:6791. — DOI: 10.1038/ncomms7791

13. Vovk V. Algorithmic Learning in a Random World / V. Vovk, A. Gammerman, G. Shafer. — Egham: Springer, 2022. — 319 p.

14. Barber R.F. Predictive inference with the jackknife+ / R.F. Barber, E.J. Candès, A. Ramdas et al. // Annals of Statistics. — 2021. — 49(1). — P. 486–507. — DOI: 10.1214/20-AOS1965

15. Bub A. The Karlsruhe Metabolomics and Nutrition (KarMeN) Study: Protocol and Methods of a Cross-Sectional Study to Characterize the Metabolome of Healthy Men and Women / A. Bub, A. Kriebel, C. Dörr et al. // JMIR Research Protocols. — 2016. — 5(3). — P. e146. — DOI: 10.2196/resprot.5792

16. Justice J.N. A framework for selection of blood-based biomarkers for geroscience-guided clinical trials: report from the TAME Biomarkers Workgroup / J.N. Justice, L. Ferrucci, A.B. Newman // GeroScience. — 2018. — 40(5–6). — P. 419–436. — DOI: 10.1007/s11357-018-0042-y

17. Moqri M. Biomarkers of aging for the identification and evaluation of longevity interventions / M. Moqri, C. Herzog, J.R. Poganik // Cell. — 2023. — 186(18). — P. 3758–3775. — DOI: 10.1016/j.cell.2023.08.003

18. Bortz J. Biological age estimation using circulating blood biomarkers / J. Bortz, A. Guariglia, L. Klaric // Communications Biology. — 2023. — 6. — P. 1089. — DOI: 10.1038/s42003-023-05456-z

19. Erema V.V. Biological Age Predictors: The Status Quo and Future Trends / V.V. Erema, A.Y. Yakovchik, D.A. Kashtanova // International Journal of Molecular Sciences. — 2022. — 23(23). — P. 15103. — DOI: 10.3390/ijms232315103

20. Wishart D.S. Metabolomics for investigating physiological and pathophysiological processes / D.S. Wishart // Physiological Reviews. — 2019. — 99(4). — P. 1819–1875. — DOI: 10.1152/physrev.00035.2018

21. Ahadi S. Personal aging markers and ageotypes revealed by deep longitudinal profiling / S. Ahadi, W. Zhou, S.M. Schüssler-Fiorenza Rose // Nature Medicine. — 2020. — 26(1). — P. 83–90. — DOI: 10.1038/s41591-019-0719-5

22. Coenen L. Markers of aging: Unsupervised integrated analyses of the human plasma proteome / L. Coenen, B. Lehallier, H.E. de Vries // Front Aging. — 2023. — 4. — P. 1112109. — DOI: 10.3389/fragi.2023.1112109