International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com OPEN ACCESS E-ISSN : 3049-2424
Volume 3 Issue 8 , August 2026 DOI: https://doi.org/10.71366/ijwos03082654306

Ageing in the Field: Interpretable Chronological Age Prediction from Functional and Physiological Markers in LASI Wave 1

Vidisha Jani[1], Ajay Goyal[2]
[1] *Prakash Higher Secondary School, Ahmedabad, Gujarat 380054 India*
[2] *Institute of Design, Nirma University, Ahmedabad, Gujarat 382481 India.*

Article Info

Article History:

Published: 16 August 2026

Publication Issue:
Volume 3, Issue 8
August-2026

Page Number:
341-359

Corresponding Author:
Vidisha Jani

Abstract:

Population ageing in India presents an urgent need for scalable, interpretable and field-compatible approaches to understand age-related functional and physiological variation. Chronological age is universally available, but it does not fully capture differences in health status, mobility, strength, metabolic condition and functional capacity among individuals of the same age. This study developed a reproducible chronological age-prediction framework from the Longitudinal Ageing Study in India. Data from 4,473 participants were integrated across Individual, Biomarker, Dried Blood Spot and Spirometry datasets. Base predictors included sex, maximum grip strength, walking speed, pulse rate, mean sleep problem score and body mass index, while the extended predictor set additionally included blood-pressure, hemoglobin, HbA1c, C-reactive protein, FVC and FEV1 measures. Four analytical datasets were constructed using complete-case and median-imputed strategies. Linear Regression, Random Forest and Gradient Boosting models were evaluated using five-fold cross-validation and an 80:20 internal holdout test design, supported by leakage audit, baseline comparison and residual diagnostics. The Base Complete-Case Linear Regression model achieved the strongest performance, with Test $R^2 = 0.86$, RMSE = 4.09 years, MAE = 3.25 years and MAPE = 6.08%. Base Complete-Case Random Forest and Gradient Boosting followed closely, with Test R^2 values of 0.85 and 0.84, respectively. Qualitatively, the findings indicate that simple field-compatible indicators captured substantial chronological age-related variation, whereas extended biomarker and spirometry variables did not improve prediction under the present missing-data conditions. The framework offers a transparent research-oriented approach for ageing studies in India, requiring longitudinal and external validation before clinical or public-health application.

Keywords: Chronological age prediction, LASI Wave 1, ageing markers, functional ageing, machine learning

1. Introduction

Population ageing has become one of the most important public-health and social transformations of the twenty-first century (Grinin et al. 2024; Ribeiro and Araújo 2025). Improvements in survival, medical care and living conditions have increased life expectancy across the world, but longer life does not necessarily mean healthier life (Jugran 2025; Gianfredi et al. 2025). The central challenge is no longer only to add years to life, but to preserve functional ability, independence and quality of life in later years (Dodds et al. 2024; Goodarzi et al. 2024). The World Health Organization defines healthy ageing in relation to the maintenance of functional ability that enables wellbeing in older age, making ageing a multidimensional issue that includes biological health, physical capacity, social participation and health-system preparedness (World Health Organization 2021; Zhang and Wang 2025). This challenge is especially relevant for India, where rapid demographic transition is occurring alongside high variability in

socioeconomic conditions, healthcare access, nutrition, chronic disease burden and environmental exposure (India State-Level Disease Burden Initiative Collaborators 2017).

Chronological age remains the most widely used marker of ageing because it is simple, objective and universally available (Kotschy et al. 2025). However, it is an incomplete representation of the ageing process. Individuals of the same chronological age may differ substantially in muscle strength, walking capacity, cardiometabolic status, inflammatory burden, respiratory function, sleep quality and ability to perform daily activities (Turkiewicz et al. 2025; Araque et al. 2026; Cheema et al. 2026). This difference reflects the distinction between chronological ageing and biological or functional ageing. Chronological age measures time lived, whereas biological and functional ageing reflect the cumulative condition of physiological systems and their capacity to withstand stress. Therefore, ageing research increasingly seeks measurable indicators that can capture variation in health status beyond calendar age (Dror 2025).

Several domains have been repeatedly associated with ageing-related decline. Grip strength is widely recognized as a practical marker of muscle function, sarcopenia risk and frailty, while gait speed or walking performance reflects neuromuscular coordination, balance, endurance and mobility reserve (Muollo et al. 2025; Kapan et al. 2025). Systematic evidence has shown that lower handgrip strength and gait speed are associated with adverse outcomes, including cardiovascular mortality and functional vulnerability (França et al. 2026). Anthropometric measures such as body mass index, height and weight provide information on nutritional status, body composition and long-term health risk (Bennett and Lim 2025; Feng 2025). Cardiovascular and metabolic indicators, including pulse, blood pressure and HbA1c, reflect circulatory regulation and glycaemic control, both of which are closely linked with morbidity in older populations (Yang et al. 2025; Fu et al. 2025; Kulecki et al. 2026; Si et al. 2026). Hemoglobin captures anaemia-related vulnerability, while C-reactive protein represents systemic inflammation; a process often implicated in chronic disease and accelerated physiological ageing (Uluç et al. 2025; Pan et al. 2025). Respiratory measures such as forced vital capacity and forced expiratory volume in one second add another important dimension because lung function declines with age even among adults without diagnosed lung disease (Bui et al. 2022; Garcia-Aymerich et al. 2025).

Recent ageing research has made substantial progress through biological age clocks, phenotypic age models and machine-learning approaches (Rutledge et al. 2022; Li et al. 2023, 2026; Qiu et al. 2023; Meng et al. 2024; Mahbub et al. 2026; Cheema et al. 2026). Epigenetic clocks, based on DNA methylation patterns, have become influential tools for estimating biological age and studying age acceleration (Teschendorff and Horvath 2025). Clinical biomarker-based models have also shown that routinely measured physiological and biochemical variables can be used to estimate biological or chronological age and reveal ageing-related health patterns (Raisi-Estabragh et al. 2024). However, important limitations remain. Many advanced ageing clocks depend on molecular, genomic, epigenetic or laboratory-intensive biomarkers that may be expensive, technically demanding and difficult to scale in low- and middle-income settings. In addition, much of the existing literature is based on Western or high-income populations, limiting generalizability to countries such as India, where ageing trajectories may be shaped by different life-course exposures, healthcare access, occupational histories, nutrition and social determinants.

Another limitation in the literature is the frequent use of single-domain models. Some studies focus mainly on blood biomarkers, others on physical performance, and others on questionnaire-based health status (Jazayeri et al. 2023; Behrenbruch et al. 2026). While each domain is useful, ageing is not confined to one biological system. It is a whole-body process involving musculoskeletal strength, mobility, cardiometabolic regulation, inflammation, respiratory function, sleep and broader functional capacity. Models that do not integrate these domains may miss important interactions and may provide an incomplete picture of ageing-related variation. There are also methodological issues related to missing data, inconsistent variable selection, inadequate preprocessing transparency and limited reporting of why specific variables were excluded or retained.

The Longitudinal Ageing Study in India (LASI) provides a valuable opportunity to address this gap. LASI is a nationally representative study of adults, designed to investigate the health, economic and social determinants and consequences of population ageing in India (Perianayagam et al. 2022). Its Wave 1 data include individual-level information, physical performance measures, anthropometry, biomarkers, dried blood spot assays and spirometry, making it suitable for developing a multimodal ageing-marker framework. Unlike narrowly defined datasets, LASI enables the integration of field-compatible indicators with laboratory and respiratory measures in a large Indian population-based cohort.

This study aimed to develop and evaluate reproducible chronological age-prediction models using field-compatible, biomarker and spirometry-derived variables from LASI Wave 1. Specifically, the study compared base and extended predictor sets, complete-case and median-imputed datasets, and three regression approaches: Linear Regression, Random Forest and Gradient Boosting. The modeling task was used to assess how different ageing-related health indicators collectively encode chronological age-related variation in an Indian cohort.

The contribution of this study is threefold. First, it develops a reproducible LASI Wave 1-based pipeline for chronological age prediction using multimodal ageing-related indicators. Second, it compares field-compatible base predictors with an extended predictor set including blood-based and spirometry variables. Third, it evaluates complete-case and median-imputed modeling strategies across linear and non-linear regression models, while interpreting predictor contributions as predictive associations rather than causal effects.

The proposed framework may support future research on functional decline, geriatric screening and community-level monitoring, but further validation against clinical and longitudinal outcomes is required. In resource-constrained settings, affordable and field-compatible indicators are especially valuable because they can inform public-health action without relying entirely on expensive molecular testing. The proposed framework is not intended to replace clinical assessment or establish causal mechanisms; rather, it offers a decision-support and research-oriented approach for understanding ageing-related health variation. In doing so, this study contributes to ageing research, health informatics and public-health decision-making by demonstrating how large-scale Indian cohort data can be transformed into a transparent, multimodal and socially relevant ageing prediction framework. This study does not estimate biological age directly and does not claim to identify accelerated ageing. Because the outcome variable is chronological age, the models should be interpreted as predictive tools for age-related variation rather than diagnostic tools for biological ageing, frailty or clinical risk. Validation against health outcomes such as disability, frailty, morbidity or mortality would be required before clinical or public-health screening applications can be claimed.

2. Methodology

The methodological workflow is shown in Figure 1. The analysis followed data integration, data cleaning, feature engineering, analytical dataset preparation, model training, model evaluation and interpretation. All analyses were implemented in MATLAB (2024b). Ethical committee of Nirma University, Ahmedabad has approved methodology of the study.

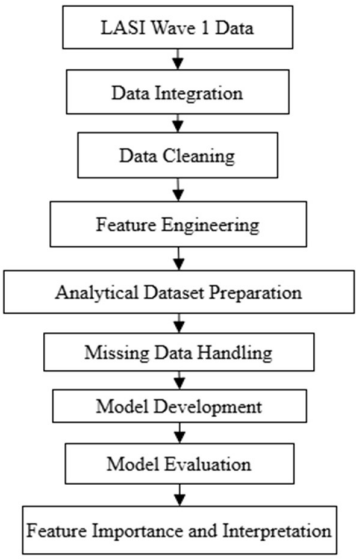

Figure 1 Methodological workflow for LASI data integration, feature engineering, analytical dataset preparation and model development.

Study design and data source

This study used data from Wave 1 of the Longitudinal Ageing Study in India (LASI) (Perianayagam et al. 2022), with written permission from International Institute for Population Sciences (IIPS), Mumbai, India. Participant data were collected from four LASI components: the Individual dataset, Biomarker dataset, Dried Blood Spot (DBS) dataset and Spirometry dataset. The Individual dataset, considered as primary data, contained the participant identifier, demographic variables, sleep-related variables, residence information and survey weights. Biomarker, DBS and Spirometry data were merged with the Individual dataset using the unique participant identifier. Table 1 shows the predictor sets used for age-prediction modeling.

Participant inclusion was documented through a stepwise data-flow procedure. The Individual dataset was used as the master file. Biomarker, Dried Blood Spot and Spirometry records were merged using the standardized participant identifier. Participants were retained in the master analytical dataset if they had a valid participant identifier and plausible chronological age. Subsequent model-ready datasets were created based on predictor availability. The base complete-case dataset included participants with complete values for all base predictors, while the extended complete-case dataset required complete values for both base and extended predictors. Median-imputed datasets retained all participants with valid chronological age and imputed missing predictor values using training-set medians.

Table 1 Predictor sets used for age-prediction modeling

Category	Variables	Reason
Base predictors	Sex, maximum grip strength, walking speed, pulse rate, mean sleep problem score, BMI	Field-compatible indicators
Extended predictors	Systolic BP, diastolic BP, hemoglobin, HbA1c, CRP, FVC, FEV1	Biomarker and respiratory indicators
Leakage variables excluded	Year of birth, month of birth, age-derived variables	Directly related to target age
Identifier variables excluded	Participant ID, household ID	Non-predictive identifiers

Survey design variables excluded	Geography variables, survey weights	Retained for audit but not used in unweighted modeling
Raw variables replaced by engineered features	Grip trials, walking trials, sleep items	Used to derive participant-level features

Data integration and cleaning

Prior to integration, the participant identifier was standardized across the Individual, Biomarker, Dried Blood Spot and Spirometry datasets. The Individual dataset was used as the master analytical file, and the remaining datasets were merged at the participant level using a left-join procedure. This approach retained the Individual dataset as the primary cohort frame while incorporating available physical-performance, biomarker and respiratory measurements. Duplicate participant identifiers were examined before merging. Exact duplicate records were removed. Where multiple records were available for the same participant, the record with the greatest completeness for the required predictors was retained. The number of duplicate records and the rule used for resolving them were documented before model development. No duplicate participant identifiers remained after this procedure.

LASI-specific missing and non-response codes were handled before feature engineering. High special codes such as 997, 998, 999, 9997, 9998, 9999 and 99999 were treated as missing values. Codes such as 97, 98 and 99 were not removed globally because they may be valid values for continuous physiological variables. These codes were treated as missing only in selected survey-coded or categorical variables.

Physiological plausibility filters were applied to age, height, weight, Body Mass Index (BMI), grip strength, walking time, walking speed, blood pressure, pulse, hemoglobin, Hemoglobin A1c (HbA1c), C - reactive protein (CRP) and spirometry variables. Implausible values were set to missing.

Feature engineering

Chronological age was calculated using Eq. (1). A_i is the chronological age of participant i, and YOB_i is the year of birth of participant i. Ages outside the plausible range of 18–110 years were set to missing. Sex was recoded using Eq. (2). S_i is the binary sex variable for i^{th} participant. Height (H_i) was recorded in meters. Body Mass Index was calculated using Eq. (3). BMI_i is the Body Mass Index of participant i, W_i is body weight in kilograms. Maximum grip strength was calculated using Eq. (4). G_i^{max} *is the maximum grip strength of participant i, and* $G_{i1}, G_{i2}, G_{i3}, G_{i4}$ are the available grip-strength trials. Mean grip strength (G_i^{mean}) was calculated using Eq. (5). G_{ir} is the r^{th} available grip-strength trial and n_i is the number of available grip-strength trials for participant i. Best walking time (T_i^{best}) was calculated using Eq. (6). T_{i1} and T_{i2} are the two walking-time trials. Walking speed (V_i) was calculated as $4/T_i^{best}$. The numerator 4 represents the 4-metre walking-test distance. The mean sleep problem score (SP_i^{mean}) and summed sleep problem score (SP_i^{sum}) were calculated as given in Eq. (7) and (8), respectively. SP_{ik} is the k^{th} available sleep item, and q_i is the number of available sleep items for participant i.

$$A_i = 2017 - YOB_i \tag{1}$$

$$S_i = \begin{cases} 1, & male \\ 0, & female \\ missing, & otherwise \end{cases} \tag{2}$$

$$BMI_i = \frac{W_i}{{H_i}^2} \tag{3}$$

$$G_i^{max} = \max(G_{i1}, G_{i2}, G_{i3}, G_{i4}) \tag{4}$$

$$G_i^{mean} = \frac{1}{n_i}\sum_{r=1}^{n_i} G_{ir} \quad (5)$$

$$T_i^{best} = \min(T_{i1}, T_{i2}) \quad (6)$$

$$SP_i^{mean} = \frac{1}{q_i}\sum_{k=1}^{q_i} SP_{ik} \quad (7)$$

$$SP_i^{sum} = \sum_{k=1}^{q_i} SP_{ik} \quad (8)$$

Analytical dataset construction

Two analytical datasets were constructed. The base dataset included field-compatible indicators: A_i, S_i, G_i^{max}, T_i^{best}, pulse rate, sleep problem score and BMI_i. The extended dataset included the base predictors and added systolic blood pressure, diastolic blood pressure, hemoglobin, HbA1c, CRP, Forced Vital Capacity, Forced Expiratory Volume in One Second and other respiratory indicators. The final model-facing base predictors were binary sex, maximum grip strength, walking speed, pulse rate, mean sleep problem score and Body Mass Index. The final extended predictors included all base predictors plus systolic blood pressure, diastolic blood pressure, hemoglobin, HbA1c, CRP, Forced Vital Capacity and Forced Expiratory Volume in One Second. Year of birth and month of birth were excluded from the predictor set to avoid direct age leakage. Participant identifiers, household identifiers, geography variables, survey weights, weak reference variables and raw trial-level variables were retained for auditability but were not used as direct predictors.

Missing data handling

Two missing-data strategies were used. First, complete-case analysis was performed by retaining only participants with valid chronological age and completes values for all predictors required by the corresponding model. Second, median imputation was used as a baseline imputation strategy. Chronological age was not imputed. For each training split, predictor medians were estimated only from the training data and then applied to validation and test data. This prevented information from the validation or test sets from influencing imputation. Because missingness in functional, biomarker and spirometry variables may be related to age, health status or test completion ability, complete-case results were interpreted cautiously and compared with median-imputed results.

Let x_{ij} denote the value of predictor j for participant i, and let x_{ij}^* denote the corresponding imputed value. Median imputation was performed using Eq. (9).

$$x_{ij}^* = \begin{cases} x_{ij}, & if\ x_{ij}\ is\ observed \\ median(x_j), & not\ observed \end{cases} \quad (9)$$

$median(x_j)$ is the median of predictor j calculated from the training subset and applied to validation or test observations when the corresponding predictor value was missing. Four model-ready datasets were therefore prepared: base complete-case, base median-imputed, extended complete-case and extended median-imputed.

Target leakage was assessed before model training. Year of birth, month of birth, chronological age, age group variables, participant identifiers, household identifiers, survey timing variables and any variables directly derived from age were excluded from the predictor matrix. Feature engineering was performed without using the outcome variable

except for defining chronological age. Imputation parameters and standardization parameters were estimated only from training data. As a negative-control analysis, the age outcome was randomly shuffled and the full modeling pipeline was repeated; model performance was expected to collapse toward zero explanatory power if no leakage was present.

Model development

Each model-ready dataset was split into training and internal holdout test subsets using an 80:20 holdout strategy. Model development was conducted using five-fold cross-validation within the training subset. All preprocessing steps, including imputation and standardization, were performed within each training fold and then applied to the corresponding validation fold. This ensured that validation data were not used during preprocessing. Where possible, folds were stratified by age group to preserve the age distribution across folds. For Linear Regression, predictors (z_{ij}) were standardized using training-set means (μ_j) and standard deviations (σ_j) using Eq. (10). The same training-set parameters were applied to validation and test data. Three regression models were developed: Linear Regression, Random Forest regression and Gradient Boosting regression. The random seed was fixed at 42 for reproducibility. Random Forest models were trained using regression trees with out-of-bag prediction and predictor-importance estimation. During cross-validation, Random Forest used 300 trees and a minimum leaf size of 5; the final Random Forest model used 500 trees. Gradient Boosting was implemented using least-squares boosting with regression-tree learners, a learning rate of 0.05 and a minimum leaf size of 5. During cross-validation, 300 boosting cycles were used; the final Gradient Boosting model used 400 boosting cycles. Linear Regression was defined as Eq. (11). Because predictors were standardised before Linear Regression, coefficients were interpreted on the standardised predictor scale. $\widehat{A}_\iota$ is the predicted age of participant i, β_0 is the intercept, β_j is the coefficient for predictor j, and p is the number of predictors. Random Forest regression was defined as Eq. (12). B is the number of trees, $T_b(x_i)$ is the prediction from the b^{th} tree, and x_i is the predictor vector for participant i. Gradient Boosting regression was expressed as Eq. (13). $F_M(x_i)$ is the final boosted prediction function, $F_0(x_i)$ is the initial prediction function, M is the number of boosting iterations, ν is the learning rate, and $f_m(x_i)$ is the regression tree fitted at iteration m. The predicted age from the Gradient Boosting model was computed using Eq. (14). $F_M(x_i)$ is the final Gradient Boosting prediction function after M boosting iterations.

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{10}$$

$$\widehat{A}_\iota = \beta_0 + \sum_{j=1}^{p} \beta_j z_{ij} \tag{11}$$

$$\widehat{A}_\iota = \frac{1}{B} \sum_{b=1}^{B} T_b(x_i) \tag{12}$$

$$F_M(x_i) = F_0(x_i) + \sum_{m=1}^{M} \nu f_m(x_i) \tag{13}$$

$$\widehat{A}_\iota = F_M(x_i) \tag{14}$$

In addition to the main models, baseline models were estimated for comparison. These included a mean-age prediction baseline, a sex-only model, a functional-only model using grip strength and walking speed, and a base predictor model

excluding sex. These comparisons were used to determine whether the full base predictor set improved performance beyond simple demographic or functional predictors.

Model evaluation and interpretation

Model performance was evaluated using Cross-Validation within the training data and internal holdout testing on the holdout test data. The evaluation metrics were Coefficient of Determination (R^2), Root Mean Squared Error ($RMSE$), Mean Absolute Error (MAE) and Mean Absolute Percentage Error ($MAPE$). n is the number of participants in the evaluated dataset. Feature importance was assessed as a model-interpretation step. For Linear Regression, feature importance was based on the absolute magnitude of standardized coefficients. For Random Forest, Out-of-Bag (OOB) permuted predictor importance was used. For Gradient Boosting, ensemble-based predictor importance was extracted from the fitted model. These importance estimates were interpreted as model-derived predictive contributions and not as causal effects.

Because complete-case and imputed datasets differed in sample size and participant composition, an additional same-sample sensitivity analysis was performed. Base and extended predictor models were compared within the same complete-case participant subset to determine whether differences in performance were due to predictor set differences or sample composition.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(A_i - \widehat{A_i}\right)^2}{\sum_{i=1}^{n}(A_i - \bar{A})^2} \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(A_i - \widehat{A_i}\right)^2} \quad (16)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|A_i - \widehat{A_i}\right| \quad (17)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{A_i - \widehat{A_i}}{A_i}\right| \quad (18)$$

Reproducibility

The workflow was designed to be reproducible and auditable. Major analytical decisions were documented, including variable selection, exclusion rules, merge checks, special-code handling, plausibility filtering, feature engineering, missing-data handling, predictor selection, model training and model evaluation.

Results

Table 2 summarizes the availability and distribution of the demographic, functional, anthropometric, physiological, blood-based and respiratory variables included in the analytical dataset. Chronological age was available for all 4,473 participants, while missingness varied across domains. Spirometry variables showed the highest missingness, with FVC missing in 23.61% and FEV1 missing in 23.56% of participants, whereas sleep-problem scores had low missingness of 0.34%. Missingness differed across measurement domains. Functional and anthropometric variables showed moderate missingness, while spirometry variables showed the highest missingness. Because missing functional or respiratory measurements may reflect participant health, frailty or test completion ability, complete-case analyses may represent a healthier or more test-compliant subgroup. Therefore, complete-case and imputed analyses were interpreted together rather than as directly equivalent samples.

Table 2 Summary of variables included in the analytical dataset.

Variable	N	Missing (%age)	Mean	SD	Median	IQR	Min	Max
age_years	4473	0.00	56.28	11.63	55.00	17.00	25.00	107.00
sex_binary	3898	12.85	0.56	0.50	1.00	1.00	0.00	1.00
grip_max	4056	9.32	23.72	7.71	23.00	10.50	0.00	57.50
grip_mean	4056	9.32	21.60	7.30	20.88	9.88	0.00	51.00
walk_time_best	3947	11.76	5.56	2.17	5.15	2.14	1.17	38.49
walk_speed_mps	3947	11.76	0.79	0.23	0.78	0.32	0.10	3.42
bm019	4091	8.54	82.32	10.62	82.00	13.88	39.00	126.50
sleep_problem_score_mean	4458	0.34	1.71	0.56	1.50	0.83	1.00	3.67
sleep_problem_score_sum	4458	0.34	10.25	3.37	9.00	5.00	5.00	22.00
bmi	4054	9.37	25.62	4.95	25.23	6.24	11.45	50.18
height_m	4054	9.37	1.55	0.09	1.55	0.12	1.00	1.89
bm067	4054	9.37	155.40	8.85	155.00	12.20	99.50	188.70
bm071	4054	9.37	61.94	13.15	61.10	16.90	24.10	130.60
bm076	4053	9.39	92.11	12.15	91.90	14.90	28.10	151.00
bm017	4091	8.54	131.20	17.85	130.00	23.50	81.50	187.00
bm018	4091	8.54	82.54	9.78	82.00	12.50	49.00	126.00
hb	3993	10.73	13.36	2.49	13.32	3.20	4.00	21.00
hba1c	3992	10.75	6.25	1.31	5.77	0.99	3.60	9.75
crp	3993	10.73	2.27	7.16	0.71	1.51	0.00	133.35
fvc	3417	23.61	2.19	0.84	2.10	0.96	0.21	11.52
fev1	3419	23.56	1.72	0.59	1.66	0.78	0.13	4.21
pef	3405	23.88	4.41	1.72	4.30	2.24	0.30	11.28
fef2575	3419	23.56	1.84	0.91	1.71	1.24	0.07	5.66

fev1fvc	3417	23.61	80.09	12.29	81.30	12.33	3.32	100.00
ref_fvc_gli	3420	23.54	2.81	0.61	2.71	0.88	1.28	5.04
ref_fev1_gli	3420	23.54	2.31	0.50	2.25	0.71	1.03	4.07

Table 3 summarizes the four model-ready analytical datasets constructed for model development. This shows the trade-off between complete-case analysis, which uses only fully observed records, and median-imputed analysis, which preserves the full participant sample.

Table 3 Model-ready analytical datasets used for chronological age prediction.

Dataset	Missing-data strategy	Predictor set	Observations	% of master dataset	Predictors
BCC	Complete-case	Base	3420	76.5%	6
B-MI	Median-imputed	Base	4473	100.0%	6
ECC	Complete-case	Extended	2862	64.0%	13
EMI	Median-imputed	Extended	4473	100.0%	13

Table 4 and Figure 2 present the ranked performance of all age-prediction models across the base complete-case, extended complete-case, base median-imputed and extended median-imputed datasets. Within the internal holdout test set, the Base complete-case Linear Regression model showed the strongest performance among the evaluated models, with Test $R^2 = 0.86$, RMSE = 4.09 years, MAE = 3.25 years and MAPE = 6.08%. However, because complete-case and imputed datasets differed in sample size and participant composition, comparisons across datasets should be interpreted cautiously. The next best models were the Base complete-case Random Forest and Base complete-case Gradient Boosting models, with Test R^2 values of 0.85 and 0.84, respectively. Overall, the base complete-case models showed stronger predictive performance than the extended and median-imputed models. Cross-validation and internal test performance were consistent for the selected model, with the BCC-LR model achieving CV $R^2 = 0.84$, CV RMSE = 4.19 years and CV MAE = 3.31 years, compared with Test $R^2 = 0.86$, Test RMSE = 4.09 years and Test MAE = 3.25 years. Baseline models showed lower predictive performance than the full base predictor model. The mean-age baseline provided no explanatory value, while the sex-only and functional-only models explained only part of the age-related variation. The full base predictor set improved performance, indicating that the combined functional, physiological, sleep-related and anthropometric measures provided stronger prediction than any single-domain baseline.

Figure 3 shows the observed versus predicted chronological age for the selected BCC-LR model. Predicted ages were closely aligned with the 45-degree reference line, supporting the model's strong internal holdout test-set performance. This agreement was consistent with the Test R^2 of 0.86, Test RMSE of 4.09 years and Test MAE of 3.25 years reported in Table 5. Most observations were concentrated around the reference line across the central age range, although prediction behavior at the youngest and oldest ages should be interpreted cautiously because fewer participants were available at these extremes. The leakage-audit analysis did not identify direct age-derived variables in the predictor matrix. Year of birth, month of birth, chronological age, age-group variables and identifiers were excluded before modeling. In the shuffled-age negative-control analysis, model performance collapsed toward zero explanatory power, supporting the absence of obvious target leakage.

Figure 4 presents the model-derived predictor importance for the best-performing Base complete-case Linear Regression model. Sex showed the highest relative importance, with an importance value of approximately 4.8,

followed by maximum grip strength at approximately 3.1. Walking speed, mean sleep problem score and pulse rate showed moderate contributions, with importance values of approximately 1.6, 1.25 and 1.0, respectively. Body Mass Index showed the lowest relative contribution among the base predictors, with an importance value of approximately 0.7. These findings indicate that demographic and functional measures contributed most strongly to prediction in the selected model, but the importance values should be interpreted as model-derived predictive contributions rather than causal effects.

Figure 5 presents residual plots for all model-dataset combinations across the observed age range. The residual distributions are broadly similar across Linear Regression, Random Forest and Gradient Boosting models, with most observations concentrated in the middle-age range where participant density was highest. No strong age-dependent curvature or systematic widening of residual spread is visible, suggesting that the models did not show a pronounced non-linear error pattern across the observed age range. The residual clouds are comparatively compact for the base complete-case models, while the extended and median-imputed models show slightly wider dispersion, which is consistent with their lower predictive performance reported in Table 4. Sparse observations at the youngest and oldest ages indicate that model performance at the age extremes should be interpreted cautiously.

Table 4 Ranked cross-validation and internal test-set performance of age-prediction models across analytical datasets.

Overall Rank	Dataset	Model	Number of observations	Number of predictors	CV R2	CV RMSE	CV MAE	Test R^2	Test RMSE	Test MAE	Test MAPE
1	BCC	LR	3420	6	0.84	4.19	3.31	0.86	4.09	3.25	6.08
2	BCC	RF	3420	6	0.83	4.32	3.42	0.85	4.24	3.38	6.33
3	BCC	GB	3420	6	0.82	4.49	3.57	0.84	4.39	3.51	6.58
4	ECC	LR	2862	13	0.77	4.82	3.82	0.79	4.69	3.74	7.06
5	B-MI	LR	4473	6	0.80	4.91	3.92	0.83	4.83	3.87	7.16
6	ECC	RF	2862	13	0.77	4.93	3.90	0.78	4.82	3.85	7.26
7	B-MI	RF	4473	6	0.80	5.06	4.04	0.82	4.98	3.99	7.38
8	ECC	GB	2862	13	0.75	5.07	4.03	0.77	4.94	3.96	7.49
9	B-MI	GB	4473	6	0.79	5.27	4.15	0.81	5.13	4.11	7.61
10	EMI	LR	4473	13	0.75	5.84	4.63	0.76	5.67	4.55	8.42
11	EMI	RF	4473	13	0.73	5.95	4.71	0.75	5.80	4.63	8.58
12	EMI	GB	4473	13	0.72	6.06	4.74	0.74	5.93	4.69	8.73

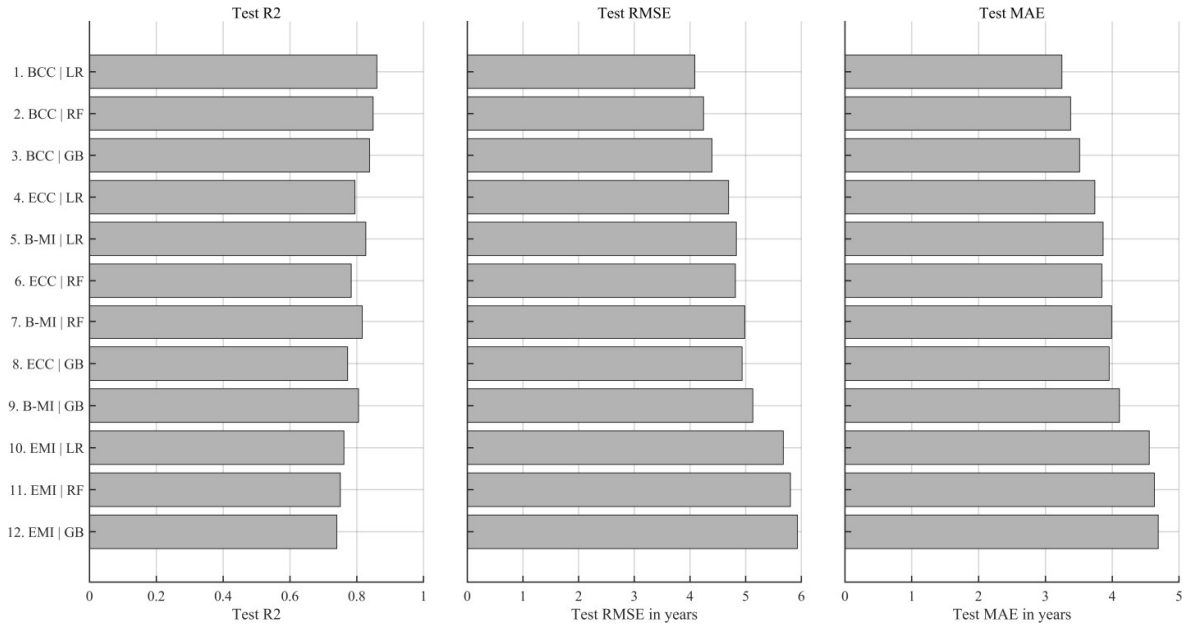

Figure 2 Internal test-set performance comparison of age-prediction models across analytical datasets.

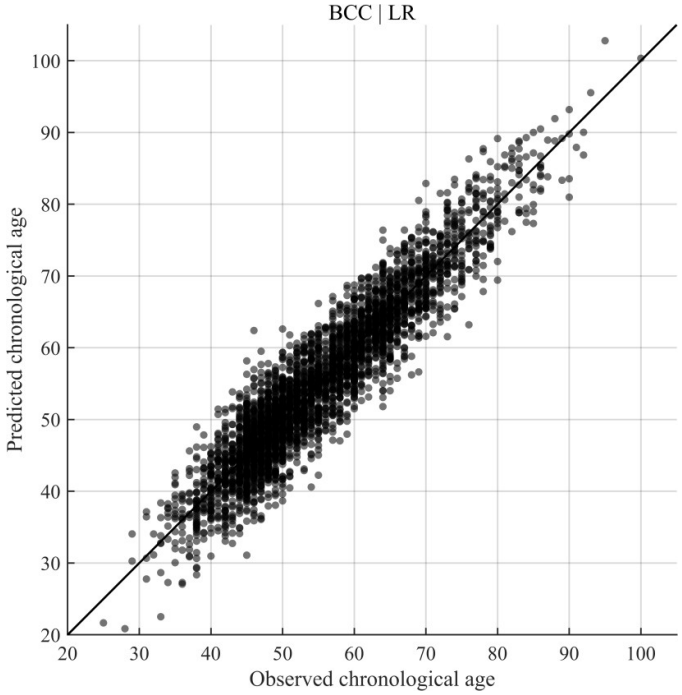

Figure 3 Observed versus predicted chronological age for the Base complete-case Linear Regression model.

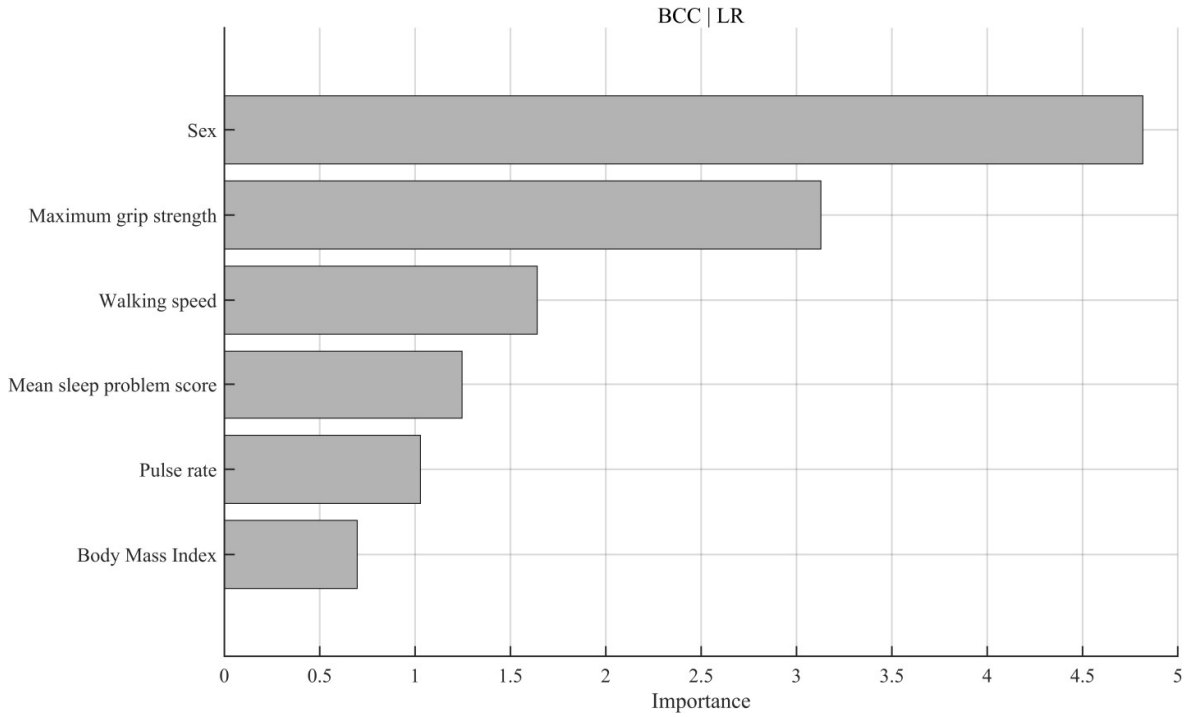

Figure 4 Model-derived predictor importance for the Base complete-case Linear Regression model.

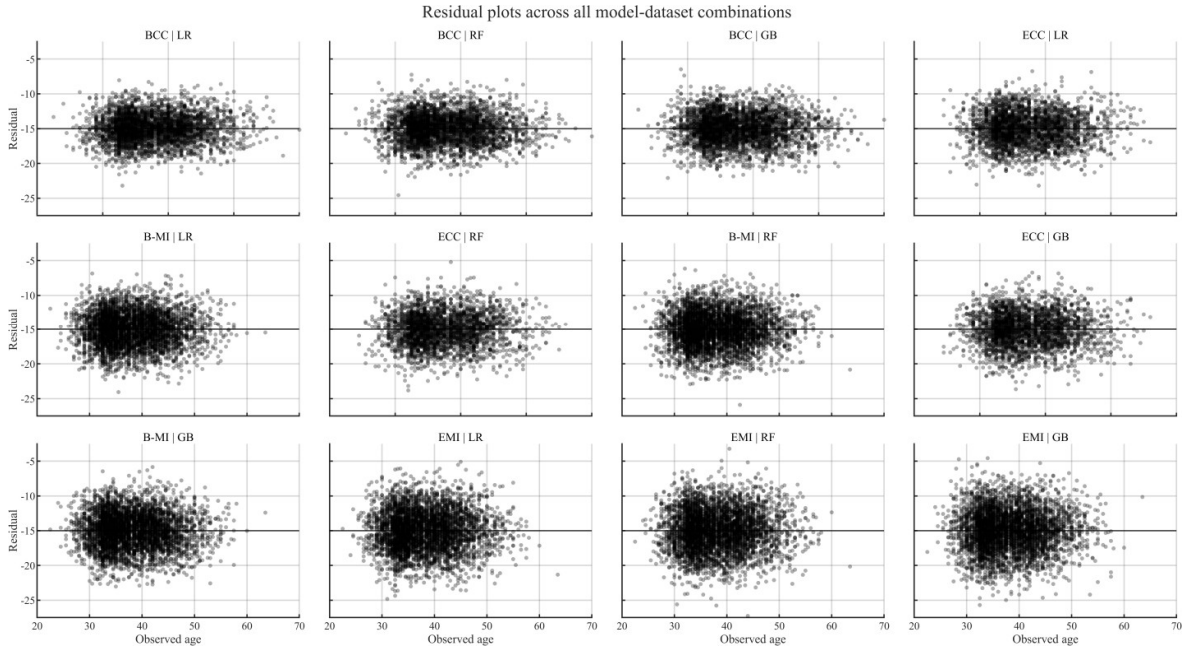

Figure 5 Residual plots across all model-dataset combinations showing the distribution of prediction errors over observed chronological age.

3. Discussion

The findings suggest that a small set of accessible, field-compatible measures can capture a substantial share of chronological age-related variation in LASI Wave 1, and that this signal is strong enough to be modeled well without resorting to more complex algorithms. The results indicates that interpretability and data quality mattered more here than expanding the predictor set or increasing model flexibility, with the Base Complete-Case Linear Regression model outperforming the extended and median-imputed alternatives.

Principal interpretation of the findings

What stands out most is not simply that the selected model performed well, but that it did so using only six predictors that are relatively easy to obtain in non-laboratory settings: sex, maximum grip strength, walking speed, sleep problem score, pulse rate and BMI. In practical terms, that pattern suggests that chronological age in this cohort was encoded strongly in broad functional and physiological signals rather than requiring a denser biomarker panel to extract age-related structure. That interpretation fits well with the healthy ageing perspective, which places functional ability and multisystem performance at the center of ageing research rather than treating age as a purely chronological construct. In other words, the model appears to be detecting the cumulative expression of ageing across mobility, strength, cardiovascular regulation, sleep-related disturbance and body composition, rather than isolating a single domain.

The stronger performance of the base complete-case models also deserves a cautious reading. One plausible interpretation is that the base variables were already sufficiently informative, so adding blood-based and spirometric variables brought relatively little incremental signal under the present analytic conditions. Another, equally plausible, interpretation is methodological: the complete-case base dataset likely contained cleaner measurements and fewer sources of heterogeneity, whereas the extended datasets absorbed more measurement variability and a markedly heavier missing-data burden, particularly for spirometry. The fact that the extended models underperformed should therefore not be read as evidence that biomarkers or respiratory measures are irrelevant to ageing. Rather, it suggests that their added value was not realized in this particular cross-sectional, partially incomplete dataset and with a relatively simple imputation strategy. That is an important distinction, because it keeps the interpretation grounded in the observed design rather than overstating a substantive conclusion about biology.

Why the base complete-case linear model likely prevailed

The superiority of Linear Regression over Random Forest and Gradient Boosting is also informative. In many ageing applications, non-linear models are attractive because physiological decline is rarely perfectly linear. Yet the present results imply that, once the predictor set was restricted to a compact group of strong field-compatible variables, much of the relevant age-related structure was captured adequately by an additive linear form. That may reflect a genuinely near-linear association between the selected predictors and chronological age across the central age range represented in the data. It may also indicate that the more flexible tree-based models had comparatively less room to improve upon the dominant signal, while remaining more sensitive to noise, sample composition and local irregularities in the training data.

This is not a weakness. If anything, it strengthens the manuscript's methodological position. Age-prediction models are often discussed as though algorithmic complexity is inherently preferable, but the broader ageing literature does not support such a simple hierarchy. In frailty research, for example, the value of a measure often lies in its ability to aggregate multisystem information in a transparent and transportable way rather than maximize technical sophistication. Searle et al. argued that frailty indices should cover multiple systems and be reproducible across datasets, while Levine et al. showed, in a different molecular context, that age-related modeling becomes most useful when it captures variation linked to morbidity and function rather than age alone (Searle et al. 2008; Levine et al. 2018). The present study aligns with that logic: a parsimonious and interpretable model may be more useful than a more opaque one when the goal is a research-oriented framework for age-related variation.

Meaning of the predictor hierarchy

The predictor ranking is broadly credible and also needs careful interpretation. Sex emerged as the strongest predictor, followed by maximum grip strength and walking speed. That ordering makes sense statistically, because sex encodes large and consistent differences in body composition, muscle strength and physical performance across adulthood. At the same time, sex is a demographic characteristic, not a modifiable ageing marker. Its dominance therefore does not mean that ageing is being "explained" by sex; it means that sex materially structures the distribution of the functional and physiological variables from which chronological age is inferred in this dataset. The manuscript should therefore

avoid reading sex as a substantive mechanism and instead treat it as a high-level stratifying variable that improves prediction.

By contrast, the prominence of grip strength and walking speed fits closely with the functional ageing and frailty literature. Fried's frailty phenotype explicitly includes weakness and slowness among its defining components, and Searle's deficit-accumulation framework similarly emphasizes the value of multisystem functional deficits in representing vulnerability in older adults (Fried et al. 1991; Searle et al. 2008). Bohannon has also described grip strength as an indispensable biomarker in older adults because it functions as a compact proxy for muscular reserve, disability risk and overall vitality (Bohannon 2019). In that sense, the present model behaves as one would expect from a cohort-based chronological age-prediction exercise: it draws heavily on markers that sit close to locomotors reserve and sarcopenic change, both of which are central expressions of ageing in later life.

The remaining predictors, walking speed, sleep problem score, pulse rate and BMI, plausibly add complementary information rather than merely duplicating what grip strength already captures. Sleep problems may reflect neurobehavioral strain, mood, and comorbidity or deregulated routines across the life course. Resting pulse can index autonomic balance, cardiorespiratory fitness and underlying disease burden. BMI is less specific, but in ageing populations it still carries information related to nutrition, adiposity and changing body composition. Importantly, none of these contributions should be interpreted causally. They are predictive associations within a chronological age-modeling task, not evidence that manipulating any single variable would alter ageing trajectories.

Positioning the study within the ageing-marker literature

The study sits in an interesting middle ground between clinical ageing markers and molecular clocks. On one side are functional frameworks such as frailty phenotype and frailty index approaches, which aim to represent vulnerability through performance, symptoms and deficit accumulation across systems (Fried et al. 1991; Searle et al. 2008). On the other side are molecular ageing measures, particularly DNA methylation clocks. Horvath's multi-tissue clock used 353 CpGs to estimate age across tissues, while Levine's DNAm PhenoAge was designed to improve prediction of mortality, morbidity and physical functioning rather than chronological age alone (Horvath 2013; Levine et al. 2018). Belsky's DunedinPoAm extended that agenda further by focusing on pace of ageing rather than static age estimation (Belsky et al. 2020). What this manuscript shows is that a much simpler, non-molecular framework can still recover meaningful chronological age-related variation in a large Indian cohort. That does not place it in competition with molecular clocks. It places it in a different part of the translational spectrum: less granular biologically, but substantially more scalable in low-resource and field settings.

This positioning matters because the added value of the manuscript is not novelty in the abstract, but context and feasibility. Much of the best-known age-clock literature comes from high-income settings and depends on assays that are not easily embedded in routine fieldwork. By contrast, the current framework is built from measures that are already familiar to population ageing studies and closer to what can realistically be implemented at scale. In that respect, the manuscript contributes a useful Indian cohort-based demonstration that routinely collected functional and physiological variables can support chronological age prediction without the infrastructure required for methylation-based models.

Residual behavior, implications and limits

The residual plots add an important layer of reassurance. The absence of strong age-dependent curvature suggests that the selected models were not missing a large, obvious non-linear structure across the observed age range. That observation is consistent with the relative success of Linear Regression. At the same time, the residual figures also show that the densest information lay in the middle of the age distribution, while the youngest and oldest age bands were sparse. That means the model should be interpreted as strongest in the age range best represented by the data, and more tentatively at the extremes. A useful next step would be subgroup diagnostics by age band and sex, because even when overall residual curvature is modest, calibration can still vary meaningfully across subpopulations.

The study also has clear strengths. It uses LASI Wave 1, a major cohort for ageing research in India; it integrates functional, physiological and selected biomarker domains; it documents a transparent modeling pipeline; and it compares complete-case and median-imputed strategies as well as linear and non-linear regressors. The leakage audit and negative-control shuffle are particularly important because they strengthen confidence that the model is learning age-related structure rather than a trivial derivation of the target. Still, the limitations remain substantial. The design is cross-sectional, so the framework cannot speak to ageing trajectories. Validation is internal rather than external. The modeling is unweighted, so the results should not be treated as nationally representative estimates. Complete-case selection may bias the analytic sample toward healthier or more test-compliant participants. Median imputation is convenient but blunt, and it may attenuate variance and flatten informative missingness. Measurement variability in field-collected functional and spirometric data is also unavoidable. Finally, because the target is chronological age, the model cannot be used to infer frailty, risk or accelerated ageing without independent validation against those outcomes.

Future directions and closing perspective

Future work should therefore move in two directions at once. One is methodological: external validation, survey-weighted modeling, multiple imputation, subgroup analysis and calibration checks across sex, region and rural-urban strata. The other is substantive: linking predicted age or age residuals to frailty, disability, morbidity, hospitalization and mortality, ideally using longitudinal LASI follow-up. That would bring the framework closer to the kind of outcome relevance that distinguishes stronger ageing measures in the broader literature, including frailty indices and molecular clocks (Searle et al. 2008; Levine et al. 2018). It would also create a more meaningful basis for comparing low-cost field models with biomarker-heavy or methylation-based approaches.

Overall, the study makes a credible and useful contribution by showing that chronological age-related variation in LASI Wave 1 can be modeled effectively using a small set of interpretable, accessible indicators. The main message is not that simple measures replace richer biological assays, but that they can provide a practical research-oriented framework for studying ageing in large population cohorts where logistics and scalability cost matter. With longitudinal validation and stronger linkage to meaningful outcomes, this line of work could become a valuable bridge between functional epidemiology and more resource-intensive ageing science.

4. Conclusion

This study developed a reproducible LASI Wave 1-based framework for predicting chronological age using multimodal ageing-related indicators. The findings show that a small set of field-compatible variables, particularly sex, grip strength, walking speed, sleep problem score, pulse rate and BMI, can capture substantial age-related variation within the analyzed cohort. The Base Complete-Case Linear Regression model performed best, suggesting that interpretable models may be sufficient when predictors are carefully selected and data quality is maintained. The weaker performance of extended and imputed models highlights the influence of missingness, measurement variability and sample composition in ageing research. Importantly, the model should not be interpreted as a biological-age estimator or clinical diagnostic tool. Instead, it offers a transparent, research-oriented approach for studying age-related functional and physiological variation in an Indian population dataset. Future work should validate the framework externally, incorporate longitudinal LASI data, apply survey-weighted methods and examine associations with frailty, disability and health outcomes.

References

1. Araque O, Sánchez-Echeverri LA, Cerón IX (2026) Cardiometabolic Risk Determinants in a University Community: Beyond Chronological Age to Anthropometric Impact. Healthcare 14:1002. https://doi.org/10.3390/healthcare14081002

2. Behrenbruch N, Schwarck S, Schumann-Werner B, et al (2026) A physically and mentally active lifestyle relates to younger brain and cognitive age. GeroScience 48:1853–1873. https://doi.org/10.1007/s11357-025-01764-w
3. Belsky DW, Caspi A, Arseneault L, et al (2020) Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. Elife 9:e54870. https://doi.org/10.7554/eLife.54870
4. Bennett JP, Lim S (2025) The Critical Role of Body Composition Assessment in Advancing Research and Clinical Health Risk Assessment across the Lifespan. J Obes Metab Syndr 34:120–137. https://doi.org/10.7570/jomes25010
5. Bohannon RW (2019) Grip Strength: An Indispensable Biomarker For Older Adults. Clin Interv Aging 14:1681–1691. https://doi.org/10.2147/CIA.S194543
6. Bui DS, Perret JL, Walters EH, et al (2022) Association between very to moderate preterm births, lung function deficits, and COPD at age 53 years: analysis of a prospective cohort study. The Lancet Respiratory Medicine 10:478–484. https://doi.org/10.1016/S2213-2600(21)00508-7
7. Cheema BS, Boztepe B, Awofolaju MO, et al (2026) Functional, molecular, and digital measurements of biological age. J Clin Invest 136:. https://doi.org/10.1172/JCI205777
8. Dodds L, Brayne C, Siette J (2024) Associations between social networks, cognitive function, and quality of life among older adults in long-term care. BMC Geriatr 24:221. https://doi.org/10.1186/s12877-024-04794-9
9. Dror DM (2025) Breaking the Mortality Curve: Investment-Driven Acceleration in Life Expectancy and Insurance Innovation. Risks 13:122. https://doi.org/10.3390/risks13070122
10. Feng LB (2025) The correlation between different lifestyles and body composition focuses on eating habits, nutritional status, and physical exercise components. Hormones 24:621–641. https://doi.org/10.1007/s42000-025-00661-3
11. França AS de, Saldanha AS, Silva AM da, et al (2026) Associations of cardiovascular risk factors with handgrip strength and gait speed among older males and females: A systematic review protocol. PLOS ONE 21:e0344309. https://doi.org/10.1371/journal.pone.0344309
12. Fried LP, Borhani NO, Enright P, et al (1991) The cardiovascular health study: Design and rationale. Annals of Epidemiology 1:263–276. https://doi.org/10.1016/1047-2797(91)90005-W
13. Fu Q, Yang Y, Liu F, et al (2025) Association between triglyceride-glucose index and early vascular aging in patients with type 2 diabetes mellitus: a retrospective cross-sectional study. BMC Cardiovasc Disord 25:828. https://doi.org/10.1186/s12872-025-05323-1
14. Garcia-Aymerich J, Heras M de las, Carsin A-E, et al (2025) General population-based lung function trajectories over the life course: an accelerated cohort study. The Lancet Respiratory Medicine 13:611–622. https://doi.org/10.1016/S2213-2600(25)00043-8
15. Gianfredi V, Nucci D, Pennisi F, et al (2025) Aging, longevity, and healthy aging: the public health approach. Aging Clin Exp Res 37:125. https://doi.org/10.1007/s40520-025-03021-8
16. Goodarzi F, Khoshravesh S, Ayubi E, et al (2024) Psychosocial determinants of functional independence among older adults: A systematic review and meta-analysis. Health Promot Perspect 14:32–43. https://doi.org/10.34172/hpp.42354
17. Grinin L, Grinin A, Korotayev A (2024) Global Aging, Adaptation to It and Future Demographic Transformations. In: Grinin L, Grinin A, Korotayev A (eds) Cybernetic Revolution and Global Aging: Humankind on the Way to Cybernetic Society, or the Next Hundred Years. Springer International Publishing, Cham, pp 269–327
18. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14:R115. https://doi.org/10.1186/gb-2013-14-10-r115

19. India State-Level Disease Burden Initiative Collaborators (2017) Nations within a nation: variations in epidemiological transition across the states of India, 1990-2016 in the Global Burden of Disease Study. Lancet 390:2437–2460. https://doi.org/10.1016/S0140-6736(17)32804-0
20. Jazayeri E, Kazemipour S, Hosseini SR, Radfar M (2023) Quality of life in the elderly: A community study. Caspian J Intern Med 14:534–542. https://doi.org/10.22088/cjim.14.3.543
21. Jugran D (2025) Too well to die; too ill to live: an update on the lifespan versus health span debate. J Glob Health 15:03022. https://doi.org/10.7189/jogh.15.03022
22. Kapan A, Ristic M, Leser A, et al (2025) Assessment of muscle fatigability using isometric repetitive handgrip strength in frail older adults. A cross-sectional study. J Transl Med 23:215. https://doi.org/10.1186/s12967-025-06239-2
23. Kotschy R, Bloom DE, Scott AJ (2025) On the Limits of Chronological Age. Annual Review of Economics 17:151–173. https://doi.org/10.1146/annurev-economics-080824-112429
24. Kulecki M, Naskret D, Mietkiewska-Dolecka M, et al (2026) Visceral fat rating, but not glycemic variability, is strongly associated with arterial stiffness in non-obese adults with type 1 diabetes. Sci Rep 16:16995. https://doi.org/10.1038/s41598-026-47105-5
25. Levine ME, Lu AT, Quach A, et al (2018) An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY) 10:573–591. https://doi.org/10.18632/aging.101414
26. Li G, Cheng L, Wong IN, et al (2026) Predicting healthspan and disease risks through biological age. Trends in Molecular Medicine 32:354–369. https://doi.org/10.1016/j.molmed.2025.10.006
27. Li Z, Zhang W, Duan Y, et al (2023) Progress in biological age research. Front Public Health 11:. https://doi.org/10.3389/fpubh.2023.1074274
28. Mahbub TBB, Safaeian P, Sohrabi S (2026) A comprehensive review of artificial intelligence as a catalyst in aging research: insights, gaps and future perspectives. Front Aging 7:. https://doi.org/10.3389/fragi.2026.1644669
29. Meng D, Zhang S, Huang Y, et al (2024) Application of AI in biological age prediction. Current Opinion in Structural Biology 85:102777. https://doi.org/10.1016/j.sbi.2024.102777
30. Muollo V, D'Emanuele S, Ghiotto L, et al (2025) Evaluating handgrip strength and functional tests as indicators of gait speed in older females. Front Sports Act Living 7:. https://doi.org/10.3389/fspor.2025.1497546
31. Pan W, Yang Y, Song X, et al (2025) C-reactive protein-albumin-lymphocyte (CALLY) index predicts overall survival in elderly Japanese patients with dysphagia: a retrospective cohort study. Front Nutr 12:. https://doi.org/10.3389/fnut.2025.1681956
32. Perianayagam A, Bloom D, Lee J, et al (2022) Cohort Profile: The Longitudinal Ageing Study in India (LASI). Int J Epidemiol 51:e167–e176. https://doi.org/10.1093/ije/dyab266
33. Qiu W, Chen H, Kaeberlein M, Lee S-I (2023) ExplaiNAble BioLogical Age (ENABL Age): an artificial intelligence framework for interpretable biological age. The Lancet Healthy Longevity 4:e711–e723. https://doi.org/10.1016/S2666-7568(23)00189-7
34. Raisi-Estabragh Z, Szabo L, Schuermans A, et al (2024) Noninvasive Techniques for Tracking Biological Aging of the Cardiovascular System. JACC: Cardiovascular Imaging 17:533–551. https://doi.org/10.1016/j.jcmg.2024.03.001
35. Ribeiro O, Araújo L (2025) The Longevity Revolution: Challenges and Opportunities for an Inclusive Ageing Society. In: Casais B (ed) Integrated Science to Achieve Sustainable Development Goals. Springer Nature Switzerland, Cham, pp 165–184
36. Rutledge J, Oh H, Wyss-Coray T (2022) Measuring biological age using omics data. Nat Rev Genet 23:715–727. https://doi.org/10.1038/s41576-022-00511-7

37. Searle SD, Mitnitski A, Gahbauer EA, et al (2008) A standard procedure for creating a frailty index. BMC Geriatr 8:24. https://doi.org/10.1186/1471-2318-8-24
38. Si W, Lai W, Wu L, et al (2026) Association between hemoglobin A1c to HDL cholesterol ratio and hypertension in middle-aged and older Chinese adults: a cross-sectional study. BMC Cardiovasc Disord. https://doi.org/10.1186/s12872-026-06174-0
39. Teschendorff AE, Horvath S (2025) Epigenetic ageing clocks: statistical methods and emerging computational challenges. Nat Rev Genet 26:350–368. https://doi.org/10.1038/s41576-024-00807-w
40. Turkiewicz A, Magnusson K, Timpka S, et al (2025) Physical health in young males and risk of chronic musculoskeletal, cardiovascular, and respiratory diseases by middle age: A population-based cohort study. PLOS Medicine 22:e1004517. https://doi.org/10.1371/journal.pmed.1004517
41. Uluç K, Akkütük Öngel E, Çolakoğlu ŞM, et al (2025) Effects of HALP Score, C-Reactive Protein/Albumin Ratio, and Platelet/Lymphocyte Ratio on Predicting Mortality in Geriatric Patients in the Respiratory Intensive Care Unit. Clinical Interventions in Aging 20:815–823. https://doi.org/10.2147/CIA.S482214
42. World Health Organization (2021) Decade of healthy ageing: baseline report. World Health Organization
43. Yang S, Liu H, Liang Y, et al (2025) Association Between Glycated Hemoglobin and Diabetic Retinopathy in Individuals with Diabetes: A Focus on the Modifying Effect of Ambulatory Blood Pressure. Clinical Interventions in Aging 20:1029–1038. https://doi.org/10.2147/CIA.S514622
44. Zhang W, Wang A (2025) Functional ability of older adults based on the World Health Organization framework of healthy ageing: a scoping review. J Public Health (Berl) 33:1513–1531. https://doi.org/10.1007/s10389-023-02121-x