Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deciphering Accelerated Biological Ageing in Breast Cancer Survivors and Cancer-Free Women Using Explainable Machine Learning

Nur Afifah Mursyida Zaujan
Universiti Kebangsaan Malaysia
Mohd Razif Shahril
razifshahril@ukm.edu.my

Universiti Kebangsaan Malaysia
Suzana Shahar
Universiti Kebangsaan Malaysia
Ezanee Azlina Mohamad Hanif
Universiti Kebangsaan Malaysia
Nur Fa'izah Ab Muin
Hospital Canselor Tuanku Muhriz (HCTM), Universiti Kebangsaan Malaysia
Razinah Sharif
Universiti Kebangsaan Malaysia
Ponnusamy Subramaniam
Universiti Kebangsaan Malaysia
Azizi Abdullah
Universiti Kebangsaan Malaysia
Michael Fenech
Adelaide University

Article

Keywords: Age acceleration, ageing phenotypes, behavioural determinants of health, telomere length, machine learning

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11074083/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

Breast cancer survivors are at increased risk of accelerated biological ageing, yet the interconnected variables driving this process remain poorly mapped. This study aimed to decipher the multidimensional factors associated with accelerated biological ageing using explainable Machine Learning (ML) architectures.

Methods

As part of the Cancer Survivors' Trajectories of Ageing Research (C*STAR) study, a cross-sectional analysis was conducted among 262 breast cancer survivors and 262 cancer-free controls. Ageing phenotypes, behavioural lifestyle factors, sociodemographic and clinical characteristics were assessed, utilising absolute telomere length as the primary ageing biomarker. An XGBoost model integrated with the Synthetic Minority Over-sampling Technique for Regression with Gaussian Noise (SMOGN) was optimised using Particle Swarm Optimisation (PSO) to map non-linear associations, while SHapley Additive exPlanations (SHAP) evaluated feature importance.

Results

The XGBoost-SMOGN pipeline substantially reduced the mean absolute error for breast cancer survivors (0.0070 vs. 0.1785) and controls (0.0074 vs. 0.1592) compared to native models. As synthetic sampling was applied globally, these metrics reflect internal feature mapping rather than generalisable predictive accuracy. SHAP analysis deciphered that accelerated ageing (shortened telomeres) in survivors was primarily associated with impaired functional status, excessive sleep, high stress, radiotherapy, and high processed meat consumption. Paradoxically, optimal cognitive status and shorter sitting durations were also linked to accelerated ageing in survivors. Decelerated ageing in survivors correlated with low sleep disturbance and specific preserved cognitive domains. For controls, accelerated ageing was linked to poor sleep efficiency, a family history of diabetes, and high confectionery/condiment intake. Conversely, decelerated ageing in controls was strongly associated with high physical fitness, Malay ethnicity, increased height, and nutrient-dense diets.

Conclusions

Explainable ML provides a robust exploratory framework for deciphering complex, interrelated variables associated with telomere attrition. Deciphering the complex biological, clinical, and behavioural profiles associated with accelerated biological ageing via machine learning can help identify highly vulnerable survivors, paving the way for targeted, personalised survivorship care to preserve telomere health.

Introduction

Breast cancer accounts for approximately 27% of all cancer cases and 16% of cancer-related deaths among females worldwide in 2025 [1]. Research has shifted beyond survival towards understanding biological ageing and long-term health outcomes among cancer survivors [2]. Genomic instability is a central hallmark of cellular ageing and oncogenesis, often driven by progressive chromosomal decay and damaged DNA repair mechanisms [3]. As a surrogate biomarker of biological ageing and cumulative cellular stress, telomere length is a well-established ageing biomarker, which helps understand accelerated biological ageing and is recognised as a sensitive biomarker for better cancer treatment [3]. While shorter telomeres are frequently associated with accelerated biological ageing, telomere dynamics are also influenced by inherited telomere length, immune cell turnover, oxidative stress, inflammation, DNA repair capacity, and telomerase regulation. At the molecular level, telomeric DNA is particularly susceptible to oxidative injury because of its high guanine content. Reactive oxygen species induce 8-oxo-dG formation and single-strand breaks within telomeric repeats, which are inefficiently repaired compared with the rest of the genome, thereby accelerating telomere shortening [4].

Consequently, cancer treatment may result in telomere attrition via DNA damage, chromosomal instability, chronic inflammation, and elevated levels of oxidative stress, leading to cancer progression and recurrence [3]. These may result in cancer and cancer treatment-related physiological stress possess risks of accelerated biological ageing and increasing long-term health issues. A meta-analysis conducted among 3,145 breast cancer survivors ranging from 4.6 to 23 years of survivorship reported that survivors had shorter telomere length compared to age-matched controls and were significantly linked with higher recurrence [6]. The cancer and its treatments synergistically caused detrimental effects on the accelerated ageing processes among cancer survivors compared to the general population. Hence, it is crucial to identify the potential modifiable factors that can minimise telomere dysfunction, particularly among Asian breast cancer survivors.

Growing evidence suggests an ageing phenotype, including cognitive, psychological, body composition, and functional status, plays a critical role in telomere length alterations [7, 8]. Cancer survivors commonly experience cognitive decline compared to the general population due to the neurotoxic effects of cancer treatments [9, 10]. Recent studies also indicate that abnormal body composition is linked to shortened telomere length [7]. However, cancer survivors commonly face challenges in maintaining an ideal body composition due to muscle depletion, weight changes, and sarcopenia. Considering these findings, an optimal ageing phenotype is crucial for enhancing telomere length among breast cancer survivors.

Behavioural determinants of lifestyle, including dietary intake, sleep quality, and physical activity, also influence telomere length [11–13]. Several studies found that nutrient-dense diets, such as the Mediterranean diet, rich with antioxidants and anti-inflammatory nutrients, fruits, and vegetables, improve the regulation of telomere length by reducing oxidative stress and inflammation [11]. Optimal exercise and quality sleep help reduce the risk of shorter telomeres by balancing circadian regulation,

increasing antioxidant defences, and increasing telomerase activity [12, 14]. Whereas cognitive impairments have been reported to activate the hypothalamic-pituitary-adrenal (HPA) axis and release glucocorticoids, promoting shortened telomeres [15]. The findings from the previous studies may underscore the critical role of ageing phenotypes and behavioural lifestyle factors in improving cancer survivorship and delaying cellular ageing among Asian breast cancer survivors.

Despite the widespread prevalence of breast cancer among the Asian community, to the best of our knowledge, the understanding of multidomain ageing phenotypes and behavioural health determinants is largely unknown. Specifically, while predicting biological ageing among cancer survivors is well-documented globally, critical gaps remain in how culture-specific lifestyle factors affect cellular ageing markers such as telomere length among Asian breast cancer survivors. Most research on predicting age-acceleration factors has focused on Western populations, leaving a gap in understanding health risks in Asian populations through an examination of factors that affect telomere length. One strategy to improve health risk assessment is the application of machine learning (ML) architectures, which provide more accurate and reliable predictions. The ML models have been implemented to predict the subtypes and cancer stages based on histopathological images, lifestyle risk factors of early gastric cancer [16], hypertension [17], and semen quality [18]. To date, these studies have developed the most accurate predictive model, which can be applied to ensure consistent and reproducible output and predictions.

However, current predictive models have yet to account for the interplay between behavioural lifestyle determinants and ageing phenotypes, including telomere length as a biological marker, in predicting age acceleration among Asian breast cancer survivors. Existing studies within Asian populations remain limited by an exclusive focus on isolated clinical parameters, such as age, tumour history, histologic type, and cancer treatment, without taking into account ageing phenotypes and behavioural lifestyle parameters [19, 20]. Hence, this study employs an ML predictive model to identify modifiable factors associated with age acceleration, incorporating ageing phenotypes and behavioural lifestyle predictors, among Asian breast cancer survivors and cancer-free controls, to offer novel insights for future interventions.

Methods

Study Design and Participants

A cross-sectional study was conducted among breast cancer survivors and cancer-free controls as part of the Cancer Survivors Trajectories of Ageing Research (C*STAR) study. The participants' recruitment was conducted using purposive sampling at the Clinic Surgery and Oncology, Hospital Chancellor Tuanku Muhriz, Universiti Kebangsaan Malaysia (HCTM UKM), located in Kuala Lumpur, from March 2023 to August 2024.

Breast cancer survivors met the eligibility criteria if they: (1) were female with a primary diagnosis of breast cancer confirmed by histopathological report (any stage at diagnosis); (2) had completed active

cancer treatment (surgery, chemotherapy, and/or radiotherapy) at least six months before recruitment; and (3) had no evidence of cancer recurrence or secondary malignancy. Cancer-free females were eligible if they had no history of cancer, as confirmed clinically. Exclusion criteria for both groups included: (1) limb deformity; (2) inability to ambulate independently; (3) pregnancy or breastfeeding; and (4) diagnosed psychiatric or neurological disorders. All participants were Malaysian females aged 18–60 years who were able to read and communicate in either Malay or English. Cancer-free controls were recruited from the same clinical settings and consisted of individuals diagnosed with benign breast conditions or those undergoing routine breast cancer screening. Cancer-free status was confirmed by medical doctors. Both groups were frequency-matched by age group (± 5 years) and ethnicity.

Sample size estimation was conducted using G*Power software version 3.1.9.7 (F test: linear multiple regression, fixed model, R^2 deviation from zero). Assuming a medium effect size ($f^2 = 0.15$), 25 predictors, a significance level of 0.05, and 80.0% statistical power, a minimum sample size of 524 participants was required. After accounting for an anticipated attrition rate of 10%, 262 participants were recruited for each group of breast cancer survivors and cancer-free controls. This study follows the guidelines of the STROBE checklist for observational studies in epidemiology [21]. The study protocol has been described in more detail elsewhere [22].

Data collection and study tools

All eligible participants received a participant information sheet and provided written informed consent prior to enrolment. Ageing phenotypes were evaluated using cognitive status, body composition, physical fitness, and functional status, whereas behavioural lifestyle factors comprised dietary intake, sleep patterns, physical activity, and perceived stress. The estimated time to complete these assessments was within the range of 45 to 60 minutes per participant. Following that, a total of 3 mL of non-fasting blood samples were collected from the participants by a trained phlebotomist through venipuncture into vacutainers. The collected data of sociodemographic profiles including age, household income (RM), ethnicity, educational level, non-communicable diseases (NCD), family history of diseases, and marital, working and alcoholic status, whereas the histopathological reports and medical records of eligible breast cancer survivors were recorded, which including patient identification numbers, cancer subtype, stage, receptor status, treatment modalities, and date of diagnosis.

The cognitive status of participants was assessed via the Malay version of Mini-Mental State Examination (M-MMSE), the Malay version of the Rey Auditory Verbal Learning Test (M-RAVLT), digit symbol test, digit symbol test, and Trail Making Test (TMT-A and TMT-B). The MMSE was administered to measure global cognitive function and visuospatial skills [23]. The cognitive cut-off for high educational level participants (secondary education or higher): 23–30 (normal), 18–22 (mild), 11–17 (moderate), < 11 (severe); meanwhile for low educational level (primary education or less: 20–30 (normal), 15–19 (mild), 10–14 (moderate), and < 7 (severe). The categorisation of “normal” is considered normal cognition, whereas “mild”, “moderate”, and “severe” are considered impaired cognition. An unauthorized version of the Malay MMSE was used by the study team without permission; however, this has now been

rectified with PAR. The MMSE is a copyrighted instrument and may not be used or reproduced in whole or in part, in any form or language, or by any means without written permission of PAR (www.parinc.com). The M-RAVLT assessed the cognitive domains of immediate memory, delayed recall, and recognition, with cognitive impairment defined as scoring 1.5 standard deviations below the control group's mean score. The digit symbol test was used to evaluate the processing speed cognitive domain, with cognitive impairment defined as scoring 1.5 standard deviations below the control group's mean score. The digit span aimed to assess the attention and concentration cognitive domains, and scores ≤ 6 indicate impaired memory [24]. The cognitive areas of motor and visual search speed were the focus TMT A, whereas TMT B was administered as a measure of executive abilities, assessing set-shifting and inhibition [25]. The cut-off scores for TMT A were > 78 seconds as cognitively impaired, whereas > 273 seconds for TMT B.

The Lawton Instrumental Activities of Daily Living (IADL) questionnaire was deployed to assess the functional status of the participants [26]. The Lawton IADL scale examined the eight domains of the capacity to use a telephone, shop, prepare food, clean house, do laundry, use a mode of transportation, be responsible for one's medications, and manage finances. With a total score of eight for independence across all domains, one point is awarded for each area of independence. A zero score is awarded if the participant requires assistance and has an impairment.

The dietary intake of 203 food items, which reflect the Malaysian diet, was assessed using the Malay version of the Malaysian Cohort Project Food Frequency Questionnaire (FFQ) [27]. The structure for responding to the items was based on the amount of food and beverages consumed daily, weekly, and monthly. The participants were required to estimate their average frequency of food item intake using the local household measurement photographs to improve estimation accuracy. The serving size (g) for each food item was calculated using a database which was developed based on the Nutrient Composition of Malaysian Food [28, 29], United States Department of Agriculture (USDA) nutrient database [30] and the Singapore Food Database FOCOS [31]. The current study examines the Nutrient Composition of Malaysian food, focusing on most local and traditional dishes and condiments that contribute to the Malaysian dietary pattern. Meanwhile, for food items not available in this reference, such as cereals and Western fast food, nutrient information was obtained from the USDA Nutrient Database and the Singapore Food Database FOCOS. For analysis purposes, the data on these nutrients and food groups were adjusted for energy intake. The plausible intake was categorised as under-reported (< 500 kcal/day), normal (500–3500 kcal/day), and over-reported (> 3500 kcal/day). Those who had under-reported and over-reported were excluded from the analysis.

The Malay-Pittsburgh Sleep Quality Index (PSQI) was used to assess sleep quality among participants [32]. Sleep duration was categorised into insufficient sleep (< 6 hrs), normal sleep (6–8 hrs), and long sleep (> 8 hrs); whereas sleep efficiency was categorised into very good (> 85%), good (75% to 84%), poor (65% to 74%), and very poor (< 65%). A global score ≥ 5 indicates poor sleep quality. Physical activity among participants was observed via the Malay-translated International Physical Activity Questionnaire (IPAQ-M) [33, 34]. Metabolic Equivalent of Task (MET-minutes per week) scores were

calculated and categorised as low (< 600 MET-minutes/week), moderate (600–3000 MET-minutes/week), and high (> 3000 MET-minutes/week). Malay version of the Perceived Stress Scale 10 (PSS-10) was employed to assess the participants' psychological stress and their ability to handle stress and act in a stressful situation [35]. The overall score was categorised into three groups: low stress (0–13), moderate stress (14–26), and high stress (27–40).

The anthropometric and body composition measurements were conducted by measuring weight (kg), height (cm), Body Mass Index (BMI) (kg/m^2), waist circumference (cm), percentage of fat mass (%), and muscle mass (kg). Each participant's height was measured using the portable SECA Stadiometer Model 213 (SECA). The waist circumference was measured using a Lufkin tape (cm) aimed to measure the abdominal fat surrounding the waist and was categorised into low (< 80 cm), intermediate (80–89 cm), and high (abdominal obesity) (≥ 90 cm). Weight, body fat percentage, and muscle mass were measured using a calibrated Tanita SC-330 (TANITA Corporation, Tokyo, Japan) to assess body weight, adiposity, and lean muscle mass, respectively. Body fat percentage was classified as low (< 20%), normal (20–35%), or obese (> 35%) according to the manufacturer's recommendations, while BMI was categorised as underweight (< 18.5 kg/m²), normal (18.5–24.9 kg/m²), overweight (25.0–29.9 kg/m²), or obese (≥ 30.0 kg/m²) based on the World Health Organisation [36].

The two-minute step test (TMST), hand grip strength test (HGS), chair stand test, chair sit-and-reach test, back scratch test, and timed up-and-go test were conducted to measure physical fitness. The TMST was adopted in 1999 as part of the Senior Fitness Test, which is frequently used to test exercise capacity and aerobic endurance [37]. The HGS test assesses upper body functionality and overall muscle strength by determining the maximum amount of voluntary pressure on the dominant hand with a hand dynamometer (Jamar Plus + Hand Dynamometer, SI Instruments Pty Ltd, SA, Australia). Lower body strength, including functional capacity, coordination, and balance, was assessed based on the 30-Second Chair Stand Test [38]. A back-scratch test, chair sit-and-reach test, and Timed Up-and-Go test were deployed to assess the upper and lower body flexibility and agility [39].

The absolute telomere length (aTL) was measured using the quantitative real-time polymerase chain reaction (qPCR) method developed by O'Callaghan and Fenech [40], which quantifies telomere length as kilobases per diploid genome (kb/genome). Briefly, the assay employed telomere and single-copy gene (36B4) amplification with synthetic oligomer standards for absolute quantification. Detailed laboratory procedures, including primer sequences, PCR conditions, and quality control procedures, have been published previously [22].

XGBoost model development

The machine learning model analysis was conducted using several Python libraries, with scikit-learn, XGBoost, Synthetic Minority Over-sampling Technique for Regression with Gaussian Noise (SMOGN), Particle Swarm Optimisation (PSO), and SHapley Additive exPlanations (SHAP). Ageing phenotypes, behavioural lifestyle factors, sociodemographic profiles and clinical characteristics were simultaneously

entered into the XGBoost model, with absolute telomere length as the outcome variable. The datasets underwent a rigorous pre-processing approach including handling missing values by replacing them with the median value, dropping duplicate features (columns) and samples (rows), outlier removal using interquartile range (IQR) and Z-score methods, features normalisation via min-max scaling. To address the data imbalance and skewed distribution of telomere lengths across the heterogeneous study population, the SMOGN technique was applied [41]. SMOGN generated a synthetic dataset for underrepresented telomere length values, minimising skewed distributions. Importantly, SMOGN was applied to the full dataset prior to model training to comprehensively map the complete feature space and enhance the detection of complex multidomain interactions. These datasets were subsequently split into a training and testing dataset (80:20), with the training dataset divided into training and validation datasets (80:20) for hyperparameter tuning and model optimisation. The XGBoost models were optimised using a cross-validated grid search via a PSO approach, which iteratively searched for the best hyperparameter tuning. Because synthetic sampling was applied globally, the resulting performance metrics reflect internal algorithmic fit rather than external predictive generalisation. Consequently, model interpretability relied heavily on SHAP analysis to evaluate the relative contribution and directional association of each feature to telomere length [42]. The SHAP bar plot ranks the features according to their importance in decreasing order. The SHAP summary plots represent the association between features and telomere length, with point colour representing the feature value (red = high; blue = low), while positive and negative SHAP values indicate contributions towards longer and shorter telomere length, respectively.

Statistical analysis

The statistical analyses were conducted using IBM SPSS Statistics for Windows, version 29.0 (IBM Corp., Armonk, NY, USA). Normality of continuous variables was assessed using the Kolmogorov–Smirnov test, with a $p > 0.05$ indicating normality. Descriptive statistics were presented as means and standard deviations for normally distributed variables, and medians with interquartile ranges for non-normally distributed variables. Categorical variables were summarised using frequencies and percentages. Between-group comparisons were conducted using independent-samples t-tests and Mann–Whitney U tests, with statistical significance set at $p < 0.05$. The ML model's performance was then evaluated and compared based on their MAE values, where lower values indicate better predictive accuracy.

Results

Participants characteristics

Both groups were comparable in most sociodemographic characteristics (Table 1). The median age was 52.5 and 51.0 years among breast cancer survivors and cancer-free controls, respectively, and the median household income was RM4,000 (approximately USD985) in both groups ($p > 0.05$). The majority of participants were Malay (75.2% vs. 75.2%), married (80.2% vs. 80.2%), had had tertiary education

(51.9% vs. 57.3%), were non-smokers (98.5% vs. 97.7%), and did not consume alcohol (93.5% vs. 94.7%), with no significant between-group differences ($p > 0.05$). Compared with cancer-free controls, survivors were significantly less likely to be unemployed (18.7% vs. 40.0%; $p < 0.05$) and had a higher prevalence of a family history of cancer (31.3% vs. 17.2%) but a lower prevalence of a family history of heart disease (11.8% vs. 22.1%) ($p < 0.05$). The prevalence of diabetes (17.4% vs. 15.1%), hypertension (23.6% vs. 23.9%), heart disease (5.4% vs. 4.2%), and family histories of diabetes (37.9% vs. 42.9%) and hypertension (37.9% vs. 42.5%) did not differ significantly between groups ($p > 0.05$).

Table 1
Socio-demographic profile among survivors of breast cancer (n = 262) and cancer-free controls (n = 262)

Sociodemographic profile	**Survivors of breast cancer**		**Cancer-free controls**		***p***
	***n* (%)**		***n* (%)**		
Ethnicity					1.000[b]
Malay	197	75.2	197	75.2	
Chinese	51	19.5	51	19.5	
Indian	14	5.3	14	5.3	
Marital status					0.647[b]
Single	23	8.8	24	9.2	
Married	210	80.2	210	80.2	
Divorce	16	6.1	20	7.6	
Widow	13	5.0	8	3.1	
Educational status					0.174[b]
None	5	1.9	3	1.2	
Primary school	11	4.2	4	1.4	
Secondary school	110	42.0	105	40.1	
College/University	136	51.9	150	57.3	
Working status					0.001[b*]
Self-employed	95	36.3	18	6.9	
Government	77	29.4	94	35.9	
Private sector	41	15.6	45	17.2	
Unemployed	49	18.7	105	40.0	
Do you or your partner smoke?					
I do not smoke	258	98.5	256	97.7	0.752[b]
My partner smokes	60	22.9	49	18.8	0.282[b]
I smoke (Current smoker)					0.391[b]

[a] Mann-Whitney test; [b] Chi-square test; *p < 0.05: showed significant differences between groups

Sociodemographic profile	Survivors of breast cancer		Cancer-free controls		*p*
	n (%)		*n* (%)		
1–4 stems per day	1	5.6	3	13.6	
5–14 stems per day	16	88.9	19	86.4	
12–24 stems per day	1	5.6	0	0.0	
Ever taken an alcoholic beverage?					
Yes	17	6.5	14	5.3	0.584[b]
No	245	93.5	248	94.7	
Do you have any of the following diseases?					
Diabetes	45	17.4	39	15.1	0.475[b]
Hypertension	61	23.6	62	23.9	
Heart attack	14	5.4	11	4.2	0.539[b]
Do your family members (e.g., mother, father, sibling) have a history of the following diseases?					
Cancer	82	31.3	45	17.2	0.001[b*]
Diabetes	97	37.9	111	42.9	0.238
Hypertension	97	37.9	110	42.5	0.275
Heart disease	28	11.8	53	22.1	0.007[b*]

[a] Mann-Whitney test; [b] Chi-square test; *p < 0.05: showed significant differences between groups

[Table 1]

Clinical characteristics of survivors of breast cancer

Survivors were diagnosed with invasive ductal carcinoma at a mean age of 46.1 years (Standard deviation: 7.97) (Table 2). Nearly half were diagnosed at stage II (48.9%), and the majority had a survival duration of five years or less (67.6%). Most survivors had undergone multimodal treatment, with surgery, radiotherapy, and chemotherapy being the most common combination (44.7%). Among the documented chemotherapy regimens, FEC was the most frequently administered (41.2% of the overall survivor cohort), followed by FEC-T (14.9%). In terms of tumour biology, most survivors were oestrogen receptor-positive (74.0%) and progesterone receptor-positive (66.4%), while approximately 30% were HER2-

positive. A smaller proportion (11.8%) were diagnosed with triple-negative breast cancer, an aggressive subtype associated with poorer prognosis.

Table 2
Clinical characteristics among survivors of breast cancer (n = 262)

Clinical characteristics	***n* (%)**
Stage	
Stage I	89 (34.0)
Stage II	128 (48.9)
Stage III	23 (8.8)
Stage IV	22 (4.4)
Type of cancer	
Invasive Ductal Carcinoma (IDC)	262 (100.0)
Survival duration	
≤ 5 years	177 (67.6)
> 5 years	85 (32.4)
Type of treatment	
Surgery	35 (13.4)
Chemotherapy	17 (6.5)
Radiotherapy	5 (1.9)
Surgery + chemotherapy	46 (17.6)
Surgery + radiotherapy	32 (12.2)
Radiotherapy + chemotherapy	10 (3.8)
Surgery + radiotherapy + chemotherapy	117 (44.7)
Types of chemotherapy drugs	
FEC	108 (41.2)
FEC-T	39 (14.9)
EC	9 (3.4)
EC-T	26 (9.9)

FEC (Fluorouracil + Epirubicin + Cyclophosphamide); FEC-T (Fluorouracil + Epirubicin + Cyclophosphamide + Docetaxel); EC (Epirubicin + Cyclophosphamide); EC-T (Epirubicin + Cyclophosphamide + Docetaxel); TC (Docetaxel + Cyclophosphamide); AC (Doxorubicin + Cyclophosphamide); AC-T (Doxorubicin + Cyclophosphamide + Docetaxel); ER (Oestrogen), PR (Progesterone), HER2 (Human Epidermal Growth Factor Receptor 2), TNBC (Triple Negative Breast Cancer)

Clinical characteristics	*n* (%)
TC	2 (0.8)
AC	1 (0.4)
AC-T	2 (0.8)
Receptor	
ER+ve	194 (74.0)
ER-ve	68 (26.0)
PR+ve	174 (66.4)
PR-ve	88 (33.6)
HER2 + ve	78 (29.8)
HER2-ve	184 (70.2)
TNBC	31 (11.8)
FEC (Fluorouracil + Epirubicin + Cyclophosphamide); FEC-T (Fluorouracil + Epirubicin + Cyclophosphamide + Docetaxel); EC (Epirubicin + Cyclophosphamide); EC-T (Epirubicin + Cyclophosphamide + Docetaxel); TC (Docetaxel + Cyclophosphamide); AC (Doxorubicin + Cyclophosphamide); AC-T (Doxorubicin + Cyclophosphamide + Docetaxel); ER (Oestrogen), PR (Progesterone), HER2 (Human Epidermal Growth Factor Receptor 2), TNBC (Triple Negative Breast Cancer)	

[Table 2]

XGBoost model performance and SHAP interpretation

Compared with the baseline XGBoost model, the XGBoost-SMOGN model substantially reduced the MAE for both breast cancer survivors (0.1785 vs. 0.0070; 96.0% reduction) and cancer-free controls (0.1592 vs. 0.0074; 95.4% reduction). Therefore, the XGBoost-SMOGN model was selected as the final predictive model.

The SHAP interpretation of the XGBoost-SMOGN models represents the selected features and their association with telomere length among survivors and controls (Fig. 1), whereas Fig. 2 illustrates the final predictive framework in this study (Red box: Features of accelerated ageing; Green box: decelerated ageing). For breast cancer survivors, low engagement in vigorous physical activity, impaired functional status, such as the need for assistance in medication and laundry management, excessive sleep duration (hours), low sitting duration (hours), high stress score, low financial management, and the presence of heart disease were associated with shortened telomere length. The SHAP analysis identified radiotherapy as one of the important features associated with telomere length shortening after simultaneously considering other clinical characteristics included in the model, rather than an

independent causal treatment effect. Optimal cognitive status, lower potassium intake, and higher consumption of fish, seafood, eggs, and processed meat are strongly correlated with accelerated telomere shortening. In contrast, the age acceleration features that were associated with decelerated telomere length shortening were having a shorter completion time (in seconds) on the TMT A), low sleep disturbance scores and cereal intake, high cognitive score on concentration (Digit Span) and global cognitive function (MMSE), and more bread spread intake.

For control groups, low sleep efficiency, high sleep disturbance, family history of diabetes, low intake of potassium, calcium, and protein, and high consumption of confectionery and condiments were associated with shortened telomere length. A high score on the attention, verbal learning performance, executive abilities, set-shifting, and inhibition, as assessed by the MMSE, RAVLT, and TMT B cognitive tests, was also associated with shortened telomere length. In contrast, high scores on the physical fitness tests, including the two-minute step test (number of steps), increased centimetres in the back-scratch test (left), and a shorter time (in seconds) to complete the up-and-go test, are correlated with longer telomere length. Similarly, increased intake of non-sugary beverages and non-starchy fruits, as well as an increase in height, associates with delaying the shortening of telomere length. The current study also observed that Malays had longer telomere length as compared to other ethnic groups, including Chinese and Indians.

[Fig. 1]

Figure 1 SHAP bar plot (left) and summary plot (right) for breast cancer survivors (A) and controls (B).

[Fig. 2]

Figure 2 The age acceleration predictive framework covered predictors of accelerated and decelerated ageing

Discussion

The current study aimed to decipher the multidimensional factors associated with accelerated biological ageing among Malaysian adult women breast cancer survivors and cancer-free controls using explainable machine learning. The XGBoost-SMOGN architecture was implemented primarily to map complex, non-linear relationships across interrelated sociodemographic profiles, clinical characteristics, ageing phenotypes, and behavioural lifestyle factors. While the model achieved a near-zero MAE compared to native XGBoost predictive models, this reflects an internally optimised algorithmic fit due to global synthetic sampling, rather than external predictive generalisability [16, 43]. Consequently, the study relies on SHAP values to estimate the relative contribution and directional association of each feature, accounting for the influence of all other variables simultaneously.

The SHAP analysis identified several variables associated with telomere length alteration. Longer sleep duration (excessive sleep) and sleep disturbances were strongly associated with shorter telomeres.

Prolonged sleep (more than 8 hours) correlates with a higher relative mortality risk and age-related diseases, likely driven by increased oxidative stress, chronic inflammation, and altered cytokine levels related to chronic stress [44–46]. Conversely, functional independence in managing medication, laundry, and finances (IADL) correlated with decelerated telomere shortening among survivors. Independently managing complex tasks reduces the psychological burden of household chores and financial hardship, which is a common stressor in cancer survivorship, and this reflects better executive function and organisational capacity [47].

Healthy dietary intake and increased height were associated with longer telomeres in the control group. Higher intakes of potassium, non-starchy fruits, non-sugary beverages, calcium, and protein were positively associated with telomere length, likely due to their antioxidant and anti-inflammatory benefits, promotion of immune resilience, and mitigation of oxidative stress [48–50]. In contrast, higher intakes of refined cereals, processed meat, fried eggs, confectionery, and high sodium or sugar condiments were associated with shorter telomeres. These dietary patterns contribute to advanced glycation end products (AGEs), systemic inflammation, and insulin resistance, negatively impacting telomere maintenance [51, 52].

Clinical correlates of accelerated ageing included perceived stress, heart disease, family history of diabetes, and radiotherapy. Psychological disorders and chronic stress elevate cortisol levels and disrupt the HPA axis [53] while cardiovascular complications and metabolic dysregulation are linked to chronic oxidative damage [54]. Although SHAP analysis identified radiotherapy as a primary feature associated with shortened telomeres, this is unlikely to be an isolated causal effect. Treatment allocation is dictated by cancer staging and overall disease severity [55]. Radiotherapy likely functions within this model as a proxy for the cumulative physiological burden of advanced disease and aggressive multimodal therapy.

Healthy dietary intake and increased height were associated with longer telomeres in the control group. Higher intakes of potassium, non-starchy fruits, non-sugary beverages, calcium, and protein were positively associated with telomere length, likely due to their antioxidant and anti-inflammatory benefits, promotion of immune resilience, and mitigation of oxidative stress [56]. In contrast, higher intakes of refined cereals, processed meat, fried eggs, confectionery, and high sodium or sugar condiments were associated with shorter telomeres. These dietary patterns contribute to advanced glycation end products (AGEs), systemic inflammation, and insulin resistance, negatively impacting telomere maintenance [51, 57]. Clinical correlates of accelerated ageing included perceived stress, heart disease, family history of diabetes, and radiotherapy. Psychological disorders and chronic stress elevate cortisol levels and disrupt the HPA axis [53] while cardiovascular complications and metabolic dysregulation are linked to chronic oxidative damage [54]. Although SHAP analysis identified radiotherapy as a primary feature associated with shortened telomeres, this is unlikely to be an isolated causal effect. Treatment allocation is dictated by cancer staging and overall disease severity [55]. Radiotherapy likely functions within this model as a proxy for the cumulative physiological burden of advanced disease and aggressive multimodal therapy.

The model yielded highly counterintuitive associations among survivors, linking vigorous physical activity, reduced sitting time, and optimal cognitive status to shorter telomere lengths. These findings contradict established literature, which identifies physical activity and optimal cognition as protective against systemic inflammation and telomere shortening [14, 58]. These inverse associations likely reflect the complex realities of post-treatment rehabilitation. The cross-sectional design captures a snapshot of current behaviours alongside accumulated biological damage. It is highly plausible that these findings reflect retroactive lifestyle corrections; survivors who endured the most severe cancer pathology, and thus suffered the most significant telomere attrition, are frequently those most heavily targeted by post-treatment rehabilitation. In an effort to reclaim their health, these survivors may aggressively adopt vigorous exercise routines and engage in cognitive rehabilitation [59]. Thus, the observed healthy behaviours currently correlate with past biological damage. Furthermore, the reliance on self-reported activity recall (IPAQ) may introduce reporting bias compared to objective accelerometers [60], while preserved cognitive function in the presence of accelerated ageing may reflect cognitive reserve, known as the ability to maintain baseline function despite substantial biological distress. An additional paradox was the association between higher bread spread intake and longer telomeres, contradicting evidence that spreads contribute to metabolic stress [61]. This discrepancy may occur if participants consumed plant-enriched, antioxidant-rich spreads, though precise interpretation requires further granular dietary data.

Implication for cancer survivorship care

The findings of this study have direct implications for survivorship care. Deciphering the specific biological, clinical, and behavioural profiles associated with telomere health can assist multidisciplinary teams in identifying at-risk breast cancer survivors early. Incorporating multidomain ageing phenotypes and behavioural assessments in follow-up consultations facilitates the early detection of accelerated ageing, allowing for targeted supportive care. For instance, survivors with poor sleep quality may be referred for cognitive behavioural therapy for insomnia (CBT-I), while those with functional limitations may benefit from structured rehabilitation. Explainable machine learning frameworks can complement conventional clinical assessments by mapping complex risk profiles, enabling highly personalised survivorship interventions.

Study limitations

This study has several limitations. Firstly, the cross-sectional study design limits the ability to draw causal inferences regarding the relationship between cancer, cancer treatment, ageing phenotypes, behavioural lifestyle factors, and biological ageing. The use of self-reported assessments of dietary intake, physical activity, sleep quality, and perceived stress may introduce social desirability and recall bias, which may affect the estimated associations and the predictive model's performance. Hence, future longitudinal studies are warranted by incorporating objective measurements to establish causal associations, and complementary biological ageing biomarkers such as DNA methylation, inflammatory

markers, and the epigenetic clock, to improve the clinical applicability of machine learning models for personalised survivorship care.

The eligibility criteria for breast cancer survivors in this study may have introduced healthy survivor bias. Recruited survivors were among those who completed active cancer treatment at least six months before, to minimise the influence of acute treatment-related effects while evaluating long-term cancer survivorship and to ensure comparability with the cancer-free controls. Participants with an inability to ambulate independently or limb deformity were also excluded, as this study required them to safely complete physical fitness assessments. Subsequently, survivors with severe treatment-related toxicities, frailty and functional impairment may have been underrepresented. Hence, the findings in this study should be generalised primarily to healthy breast cancer survivors in the post-treatment survivorship phase. Future studies remain needed to include survivors across a broader spectrum of health conditions, including those with limb deformities, greater functional limitations or diagnosed psychiatric disorders, thereby improving the representativeness and generalisability of the findings.

One of the major limitations of the study is that telomere length alone was the main measure used of cellular ageing. However, shorter telomeres only result in chromosomal instability if the telomeres of a cell are too short to form the T-loop on which the shelterin complex is assembled. This is critical because the T-loop and a functional shelterin complex assembled upon it are needed to prevent telomere end fusion, which is the mechanism by which dicentric chromosomes are formed. Furthermore, dicentric chromosome formation leads to anaphase bridging of the dicentric chromosomes, leading to their breakage and telomere end fusions. In turn, this results in a persistent cycling of chromosomal instability due to breakage-fusion-bridge cycles [4, 62]. More robust data could be obtained in future studies by including measures of telomere dysfunction, including complete telomere loss due to terminal chromosome breakage, anaphase bridges of dicentric chromosomes, guanine base oxidation of the TTAGGG repeat of telomeres and mutations in the shelterin complex proteins.

A methodological limitation of the current machine learning pipeline is the application of the SMOGN technique to the full dataset prior to the train-test split. While this approach successfully addressed the skewed distribution of telomere lengths and allowed the XGBoost model to comprehensively map non-linear relationships across the entire feature space, it inherently introduces data leakage into the testing subset. As a result, the near-zero mean absolute error (MAE) observed in both groups reflects an internally optimised algorithmic fit rather than a generalizable predictive accuracy for unseen clinical data. Therefore, the findings of this study should be interpreted as an exploratory, hypothesis-generating framework driven by SHAP feature associations, rather than a deployable clinical predictive tool.

Declarations

Ethical approval

The study received ethical approval from the Medical Research and Ethics Committee, Universiti Kebangsaan Malaysia (JEP-2022-700), and was conducted according to the Declaration of Helsinki

(1964) and the Malaysian Clinical Practice guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

This research was supported by the Ministry of Higher Education (MOHE) Malaysia through the Fundamental Research Grant Scheme (FRGS/1/2021/SKK06/UKM/02/10).

Author Contribution

Nur Afifah Mursyida Zaujan: Conceptualisation, methodology, formal analysis, investigation, writing - original draft, visualisation. Mohd Razif Shahril: Conceptualisation, methodology, writing - review and editing, supervision, funding acquisition. Suzana Shahar: Conceptualisation, writing - review and editing, supervision. Ezanee Azlina Mohamad Hanif: Conceptualisation, writing - review and editing. Nur Fa'izah Ab Muin: Conceptualisation, writing - review and editing. Razinah Sharif: Conceptualisation, writing - review and editing. Ponnusamy Subramaniam: Conceptualisation, writing - review and editing. Aziz Abdullah: Conceptualisation, methodology, writing - review and editing, supervision. Michael Fenech: Conceptualisation, writing - review and editing.

Acknowledgement

The authors expressed gratitude to all study participants and the healthcare professionals who assisted in participant recruitment and data collection at the hospital. Special thanks to the Universiti Kebangsaan Malaysia Medical Molecular Biology Institute (UMBI) for providing facilities and support for storing the blood samples, which were used to determine ageing biomarkers of telomere length.

Data Availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

References

1. Jemal, A., Sung, H., Kelly, K., Soerjomataram, I. & Bray, F. The Cancer Atlas. (2025). Available at: https://canceratlas.cancer.org/
2. Guida, J. L. et al. Measuring aging and identifying aging phenotypes in cancer survivors. *Natl. Cancer Inst.* **111** (12), 1245–1253 (2019).
3. Iskandar, M. et al. A Review of Telomere Attrition in Cancer and Aging: Current Molecular Insights and Future Therapeutic Approaches. *Cancers* **17** (257), 1–55. https://doi.org/https://doi.org/10.3390/cancers17020257 (2025).
4. Fouquerel, E. et al. Targeted and Persistent 8-Oxoguanine Base Damage at Telomeres Promotes Telomere Loss and Crisis. *Mol. Cell*. **75**, 1–14 (2019).
5. Zaujan, N. A. M. et al. Comparative Analysis of Aging-related Metabolic and Biological Markers between Cancer Survivors and Healthy Populations: A Systematic Review. *J. Cancer Prev.* **30** (3), 121–137. https://doi.org/https://doi.org/10.15430/JCP.25.012 (2025).
6. Sasmita, D. M. A., Permana, K. G. & Heriyanto, Anwar, S. L. Shorter telomere length as a prognostic marker for survival and recurrence in breast cancer: a systematic review and meta-analysis. *Explor. Target. Anti-tumor Therapy*. **6**, 1002289. https://doi.org/https://doi.org/10.37349/etat.2025.1002289 (2025).
7. Davis, S. K., Xu, R., Khan, R. J. & Gaye, A. Adiposity and Leukocyte Telomere Length in US Adults by Sex-Specific Race/Ethnicity: National Health and Nutrition Examination Survey. *Ethn. Dis.* **30** (3), 441–450. https://doi.org/10.18865/ed.30.3.441 (2020).
8. Rojas, D. M. et al. Short Telomere Length Is Related to Limitations in Physical Function in Elderly European Adults. *Front. Physiol.* **9** (1110), 1–6. https://doi.org/10.3389/fphys.2018.01110 (2018).
9. Janelsins, M. C. et al. The National Cancer Institute clinical trials planning meeting to address gaps in observational and intervention trials for cancer-related cognitive impairment. *J. Natl. Cancer Institution*. **117** (2), 217–228. https://doi.org/https://doi.org/10.1093/jnci/djae209 (2025).
10. Levstek, T., Kozjek, E., Dolˇzan, V. & Podkrajˇsek, K. T. Telomere Attrition in Neurodegenerative Disorders. *Front. Cell. Neurosci.* **14** (219), 1–10. https://doi.org/10.3389/fncel.2020.00219 (2020).
11. Galiè, S. et al. Impact of Nutrition on Telomere Health: Systematic Review of Observational Cohort Studies and Randomized Clinical Trials. *Adv. Nutr.* **11**, 576–601. https://doi.org/https://doi.org/10.1093/advances/nmz107 (2020).
12. Sabot, D., Lovegrove, R. & Stapleton, P. *The association between sleep quality and telomere length: A systematic literature review.* Brain, Behavior, & Immunity - Health, ;28(2023):100577. (2023).
13. Sabot, D., Lovegrove, R., Stapleton, P., Brain & Behavior The association between sleep quality and telomere length: A systematic literature review. & Immunity – Health. ;28(2023):100577. (2023). https://doi.org/https://doi.org/10.1016/j.bbih.2022.100577
14. Song, S., Lee, E. & Kim, H. Does Exercise Affect Telomere Length? A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Medicina* **58** (242), 1–10. https://doi.org/https://doi.org/10.3390/medicina58020242 (2022).

15. Min, J., Kim, J. Y., Choi, J. Y. & Kong, I. D. Association between Physical Activity and Telomere Length in Women with Breast Cancer: A Systematic Review. *J. Clin. Med.* **11** (2527), 1–10. https://doi.org/https://doi.org/10.3390/jcm11092527 (2022).
16. Spiers, J. G., Chen, H. J. C., Sernia, C. & Lavidis, N. A. Activation of the hypothalamic-pituitary-adrenal stress axis induces cellular oxidative stress. *Front. NeuroSci.* **8** (456), 1–6. https://doi.org/10.3389/fnins.2014.00456 (2015).
17. Afrash, M. R., Shafiee, M. & Kazemi–Arpanahi, H. Establishing machine learning models to predict the early risk of gastric cancer based on lifestyle factors. *BMC Gastroenterol.* **23** (6), 1–13. https://doi.org/https://doi.org/10.1186/s12876-022-02626-x (2023).
18. Zhou, M., Yao, T., Fan, W., Guan, Y. & Xu, B. Preliminary prediction of semen quality based on modifiable lifestyle factors by using the XGBoost algorithm. *Front. Med.* **9** (811890), 1–14. https://doi.org/10.3389/fmed.2022.811890 (2022).
19. Li, D. et al. Development and validation of a machine learning-based survival prediction model for Asian glioblastoma patients using the SEER database and Chinese data. *Sci. Rep.* **15**, 31114. https://doi.org/https://doi.org/10.1038/s41598-025-15553-0 (2025).
20. Kadir, M. N. N. A. et al. Development of Predictive Models for Survival among Women with Breast Cancer in Malaysia. *Int. J. Environ. Res. Public Health*. **9**, 15335. https://doi.org/https://doi.org/10.3390/ijerph192215335 (2022).
21. Elm, E. V. et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Ann. Intern. Med.* **147** (8), 573–161. https://doi.org/https://doi.org/10.7326/0003-4819-147-8-200710160-00010 (2007).
22. Zaujan, N. A. M. et al. Ageing phenotypes and behavioural determinants of health on age acceleration: the Cancer Survivors' Trajectories of Ageing Research (C*STAR) study protocol. *BMJ Open.* **16**, e102594. https://doi.org/10.1136/bmjopen-2025-102594 (2026).
23. Zarina, Z., Zahiruddin, O. & Wan, A. C. Validation of Malay Mini Mental State Examination. *Malaysian J. Psychiatry*. **16** (1), 16–20 (2007).
24. Ah, D. V. et al. Association of cognitive impairment and breast cancer survivorship on quality of life in younger breast cancer survivors. *J. Cancer Surviv.* **00** (00), 1–9 (2021).
25. MacPherson, S. E., Allerhand, M., Cox, S. R. & Deary, I. J. Individual differences in cognitive processes underlying Trail Making Test-B performance in old age: The Lothian Birth Cohort 1936. Journal of Intelligence. ;75(2019):23-3226. (2019).
26. Lawton, M. P. & Brody, E. M. Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. *Gerontologist* **9** (3), 179–186 (1969).
27. Shahar, S. et al. Development and Relative Validity of a Semiquantitative Food Frequency Questionnaire to Estimate Dietary Intake among a Multi-Ethnic Population in the Malaysian Cohort Project. *Nutrients* ;**13**. (2021).
28. Siong, T. E., Idris, K., Azudin, M. N. & Noor, M. I. Nutrient composition of Malaysian Foods. 4th edition ed. : Malaysia Ministry of Health. (1997).

29. Ministry of Health Malaysia (MOH). Malaysian Food Composition Database (MYFCD). [Cited 27 October 2024]; (1997). Available from: https://myfcd.moh.gov.my/
30. United States Department of Agriculture (USDA). Food data central USDA. [Cited 12 May 2025]; (2021). Available from: https://fdc.nal.usda.gov/
31. Singapore Food Database Board. Energy and nutrient composition of food: Food Composition System Singapore (FOCOS). [Cited 16 May 2025]; (2021). Available from: https://focos.hpb.gov.sg/eservices/ENCF/
32. Farah, N. M., Yee, T. S. & Rasdi, H. F. M. Self-Reported Sleep Quality Using the Malay Version of the Pittsburgh Sleep Quality Index (PSQI-M) In Malaysian Adults. *Int. J. Environ. Res. Public Health*. **16** (23), 1–10. https://doi.org/10.3390/ijerph16234750 (2019).
33. Craig, C. L. et al. International Physical Activity Questionnaire: 12-Country Reliability and Validity. *Med. Sci. Sports Exerc.* **35** (8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB (2003).
34. IPAQ. Malaysian short International Physical Activity Questionnaire (IPAQ). [cited 25 June 2024]; (2021). Available from: https://sites.google.com/view/ipaq/download?authuser=0
35. Sandhu, S. S., Ismail, N. H. & Rampal, K. G. The Malay Version of the Perceived Stress Scale (PSS)-10 is a Reliable and Valid Measure for Stress among Nurses in Malaysia. *Malaysian J. Med. Sci.* **22** (6), 26–31 (2015).
36. Salleh, Y., Lian, C. W., Shii, L. L., Stephen, J. & Jool, T. W. Demographic, socioeconomic and dietary factors are associated with double burden of malnutrition among indigenous communities in Sarawak. *Malaysian J. Nutr.* **30** (3), 387–402. https://doi.org/https://doi.org/10.31246/mjn-2024-0056 (2024).
37. Bohannon, R. W. & Crouch, R. H. Two-Minute Step Test of Exercise Capacity: Systematic Review of Procedures, Performance, and Clinometric Properties. *J. Geriatr. Phys. Ther.* **0** (0), 1–8. https://doi.org/10.1519/JPT.0000000000000164 (2017).
38. Ensink, I., Rothbauer, M. J. A., Röhlinger, S., Merry, A. H. H. & Sipers, W. M. W. H. The Modified-30-Seconds-Chair-Stand-Test: A Practical and Reproducible Tool to Assess Muscle Strength in Acutely Ill Hospitalized Geriatric Patient. *J. Frailty Sarcopenia Falls*. **10** (1), 8–14. https://doi.org/10.22540/JFSF-10-008 (2025).
39. Shahar, S. et al. Approaches in methodology for population-based longitudinal study on neuroprotective model for healthy longevity (TUA) among Malaysian Older Adults. *Ageing Clin. Experimental Res.* **28** (6), 1089–1104. https://doi.org/10.1007/s40520-015-0511-4 (2015).
40. O'Callaghan, N. J. & Fenech, M. A quantitative PCR method for measuring absolute telomere length. *Biol. Procedures Online*. **13** (3), 1–10 (2011).
41. Branco, P. L. & Torgo, Ribeir, R. P. SMOGN: a Pre-processing Approach for Imbalanced Regression. in Proceedings of Machine Learning Research. University of Porto Porto, Portugal. (2017).
42. Salih, A. M. et al. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Adv. Intell. Syst.* **7** (2400304), 1–8. https://doi.org/10.1002/aisy.202400304 (2025).

43. Cao, K. et al. Prediction of cardiovascular disease based on multiple feature selection and improved PSO-XGBoost model. *Sci. Rep.* **15** (12406), 1–12. https://doi.org/https://doi.org/10.1038/s41598-025-96520-7 (2025).
44. Grandner, M. A. & Drummond, S. P. A. Who Are the Long Sleepers? Towards an Understanding of the Mortality Relationship. *Sleep. Med. Rev.* **11** (5), 341–360. https://doi.org/10.1016/j.smrv.2007.03.010 (2007).
45. Tempaku, P. et al. Long Sleep Duration, Insomnia, and Insomnia with Short Objective Sleep Duration Are Independently Associated with Short Telomere Length. *J. Clin. Sleep Med.* **14** (12), 2037–2045 (2018).
46. Irwin, M. R., Olmstead, R., Carroll, J. E. & Sleep Disturbance Sleep Duration, and Inflammation: A Systematic Review and Meta-Analysis of Cohort Studies and Experimental Sleep Deprivation. *Biol. Psychiatry*. **80** (1), 40–52. https://doi.org/10.1016/j.biopsych.2015.05.014 (2016).
47. Gupta, S., Mazza, M. C., Hoyt, M. A. & Revenson, T. A. The Experience of Financial Stress among Emerging Adult Cancer Survivors. *J. Psychosoc Oncol.* **38** (4), 435–448. https://doi.org/10.1080/07347332.2019.1707928 (2020).
48. Singh, M. K. et al. Fruit Carbohydrates and Their Impact on the Glycemic Index: A Study of Key Determinants. *Foods* **14** (646), 1–22. https://doi.org/https://doi.org/10.3390/foods14040646 (2025).
49. Sabuncular, G., Akbulut, G. & Yaman, M. Effects of beverages and salads on lipid oxidation and glycation during in vitro meat digestion. *Sci. Rep.* **15**, 17005. https://doi.org/https://doi.org/10.1038/s41598-025-98977-y (2025).
50. Noer, E. R., Dewi, L., Indraswari, D. A. & Suroto, Ardiaria, M. High-protein Diet and Exercise Improve Inflammatory Marker and Oxidative Stress After Weight Loss Program in Adolescents Obesity: A Randomized Control Trial. *Malaysian J. Med. Health Sci.* **16** (SUPP14), 1–5 (2020).
51. Rafie, N., Hamedani, S. G., Barak, F., Safavi, S. & Miraghajani, M. Dietary patterns, food groups and telomere length: a systematic review of current studies. *Eur. J. Clin. Nutr.* **71**, 151–158 (2017).
52. Vignesh, A., Amal, T. C., Shanmugam, A., Vasanth, K. & Selvakumar, S. Effects of dietary approaches to prevent hypertension and enhance cardiovascular health. *Discover Food*. **5** (9), 1–14. https://doi.org/https://doi.org/10.1007/s44187-025-00278-w (2025).
53. Mbiydzenyuy, N. E. & Qulu, L. A. Stress, hypothalamic-pituitary-adrenal axis, hypothalamic-pituitary-gonadal axis, and aggression. *Metab. Brain Dis.* **39**, 1613–1636. https://doi.org/https://doi.org/10.1007/s11011-024-01393-w (2024).
54. Almeida, A. J. P. O. et al. Unveiling the Role of Inflammation and Oxidative Stress on Age-Related Cardiovascular Diseases. *Oxidative Med. Cell. Longev. 2020;May* **8**(2020):1–12. https://doi.org/https://doi.org/10.1155/2020/1954398
55. Granata, V. et al. Treatments and cancer: implications for radiologists. *Front. Immunol.* **16**, 1564909. https://doi.org/10.3389/fimmu.2025.1564909 (2025).

56. Baliou, S. et al. The Impact of the Mediterranean Diet on Telomere Biology: Implications for Disease Management—A. *Narrative Rev. Nutrients*. **16** (2525), 1–28. https://doi.org/https://doi.org/10.3390/nu16152525 (2024).
57. Ma, X. et al. Excessive intake of sugar: An accomplice of inflammation. *Front. Immunol.* **13**, 988481. https://doi.org/10.3389/fimmu.2022.988481 (2022).
58. Feng, J. et al. Oxidative stress, the blood brain barrier and neurodegenerative diseases: The critical beneficial role of dietary antioxidants. *Acta Pharm. Sinica B*. **13** (10), 3988–4024 (2023).
59. Fernandes, H. A., Richard, N. M. & Edelstein, K. Cognitive rehabilitation for cancer-related cognitive dysfunction: a systematic review. *Support. Care Cancer*. **23**, 3253–3279 (2019).
60. Rääsk, T. et al. Comparison of IPAQ-SF and Two Other Physical Activity Questionnaires with Accelerometer in Adolescent Boys. *Plos One*. **12** (1), e0169527. https://doi.org/10.1371/journal.pone.0169527 (2017).
61. Woodman, R. et al. Patterns of Dietary Fatty Acids and Fat Spreads in Relation to Blood Pressure, Lipids and Insulin Resistance in Young Adults: A Repeat Cross-Sectional Study. *Nutrients* **17** (869), 1–18. https://doi.org/https://doi.org/10.3390/nu17050869 (2025).
62. Dhillon, V., Bull, C. & Fenech, M. Chapter 10: Telomeres, Aging, and Nutrition, in Molecular Basis of Nutrition and Aging, A Volume in the Molecular Nutrition Series. *Chap* **10**, 129–140. http://dx.doi.org/10.1016/B978-0-12-801816-3.00010-8 (2016).

Figures

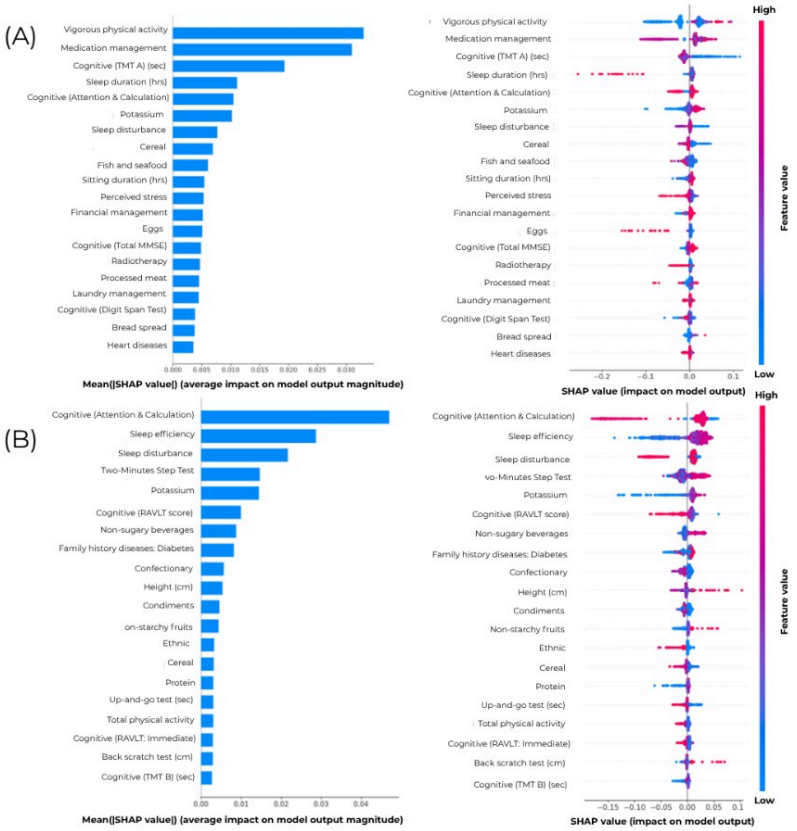


Figure 1

SHAP bar plot (left) and summary plot (right) for breast cancer survivors (A) and controls (B).

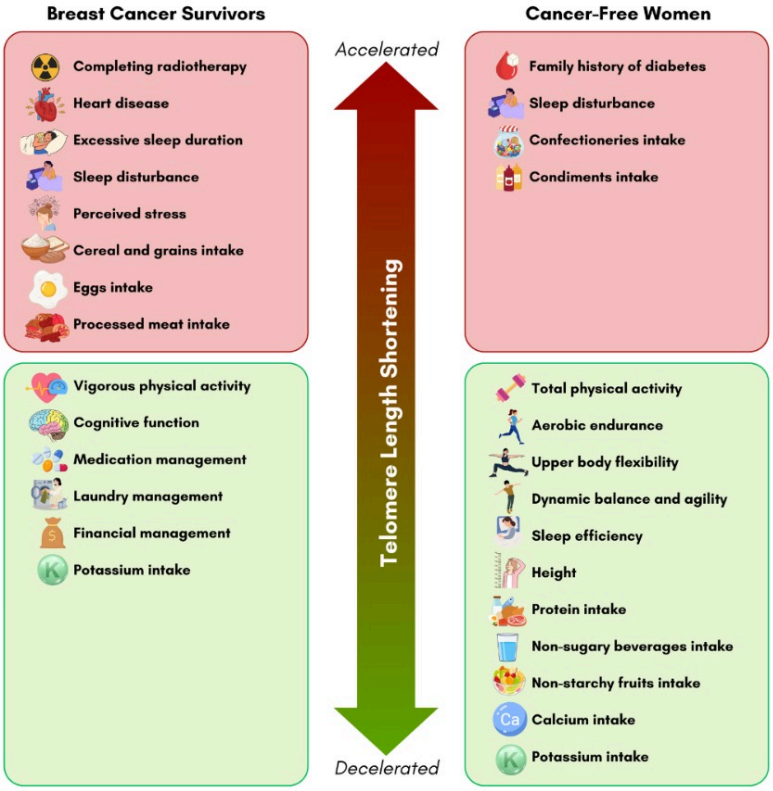


Figure 2

The age acceleration predictive framework covered predictors of accelerated and decelerated ageing