Medical Images as Biomarkers of Ageing - From Global and Local Patterns to Digital Twins

Tamara T. Mueller[1], Sophie Starck[1], Rozafe Llalloshi[1], Georgios Kaissis[1], Alexander Ziller[1], Daniel Rueckert[1,3,*], Rickmer Braren[2,4,*]

[1]Lab for AI in Medicine and Healthcare, Technical University of Munich, Germany.
[2]Institute for diagnostic and interventional radiology, Klinikum rechts der Isar, Technical University of Munich, Germany.
[3]Department of Computing, Imperial College London, United Kingdom.
[4]German Cancer Consortium (DKTK), Munich partner site, Heidelberg, Germany.
[*]equal contributions.

Contributing authors: tamara.mueller@tum.de;

Abstract

Understanding the process of ageing has become a highly desirable goal in human health and disease research. A multitude of factors impact the way we age and understanding their connection can help with early detection of age-related diseases and give insights into what causes abnormal ageing. In this study, we analyse how medical images can be used as biomarkers for ageing via deep learning techniques. We evaluate ageing in different local body systems (liver, lungs, spine, intestine, muscle, heart), the brain, and across the whole body using 70 000 subjects from the UK Biobank population study. We analyse correlations between lifestyle factors –such as smoking– or diseases and accelerated ageing. Furthermore, we show differences in survival between subjects with accelerated and decelerated ageing and study the impact of individual body regions on the whole body age by generating a Digital Twin that can aid personalised medicine. Our work gives insights into how medical imaging can be used to identify abnormal ageing both on a local and a global scale and investigates the interplay between different body regions. It can help understand ageing patterns and lay a foundation for future applications of age predictors for risk assessment.

Keywords: Medical Imaging, Ageing Biomarkers, Accelerated Ageing, Population Studies

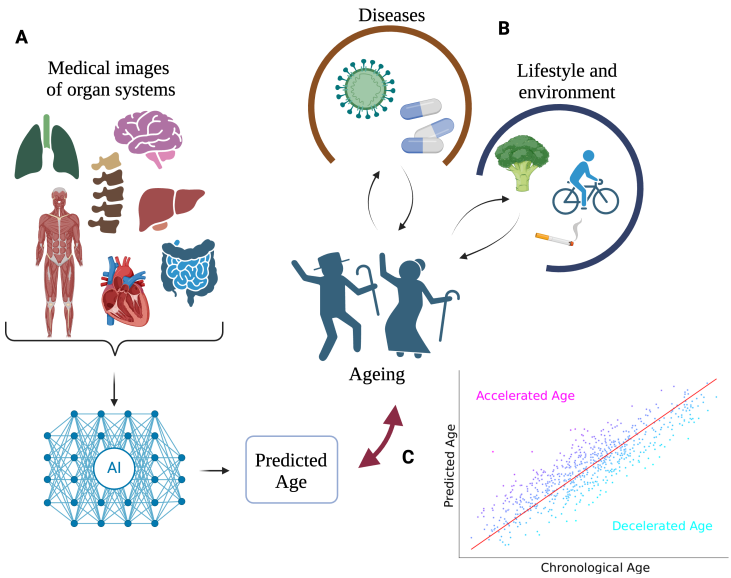


Fig. 1 **A:** We investigate how medical imaging can be used to predict the ageing of the whole body or individual organ systems and **B:** correlate resulting age gaps with diseases and lifestyle features. **C:** We refer to accelerated age if the predicted age is larger than the chronological age and decelerated age if it is the other way around.

1 Introduction

Ageing is an inevitable aspect of life and the most prominent common risk factor for several diseases [1, 2]. It is a complex interplay between the passing of time, genetic factors, lifestyle characteristics, and environmental impacts [3]. It comes with a decline in functional capacity and stress resilience and affects most or all tissues in the body [4]. One goal of ageing research is to distinguish between the *chronological age*, which refers to the time that has passed since birth, and the *biological age*, which is hoped to describe the actual degree of ageing that took place in the cells, tissues, and organs of the body. The aim of estimating a person's biological age is to obtain a more accurate estimator for risk factors for diseases and mortality [5, 6]. The difference between biological age (extracted from the ageing biomarkers) and chronological age is often referred to as the *age gap*. With this distinction, one can analyse accelerated (biological age > chronological age) or decelerated (biological age < chronological age) ageing (see Figure 1). Accelerated ageing can indicate that a person's health, environment or lifestyle negatively impact the natural ageing process and has led to more advanced age-related changes in their body compared to peers. It is challenging to identify reliable biomarkers for ageing that indeed represent the biological age of a person. Therefore, multiple studies have investigated ageing biomarkers in different contexts, including cellular, molecular, and physiological biomarkers [7].

Previous works have defined criteria that a suitable ageing biomarker should meet: [8], for instance, define that ageing biomarkers should be able to predict the outcome of age-sensitive tests better than chronological age and predict longevity at an age at which 90% of the population is still alive. A variety of such ageing biomarkers have been studied and shown promising results in detecting biological age: Burkle et al. [4] introduce MARK-AGE, a project that aimed at identifying a set of ageing biomarkers that measure biological age better than the individual components. They collect physiological data, genomic and protein data, immunological markers, hormones, and metabolic data. Some other works investigate age prediction for individual organs, such as the kidney [9], the lungs [10], or the heart [11], showing promising directions of analysing ageing in more local regions of the human body. Tian et al. [12] were the first to develop a whole-body multi-organ characterisation of ageing. The authors utilise both, non-imaging and image-derived phenotypes and support vector machines.

However, end-to-end deep learning approaches with medical imaging have been used less frequently for age prediction and especially whole-body images are under-explored in this domain. Brain imaging has been, for instance, used to predict brain age and several works have shown how brain age predictions correlate with mortality and (brain-related) diseases [13–17]. Ecker et al. [18] are one of few works using imaging data of different organ systems and deep learning techniques to assess biological age and have shown promising first results in this direction. The authors use different organ systems compared to our study and do not investigate the utility of their age predictions as biomarkers for ageing and the interplay of the different body systems with whole-body ageing.

In this work, we investigate how medical imaging can be used directly as ageing biomarkers and employ end-to-end deep learning techniques to quantify accelerated ageing. We study both local and global components, their interplay, and their connection to lifestyle factors, diseases, and mortality. Furthermore, we build a "Digital Twin", which can be used to assess individual ageing patterns and potentially guide personalised recommendations and interventions. With guided interventions, a slower age-related decline can be targeted, which can lead to more healthy longevity [19]. The source code will be made public upon publication of this work.

2 Results

We use (1) 73 094 whole-body and (2) 45 058 brain MR images from the UK Biobank [20] to estimate the biological age of (a) the whole body, (b) the brain, and (c) different localised body areas: muscle, spine, liver, lungs, heart, and intestine. We train convolutional neural networks on a "healthy" subset of the population to predict their chronological age (see Methods Section) and train individual models of chronological age for each body part (whole-body, brain, muscle, spine, liver, lungs, heart, intestine). We refer to the bias-corrected (Section 4.3) difference between the predicted age and the chronological age as the *age gap* of a subject.

2.1 Age Predictors

For each imaging modality, we train five 3D ResNet-18 models using different randomly initialised networks on the "healthy" sub-population (no ICD-10 records and no self-reported diseases) to predict chronological age. All subjects that have a record of a disease (ICD-10

Table 1 Overview of the age distribution of the specific dataset, the number of subjects in each dataset (train, test, and val), the baseline mean absolute error (MAE) of a naive mean age predictor over the whole dataset ("Mean Pred. MAE"), and the corresponding MAE in years of our models on the validation and test sets, averaged over five random seeds.

Dataset	Age Distribution	Train Samples	Val. Samples	Test Samples	Mean Pred. MAE	Val MAE	Test MAE
Whole Body	66.37 ± 7.95	2841	712	69541	6.62	2.492 ± 1.37	2.564 ± 1.45
Brain	64.88 ± 7.70	1908	466	42684	6.44	2.338 ± 1.35	2.496 ± 1.75
Heart	66.24 ± 7.96	2841	712	33135	6.63	2.634 ± 2.38	3.022 ± 3.34
Spine						3.028 ± 1.71	3.164 ± 2.44
Liver						3.842 ± 1.87	3.929 ± 1.87
Lungs						3.071 ± 1.57	3.152 ± 1.60
Muscle						2.627 ± 1.30	2.722 ± 1.36
Intestine						3.639 ± 1.93	3.668 ± 2.02

code or self-reported diseases) are used as test subjects. The concrete numbers of subjects for each imaging modality can be found in Table 1. Different parts of the human body can potentially age differently [21]. We are therefore interested in *global* ageing patterns –in the whole body and the brain– as well as *local* ageing patterns in different body regions. To investigate the latter, we segment [22] six different body regions in the whole-body MR images and use these regions to predict heart, liver, spine, muscle, intestine and lung age. We here use the same training and validation set as for the whole-body dataset, and a subset of the test set of more than 33 000 subjects.

Brain Age Predictor

The model trained on 1908 T1-weighted brain MR images (48% female, 52% male, 66.37 ± 7.95 years) achieves a mean absolute error (MAE) of 2.50 years on the test set (subjects with disease records (ICD-10 code or self-reported)) and 2.34 years on the validation set (subjects without disease records). The MAE of a baseline mean age predictor (which always predicts the mean of the dataset) is 6.62.

Whole-body Age Predictor

The mean absolute error (MAE) of the whole-body age prediction model (using the entire neck-to-knee images) is 2.49 years on the validation set and 2.56 years on the test set. The MAE of a baseline naive mean age predictor is 6.44 (47% female, 53% male, 64.88 ± 7.70 years).

Body Region Age Predictors

The individual body-region models show generally a slightly higher MAE ranging from 2.72 to 3.93 years than the whole-body (2.56) or the brain age (2.50) predictors since they contain highly local information only. The naive mean age prediction baseline has an MAE of 6.63 years. We note that while we constrain the input of these models to only contain the region of interest in the body, it does not portray a detailed representation of the region of interest and also contains a small portion of the surrounding tissue, due to segmentation inaccuracies and post-processing methods. Further results can be found in the Methods section and the Appendix, including a differentiation between model performances on male and female subjects (Figures 8, A1).

2.2 Interplay of Body Region Ages

Our age predictions indicate that different areas in the body can age differently. In order to investigate the interplay of the ageing of different body regions, we evaluate their correlation across the whole dataset, which is visualised in Figure 2. We observe a higher correlation between age gaps in the localised body regions (spine, intestine, heart, liver and lungs). We furthermore note the high correlation between muscle and whole-body age gaps. We suspect that this is because muscle tissue covers large parts of the body, which can thus be used for the prediction. Additionally, the lung age and the heart age correlate highly. A technical explanation for this is that the heart is partially also visible in the lung images due to their local proximity. Moreover, heart and lungs are closely interacting systems, where deoxygenated blood flows from the heart to the lungs to get enriched with oxygen and then back to the heart, from where it is distributed in the body. These two organs are therefore in constant exchange and strongly influence each other. We observe that the brain age gaps are less correlated (≤ 0.34) to all age gaps from the whole body images. This indicates that the brain might age differently from the body and a subject can have an accelerated brain age and a decelerated body age or the other way around. This aligns with related works such as Elliott et al. [23], where the authors have shown that accelerated brain ageing is correlated with accelerated body ageing, but also only loosely. Furthermore, we observe this also in real-world experiences, where discrepancies between mental and physical abilities can occur.

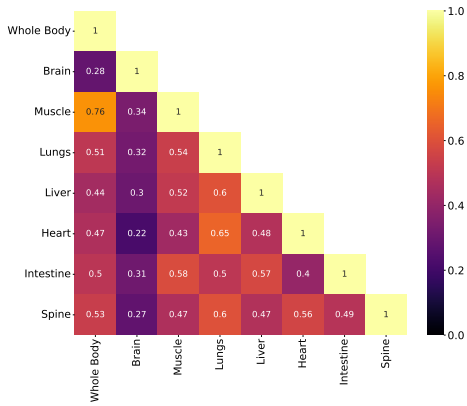


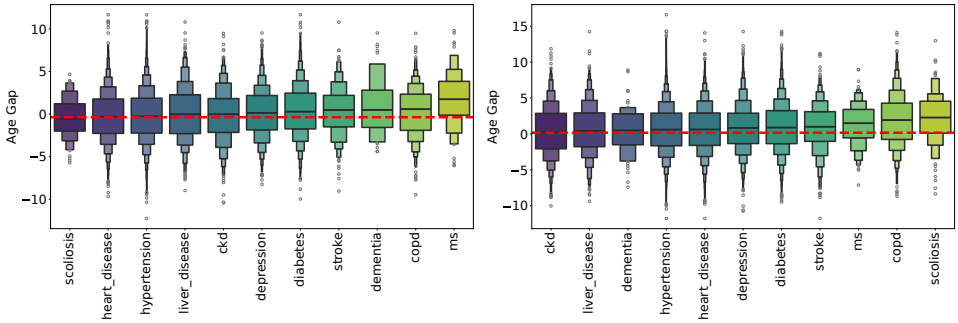
Fig. 2 Visualisation of the correlations between the age gaps of the different body regions. The brain age gaps are less correlated to the body age gaps than the latter among each other, indicating more independent ageing patterns between the brain and the body.

Additionally, we investigate how the different body-region age predictions are connected to whole-body ageing. We, therefore, fit a random forest and a linear regression, as very simple but interpretable models. We use all age predictions apart from the whole body (heart, brain, spine, liver, lungs, muscle, intestine) to predict the whole-body age. Subsequently, we evaluate the impurity-based feature importance values of the body-region ages for the random forest and the coefficients for the linear regression to determine the importance of the individual body regions for the whole-body age. The random forest achieves an R^2 score of 0.917 on the test set and the linear regression an R^2 score of 0.937. Muscle and spine

Table 2 Impurity-based feature importance values of the random forest and the coefficients of the linear regression for the individual body regions that are used to predict the whole-body age. Aligning with the correlation results, the muscle and the spine ages have the strongest impact on the whole-body age prediction. The highest importance value is **bold**, the second-highest underlined.

Method	R2 Score	Heart	Brain	Spine	Liver	Lungs	Muscle	Intestine
Linear Regression	0.937	0.068	0.055	0.139	-0.033	0.053	**0.678**	0.063
Random Forest	0.917	0.007	0.008	0.057	0.007	0.008	**0.903**	0.010

ages have the highest importance values and coefficients in predicting the whole-body age from the individual body-region ages (Table 2). Interestingly, the linear regression assigns the liver age a slightly negative coefficient. This might be because the liver is a comparable small body region, which might not have a strong impact on the here-evaluated whole-body age or because the liver shows different ageing patterns compared to the whole body. This experiment shows that the prediction of the whole-body age from the individual body-region ages is possible with simple methods and high accuracy. However, we note that ageing is highly complex and that this is only an exemplary way of investigating the interplay between the evaluated region ages in this work on a population level.



(a) Brain age gaps of disease groups

(b) Whole-body age gaps of disease groups

Fig. 3 Distribution of age gaps for different diseases for (a) the **brain** age predictions and (b) the **whole-body** age predictions. The dotted horizontal (red) line indicates the average age gap of all subjects in the validation set ("healthy" subjects). The diseases are ordered by ascending median age gap.

2.3 Chronic Diseases

We investigate the interplay between different diseases and accelerated ageing. We select eleven different diseases or disease groups: chronic kidney disease (CKD), multiple sclerosis (MS), diabetes, dementia, stroke, depression, hypertension, ischemic heart disease, scoliosis, chronic obstructive pulmonary disease (COPD), and liver disease (using ICD-10 codes and self-reported diseases from the UK Biobank dataset). For this analysis, we limit the population to be within the 10% and 90% percentile of the age distribution of the training and validation set to reduce the impact of the model bias on the disease-related age gaps. We note that the exact cutoff thresholds are chosen arbitrarily and are only a means to limit the analysis

to a range with sufficient data and less model bias. Age prediction models tend to over-estimate younger subjects and under-estimate older subjects [24, 25]. Specific diseases such as dementia and MS are especially prevalent in older or younger subgroups of the population.

Figure 3(a) shows the *brain age gap* distributions of named diseases. Subjects with MS show the highest median brain age gap of 1.7 years ($n = 76$). This aligns with literature showing that MS patients often show accelerated brain ageing [26, 27]. Other brain-related diseases, such as dementia ($n = 36$), stroke ($n = 236$), and depression ($n = 885$) also show an increased median age gap compared to the validation set, while subjects with body-centred diseases such as scoliosis ($n = 67$), hypertension ($n = 5176$), or heart disease ($n = 1924$) do not show strongly increased brain age gaps. We also observe COPD ($n = 397$) subjects to show accelerated brain ageing, which might be due to a high prevalence of smoking in COPD subjects [28], which is a known risk factor for Alzheimer's disease [29].

Figure 3(b) shows the age gap distributions of the *whole-body age* of the selected diseases. Subjects with scoliosis ($n = 109$) are identified as the sub-group with the highest median whole-body age gap of 2.3 years, followed by COPD subjects ($n = 613$) and MS subjects ($n = 137$). Other works have shown that ML models trained on whole-body images for age prediction have a strong focus on the spine as a region of interest [30]. This might be a factor for scoliosis subjects showing the "strongest" accelerated ageing profile. When analysing disease-specific age gaps for the smaller body regions (lungs, liver, spine, intestine, heart, and muscle), we do not observe a clear correlation between body-region-specific diseases and their age gaps (Section A.2). While subjects with scoliosis show accelerated spine ageing and COPD subjects show accelerated lung ageing, liver-disease subjects do not show accelerated liver ageing and heart-disease and hypertension subjects do not show accelerated heart ageing. There is a multitude of potential reasons for this. For instance, specific diseases might show an impact on the visual appearance of the body region, but this change might not correspond to healthy ageing patterns. Furthermore, the body regions are not independent and while COPD manifests mostly in the lungs, it is often caused by smoking [28], which can affect multiple body regions [29, 31].

2.4 Lifestyle and Environment

We analyse the interplay between different lifestyle and environmental factors (non-imaging) with age gaps of all image modalities by performing a phenome-wide association study (PheWAS). We select a set of 264 non-imaging phenotypes of the UK Biobank of 17 different categories: alcohol, body composition, diet, employment, family history, female- and male-specific factors, general health, medical conditions, mental health, pain, physical activity, sexual factors, sleep, smoking, social support, and sun exposure (see Supplementary Material). Figure 4A shows the Manhattan plot of the negative log of the P-values (two-sided t-test, $P < 2.37 \times 10^{-4}$ Bonferroni-corrected for 264 features $\times$ 8 image modalities) of the non-imaging phenotypes and the modality-based predicted age gaps. The PheWAS results show several statistically significant differences between non-imaging phenotypes and the predicted age gaps of different body parts. Some examples are "fat-free muscle mass" and "muscle fat infiltration" from the body composition features, which show significant differences for all investigated ageing types. Several physical activity features show a significant effect for whole-body, lung, liver, muscle, spine, and brain age gaps. Features indicating general health show a significant difference for the age gaps of all body regions. The dietary features that

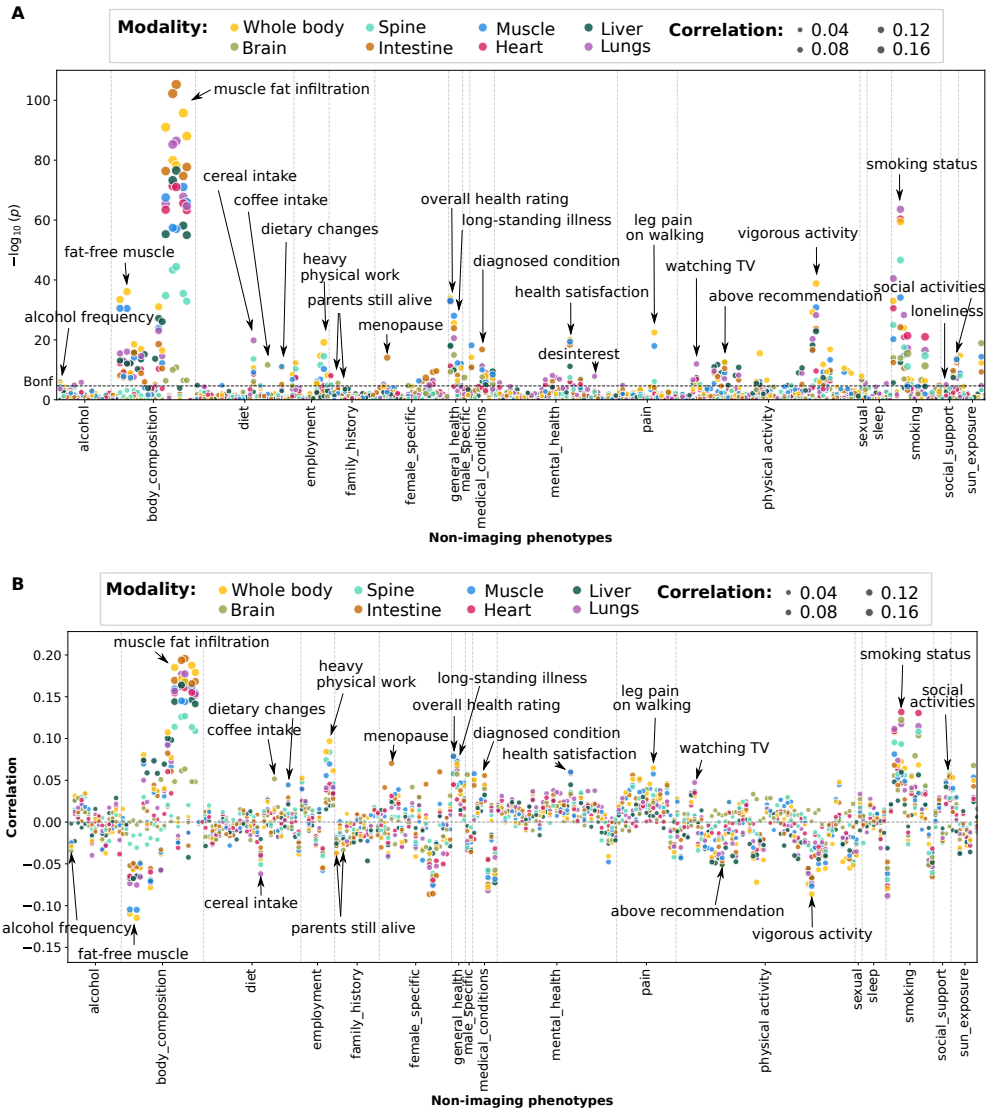


Fig. 4 Results of PheWAS study on non-imaging phenotypes and the age gap for different body parts. **A:** Manhattan plot of the negative logarithm of the *P*-values. The Bonferroni correction is indicated by the horizontal dashed line. **B:** The positive and negative correlation values between the non-imaging features and the age gaps of the different body parts.

show a significant difference are whether a dietary change has happened within the last five years, cereal intake, and coffee intake. Some social factors (social activities and loneliness) show significant differences for accelerated ageing in almost all modalities. This aligns with findings from studies such as [32].

The corresponding correlations (positive and negative) are visualised in Figure 4B. We note that a high value in the UK Biobank feature "alcohol frequency" (ID 1558) indicates rare alcohol consumption. The negative correlation therefore indicates that *frequent* alcohol consumption is *positively* correlated with accelerated ageing. A high fat-free muscle mass is negatively correlated with accelerated ageing, while a high muscle fat infiltration is positively correlated with accelerated ageing. We note that possibly some of these non-imaging features are confounders. People who have an overall low score on "health satisfaction", e.g. might also tend to not be able to exercise a lot or experience more pain.

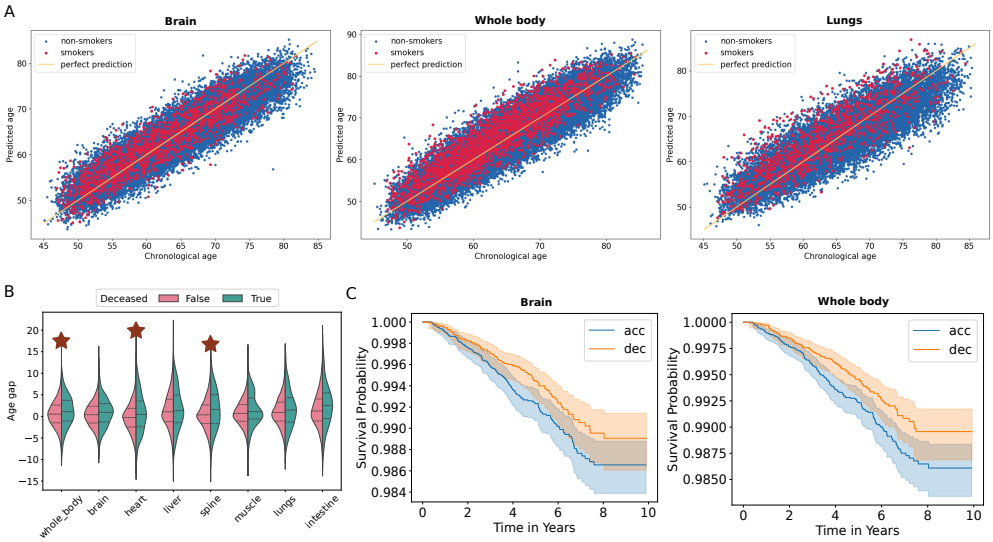


Fig. 5 **A:** Age predictions (y-axis) vs. chronological age (x-axis) of smokers (red) and non-smokers (blue) for the brain, the whole body, and the lung age (left to right). We see a stronger shift for the whole body and lung age predictions compared to the brain age. **B:** Difference in distributions for subjects which are deceased and ones that are still alive of the population younger than 70 at imaging time point. Statistically significant differences are indicated by the red asterisk ($P < 0.00625$, two-sided t-test, Bonferroni-corrected). **C:** Kaplan Meier Survival Curve for brain age prediction (left) and the whole-body age prediction (right) for the accelerated agers (blue) and the decelerated agers (orange).

We identify smoking as one of the most correlated lifestyle factors with higher age gaps (Figure 4B). Figure 5A visualises the age predictions of smokers and non-smokers and their chronological age for the brain, the whole body, and the lungs. Subjects who are current smokers (red) tend to have a higher age gap than non-smokers. This trend is especially visible for the whole body and the lung predictions while being weaker for the brain age prediction. This aligns with the PheWAS analysis, where the brain age gap is less correlated to the smoking status than the age gaps of the other body regions. The mean whole-body age gap of smokers is +1.728 years (n = 2254), while the mean age gap of non-smokers is +0.029 years (n = 42047). Subjects that have previously smoked, show a mean age gap of +0.380 years (n = 23026), supporting clinical findings that cessation of smoking leads to normalisation of e.g. risk for lung cancer [33].

2.5 Survival Analysis

To investigate the interplay between the obtained age gaps for different body regions and survival, we utilise data linkages to national death registries in the UK, which are provided by the UK Biobank. Butler et al. [8] define a useful age biomarker as one that predicts longevity at an age at which 90% of the population is still alive. Given the small number of events (n = 79 females, n = 42 males for the whole body; n = 91 females, n = 35 males for the brain), we slightly increase this to an age where 85% of the population is still alive for this analysis (n = 150 females, n = 99 males for the whole body; n = 140 females, n = 87 males for the brain). Individuals without an entry in the death registry are right-censored (whole body: n = 25546 females, n = 13739 males; brain: n = 17511 females, n = 9387 males).

We first compare the age gap distributions between deceased subjects and subjects that are still alive for all body regions. This allows us to study whether the predicted age gaps are correlated to early death. Figure 5B visualises the different age gap distributions for all image modalities. We see an upward shift in the median age gap for the deceased subjects in all image regions and the differences for the whole-body, the heart, and the spine age gaps are significant (indicated by the asterisk) using a two-sided t-test, $P < 0.00625$, Bonferroni-corrected for 8 imaging modalities. While the difference between distributions is non-significant for some image regions (brain, liver, muscle, lungs, intestine), the Cohen's d effect sizes are similar for all body regions apart from the liver (brain: 0.17, whole-body: 0.20, heart: 0.25, liver: 0.07, spine: 0.26, muscle: 0.21, lungs: 0.13, intestine: 0.19).

Secondly, we visualise the Kaplan Meier curves for accelerated ("acc", blue) and decelerated ("dec", orange) agers in Figure 5 C for the brain (accelerated age: n = 10556, decelerated age: n = 7156) and the whole-body age (accelerated age: n = 11783, decelerated age: n = 8383). For both image modalities, the accelerated agers constantly have a lower survival probability and the shaded confidence interval overlaps less for the whole-body age, aligning with the t-test results from above. Survival durations were right-censored for individuals without an entry in the death registry. Corresponding visualisations for the other image modalities can be found in the Appendix. Finally, we fit a Cox proportional hazard model to the dataset using age, sex, and a binary variable indicating whether a person shows accelerated or decelerated (brain or whole-body) ageing as confounders. The Hazard Ratios (HRs) for the binary attribute indicating accelerated or decelerated ageing is 1.24 for the brain and 1.23 for the whole-body age ($P < 0.005$). Both HRs indicate an increased hazard if the subjects show accelerated ageing.

2.6 Digital Twin

In order to investigate ageing patterns of different regions in the body for individuals, we generate a "Digital Twin" [34] – a digitised representation of a person using the MR images of the different body regions and their age predictions. This Digital Twin allows (a) a detailed investigation of existing ageing patterns and (b) artificially changing the appearance of the person and studying the impact of local changes on the whole body on an individual level.

Figure 6 visualises an example workflow of a female subject (a random subject that shows accelerated whole-body age). For a subject that obtains MR images at a medical screening, we can create their Digital Twin. This subject shows an overall whole-body accelerated ageing of +5.9 years. A further investigation of the body regions shows almost *no accelerated ageing*

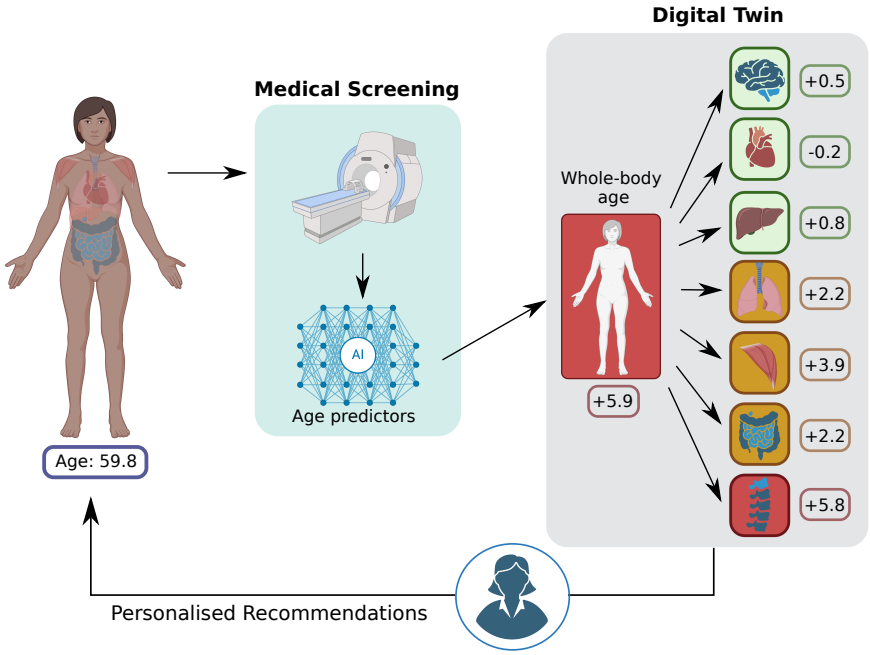


Fig. 6 Screening Scenario: A female subject (left) of age 59.8 years obtains MR images at a screening. The digital twin of this subject shows accelerated whole-body age (+5.9 years) with different accelerations in different body regions. The spine shows the biggest acceleration. This information could be used for personalised recommendations to the patient to consider their region-specific ages in daily life.

(age gap $\leq$ 1 year) in the brain, the heart, and the liver, and *mildly accelerated ageing* (age gap $\leq$ 4 years) in the lungs, the muscles, and the intestine. The most accelerated body-region of this subject is the spine with an age gap of +5.8 years. These insights could be used by physicians to formulate personalised recommendations or interventions that can aid a person in supporting their most accelerated body regions. In Section 2.2, we have discovered that the spine age has a big impact on the whole-body age. In order to further inspect the impact of the spine age at the individual level of this subject, we artificially alter the spine morphology of this subject. We hereby harness the flexibility of the created digitised representation of the subject and alter the appearance of the spine in the whole-body image of this subject. We register an MR image of a subject with the *same chronological age* (to ensure comparability) but *decelerated spine age* to the MR image of the subject of interest. By now replacing the original spine with the spine of the subject with decelerated spine age, we can investigate how much impact the spine has on the subject's accelerated whole-body age. After replacing the subject's spine with a decelerated spine, the Digital Twin shows a less accelerated whole-body age of +4.1 years - almost two years younger than the initial whole-body age. After replacing all body regions that show accelerated ageing, the Digital Twin shows a similar whole-body age gap of +4.2 years, indicating that this subject's accelerated spine age might have the biggest impact on the accelerated whole-body age - considering the body regions we investigate in this work. More results, including artificially changing the appearance of other body regions, are in the Appendix, Section A.5 and Table A1.

These individual ageing patterns could subsequently be used by physicians to formulate informed personalised recommendations. In the PheWAS analysis (Section 2.4), we have, for instance, detected that the frequency of vigorous physical activity is negatively correlated with accelerated spine age. This subject could potentially try to increase their activity and therefore support their spine ageing. On the other hand, some lifestyle factors such as heavy physical work (which is positively correlated with accelerated spine age) might not be adjustable and the subject could be assured that their accelerated whole-body age mostly stems from an accelerated spine age and that the other body regions show more comparable ageing to peers. We note that this is a hypothetical scenario that requires further and more detailed investigations of causality and the impact of specific interventions on the ageing of a subject. We do not perform a de-ageing on the Digital Twin but investigate the impact of individual body regions on the whole-body age on an individual level, compared to the population-level analysis in Section 2.2. As a course of personalised medicine, this analysis would need to be performed for each individual separately and each recommendation needs to be tailored to the subject circumstances, such as lifestyle and ageing patterns. Furthermore, we note that diseases or injuries could also potentially impact the accelerated age. However, such findings could open opportunities for a wide range of interventions, including specific adaptations of working conditions or lifestyle elements.

3 Discussion

Several works have previously investigated ageing biomarkers, using various modalities [4, 12, 14, 15]. While brain age has been studied in more detail [13–17], the interplay between brain ageing, whole-body ageing, and the ageing of local regions in the body is underrepresented. Tian et al. [12] show how individual organ systems can shed light on the complex and multi-faceted ageing process in the human body. We extend their method to utilising medical imaging and convolutional neural networks. With the full 3D volume of medical MR images, we achieve accurate age gap predictions and therefore reliable results regarding accelerated and decelerated ageing. Using images eliminates the necessity of selecting features and relying on e.g. self-reported features, both prone to introduce human bias. We show that brain age is less correlated to the ageing of the whole body and the selected local regions, indicating a more detached ageing process between body and brain. Finally, by creating a Digital Twin of a subject, we can detect their most accelerated body regions, which can guide recommendations for interventions and lifestyle changes. We envision this to be incorporated into medical workflows and a step towards more personalised medicine.

We note that the age gaps presented in this work are in general smaller than the ones in non-imaging studies such as [12]. This might be because the ML models trained on the images show better performance and therefore do not have very strong deviations from the actual age, even if a disease is present. We opt to interpret the actual number of accelerated years more as a relative measure for "how strongly" someone shows accelerated ageing than for a concrete measure of how many years this person might die earlier. Our models have only seen individuals without disease records during training and have therefore learned to identify healthy ageing patterns. Some diseases might not show the same ageing patterns and therefore might not show more accelerated ageing compared to healthy subjects. Circling back to the defined characteristics of [8] that ageing biomarkers should meet, we can conclude

that medical images can be used as biomarkers for ageing but they are no perfect biological age indicators. The discovered correlations between age gaps and lifestyle features highlight several correlated phenotypes. For instance, physical activity, body composition, diet, health status, and social components show significant differences ($P < 2.37 \times 10^{-4}$, two-sided t-test, Bonferroni corrected for 265 features ×8 images).

While our work renders interesting results about global and local ageing patterns in medical images, we note some limitations of this work. Firstly, the dataset has a limited age range, which makes age prediction especially challenging. An investigation of a dataset with a wider age range would be highly interesting. Our approach is highly general and can be applied to arbitrary input data, including higher resolution or functional imaging data as well as more routine imaging data such as computed tomography. We also highlight the complexity of the ageing process, with pathological conditions superimposed and resulting from a multitude of impact factors that can change over time. Disentangling these components in medical images and modelling the whole complexity of ageing may not be feasible but our approach could for example raise warnings and trigger awareness of future medical conditions. Other components of the ageing process, such as genetics, the immune system, or behavioural components should be taken into account in future modelling approaches, to frame a more diverse ageing profile for individuals. Additionally, pathological and physiological ageing phenotypes cannot be distinguished with this method. Disentangling these presents a major challenge and requires future work, including long-term follow-up data and targeted studies. We define our "healthy" cohort as subjects that do not have an ICD-10 record and no self-reported diseases, which might not be a sufficient characteristic to ensure their ageing is "healthy". However, this is the most commonly used approach in current studies [12].

Lastly, we note that while imaging techniques allow for a global and general investigation of ageing, obtaining the images is expensive and sometimes difficult to include in standard medical procedures. Still, medical images are used in practice for the assessment of patients on several occasions and show promising results to function as biomarkers for ageing. Region-specific imaging modalities with a higher resolution could be beneficial for such analyses. While we also investigate local ageing patterns by using specific body regions and evaluating their ages, the strong correlation between the age gaps of all body regions shows that with this method, the ageing process cannot be completely separated into different smaller areas in the body. This is expected because most lifestyle factors, as well as diseases, have an impact on not only a single organ. We envision the investigation of more body regions using pipelines such as [22] to get an even better understanding of locally different structures of ageing as interesting future work. We furthermore see the investigation of longitudinal data and the development of ageing within one subject over time as an interesting next step. This could help identify early markers of ageing, inform personalised interventions and monitor the efficacy of such targeted therapies. The routine assessments of the ageing patterns of a subject as discussed in form of the here presented Digital Twin can shed light on potential risk factors, guide medical examination towards which regions show especially large age gaps, and help in guiding interventions towards a more healthy ageing process.

4 Methods

We here summarise the methods utilised in this work, including details about the dataset, the deep learning methods, the PheWAS study, the Digital Twin experiments, and the Grad-CAM maps. Furthermore, we provide more details of the selected diseases and non-imaging phenotypes.

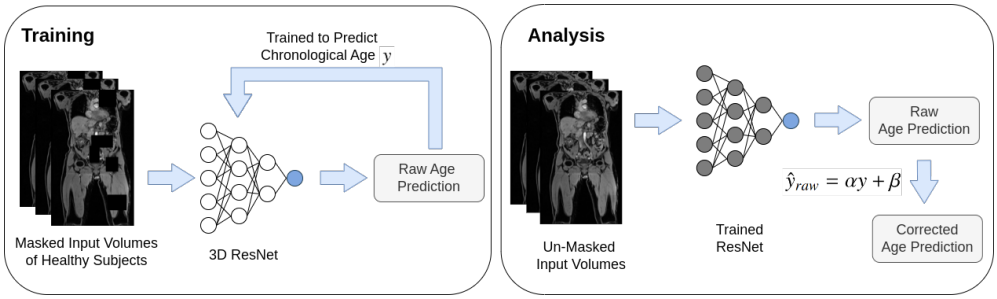


Fig. 7 We train a 3D ResNet18 model for each imaging modality to predict the chronological age of the healthy training subjects. Here, we use the masked input images. Subsequently, using these raw age predictions, we fit a linear regression model to correct the regression-toward-the-mean bias and obtain the corrected age gap for all subjects. We speak of accelerated ageing, if the age gap is positive and decelerated ageing if it is negative.

4.1 Dataset

We utilise (a) 73 094 T1-weighted, dual-echo gradient whole-body MR images with a size of $224 \times 168 \times 363$ voxels and a resolution of $2,23 \times 3 \times 2,23$mm. The whole-body MR images are acquired in several stages, which were stitched using the pipeline of [35]. Furthermore, we use (b) 45 058 T1-weighted scull-stripped brain images of the UK Biobank with an isotropic spacing of 1mm^3 and a size of $160 \times 225 \times 160$ voxels. We split the dataset into healthy and unhealthy subjects. Healthy subjects are selected as ones that do not have ICD-10 entries or self-reported diseases. These subjects are split into training (80%) and validation (20%) sets. The remaining "unhealthy" subjects are used as a test set. The detailed numbers of samples for all sets and modalities can be found in Table 1. We aim for the model to learn to predict the chronological age of healthy subjects and, therefore, be able to identify visual changes in the "unhealthy" subjects of the test set.

In order to investigate images of specific body regions as ageing biomarkers, we use the segmentation pipeline of [22] and select six body regions of interest: the heart, the liver, the spine, the muscle, the lungs, and the intestine. We segment all training and validation samples and 33 135 test samples. We filter out subjects for which the segmentation did not yield any results for specific body regions and exclude these subjects from the pipeline. Furthermore, we dilate the original segmentation mask and select the largest connected component only in order to exclude any faulty segmentations. We then crop the image to the region of interest, containing the region of interest, and mask out all voxels that do not represent the region of interest plus the dilation. The concrete region-wise labels we used are listed in Table 3, column "*Segmentation*".

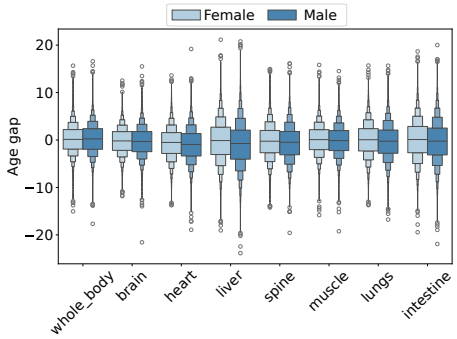


Fig. 8 Age gap distributions for all image modalities for male and female subjects.

4.2 Deep Learning Models

We train separate 3D ResNet18 models on (1) the neck-to-knee (whole-body) magnetic resonance images (MRIs) of the UKBB, (2) the T1-weighted skull-stripped brain MRIs, and (3) different parts of the whole-body MR images to investigate the ageing of local body regions. For all models, we use the pre-trained weights from ImageNet that are provided by `PyTorch` [36]. Several works [30, 37] have shown that one can predict a person's age successfully from whole-body images. The authors also show that the model specifically focuses on three main regions of interest in the human body: the spine, the aortic region, and the back muscles. In order to prevent the model from focusing on only highly predictive regions in the body, such as the spine and the cardiac area, we randomly mask out parts of the images during training to force the model to use the whole available information to predict a subject's age. We use this "masking" for the whole-body and the brain age predictors. Furthermore, we use data augmentation (addition of noise, rotation) for all models to prevent over-fitting. For the region-specific deep learning models, we do not apply any masking, since the areas of interest are already limited to small parts of the whole image and we expect the field of view to be large enough to prevent the model from focusing on highly specific sub-regions. All code is implemented using `PyTorch` [36] and `PyTorch Lightning` [38].

We always use five models to make the predictions more stable since individual models can disagree on predictions (model multiplicity [39]) and thus construct an ensemble method [40]. The model performances (averaged over five random seeds), the number of data samples, and the age distributions are summarised in Table 1.

Table 3 Hyperparameters of trained models for all modalities and their mean performance over 5 different random seeds (mean absolute error in years) with the corresponding standard deviation. The segmentation IDs correspond to the segmentations of [22].

Modality	Batch size	Learning rate	Segmentation	Augmentation	Train MAE	Val MAE	Test MAE
Whole body	4	0.0001	-	blur, noise, flips, masking	1.706 ± 1.31	2.492 ± 1.37	2.564 ± 1.45
Brain	6	0.00005	-	blur, noise, flips, masking	1.500 ± 1.62	2.338 ± 1.35	2.496 ± 1.75
Heart	8	0.00005	24,25,26,27,28,29,30,31	blur, noise, flips	0.980 ± 1.67	2.634 ± 2.38	3.022 ± 3.34
Spine	32	0.0001	69	blur, noise, flips	0.750 ± 1.32	3.028 ± 1.71	3.164 ± 2.44
Liver	64	0.0001	5	blur, noise, flips	0.841 ± 1.42	3.842 ± 1.87	3.929 ± 1.87
Lungs	32	0.0001	10,11,12,13,14	blur, noise, flips	0.784 ± 1.15	3.071 ± 1.57	3.152 ± 1.60
Muscle	4	0.00005	66	blur, noise, flips	0.939 ± 1.20	2.627 ± 1.30	2.722 ± 1.36
Intestine	16	0.0001	18	blur, noise, flips	1.033 ± 1.51	3.639 ± 1.93	3.668 ± 2.02

4.3 Bias Correction

The age distribution of the utilised dataset is slightly unbalanced, with more subjects between 60 and 70 and fewer younger and older subjects. Our models show a regression-toward-the-mean bias [24, 25], underestimating the older subjects' age and overestimating the younger ones' age. We, therefore, perform a bias correction on the predicted ages by fitting a linear regression model to predict the raw age gap ag_{raw} from the DL model from the chronological age y of the subject:

$$\hat{ag}_{raw} = \alpha y + \beta. \tag{1}$$

The corrected age gap $\hat{ag}_{corrected}$ is then defined as the original model prediction $\hat{y}$ minus the predicted age gap $\hat{ag}_{raw}$:

$$\hat{ag}_{corrected} = \hat{y} - \hat{ag}_{raw}. \tag{2}$$

All reported results in this work are based on the corrected age gaps.

4.4 Prediction of Whole-body Age from Body Region Ages

We investigate the interplay between individual body region ages and the whole-body age by fitting (a) a linear regression and (b) a random forest on the region-specific age predictions to estimate the whole-body age prediction. We utilise the same training and test sets as for the deep learning models and use `Sklearn`'s [41] linear regression and random forest models. The random forest uses a maximum number of 100 estimators and the Poisson criterion. Reported results are coefficient of determination R^2 scores for both the linear regression and the random forest. The R^2 score is defined as $1 - \frac{u}{v}$, where u is the residual sum of squares $\sum(y - \hat{y})^2$ and v is the total sum of squares $\sum(y - \tilde{y})^2$, where y is the true target value, $\hat{y}$ the predicted value and $\tilde{y}$ the mean over the true target values. The importance values of the individual input ages (heart age, brain age, spine age, liver age, lung age, muscle age, intestine age) of the linear regression are the coefficients of the fit model. The importance values of the individual input ages (heart age, brain age, spine age, liver age, lung age, muscle age, intestine age) of the random forest are the impurity-based feature importance values.

4.5 Lifestyle and Environment

We follow the approach of Bai et al. [42] in performing a phenome-wide association study (PheWAS) with selected non-imaging phenotypes. As confounding variables we use sex, age, sex*age, weight, and height and we remove negative values from the phenotypes since they usually indicate irrelevant responses in the UK Biobank dataset. We furthermore discard features with more than 90% missing data, ones that show more than 95% overlap in entries, and features that correlate higher than 0.9999 with any other utilised feature to avoid redundancies. All features are normalised using mean-standard deviation normalisation for continuous values and rank normalisation for categorical features. We use $P < 0.00625$, Bonferroni-corrected for 264 features $\times 8$ images, and identify the significant features. We use the UK Biobank feature 20116 at the time point of the second assessment (first imaging assessment) to determine whether a subject is a smoker, a non-smoker, or a previous smoker. 0 indicates a non-smoker, 1 a previous smoker, and 2 a current smoker.

4.6 Chronic Diseases

We use the UK Biobank fields 41270 for the ICD-10 records and 20002 for the self-reported non-cancer illness codes to determine, whether a subject is counted to be in a specific disease sub-group. We furthermore use the UK Biobank field 41280 to determine the date of diagnosis of the ICD-10 code. We only select subjects that have a record of the specific disease before the time of the imaging assessment or within one year after. Our goal is to evaluate ageing profiles for different diseases, not predicting disease development. Therefore, we are primarily interested in subjects that already show signs of named diseases in their MR images.

4.7 Survival Analysis

We use the UK Biobank death registry to perform the survival analysis. More information about the death registry can be found at: https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/DeathLinkage.pdf. We filter for male subjects younger than 65 years old and female subjects younger than 70, following the statistics from the Office of National Statistics [43], using the tables for England and the survival rates in the years 2021-2023. Furthermore, we only select subjects for which we have a data entry for their month and year of birth, to allow for an accurate calculation of the survival. With this, for the brain dataset, 227 individuals have a confirmed death date, for the whole-body dataset, 249 individuals have a confirmed death date, and for the body-region datasets, 106 individuals have a confirmed death date. Living individuals were right-censored, where survival duration was calculated as days between the date of assessment, where the images were recorded, and the date of data download (1st April 2024). The Cox proportional hazards models were performed assuming that the mortality HRs in relation to the age gaps of all modalities stay constant over time. We use the `Python` package `lifelines` [44] for the analysis and plotting. We utilise the training, validation, and test sets for this analysis.

4.8 Digital Twin

In order to investigate the impact of individual accelerated body regions on the overall whole-body age prediction, we assess how region-specific changes in the digital twin of a subject affect their whole-body ageing. We here replace highly accelerated regions from medical images with decelerated regions of other subjects. This replacement of the regions is performed in two steps: (1) the images are registered to each other and subsequently (2) the new region is cut out and transferred to the image of the initial subject.

We follow the approach of Starck et al. [45] for whole-body MR image registration and only register subjects of the same sex to one another. For registration, a (a) *reference image* is selected, to which several (b) *moving images* can be registered. We (a) select a random (female) subject that shows accelerated whole-body ageing and evaluate their region-specific age gaps. This subject has an age of 59.8 years. We then (b) select female subjects of the same age that show an overall decelerated age and a decelerated region age. Table 4 summarises the predicted whole-body and region-specific ages of the selected subjects, which are used to replace the original region of interest in the reference subject. All selected subjects have the same chronological age of 59.8 years to ensure comparability. In order to adapt the appearance of the Digital Twin, we then register the whole-body images of the subjects that will function as the new substituted body regions (moving images) to the selected reference image. We use

Deepali [46] as the registration software and register in two steps: First, we affinely align the images and then add a deformable registration step to fine-tune the registration. We note that we cannot replace the brain images but only use the body region from the whole-body images. Subsequently, we cut out the region of interest from the respective moving image and replace the corresponding region in the reference image. We can do this for an arbitrary combination of body regions. In Figure 6, we show an example of artificially altering all accelerated body regions (spine, intestine, muscle, lungs) and compare the whole-body age predictions of the resulting prediction of the Digital Twin with the original whole-body age prediction of the reference subject. More detailed results can be found in the Appendix, Section A.5.

Table 4 Region-specific and whole-body ages of selected subjects that are used to replace the reference's body regions. All subjects utilised have a chronological age of 59.8 years. Colours indicate the severity of the age gap for the reference subject: Green for age gaps smaller than one year, orange for age gaps smaller than 4 years and red for age gaps larger than 4 years. We note that we do not replace the whole-body image or the brain.

Modality	Organ Age of Reference	Organ Age of Replacement	Whole-body Age of Replacement
Whole body	65.69	-	-
Spine	65.63	53.49	51.45
Heart	59.61	52.99	59.46
Lungs	62.01	55.15	57.59
Liver	60.57	54.48	58.88
Intestine	62.02	53.49	51.45
Muscle	63.73	53.49	51.45
Brain	60.34	-	-

Supplementary information. We provide Supplementary information about the specific ICD-10 codes and self-reported disease codes used to identify sub-populations with specific diseases (`config_diseases.py`) as well as a list of all non-imaging phenotypes used for the PheWAS analysis (`phewas_table2.csv`). Furthermore, we provide an interactive visualisation of the PheWAS results as the Manhattan plot (Figure 4A) and the source code for running the experiments. All files can be found in the corresponding source code repository at: https://anonymous.4open.science/r/age_biomarkers-0AE9.

Acknowledgements. Figure 1 was created in BioRender. Mueller, T. (2025) https://BioRender.com/8ryioo7. Figure 6 was created in BioRender. Mueller, T. (2025) https://BioRender.com/8rm2rn2.

References

[1] Niccoli, T. & Partridge, L. Ageing as a risk factor for disease. *Current biology* **22**, R741–R752 (2012).

[2] Hou, Y. *et al.* Ageing as a risk factor for neurodegenerative disease. *Nature Reviews Neurology* **15**, 565–581 (2019).

[3] Kirkwood, T. B. Understanding the odd science of aging. *Cell* **120**, 437–447 (2005).

[4] Bürkle, A. *et al.* Mark-age biomarkers of ageing. *Mechanisms of ageing and development* **151**, 2–12 (2015).

[5] Ferrucci, L. *et al.* Measuring biological aging in humans: A quest. *Aging cell* **19**, e13080 (2020).

[6] Khan, S. S., Singer, B. D. & Vaughan, D. E. Molecular and physiological manifestations and measurement of aging in humans. *Aging cell* **16**, 624–633 (2017).

[7] Chen, R. *et al.* Biomarkers of ageing: Current state-of-art, challenges, and opportunities. *MedComm–Future Medicine* **2**, e50 (2023).

[8] Butler, R. N. *et al.* Aging: the reality: biomarkers of aging: from primitive organisms to humans. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **59**, B560–B567 (2004).

[9] Nagaraj, S. B., Kieneker, L. M. & Pena, M. J. Kidney age index (kai): A novel age-related biomarker to estimate kidney function in patients with diabetic kidney disease using machine learning. *Computer Methods and Programs in Biomedicine* **211**, 106434 (2021).

[10] Morris, J. F. & Temple, W. Spirometric "lung age" estimation for motivating smoking cessation. *Preventive medicine* **14**, 655–662 (1985).

[11] Wells, S., Kerr, A., Eadie, S., Wiltshire, C. & Jackson, R. 'your heart forecast': a new approach for describing and communicating cardiovascular risk? *Heart* **96**, 708–713 (2010).

[12] Tian, Y. E. *et al.* Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nature medicine* **29**, 1221–1231 (2023).

[13] Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A. & Smith, S. M. Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis* **68**, 101871 (2021). URL https://www.sciencedirect.com/science/article/pii/S1361841520302358.

[14] Cole, J. H. *et al.* Brain age predicts mortality. *Molecular psychiatry* **23**, 1385–1392 (2018).

[15] Bashyam, V. M. *et al.* Mri signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain* **143**, 2312–2324 (2020).

[16] Kaufmann, T. *et al.* Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature neuroscience* **22**, 1617–1623 (2019).

[17] Cole, J. H. & Franke, K. Predicting age using neuroimaging: innovative brain ageing biomarkers. *Trends in neurosciences* **40**, 681–690 (2017).

[18] Ecker, V., Früh, M., Yang, B., Gatidis, S. & Küstner, T. *Deep regression for biological age estimation in multiple organs: Investigations on 40,000 subjects of the uk biobank*, 2255–2259 (IEEE, 2024).

[19] Kaeberlein, M., Rabinovitch, P. S. & Martin, G. M. Healthy aging: the ultimate preventative medicine. *Science* **350**, 1191–1193 (2015).

[20] Sudlow, C. *et al.* Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**, e1001779 (2015).

[21] Slieker, R. C., Relton, C. L., Gaunt, T. R., Slagboom, P. E. & Heijmans, B. T. Age-related dna methylation changes are tissue-specific with elovl2 promoter methylation as exception. *Epigenetics & chromatin* **11**, 1–11 (2018).

[22] Graf, R. *et al.* Totalvibesegmentator: Full torso segmentation for the nako and uk biobank in volumetric interpolated breath-hold examination body images. *arXiv preprint arXiv:2406.00125* (2024).

[23] Elliott, M. L. *et al.* Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. *Molecular psychiatry* **26**, 3829–3838 (2021).

[24] Le, T. T. *et al.* A nonlinear simulation framework supports adjusting for age when analyzing brainage. *Frontiers in aging neuroscience* **10**, 317 (2018).

[25] Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E. & Miller, K. L. Estimation of brain age delta from brain imaging. *NeuroImage* **200**, 528–539 (2019). URL https://www.sciencedirect.com/science/article/pii/S1053811919305026.

[26] Høgestøl, E. A. *et al.* Cross-sectional and longitudinal mri brain scans reveal accelerated brain aging in multiple sclerosis. *Frontiers in neurology* **10**, 450 (2019).

[27] Cole, J. H. *et al.* Longitudinal assessment of multiple sclerosis with the brain-age paradigm. *Annals of neurology* **88**, 93–105 (2020).

[28] Viegi, G., Scognamiglio, A., Baldacci, S., Pistelli, F. & Carrozzi, L. Epidemiology of chronic obstructive pulmonary disease (copd). *Respiration* **68**, 4–19 (2001).

[29] Cataldo, J. K., Prochaska, J. J. & Glantz, S. A. Cigarette smoking is a risk factor for alzheimer's disease: an analysis controlling for tobacco industry affiliation. *Journal of Alzheimer's disease* **19**, 465–480 (2010).

[30] Starck, S. *et al.* Atlas-based interpretable age prediction. *arXiv preprint arXiv:2307.07439* (2023).

[31] Mainali, P., Pant, S., Rodriguez, A. P., Deshmukh, A. & Mehta, J. L. Tobacco and cardiovascular health. *Cardiovascular toxicology* **15**, 107–116 (2015).

[32] Freilich, C. D., Markon, K. E., Cole, S. W. & Krueger, R. F. Loneliness, epigenetic age acceleration, and chronic health conditions. *Psychology and aging* (2024).

[33] Godtfredsen, N. S., Prescott, E. & Osler, M. Effect of smoking reduction on lung cancer risk. *Jama* **294**, 1505–1510 (2005).

[34] Laubenbacher, R., Mehrad, B., Shmulevich, I. & Trayanova, N. Digital twins in medicine. *Nature Computational Science* **4**, 184–191 (2024).

[35] Lavdas, I. *et al.* Machine learning in whole-body mri: experiences and challenges from an applied study using multicentre data. *Clinical radiology* **74**, 346–356 (2019).

[36] Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *CoRR* **abs/1912.01703** (2019). URL http://arxiv.org/abs/1912.01703.

[37] Langner, T., Wikström, J., Bjerner, T., Ahlström, H. & Kullberg, J. Identifying morphological indicators of aging with neural networks on large-scale whole-body mri. *IEEE transactions on medical imaging* **39**, 1430–1437 (2019).

[38] Falcon, W. *et al.* Pytorch lightning (2020). URL https://doi.org/10.5281/zenodo.3828935.

[39] Black, E., Raghavan, M. & Barocas, S. *Model multiplicity: Opportunities, concerns, and solutions*, 850–863 (2022).

[40] Dong, X., Yu, Z., Cao, W., Shi, Y. & Ma, Q. A survey on ensemble learning. *Frontiers of Computer Science* **14**, 241–258 (2020).

[41] Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

[42] Bai, W. *et al.* A population-based phenome-wide association study of cardiac and aortic structure and function. *Nature medicine* **26**, 1654–1662 (2020).

[43] Office of National Statistics. Number surviving at exact age (lx), by single year of age, https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/numbersurvivingatexactagelxbysingleyearofage, access date: 2024-11-05. (2024). URL https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/numbersurvivingatexactagelxbysingleyearofage.

[44] Davidson-Pilon, C. lifelines: survival analysis in python. *Journal of Open Source Software* **4**, 1317 (2019). URL https://doi.org/10.21105/joss.01317.

[45] Starck, S. *et al.* Using uk biobank data to establish population-specific atlases from whole body mri. *Communications Medicine* **4** (2024).

[46] Schuh, A. deepali: Image, point set, and surface registration in pytorch (2023). URL https://biomedia.github.io/deepali/.

Appendix A Additional Results

A.1 Model Performance

All model performances are visualised in Figure A1. The left column shows the performance on the training set, the middle column on the validation set, and the most-right column on the test set. We note that all models tend to over-fit on the training set but still generalise well to the validation and the test set. The training and validation sets contain subjects without any records of ICD-10 codes and self-reported diseases and are therefore meant to represent healthy ageing patterns.

A.2 Chronic Diseases

We summarise the results of our analysis comparing accelerated ageing and specific diseases for all body regions in Figure A2. We note that corresponding visualisations for the whole-body age and the brain age are contained in the main part of this manuscript.

A.3 Lifestyle Features

Figure A3 shows the differences in predicted age versus chronological age for smokers (red) and non-smokers (blue) for different body region ages (heart, liver, spine, muscle, intestine). Smokers tend to show more accelerated ageing compared to non-smokers. Their age prediction lies mostly above the line of perfect prediction (orange), while non-smokers are more evenly distributed showing accelerated and decelerated ageing.

A.4 Survival Analysis

We visualise additional Kaplan Meier curves for the remaining body regions that show the difference in survival probability between accelerated agers ("acc", orange) and decelerated agers ("dec", blue) in Figure A4. The shaded area indicates the confidence interval. For all visualised body regions (lungs, liver, spine, muscle, intestine, heart), 5106 subjects show accelerated ageing and 5234 decelerated ageing.

A.5 Digital Twin

We here summarise additional results of the Digital Twin, where we simulate younger body region ages for all regions individually as well as for all accelerated regions and summarise the respective whole-body ages and the corresponding age gaps in Table A1. We highlight that artificially reducing the spine age has the same impact on the whole-body age as simulating decelerated ages for all accelerated body regions. With these experiments, it always needs to be kept in mind that even though a simulated younger region does not impact the overall whole-body age prediction, this can still mean that it is important for the subject to intervene on their body-region ages and, for example, make corresponding lifestyle changes. In this experiment, replacing the spine in the digital twin led to the least accelerated whole-body age with an age gap of +4.06.

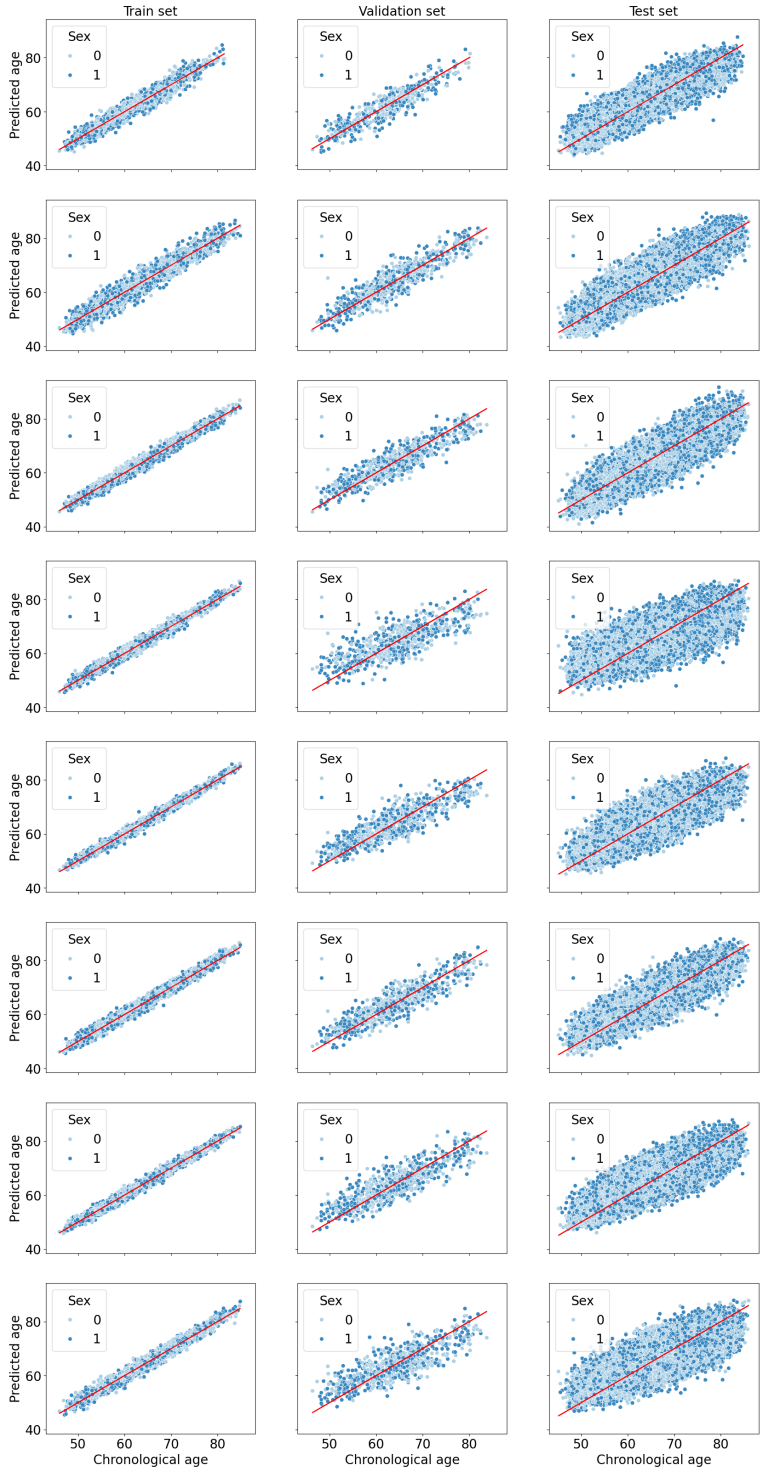
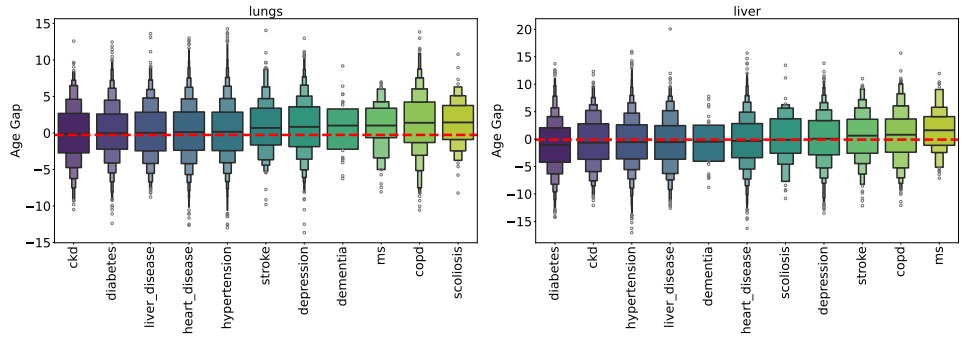


Fig. A1 Model performance on the train set (left), the validation set (middle), and the test set (right) for the brain (first row) and the whole body (second row). The red line indicates a perfect prediction and the colours indicate sex (0: female, 1: male).

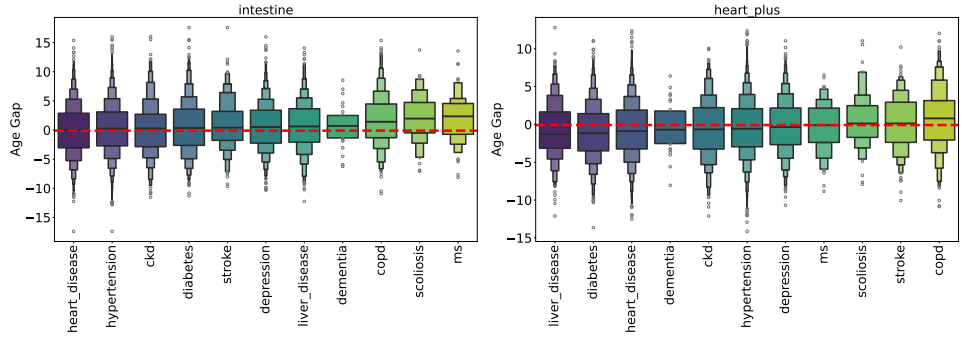
Table A1 Results of the Digital Twin for simulating younger ages of local regions in the body. The whole-body age and the age gap indicate the age prediction/age gap of the digital twin after altering the respective region(s). The subject of reference is 59.8 years old and the smallest age gap is achieved by decreasing their spine age (highlighted in **bold**).

Modality	Whole-body Age	Age Gap
Spine	63.86	**+4.06**
Heart	67.18	+7.38
Lungs	65.87	+6.07
Liver	65.07	+5.27
Intestine	66.48	+6.68
Muscle	64.97	+5.17
All	64.83	+5.03
Spine, Intestine, Muscle, Lungs	63.99	+4.19



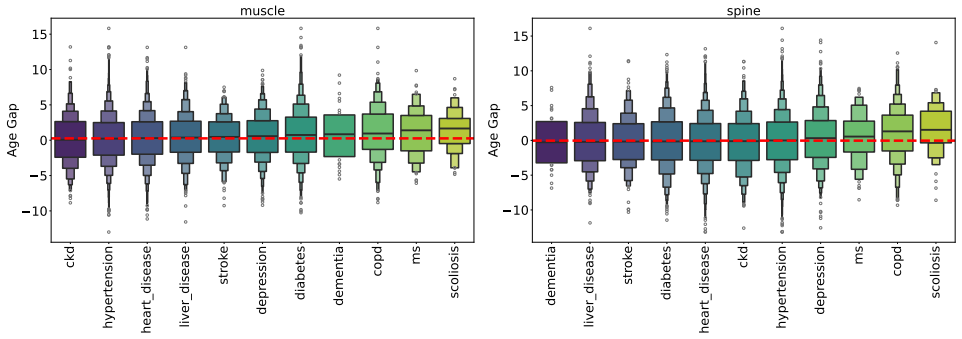
(a) *Lung* age gaps of disease groups

(b) *Liver* age gaps of disease groups



(c) *Intestine* age gaps of disease groups

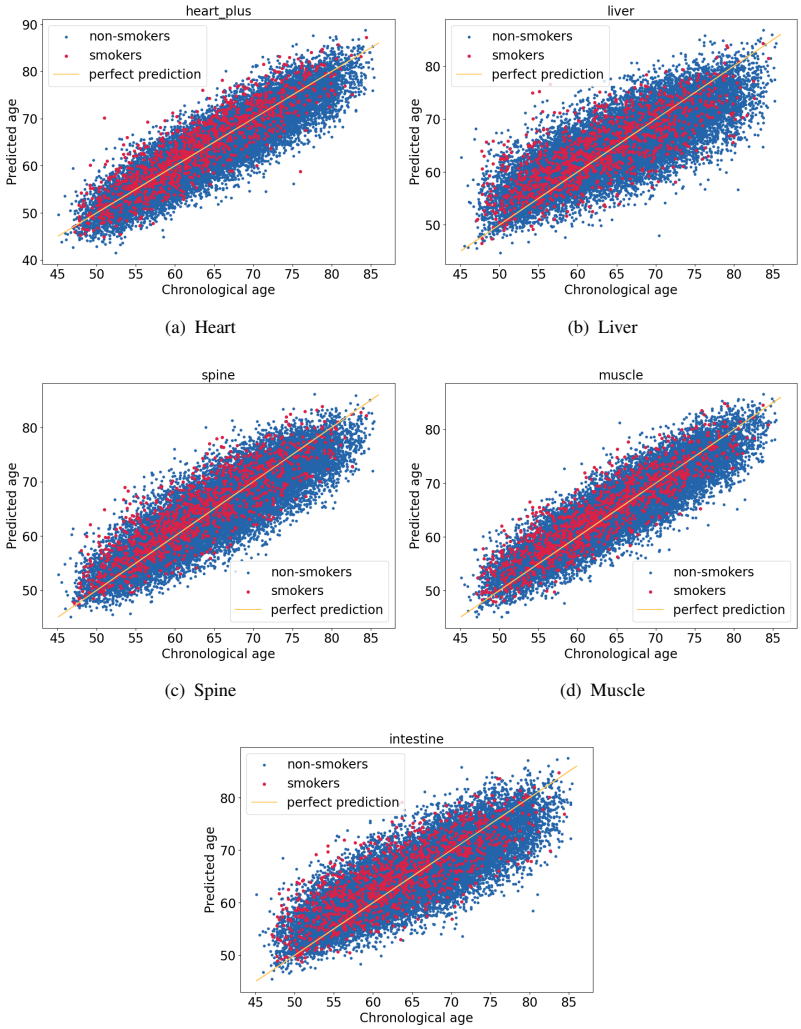
(d) *Heart* age gaps of disease groups



(e) *Muscle* age gaps of disease groups

(f) *Spine* age gaps of disease groups

Fig. A2 Distribution of age gaps for different diseases for (a) the lung, (b) the liver, (c) the intestine, (d) the heart, (e) the muscle, and (f) the spine age predictions. The dotted horizontal (red) line indicates the average age gap of all subjects with diseases (test set). The diseases are ordered by ascending median age gap.



(a) Heart

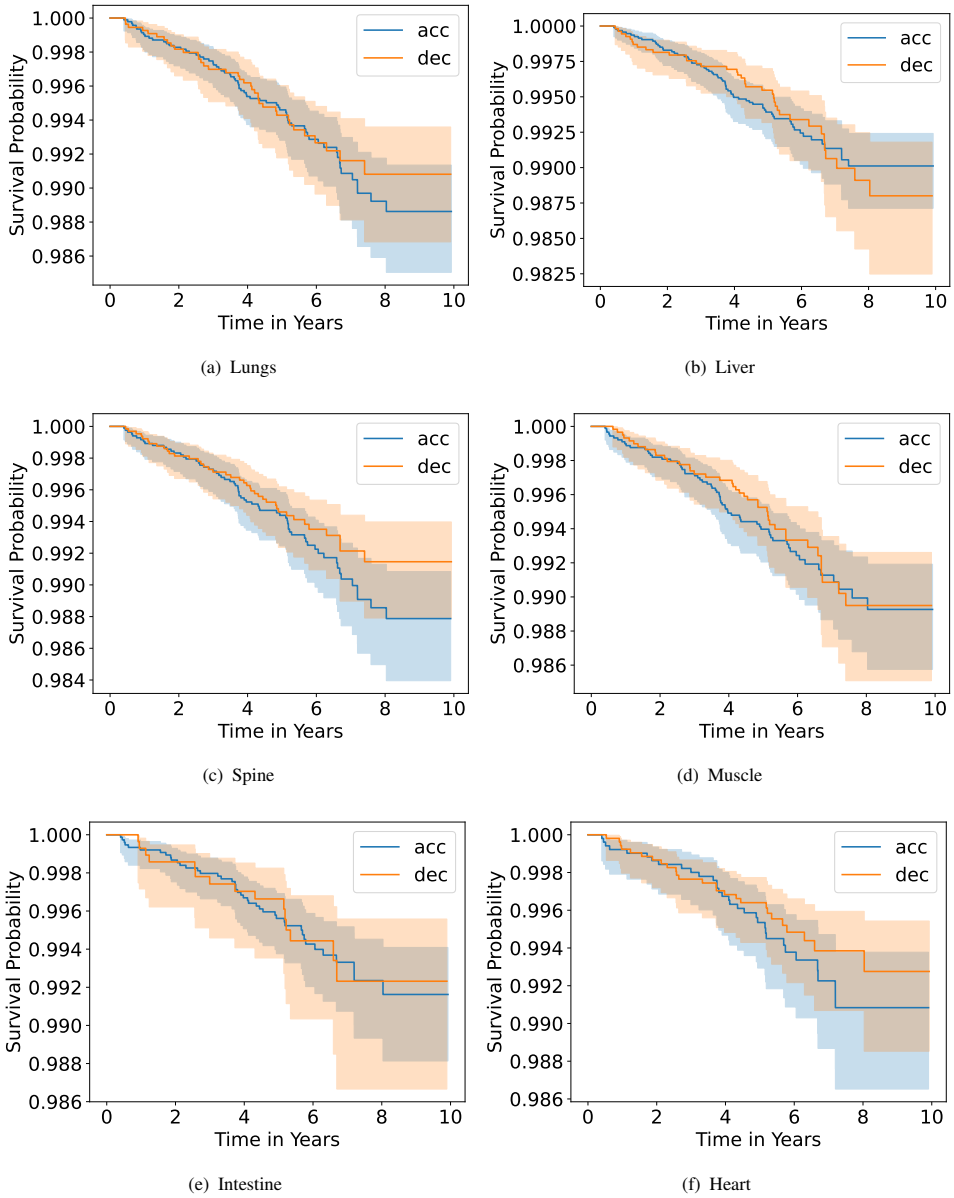
(b) Liver

(c) Spine

(d) Muscle

(e) Intestine

Fig. A3 Visualisation of the predicted age vs. the chronological age of smokers (red) and non-smokers (blue) for (a) the heart, (b) the liver, (c) the spine, (d) the muscle, and (e) the intestine images. Smokers tend to show more accelerated ageing in all image modalities.



(a) Lungs

(b) Liver

(c) Spine

(d) Muscle

(e) Intestine

(f) Heart

Fig. A4 Kaplan Meier curves for the remaining body-region ages, comparing accelerated (acc, blue) and decelerated (dec, orange) agers.