Deep learning predicts onset acceleration of 38 age-associated diseases from blood and body composition biomarkers in the UK Biobank

Mica Xu Ji[1*], Marjola Thanaj[1], Léna Nehale-Ezzine[2], Brandon Whitcher[1], E. Louise Thomas[1], Jimmy D. Bell[1]

[1*]Research Centre for Optimal Health, School of Life Sciences, University of Westminster, London, UK.
[2]Mila – Quebec AI Institute, Montreal, Canada.

*Corresponding author. E-mail: xu.ji.post@gmail.com

Abstract

A major challenge in multimorbid aging is understanding how diseases co-occur and identifying high-risk groups for accelerated disease development, but to date associations in the relative onset acceleration of disease diagnoses have not been used to characterise disease patterns. This study presents the development and evaluation of a neural network Cox model for predicting onset acceleration risk for age-associated conditions, using demographic, anthropomorphic, imaging and blood biomarker traits from 60,396 individuals and 218,530 outcome events from the UK Biobank. Risk prediction was evaluated with Harrell's concordance index (C-index). The model performed well on internal (C-index $\mathbf{0.6830 \pm 0.0904}$, $\boldsymbol{n = 8,931}$) and external (C-index $\mathbf{0.6461 \pm 0.1264}$, $\boldsymbol{n = 855}$) test sets, attaining C-index $\geq \mathbf{0.6}$ on 38 out of 47 (**80.9%**) conditions. Inclusion of body composition and blood biomarker input traits were independently important for predictive performance. Kaplan-Meier curves for predicted risk quartiles (log-rank $\boldsymbol{p \leq 1.16E-16}$) indicated robust stratification of individuals into high and low risk groups. Analysis of risk quartiles revealed cardiometabolic, vascular-neuropsychiatric, and digestive-neuropsychiatric disease clusters with strong statistically significant inter-correlated onset acceleration ($\boldsymbol{r \geq 0.6}$, $\boldsymbol{p \leq 3.46E-5}$), while 13 and 19 conditions were strongly associated with onset acceleration of all cause mortality and all cause morbidity respectively.

In prognostic survival analysis, the proportional hazards assumption was met (Schoenfeld residual $\boldsymbol{p > 0.05}$) in 435 out of 435 or 100% (1238 out of 1334 or 92.8%) of cases across outcomes, $\boldsymbol{aHR = 6.11 \pm 9.00}$ ($\boldsymbol{aHR = 3.67 \pm 5.78}$) with (without) Bonferroni correction. The neural architecture of OnsetNet was interpreted with saliency analysis and several significant body composition and blood biomarkers were identified. The results demonstrate that neural network survival models are able to estimate prognostically informative onset acceleration risk, which could be used to improve understanding of synchronicity in the onset of age-associated diseases and reprioritize patients based on disease-specific risk.

Keywords: multimorbidity, deep learning, healthy aging, UK Biobank

1 Introduction

As we age, the probability of developing more than one condition greatly increases [1]. Multimorbidity, which can be defined as co-occurrence of more than one condition in the same individual, is thus an important problem in aging societies [2, 3], being associated with lower quality of life [4], higher mortality risk [5], and greater usage of healthcare resources than healthy people or those with just one chronic condition [6, 7]. There is a high and increasing prevalence of multimorbidity in the general population, with elderly and socioeconomically deprived groups particularly at risk [1, 8–10]. Chowdhury et al. [11] determined that the global prevalence of multimorbidity was 37.2%, rising to 51.0% in people over 60 years old, while Nunes et al. [12] found that mortality risk of individuals with multimorbidity was 1.73 times greater than those without multimorbidity. Despite these concerns, healthcare delivery systems and quality evaluation remain focused on the treatment of single diseases [13, 14]. An important part of generating an evidence base for improvements in clinical practice is ascertaining patterns of disease and inter-disease associations beyond chance [15].

While epidemiological studies of multimorbidity vary widely in sample size, age settings and statistical methodology, there is a growing consensus that age-associated chronic diseases do not co-occur randomly but in associated fashion,

including conditions across the physical and mental health divide [1, 16]. A 2014 meta-review of 14 studies [15] identified cardiometabolic, neuropsychiatric-gastrointestinal, and musculoskeletal-vascular-neuropsychiatric-gastrointestinal co-occurrence clusters, while more recently [17] identified cardiometabolic, neuropsychiatric, and allergic disease co-occurrence clusters as being commonly reported among 41 studies. The majority of reviewed studies used agglomeration methods such as agglomerative hierarchical clustering and explanatory factor analysis to link diseases on the basis of co-occurrence in individuals, and were not restricted to age-associated conditions. While co-occurrence is important for characterising disease associations, another significant factor is age of disease onset [1]. For example, whether an individual develops cardiovascular disease or hearing loss may be less informative for their health status than age of onset, given that prevalence is estimated at 79-86% and 80.6% in those older than 80 and 85 respectively in the United States [18, 19]. The timing of disease onset has been shown to be informative for risk of secondary diagnoses; early onset cancer is associated with increased risk of cardiovascular disease and secondary cancers compared to later onset cancer [20, 21]. Accelerated biological age estimated using DNA methylation or phenotypic aging clocks has been associated with increased risk of mortality and adverse health outcomes compared to less accelerated biological age [22, 23]. [24] found that certain lifestyle factors affected only age of onset and not odds of diagnosis for 10 cardiometabolic, autoimmune and respiratory conditions. Taken together, this suggests that the temporal dimension of multimorbidity may be important for characterising disease patterns.

The present study aims to determine (1) the extent to which onset acceleration risk of age-associated diseases can be accurately predicted from individual traits, and (2) how strongly early onset risk of one disease is associated with early onset risk of another disease beyond chance. To this end, Cox models were trained and evaluated for prediction of individual risk for early onset of diverse age-associated diseases. Using

data from 60,396 individuals in the UK Biobank imaging cohort [25], we trained a single multimodal neural network, OnsetNet, to predict time-to-event log-risk of the first diagnosis of 47 conditions with birth as temporal baseline, including two composite conditions of all-cause mortality and all-cause morbidity.

First, we evaluated performance of neural network models across disease conditions using C-index measured on internal center and external center held-out test sets, where model hyperparameters were selected by fivefold cross-validation and internal test and C-index ≥ 0.6 was considered acceptable prediction [26]. We compared performance of neural network models on eight different combinations of five input trait groups (demographic, anthropomorphic, summary body composition, spatial body composition, and blood biomarkers) totalling 43 input variables. We assessed performance differences between neural network and linear architectures to evaluate the benefit of using a neural network, which is state-of-the-art and may be a more appropriate choice where it is too simplistic to assume linearity of the log-risk function [27]. OnsetNet score quartiles were computed to illustrate within-disease heterogeneity across onset acceleration risk quartiles. Between-disease risk relationships were analysed by computing all-to-all correlations in OnsetNet risk quartiles between diseases, and clusters containing disease conditions with high inter-correlated onset acceleration risk were identified with agglomerative hierarchical clustering. We trained standard prognostic linear adjusted Cox models (baseline as measurement time, pre-existing diagnoses excluded) with OnsetNet score as an input trait, yielding adjusted hazard ratios across disease pairs. The proportional hazards assumption was tested by examination of Schoenfeld residuals. Finally, the model was analysed with gradient-based saliency to infer relative importance of individual biomarkers and to assess for biological validity of learned associations between biomarkers and disease risk.

OnsetNet's output is interpretable as a metric of accelerated biological aging, given that accelerated biological age can be characterized as event acceleration risk of

age-associated disease from birth [28]. While the present study is purely associative, OnsetNet's ability to stratify individuals into high and low risk groups may be useful for clinical settings, where important problems include early identification of high-risk individuals in order to effectively reduce disease burden on the general population [29] and characterising within-disease heterogeneity due to diversity in pathology [30, 31].

2 Methods

2.1 Study participants

This study was a secondary analysis of longitudinal data from the UK Biobank resource [32] under application number 44584. Data was obtained from the UK Biobank imaging cohort [33] for 60,396 individuals who were initially assessed between April 2007 and October 2010 and had outcomes recorded up to November 30th 2022 (follow-up 13.9 ± 0.819 years, min 12.2 years, max 15.6 years). Eligible individuals had full records of all covariates (29,873 male, 30,523 female, age at imaging 66.2 ± 8.01 years). Where multiple measurements were made, the latest measurement for each individual was taken. Data was split into a training and validation dataset (train-val) containing individuals assessed at Cheadle, Reading, Newcastle and unspecified centers (24,984 male, 25,626 female, age at imaging 66.2 ± 8.00 years), an internal test set containing individuals assessed at train-val centers (4,455 male, 4,476 female, age at imaging 66.0 ± 8.10 years), and an external test set contained individuals assessed at the Bristol center (434 male, 421 female, age at imaging 67.9 ± 7.54 years).

2.2 Covariates

Input variables were selected based on risk factors identified in literature review and were separated into 5 groups. Demographic ("demog") contained sex, race, education level, and Townsend deprivation index. Basic body ("bb") contained weight, height, body mass index (BMI), waist circumference, grip strength, systolic blood pressure,

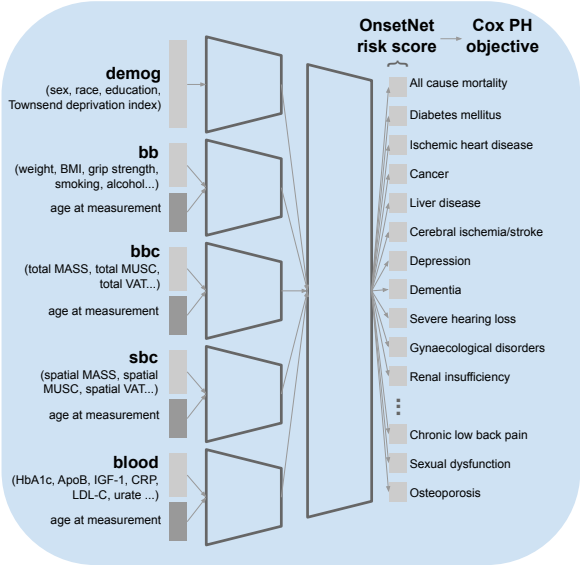

Fig. 1: Architecture of OnsetNet, a neural network Cox model trained to predict onset acceleration as log-risk of first incidence of 47 disease conditions from individual traits, using birth as temporal baseline. Abbreviations: body mass index (BMI), body mass (MASS), muscle (MUSC), visceral adipose tissue (VAT), glycated haemoglobin (HbA1c), apolipoprotein B (ApoB), insulin-like growth factor 1 (IGF-1), c-reactive protein (CRP), low-density lipoprotein cholesterol (LDL-C).

alcohol consumption, and smoking history. Basic body composition computed from neck-to-knee magnetic resonance imaging scans ("bbc") contained total mass (MASS), total subcutaneous adipose tissue volume (SAT), total muscle volume (MUSC), total visceral adipose tissue volume (VAT), and total thigh intramuscular and intermuscular adipose tissue mass (TMAT). Multidimensional spatial body composition ("sbc") contained MASS, SAT, MUSC, VAT, and TMAT measured along 370 slices across the vertical height dimension. Blood-based biomarkers ("blood") contained low-density lipoprotein cholesterol (LDL-C), triglycerides, apolipoprotein A (ApoA), apolipoprotein B (ApoB), cholesterol, high-density lipoprotein cholesterol (HDL-C), glucose, glycated haemoglobin (HbA1c), insulin-like growth factor 1 (IGF-1), urate, urea, total protein, alanine aminotransferase (ALT), albumin (ALB), alkaline phosphatase (ALP),

aspartate aminotransferase (AST), c-reactive protein (CRP), calcium (Ca), creatinine, cystatin C, gamma glutamyltransferase (GGT), phosphate, sex hormone binding globulin (SHBG), total bilirubin, and vitamin D.

For non-demographic variables, age at trait measurement was provided as an additional input variable with the measurement (Figure 1). Continuous input variables were scaled to zero mean and unit standard deviation using statistics from the train-val dataset, and discrete input variables were represented with one-hot encodings. Body composition variables were computed by automated segmentation of MRI instances by a previously published deep learning system that was used to produce publicly available UK Biobank body composition variables [34–37]. In brief, UNet [38] based deep learning models were trained to segment body composition tissues from UKBiobank neck-to-knee Dixon acquisitions using manual annotations. The UKBiobank protocol utilized the Siemens Aera 1.5T scanner (Syngo MR D13) and is detailed in [33].

Outcomes for 47 age-related conditions were chosen using literature on aging and multimorbidity [39, 40]. We obtained date of first occurrence for 45 non-composite disease categories defined by ICD10 codes (Appendix A.4) from the UK Biobank first occurrence register, which incorporates primary care, hospital inpatient, death register and self-reported records. Date of death was taken from the death register and date for all-cause morbidity was taken as first occurrence date for any of the 45 non-composite disease categories.

2.3 Deep learning estimation of onset acceleration risk

Neural network models trained to predict onset acceleration risk consisted of separate encoders per trait group each producing a feature vector of fixed length $k \in \{128, 256, 512\}$, followed by feature concatenation and a shared predictor head with one scalar output for each disease condition (Figure 1). Neural networks had an average of $2.31E6 \pm 1.00E6$ parameters (min 8.54E5, max 4.92E6, Appendix B).

Fivefold cross-validation was used to train models on the train-val dataset by optimizing objective $l(\theta)$, the Cox proportional hazard (Cox PH) partial log-likelihood objective function with weight decay regularization [41] over 47 disease conditions $d \in \mathcal{D}$:

$$g^d(\theta) = -\frac{1}{N_{E^d=1}} \sum_{i:E_i^d=1} \left(h_\theta^d(x_i) - \log \sum_{j \in \mathcal{R}^d(T_i^d)} e^{h_\theta^d(x_j)} \right),$$
$$l(\theta) = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} g^d(\theta) + \lambda \|\theta\|_2^2,$$

where θ are model parameters, λ is the weight decay coefficient, x_i are input traits for individual i, log-risk $h_\theta^d(x_i)$ is the scalar output of the model for disease condition d and individual i, E^d is the censoring vector for d where $E_i^d = 1$ (uncensored) if individual with index i received a diagnosed for d and $E_i^d = 0$ (censored) otherwise. $N_{E^d=1}$ is the number of uncensored individuals for d. Birth was used as temporal baseline and time to event was measured in years (Figure 2). T_i^d is age at first diagnosis for individual i with a diagnosis of d and otherwise age at censor date (November 30th 2022), risk set $\mathcal{R}^d(t) = \{i : T_i^d \geq t\}$ is the set of individuals undiagnosed with d at age t.

To evaluate performance, test sets were unseen during model selection of neural network Cox models and linear Cox comparison baselines. Hyperparameter values (λ, k, learning rate, dropout rate) for each input trait group combination were chosen by average validation fold C-index. Computation of C-index was batched with batch size of 2048 for tractability [27]. Neural networks were implemented in Python with GPU-accelerated PyTorch 2.4.1 and linear models in R 4.1.2. The code will be made available upon publication.

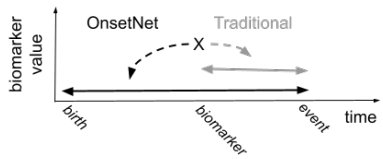

Fig. 2: Cox models commonly predict risk of event earliness from measurement baseline while OnsetNet models event earliness from birth.

2.4 Statistical analysis

To investigate onset acceleration risk correlation between conditions, Pearson's r was computed between OnsetNet score quartiles $q \in [1, 4]$ for all pairs of conditions with internal test C-index performance ≥ 0.6. Correlation of $r \geq 0.6$ was considered strong and $p \leq 3.46E-5$ was considered significant. Using the all-pairs correlation matrix, a clustering dendrogram was computed using unweighted pair group method with arithmetic mean (UPGMA) hierarchical agglomerative clustering for the linkage function (SciPy *cluster.hierarchy.linkage*) with cosine distance between correlation vectors as the clustering metric. Statin users (12,157 out of 60,396 in the study cohort reporting statin medication usage in verbal interview) were removed, and sensitivity analyses were conducted for removal of Bonferroni correction and inclusion of statin users.

To evaluate the prognostic ability of onset acceleration scores, a prognostic Cox model was trained for each pair of disease conditions (A, B), with OnsetNet risk quartile for condition A as input and individual log-risk of time-to-event for first occurrence of condition B as output. Baseline time was taken as imaging assessment time, corresponding to the latest date across input trait measurements, outcome events were within 10 years of baseline, individuals with pre-existing diagnoses were excluded, and time-to-event was measured in months. Demographic traits were included as input covariates for primary adjustment of hazard ratios, and secondary adjustment for both demographic and basic body traits was also performed to test for sensitivity. 6 outcome conditions with greater than 75% allocation to either sex (female: gynecological problems, thyroid diseases, osteoporosis; male: prostatic hyperplasia, hyperuricemia/gout, sexual dysfunction) were trained for the dominant sex only. Outcome events for 3 conditions with fewer than 50 uncensored individuals after removal of pre-imaging diagnoses were removed (hemorrhoids, Parkinson's disease, and somatoform disorders).

Two-tailed tests were used to compute p-values and Bonferroni corrections were made at $p = 0.05$ level. The study adhered to the EQUATOR TRIPOD+ AI [42] guidelines for transparent reporting of an AI multivariable prediction model and STROBE [43] guidelines for reports of observational studies. Completed checklists are given in Appendix D.

3 Results

3.1 Predictive performance

A summary of participant characteristics for training and test datasets is given in Table 1. Over eight combinations of trait group inputs for the neural network models, the full traits model (demog+bb+bbc+sbc+blood) had the highest average validation C-index (0.6856 ± 0.0908) and highest average internal test C-index (0.6830 ± 0.0904), on average over cross-validation folds and conditions (Figure 3). Diabetes mellitus non-T1 ($0.940 \pm 3.06E-3$) and severe vision reduction ($0.550 \pm 2.54E-3$) performed best and worst on internal test respectively. On the external test set, demog+bb+bbc+blood performed best out of neural network models (0.6472 ± 0.1233) followed by full traits (0.6461 ± 0.1264). The largest gains on average internal test set C-index from adding traits to model inputs were adding bb to demog ($\Delta 0.0542$, $p = 9.68E-15$), sbc to demog+bb ($\Delta 0.0295$, $p = 4.81E-5$), and blood to demog+bb ($\Delta 0.0272$, $p = 4.48E-4$) (Appendix C).

On internal and external test sets, utilizing body composition traits increased performance and was approximately twice as important in the absence of blood biomarkers (internal test, adding bbc to demog+bb: $\Delta 0.0244$, $p = 6.85E-4$, sbc to demog+bb: $\Delta 0.0295$, $p = 4.81E-5$, bbc to demog+bb+blood: $\Delta 0.0122$, $p = 1.42E-1$, sbc to demog+bb+blood: $\Delta 0.0143$, $p = 8.59E-2$). For models with blood traits included, small gains were observed on internal test C-index from utilizing increasingly complex body composition inputs (demog+bb+bbc+blood:

Table 1: Characteristics of UK Biobank participants used for the present study. *p*-values were computed with two-tailed t-test for continuous variables and χ^2-test for categorical variables.

Trait	Train-val	Internal test		External test	
	Value (ref.)	Value	*p*-value	Value	*p*-value
n	50,610	8,931	–	855	–
Age at baseline (yrs)	54.89 ±7.57	54.74 ±7.68	9.80E-02	54.29 ±7.53	2.13E-02
Sex			9.18E-01		7.80E-01
Female	50.63	50.12		49.24	
Race			9.99E-01		9.64E-01
White	96.66	96.93		97.31	
Education			1.00E+00		9.99E-01
College or University degree	45.66	45.45		46.20	
Smoking			9.99E-01		4.71E-01
Never	61.57	60.83		58.83	
Alcohol			1.00E+00		9.67E-01
Three or four times a week	27.84	27.16		25.50	
BMI (Kg/m^2)	26.55 ±4.35	26.61 ±4.42	2.20E-01	26.09 ±4.38	2.51E-03
Grip strength (Kg)	30.43 ±10.61	30.68 ±10.65	4.64E-02	30.23 ±10.28	5.76E-01
HbA1c ($mmol/mol$)	35.04 ±5.09	35.07 ±5.16	7.22E-01	34.62 ±4.70	1.55E-02
HDL-C ($mmol/L$)	1.47 ±0.37	1.47 ±0.37	6.20E-01	1.46 ±0.36	3.37E-01
ApoB (g/L)	1.01 ±0.21	1.01 ±0.21	7.74E-01	1.01 ±0.22	9.84E-01
Hypertension (%)	31.03	31.51	9.18E-01	32.98	6.74E-01
Diabetes mellitus non-T1 (%)	5.56	5.78	9.26E-01	4.80	7.37E-01
Depression (%)	11.03	11.29	9.34E-01	11.93	7.73E-01

0.6799±0.0908, demog+bb+sbc+blood: 0.6820±0.0904, demog+bb+bbc+sbc+blood: 0.6830 ± 0.0904), but performance on the external test set was highest for basic body composition (demog+bb+bbc+blood: 0.6472 ± 0.1233, demog+bb+sbc+blood: 0.6446 ± 0.1269, demog+bb+bbc+sbc+blood: 0.6461 ± 0.1264). C-index was lower ($\Delta - 0.0306 \pm 0.0080$) with higher variance on the external test set compared to the internal test set for all input trait combinations.

We analysed performance differences between linear and neural network architectures on full univariate traits (demog+bb+bbc+blood) and compared performance

with the full traits neural network model (demog+bb+bbc+sbc+blood). High-dimensional spatial body composition was utilized for neural network models but not linear models due to correlation in body composition along the height dimension. On the internal test set, neural network demog+bb+bbc+sbc+blood had the highest average C-index on average over conditions (0.6830 ± 0.0904), outperforming neural network demog+bb+bbc+blood and linear demog+bb+bbc+blood on 26 out of 47 conditions, and neural network demog+bb+bbc+blood (0.6799 ± 0.0908) was second best, despite the neural networks being trained on 20% less data than the linear model due to fivefold cross-validation (Figure 3). On the external test set, linear demog+bb+bbc+blood had the highest average C-index on average over conditions (0.6724 ± 0.0897), outperforming neural networks on 28 out of 47 conditions, and neural network demog+bb+bbc+blood was second best (0.6472 ± 0.1233) (Appendix C). The results indicate that neural network architectures attained the highest performance on in-distribution test data drawn from the same assessment centers used in training, but generalized worse than linear models on out-of-distribution test data.

Following model selection and performance evaluation, the full dataset was utilized for post-hoc analysis. The best performing neural network by average validation C-index was demog+bb+bbc+sbc+blood which provided OnsetNet scores for analysis. Onset acceleration score quartiles were computed by discretizing OnsetNet scores using sex and condition specific quartile bins computed on the train-val dataset. Kaplan-Meier estimates for onset acceleration score quartile illustrated within-condition heterogeneity (Figure 4).

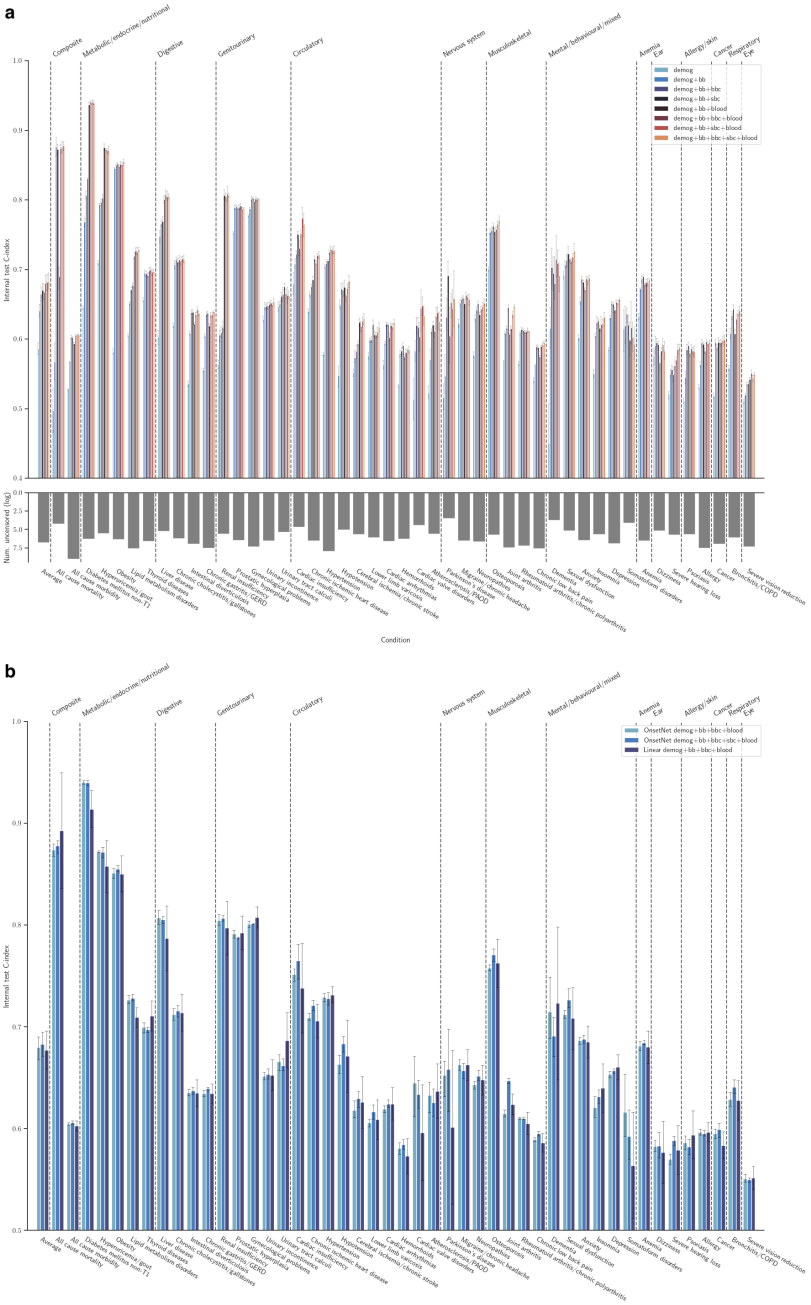

Fig. 3: (a) Internal test C-index across disease conditions for different input trait combinations. The best neural network per trait combination was selected by average validation fold performance. (b) Comparison of internal test C-index for neural network and linear architectures. Grey error bars illustrate 95% confidence intervals.

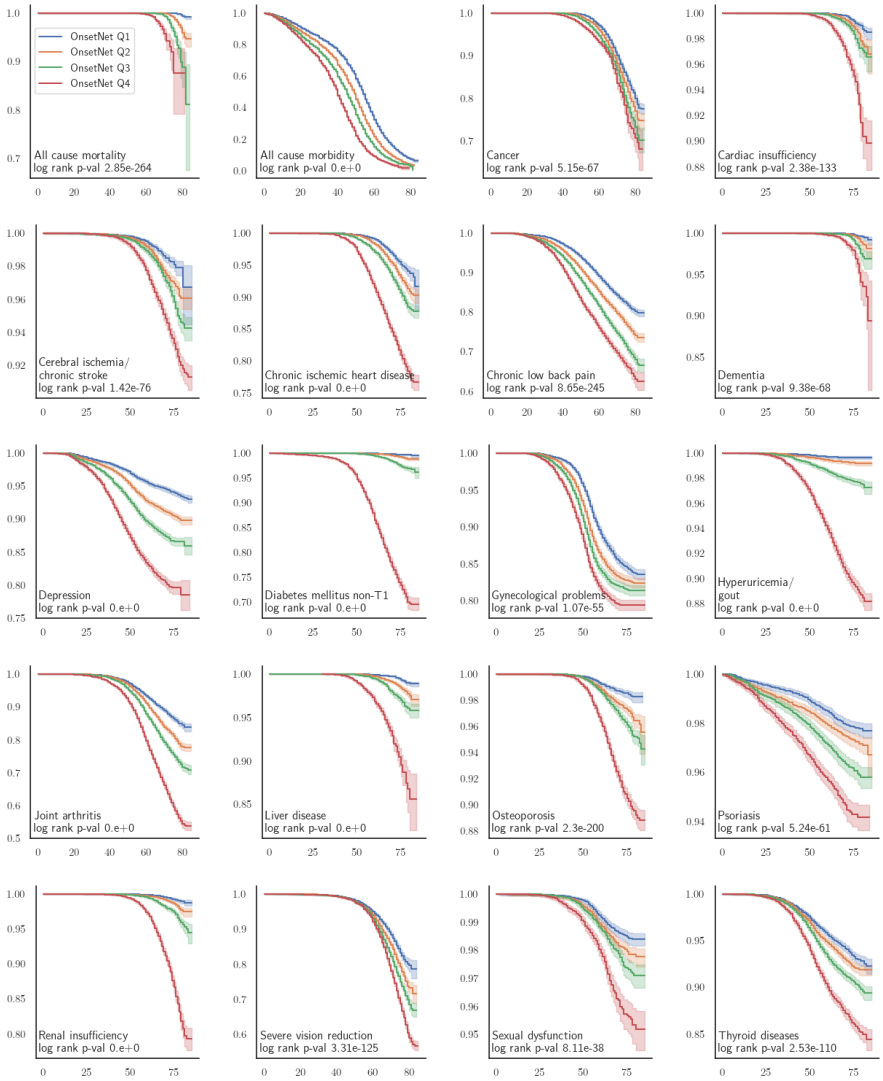

Fig. 4: Kaplan-Meier survival curves of age in years (horizontal axis) against survival probability (vertical axis) for OnsetNet risk quartiles. All conditions and at-risk counts are given in Appendix C.

3.2 Inter-disease onset acceleration analysis

Pearson's r was computed between OnsetNet score quartiles for each pair of 38 conditions where both conditions had internal test C-index performance ≥ 0.6 (Figure 5,

Figure 6). 1420 out of 1444 correlations were significant at $p \leq 3.46E-5$ level (Bonferroni corrected, $p = 5.87E-9 \pm 9.81E-8$) including all strong correlations ($r \geq 0.6$). Disease clusters and strong correlations were reproduced in sensitivity analysis without Bonferroni correction ($p \leq 0.05$) or with inclusion of statin users (Appendix C).

Onset acceleration risk quartile for all cause mortality and all cause morbidity were strongly correlated with each other ($r = 0.78$) and jointly strongly correlated with anxiety ($r = 0.86, 0.82$), dementia ($r = 0.84, 0.82$), depression ($r = 0.80, 0.83$), migraine/chronic headache ($r = 0.77, 0.69$), cardiac valve disorders ($r = 0.75, 0.65$), sexual dysfunction ($r = 0.73, 0.67$), hypotension ($r = 0.69, 0.63$), cardiac arrhythmias ($r = 0.64, 0.63$), and rheumatoid arthritis/chronic polyarthritis ($r = 0.63, 0.79$). Insomnia ($r = 0.81$), chronic gastritis/GERD ($r = 0.78$), intestinal diverticulosis ($r = 0.75$), liver disease ($r = 0.72$), cardiac insufficiency ($r = 0.69$), neuropathies ($r = 0.68$), urinary incontinence ($r = 0.68$), obesity ($r = 0.67$), and urinary tract calculi ($r = 0.63$) were strongly correlated with all cause morbidity but not with all cause mortality. Correlation between onset acceleration risk in obesity and all cause mortality was $r = 0.43$. Conversely, all cause mortality was strongly correlated with Parkinson's disease ($r = 0.72$), anemia ($r = 0.68$), and atherosclerosis/PAOD ($r = 0.62$), which were not strongly correlated with all cause morbidity. The two strongest onset acceleration correlations with Parkinson's disease were dementia ($r = 0.73$) and all cause mortality ($r = 0.72$).

Within non-composite conditions, onset acceleration risk clusters were identified by cosine distance between correlation vectors. A cardiometabolic cluster (corners of Figure 5) was identified where lipid metabolism disorders were associated with chronic ischemic heart disease ($r = 0.79$), hypertension ($r = 0.79$), cerebral ischemia/chronic stroke ($r = 0.70$), diabetes mellitus non-T1 ($r = 0.63$), hyperuricemia/gout ($r = 0.61$), and chronic cholecystitis/gallstones ($r = 0.60$) in a cluster that did not include obesity ($r = 0.35$).

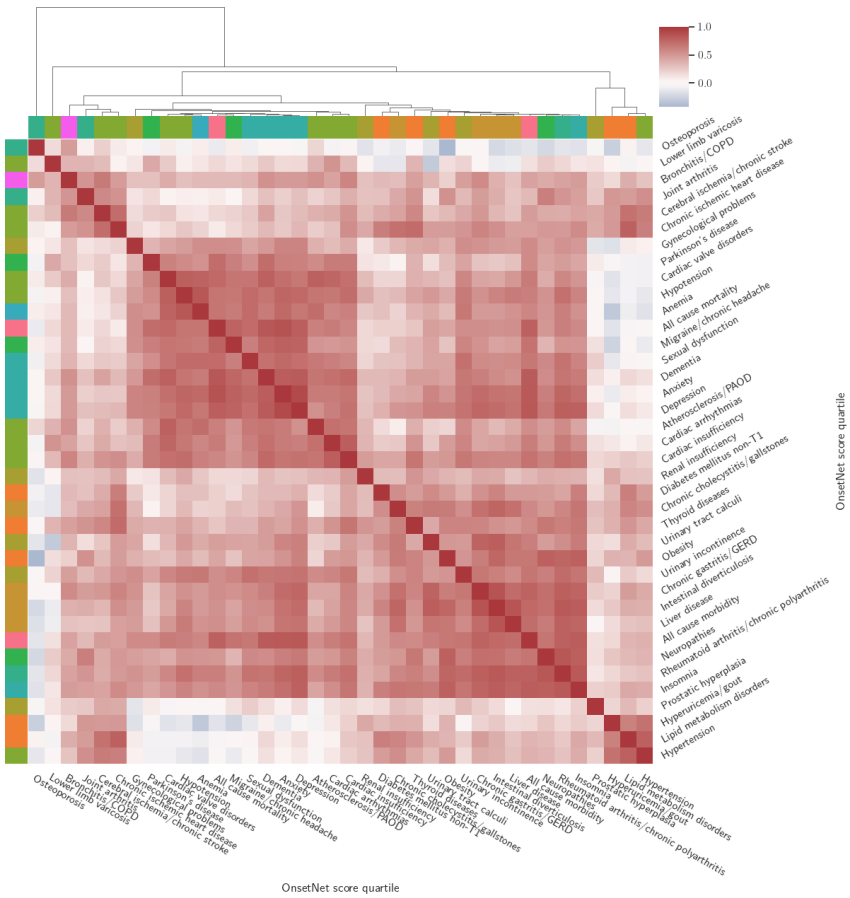

Fig. 5: Clusterplot illustrating structure in the onset acceleration risk of age-associated diseases. Onset acceleration risk quartile correlations for conditions with internal test C-index ≥ 0.6 were clustered by hierarchical agglomerative clustering with cosine distance between correlation vectors as the clustering metric. Statin users were removed and 24 out of 1444 entries with $p > 3.46E-5$ (Bonferroni corrected) were set to 0 for cosine distance. Row and column colours indicate condition types.

Second, chronic gastritis/GERD was associated with a mixed digestive and neuropsychiatric cluster (lower right, Figure 5) containing intestinal diverticulosis ($r = 0.85$), rheumatoid arthritis/chronic polyarthritis ($r = 0.79$), all cause morbidity ($r = 0.78$), urinary tract calculi ($r = 0.77$), insomnia ($r = 0.77$), depression ($r = 0.75$),

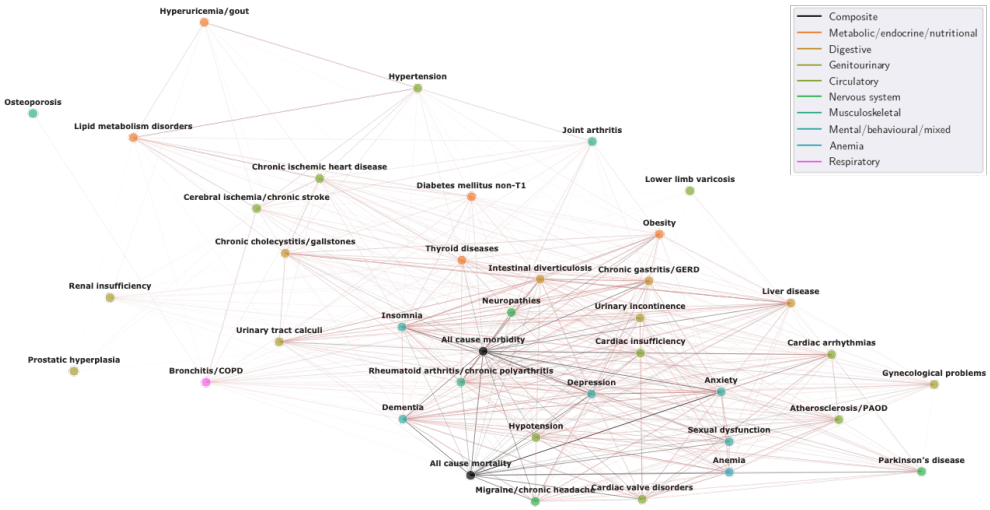

Fig. 6: Graph of inter-disease OnsetNet risk quartile correlations for conditions with internal test C-index ≥ 0.6 and correlations with $r \geq 0.3$ and $p \leq 3.46E{-}5$ (Bonferroni corrected). Statin users were removed and r scales with edge width.

anxiety ($r = 0.73$), liver disease ($r = 0.73$), neuropathies ($r = 0.71$), chronic cholecystitis/gallstones ($r = 0.71$), urinary incontinence ($r = 0.66$), obesity ($r = 0.63$), dementia ($r = 0.63$), chronic ischemic heart disease ($r = 0.60$), cardiac insufficiency ($r = 0.60$), and thyroid diseases ($r = 0.60$).

Third, a vascular-neuropsychiatric cluster was identified (upper left, Figure 5), where onset acceleration risk in cardiac insufficiency was strongly associated with cardiac arrhythmias ($r = 0.80$), dementia ($r = 0.75$), hypotension ($r = 0.73$), rheumatoid arthritis/chronic polyarthritis ($r = 0.71$), cardiac valve disorders ($r = 0.71$), depression ($r = 0.70$), all cause morbidity ($r = 0.69$), atherosclerosis/PAOD ($r = 0.69$), insomnia ($r = 0.68$), thyroid diseases ($r = 0.68$), anxiety ($r = 0.68$), urinary incontinence ($r = 0.67$), parkinson's disease ($r = 0.64$), anemia ($r = 0.62$), neuropathies ($r = 0.61$), chronic gastritis/GERD ($r = 0.60$), and bronchitis/COPD ($r = 0.60$).

Notable associations within the vascular-neuropsychiatric cluster included anxiety and depression ($r = 0.9$) which demonstrated the strongest correlation in onset acceleration risk overall, with top shared correlations for all cause mortality ($r = 0.86, 0.80$),

all cause morbidity ($r = 0.82, 0.83$), dementia ($r = 0.81, 0.81$), migraine/chronic headache ($r = 0.78, 0.77$), rheumatoid arthritis/chronic polyarthritis ($r = 0.77, 0.79$), hypotension ($r = 0.76, 0.73$), sexual dysfunction ($r = 0.74, 0.71$), anemia ($r = 0.74, 0.67$), chronic gastritis/GERD ($r = 0.73, 0.75$), cardiac valve disorders ($r = 0.71, 0.67$), insomnia ($r = 0.71, 0.79$), intestinal diverticulosis ($r = 0.69, 0.73$), liver disease ($r = 0.69, 0.69$), cardiac insufficiency ($r = 0.68, 0.70$), urinary incontinence ($r = 0.67, 0.69$), cardiac arrhythmias ($r = 0.64, 0.63$), neuropathies ($r = 0.63, 0.66$), parkinson's disease ($r = 0.62, 0.62$), gynecological problems ($r = 0.61, 0.60$), and urinary tract calculi ($r = 0.60, 0.62$). The strongest association with anemia was hypotension and vice versa ($r = 0.84$).

3.3 Prognostic survival analysis

Adjusted hazard ratios (aHR) for OnsetNet Q4 (Q1 as reference) for 47 input conditions and 44 outcome event conditions were rendered (Figure 7). 1334 out of 2068 (64.5%) hazard ratios were significant at $p = 0.05$ level after primary adjustment without Bonferroni correction, of which Schoenfeld residuals were insignificant ($p > 0.05$) for 1238 or 92.8% ($aHR = 3.67 \pm 5.78$), and 1037 or 77.7% remained statistically significant after secondary adjustment. 435 out of 2068 (21.0%) hazard ratios were significant at $p = 9.05E-7$ level (Bonferroni corrected $p = 0.05$) after primary adjustment ($aHR = 6.11 \pm 9.00$), where all Schoenfeld residuals were insignificant ($p > 0.05$), and 184 (42.3%) remained statistically significant after secondary adjustment. With Bonferroni correction, 25 out of 44 (56.8%) input conditions yielded statistically significant Q4 hazard ratios for the same outcome condition ($aHR = 11.09 \pm 17.8$), and 27 statistically significant Q4 hazard ratios were observed for all cause mortality ($aHR = 6.26 \pm 9.33$) of which 23 were strong with at least threefold increase in mortality risk compared to Q1 ($aHR \geq 3$, $aHR = 6.85 \pm 9.98$). Using binarized OnsetNet score instead of quartiles (above median vs. below median, Appendix C) yielded 23

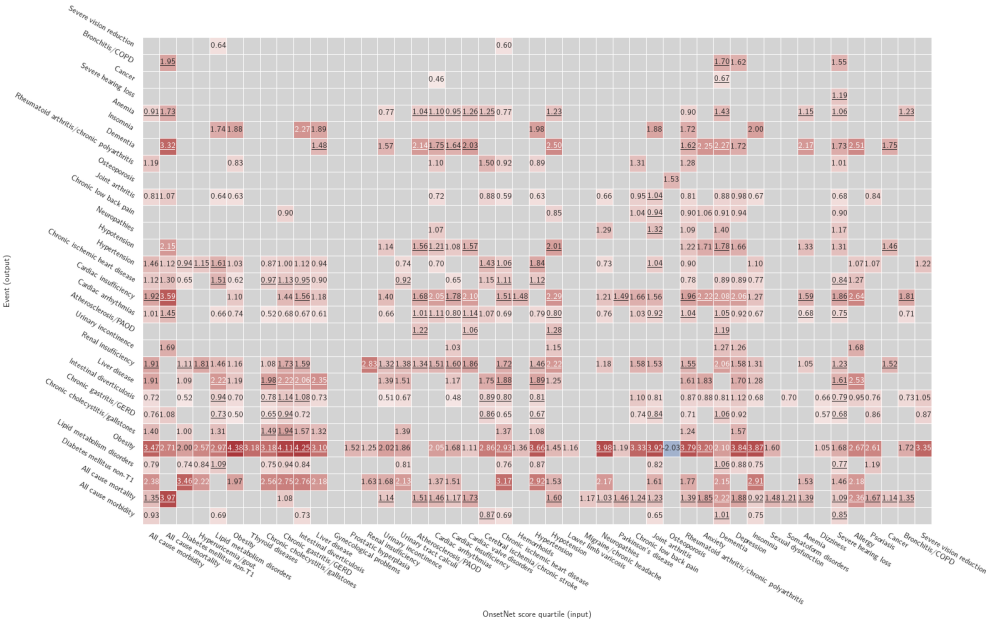

Fig. 7: Adjusted log hazard ratios for OnsetNet risk Q4 (Q1 as reference) given primary adjustment with demographic variables. Statistically insignificant adjusted hazard ratios with $p > 9.05E{-}7$ (Bonferroni corrected) or Schoenfeld residual $p \leq 0.05$ are omitted in grey. Entries that remained statistically significant after secondary adjustment with basic body traits are underlined.

statistically significant adjusted hazard ratios for all cause mortality with Bonferroni correction ($aHR = 2.24 \pm 0.408$) of which only 2 were strong (all cause mortality and dementia, $aHR = 3.39 \pm 0.222$), suggesting that onset acceleration risk Q4 was disproportionately associated with mortality hazard compared to lower quartiles.

3.4 Biomarker saliency

To investigate the relationship learned by the neural network between input traits and onset acceleration risk of conditions, we computed sex-specific normalized absolute partial derivative of network input variables with respect to each output [44, 45]. Univariate traits (Figure 8) and multivariate spatial body composition (Figure 9) were normalized separately.

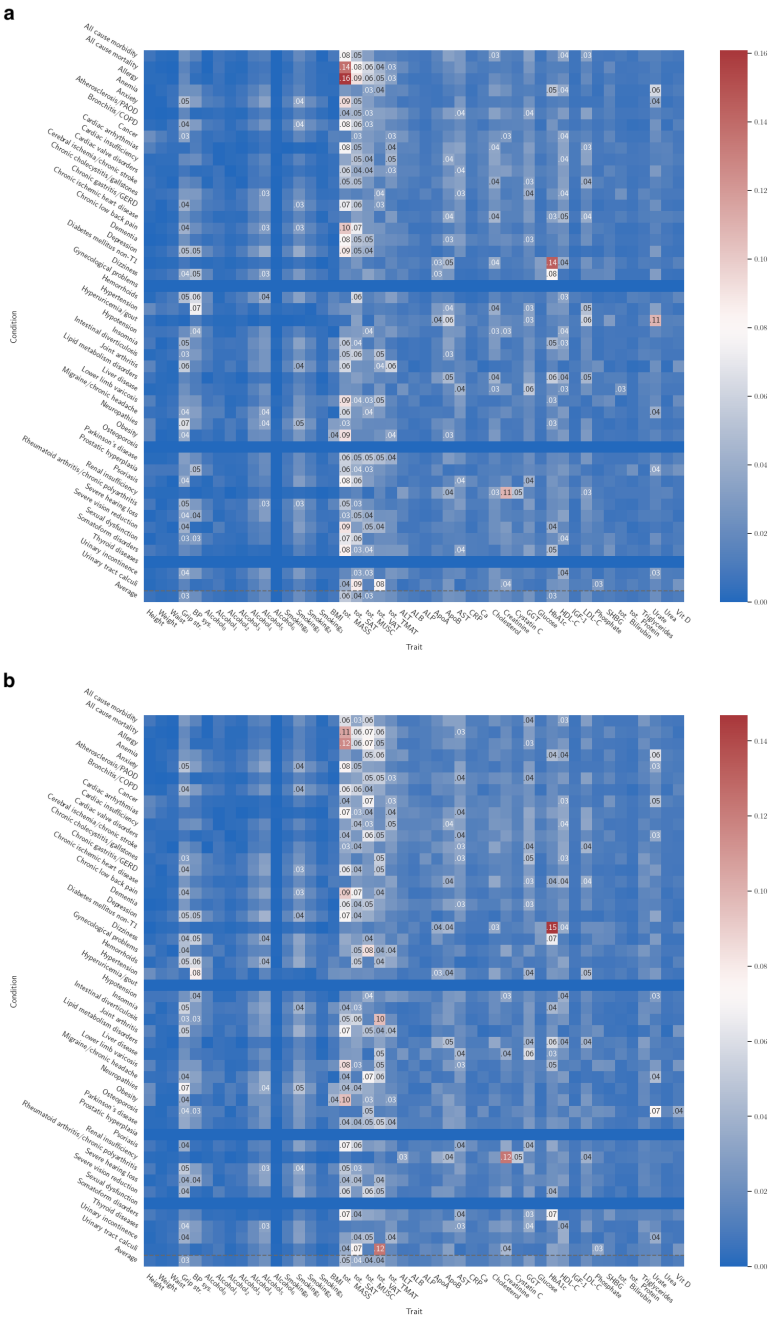

Fig. 8: Saliency map for univariate traits for (a) male and (b) female participants. Top 5 salient traits per condition are labelled. Average saliency across conditions given in the bottom row.

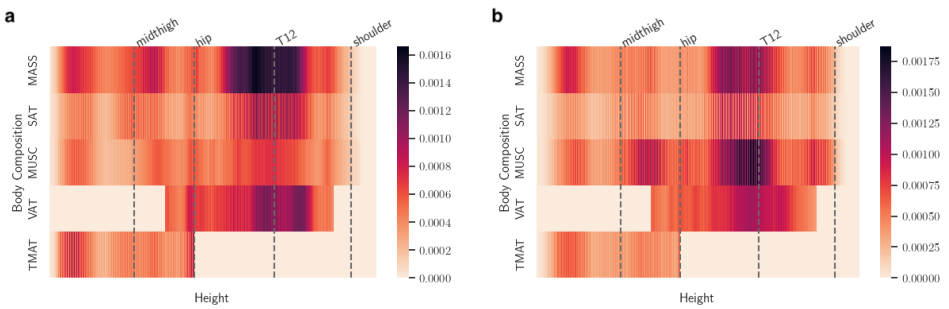

Fig. 9: Spatial body composition saliency maps averaged over 47 conditions for (a) male and (b) female participants. Condition-specific maps are given in Appendix C.

Known relationships inferred by saliency without a-priori knowledge include creatinine and cystatin C for renal insufficiency, HbA1c for diabetes mellitus non-T1, urate for hyperuricemia/gout, GGT for liver disease, total MASS for obesity, systolic blood pressure for hypertension and hypotension. Notably, BMI and total MASS were not in the top 5 traits for onset acceleration risk of diabetes mellitus non-T1, which was dominated by HbA1c. The 10 most important individual traits on average over conditions for the full traits OnsetNet model were total MASS, total SAT, total MUSC, total VAT, grip strength, GGT, HbA1c, HDL-C for both males and females, ApoB and cholesterol for males, and AST and total TMAT for females. Additional traits in the top 10 for the full univariate traits model included smoking, total TMAT, and urate for males, and alcohol, smoking and urate for females (Appendix C). The most salient spatial body composition regions were VAT and MASS located at the abdomen for males, and muscle and MASS around the abdomen for females on average over conditions. Less overt relationships inferred for both sexes included HBA1c, HDL-C and urate for anemia; GGT, ApoB, AST and body composition for dementia; urate for osteoporosis; total MASS, HBA1c and grip strength for insomnia. Total MASS was highly salient for chronic low back pain, and total TMAT was highly salient for cardiac conditions (arrhythmias, insufficiency, valve disorders). For both sexes, body composition traits were more salient than blood biomarkers for predicting onset acceleration

of dementia and Parkinson's disease; the top five univariate biomarkers for Parkinson's disease were basic body composition traits. ALT was salient for liver disease but not in the top 5 traits, which featured AST and GGT.

4 Discussion

For the methodological contribution of the study, we validated that onset acceleration risk of diverse age-associated diseases can be modelled using neural network and linear Cox architectures, with neural networks outperforming linear models on in-distribution test data and vice versa on out-of-distribution test data. While there was high variability in predictive performance across conditions, the final model selected by cross-validation utilized all input traits and attained internal test C-index ≥ 0.6 on 38 out of 47 conditions (80.9%). Model applicability was not limited by temporal differences in trait measurement times (11.3 ± 3.03 years between blood samples and imaging scans for individuals in this study) since the relevant period is lifespan instead of post-measurement time. Inclusion of body composition was found to be important for predictive performance across disease conditions, and utilizing spatial body composition generally yielded comparable performance to using summary body composition metrics across disease conditions (Figure 3), with performance gain from adding body composition metrics to basic demographic and anthropomorphic traits being comparable to the performance gain from adding blood biomarkers.

In post-hoc analysis, a number of connections between diseases with existing support in the literature were inferred without a-priori knowledge, such as associations in onset correlation risk between Parkinson's disease, dementia and all cause mortality, cardiovascular and cerebrovascular conditions [16, 39, 46]. Strong correlations in onset acceleration risk were inferred for cardiometabolic, digestive-neuropsychiatric, and vascular-neuropsychiatric condition groups, partially corroborating a previous review

of cross-sectional multimorbidity studies across different populations which identified cardiometabolic, neuropsychiatric-gastrointestinal, and musculoskeletal-vascular-neuropsychiatric-gastrointestinal co-occurrence clusters [15] despite the use of different methodology that associated diseases by co-occurrence. Automated inference of associations between biomarkers and disease identified pre-supported associations in a precise manner, including biomarkers for renal insufficiency, diabetes and liver disease, adding to previous work that used partial derivative based saliency to identify biologically salient factors for disease [45]. Saliency analysis validated the importance of grip strength and body composition traits [47].

This study is subject to several limitations. Firstly, existing work showed that the UK Biobank cohort contains a healthy participant bias and is predominantly composed of individuals from White ethnic backgrounds [48]. Although models adjusted for sociodemographic factors, conclusions drawn may not hold for more representative or ethnically diverse populations. All conditions except all cause mortality were reliant on the timing and definition of diagnoses, which depends on factors such as access to healthcare services and diagnostic criteria. Underestimation of disease prevalence in the UK Biobank is known to occur secondary to underdiagnosis [49, 50]. While Cox models are the dominant survival analysis approach in medical research and clinical trials [51, 52] and have been used for time-to-death analysis of aging clocks [22], it is an approximate model of survival as the proportional hazards assumption generally does not hold perfectly. In this exploratory analysis, results from evaluation against ground truth event times (Figures 3, 4, 7) indicated that the Cox PH objective was capable of training models to predict onset acceleration risk, but alternative models may also be applied in future work. Additionally, models were assessed on held-out in-distribution and out-of-distribution test sets but are yet to be evaluated on an entirely independent cohort, and direct clinical utility and explicit causal relationships were

not assessed. Prospective and intervention based studies are required to show clinical utility and support the validity of associations inferred in this study.

In summary, we introduced and evaluated the use of Cox models for modelling onset acceleration risk in the context of multimorbid aging, and performed analysis on inter-disease relationships, within-disease heterogeneity across individuals, and biomarker saliency for onset acceleration risk prediction. This work opens up novel approaches for the analysis of age-related diseases and associated biomarker predictor traits, with potential clinical applicability to early identification of high-risk groups in the population.

Declarations

Acknowledgements

The authors would like to thank the UK Biobank Reading team, Madeleine Cule, Ramprakash Srinivasan, Adam Kosiorek, and Martin Tobin for their support.

Sources of Funding

Institutional funding for the study was provided by Calico Life Sciences LLC.

Disclosures

The authors report no conflicts of interest.

Author statement

M. X. Ji conducted literature search, study design, data curation and verification, model training and data analysis, results interpretation, figures, writing first draft, and writing review and edits. M. Thanaj conducted literature search, study design, data curation and verification, results interpretation, and writing review and edits. L. Néhale-Ezzine, B. Whitcher, E. L. Thomas, and J. D. Bell conducted results interpretation, and writing review and edits.

Data access statement

Participant data from the UK Biobank cohort was obtained through UK Biobank Access Application number 44584. The UK Biobank has approval from the North West Multi-center Research Ethics Committee (REC reference: 11/NW/0382). All methods were performed in accordance with the relevant guidelines and regulations, and informed consent was obtained from all participants. Researchers may apply to use the UK Biobank data resource by submitting a health-related research proposal that is in the public interest. More information may be found on the UK Biobank researchers and resource catalogue pages (www.ukbiobank.ac.uk). No participant or public involvement occurred during the design, conduct, reporting, interpretation, or dissemination of this study.

References

[1] Barnett, K., Mercer, S.W., Norbury, M., Watt, G., Wyke, S., Guthrie, B.: Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. The Lancet **380**(9836), 37–43 (2012)

[2] Pearson-Stuttard, J., Ezzati, M., Gregg, E.W.: Multimorbidity—a defining challenge for health systems. The Lancet Public Health **4**(12), 599–600 (2019)

[3] Valderas, J.M., Starfield, B., Sibbald, B., Salisbury, C., Roland, M.: Defining comorbidity: implications for understanding health and health services. The Annals of Family Medicine **7**(4), 357–363 (2009)

[4] Williams, J.S., Egede, L.E.: The association between multimorbidity and quality of life, health status and functional disability. The American journal of the medical sciences **352**(1), 45–52 (2016)

[5] Di Angelantonio, E., Kaptoge, S., Wormser, D., Willeit, P., Butterworth, A.S.,

Bansal, N., O'Keeffe, L.M., Gao, P., Wood, A.M., Burgess, S., *et al.*: Association of cardiometabolic multimorbidity with mortality. Jama **314**(1), 52–60 (2015)

[6] Frølich, A., Ghith, N., Schiøtz, M., Jacobsen, R., Stockmarr, A.: Multimorbidity, healthcare utilization and socioeconomic status: a register-based study in denmark. PloS one **14**(8), 0214183 (2019)

[7] Struijs, J.N., Baan, C.A., Schellevis, F.G., Westert, G.P., Van Den Bos, G.A.: Comorbidity in patients with diabetes mellitus: impact on medical health care utilization. BMC health services research **6**, 1–9 (2006)

[8] Uijen, A.A., Lisdonk, E.H.: Multimorbidity in primary care: prevalence and trend over the last 20 years. The European journal of general practice **14**(sup1), 28–32 (2008)

[9] Kone, A.P., Mondor, L., Maxwell, C., Kabir, U.S., Rosella, L.C., Wodchis, W.P.: Rising burden of multimorbidity and related socio-demographic factors: a repeated cross-sectional study of ontarians. Canadian Journal of Public Health **112**, 737–747 (2021)

[10] Moreno-Juste, A., Gimeno-Miguel, A., Poblador-Plou, B., Calderón-Larrañaga, A., Del Pozo, M.C., Forjaz, M.J., Prados-Torres, A., Gimeno-Feliú, L.A.: Multimorbidity, social determinants and intersectionality in chronic patients. results from the epichron cohort. Journal of Global Health **13** (2023)

[11] Chowdhury, S.R., Das, D.C., Sunna, T.C., Beyene, J., Hossain, A.: Global and regional prevalence of multimorbidity in the adult population in community settings: a systematic review and meta-analysis. EClinicalMedicine **57** (2023)

[12] Nunes, B.P., Flores, T.R., Mielke, G.I., Thume, E., Facchini, L.A.: Multimorbidity and mortality in older adults: a systematic review and meta-analysis. Archives of

gerontology and geriatrics **67**, 130–138 (2016)

[13] Tinetti, M.E., Fried, T.R., Boyd, C.M.: Designing health care for the most common chronic condition—multimorbidity. Jama **307**(23), 2493–2494 (2012)

[14] Ong, K.Y., Lee, P.S.S., Lee, E.S.: Patient-centred and not disease-focused: a review of guidelines and multimorbidity. Singapore medical journal **61**(11), 584 (2020)

[15] Prados-Torres, A., Calderón-Larrañaga, A., Hancco-Saavedra, J., Poblador-Plou, B., Akker, M.: Multimorbidity patterns: a systematic review. Journal of clinical epidemiology **67**(3), 254–266 (2014)

[16] Whitty, C.J., Watt, F.M.: Map clusters of diseases to tackle multimorbidity. Nature **579**(7800), 494–496 (2020)

[17] Ng, S.K., Tawiah, R., Sawyer, M., Scuffham, P.: Patterns of multimorbid health conditions: a systematic review of analytical methods and comparison analysis. International journal of epidemiology **47**(5), 1687–1704 (2018)

[18] Members, W.G., Lloyd-Jones, D., Adams, R., Carnethon, M., De Simone, G., Ferguson, T.B., Flegal, K., Ford, E., Furie, K., Go, A., *et al.*: Heart disease and stroke statistics—2009 update: a report from the american heart association statistics committee and stroke statistics subcommittee. Circulation **119**(3), 21–181 (2009)

[19] Lin, F.R., Thorpe, R., Gordon-Salant, S., Ferrucci, L.: Hearing loss prevalence and risk factors among older adults in the united states. Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences **66**(5), 582–590 (2011)

[20] Lee, J.S., DuBois, S.G., Coccia, P.F., Bleyer, A., Olin, R.L., Goldsby, R.E.: Increased risk of second malignant neoplasms in adolescents and young adults

with cancer. Cancer **122**(1), 116–123 (2016)

[21] Chao, C., Xu, L., Bhatia, S., Cooper, R., Brar, S., Wong, F.L., Armenian, S.H.: Cardiovascular disease risk profiles in survivors of adolescent and young adult (aya) cancer: the kaiser permanente aya cancer survivors study. Journal of Clinical Oncology **34**(14), 1626–1633 (2016)

[22] Lu, A.T., Quach, A., Wilson, J.G., Reiner, A.P., Aviv, A., Raj, K., Hou, L., Baccarelli, A.A., Li, Y., Stewart, J.D., *et al.*: Dna methylation grimage strongly predicts lifespan and healthspan. Aging (albany NY) **11**(2), 303 (2019)

[23] Warner, B., Ratner, E., Datta, A., Lendasse, A.: A systematic review of phenotypic and epigenetic clocks used for aging and mortality quantification in humans. Aging (Albany NY) **16**(17), 12414 (2024)

[24] Rohmann, N., Geese, T., Nestel, S., Schlicht, K., Geisler, C., Türk, K., Brix, F., Jensen-Kroll, J., Demetrowitsch, T., Bang, C., *et al.*: Metabolic and lifestyle factors accelerate disease onset and alter gut microbiome in inflammatory non-communicable diseases. BMC medicine **22**(1), 493 (2024)

[25] Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., *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**(3), 1001779 (2015)

[26] Zhang, Y., Wong, G., Mann, G., Muller, S., Yang, J.Y.: Survbenchmark: comprehensive benchmarking study of survival analysis methods using both omics data and clinical data. GigaScience **11**, 071 (2022)

[27] Wiegrebe, S., Kopper, P., Sonabend, R., Bischl, B., Bender, A.: Deep learning for survival analysis: a review. Artificial Intelligence Review **57**(3), 65 (2024)

[28] Aging Consortium, B., Herzog, C.M., Goeminne, L.J., Poganik, J.R., Barzilai, N., Belsky, D.W., Betts-LaCroix, J., Chen, B.H., Chen, M., Cohen, A.A., et al.: Challenges and recommendations for the translation of biomarkers of aging. Nature Aging, 1–12 (2024)

[29] Organization, W.H., et al.: Global status report on non-communicable diseases 2010 (2011)

[30] Torkamani, A., Andersen, K.G., Steinhubl, S.R., Topol, E.J.: High-definition medicine. Cell **170**(5), 828–843 (2017)

[31] Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V., Madabhushi, A.: Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nature reviews Clinical oncology **16**(11), 703–715 (2019)

[32] Collins, R.: Uk biobank: protocol for a large-scale prospective epidemiological resource (2007)

[33] Littlejohns, T.J., Holliday, J., Gibson, L.M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J.D., Boultwood, C., Collins, R., Conroy, M.C., *et al.*: The uk biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. Nature communications **11**(1), 2624 (2020)

[34] Liu, Y., Basty, N., Whitcher, B., Bell, J.D., Sorokin, E.P., Bruggen, N., Thomas, E.L., Cule, M.: Genetic architecture of 11 organ traits derived from abdominal mri using deep learning. Elife **10**, 65554 (2021)

[35] Fitzpatrick, J.A., Basty, N., Cule, M., Liu, Y., Bell, J.D., Thomas, E.L., Whitcher, B.: Large-scale analysis of iliopsoas muscle volumes in the uk biobank. Scientific reports **10**(1), 20215 (2020)

[36] Thanaj, M., Basty, N., Whitcher, B., Sorokin, E.P., Liu, Y., Srinivasan, R., Cule, M., Thomas, E.L., Bell, J.D.: Precision mri phenotyping of muscle volume and quality at a population scale. Frontiers in Physiology **15**, 1288657 (2024)

[37] Whitcher, B., Thanaj, M., Cule, M., Liu, Y., Basty, N., Sorokin, E.P., Bell, J.D., Thomas, E.L.: Precision mri phenotyping enables detection of small changes in body composition for longitudinal cohorts. Scientific Reports **12**(1), 3748 (2022)

[38] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-assisted intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241 (2015). Springer

[39] Guo, J., Huang, X., Dou, L., Yan, M., Shen, T., Tang, W., Li, J.: Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. Signal Transduction and Targeted Therapy **7**(1), 391 (2022)

[40] Koller, D., Schön, G., Schäfer, I., Glaeske, G., Bussche, H., Hansen, H.: Multimorbidity and long-term care dependency—a five-year follow-up. BMC geriatrics **14**, 1–9 (2014)

[41] Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC medical research methodology **18**, 1–12 (2018)

[42] Collins, G.S., Moons, K.G., Dhiman, P., Riley, R.D., Beam, A.L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J.B., Van Smeden, M., et al.: Tripod+ ai statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. bmj **385** (2024)

[43] Von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P.: The strengthening the reporting of observational studies in epidemiology (strobe) statement: guidelines for reporting observational studies. The lancet **370**(9596), 1453–1457 (2007)

[44] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)

[45] Hao, J., Kim, Y., Mallavarapu, T., Oh, J.H., Kang, M.: Interpretable deep neural network for cancer survival analysis by integrating genomic and clinical data. BMC medical genomics **12**, 1–13 (2019)

[46] Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., Meinow, B., Fratiglioni, L.: Aging with multimorbidity: a systematic review of the literature. Ageing research reviews **10**(4), 430–439 (2011)

[47] Jung, M., Raghu, V.K., Reisert, M., Rieder, H., Rospleszcz, S., Pischon, T., Niendorf, T., Kauczor, H.-U., Völzke, H., Bülow, R., et al.: Deep learning-based body composition analysis from whole-body magnetic resonance imaging to predict all-cause mortality in a large western population. EBioMedicine **110** (2024)

[48] Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., Allen, N.E.: Comparison of sociodemographic and health-related characteristics of uk biobank participants with those of the general population. American journal of epidemiology **186**(9), 1026–1034 (2017)

[49] Freer, C.L., George, E.S., Tan, S.-Y., Abbott, G., Scott, D., Daly, R.M.: Prevalence of sarcopenia and its defining components in non-alcoholic fatty liver disease

varies according to the method of assessment and adjustment: Findings from the uk biobank. Calcified Tissue International, 1–11 (2024)

[50] Dickey, A.K., Quick, C., Ducamp, S., Zhu, Z., Feng, Y.-C.A., Naik, H., Balwani, M., Anderson, K.E., Lin, X., Phillips, J.E., *et al.*: Evidence in the uk biobank for the underdiagnosis of erythropoietic protoporphyria. Genetics in Medicine **23**(1), 140–148 (2021)

[51] Spruance, S.L., Reid, J.E., Grace, M., Samore, M.: Hazard ratio in clinical trials. Antimicrobial agents and chemotherapy **48**(8), 2787–2792 (2004)

[52] Chen, H.-C., Kodell, R.L., Cheng, K.F., Chen, J.J.: Assessment of performance of survival prediction models for cancer prognosis. BMC medical research methodology **12**, 1–11 (2012)