

Deep longitudinal phenotyping of wearable sensor data reveals independent markers of longevity, stress, and resilience

Timothy V. Pyrkov¹, Ilya S. Sokolov¹, Peter O. Fedichev^{1,2}

¹Gero PTE. LTD., Singapore 409051, Singapore

²Moscow Institute of Physics and Technology, Moscow Region 141700, Russia

Correspondence to: Timothy V. Pyrkov, Peter O. Fedichev; email: tim.pyrkov@gero.ai, peter.fedichev@gero.ai

Keywords: biological age acceleration (BAA), public health, personalized interventions, Gompertz law, resilience

Received: January 6, 2021

Accepted: March 3, 2021

Published: March 14, 2021

Copyright: © 2021 Pyrkov et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Biological age acceleration (BAA) models based on blood tests or DNA methylation emerge as a *de facto* standard for quantitative characterizations of the aging process. We demonstrate that deep neural networks trained to predict morbidity risk from wearable sensor data can provide a high-quality and cheap alternative for BAA determination. The GeroSense BAA model was trained and validated using steps per minute recordings from 103,830 one-week long and 2,599 of up to 2 years-long longitudinal samples and exhibited a superior association with life-expectancy over the average number of steps per day in, e.g., groups stratified by professional occupations. The association between the BAA and effects of lifestyles, the prevalence of future incidence of diseases was comparable to that of BAA from models based on blood test results. Wearable sensors let sampling of BAA fluctuations at time scales corresponding to days and weeks and revealed the divergence of organism state recovery time (resilience) as a function of chronological age. The number of individuals suffering from the lack of resilience increased exponentially with age at a rate compatible with Gompertz mortality law. We speculate that due to the stochastic character of BAA fluctuations, its mean and auto-correlation properties together comprise the minimum set of biomarkers of aging in humans.

INTRODUCTION

Any advances in personalized and informed lifestyle interventions to promote longevity and health will require reliable and immediate feedback on health status changes in response to treatments. Such capabilities have just recently become available in the form of biological clocks and are increasingly used in the field of quantitative aging research. State-of-art implementations involve machine learning of the associations of the DNA methylation patterns [1] or blood variables [2–4] with either the chronological age or risks of death and diseases. The aging clocks have been used in clinical trials of anti-aging interventions [5].

Large-scale biochemical or genomic profiling of Biological age acceleration (BAA) is, however, still

logistically difficult and expensive. Mobile technology holds a great promise for the democratization of population health studies. It already provides engagement tools to help customers maintain physical activity levels, body weight, and adhere to lifestyles known to promote a healthy lifespan. In 2019, one-in-five U.S. adults (21%) reported they regularly used a wearable fitness tracker or smartwatch [6]. The health and home fitness app downloads grew by 46% during COVID-19 lockdown [7].

In fact, only mobile technology can support large-scale studies involving monitoring of early signs of a disease or measuring recovery rates, all requiring sampling more often than once per week. Recent examples include the analysis of the worldwide distribution of physical activity [8], changes in physical activity levels in response to COVID-19 lockdown [9], and the

associations of physical activity and the risks of COVID-19 mortality [10, 11]. There are, however, multiple unresolved issues, such as inaccuracies of sensor data, missing data, outliers, varying measurements between devices of different manufacturers, and seasonal variation of physical activity [12, 13] -all precluding from wider acceptance of the wearables signal in population studies.

We applied deep learning technology to systematically address these challenges. We trained and characterized a simple model that learns physical activity patterns from wearable devices, which are directly associated with morbidity risks on the population level. Accordingly, the organism state representation output by this model is a single dynamic variable closely related to BAA. The neural network architecture included components specifically designed to resolve the missing data and solve transferability across platforms. We found that both blood-based and wristband step-counter-based models demonstrated surprisingly similar levels of sensitivity in applications involving BAA associations with diseases and lifestyles. Moreover, the activity-based models' signal-to-noise ratio could be improved by averaging over longer motion tracks. After just a few months of averaging, the activity-based model applied to a wristband signal may detect the effects of chronic diseases and smoking at the same level of significance as blood-based PhenoAge from [2] and Dynamic Organism State Indicator (DOSI) from [4]. The same finding held for the association of BAA with the incidence and severity of seasonal infectious diseases (including COVID-19).

Finally, we investigated the auto-correlation properties of the BAA fluctuations. The diverging autocorrelation times are typical for systems approaching tipping or disintegration point [14] and a hallmark of aging [15, 4]. Accordingly, we observed vanishing recovery rate and the exponentially increasing fraction of individuals with long recovery times in subsequent age cohorts. The number of non-resilient individuals doubled every 8 years, which is compatible with the mortality rate doubling time characteristic to the Gompertz mortality law [16]. We conclude that due to the inherent stochastic character of BAA fluctuations, the BAA mean and the BAA autocorrelation time (the resilience) are the two most basic and independent health indicators, closely related to aging and human mortality.

RESULTS

Biological age predicts morbidity and mortality

We trained the GeroSense system, a deep artificial neural network (Figure 1) to extract health-associated features from the physical activity recordings. The

system included the encoder part, which took the input in the form of a series of step count per minute measurements for at least as long as one week and compressed the signal into 4-dimensional representations (embeddings). During the training and test procedure, we used one week-long samples of steps per minute recordings for 97,320 UK Biobank and 6,510 NHANES participants along with recordings samples from longitudinal data obtained for 1,876 smartphone and 723 smartwatch users. The embedding vectors were further fed into the domain-adaptation network, trained to reduce the difference between the feature sets distribution in samples originating from different devices. In such a way, we were able to produce the most common features present in the motion data.

At the top layer, log-linear proportional hazards models of all-cause mortality are natural tools to build the biological age acceleration models, see, e.g., the PhenoAge model [2, 17]. If, however, the number of observed events is small, a simple logistic regression model provides an excellent approximation to the solution of the corresponding proportional hazards [18, 19]. Therefore, in the present study, we trained the neural network using cross-entropy loss to predict binary labels: the prevalence of at least one chronic disease. Overall, we labeled events for 23% and 29% samples in NHANES and UKB, respectively (see Materials and Methods section "Morbidity status" for the precise definition).

The model's output was the Biological Age Acceleration (BAA), estimated once per each seven days and calculated as the linear combination of the physical activity signal embeddings and biological sex label. During the training procedure, BAA was added to the chronological age of each participant to produce biological age followed by sigmoid activation layer and cross-entropy loss on the prediction of morbidity status.

To control for over-fitting, we split all data into training and test subsets. The quality of GeroSense BAA for predicting the morbidity status was similar in training and test subsets in both NHANES (Figure 2A) and UK Biobank (Figure 2C) with ROC AUC 0.60–0.61 in test subsets.

We also expected the high concordance between the mortality and morbidity predictors [20]. Accordingly, we tested the ability of the model to predict future mortality events (see Figure 2B, 2D for the summary of the GeroSense BAA model performance in NHANES and UKB datasets, respectively). The scoring performance was similar to that of morbidity status and yielded ROC AUC 0.60–0.62 in test subsets.

BAA and the life expectancy in professional occupation groups

BAA from the network was superior to average daily physical activity-based BAA in scoring life expectancy in various professional occupations. The number of

steps per day averaged over a sufficiently long period is an easy-to-understand and adjustable parameter that predicts mortality and morbidity [20]. This can be readily seen in Figure 2, where the negative logarithm of the number of daily steps (nloga) has all the properties required of BAA. However, the average

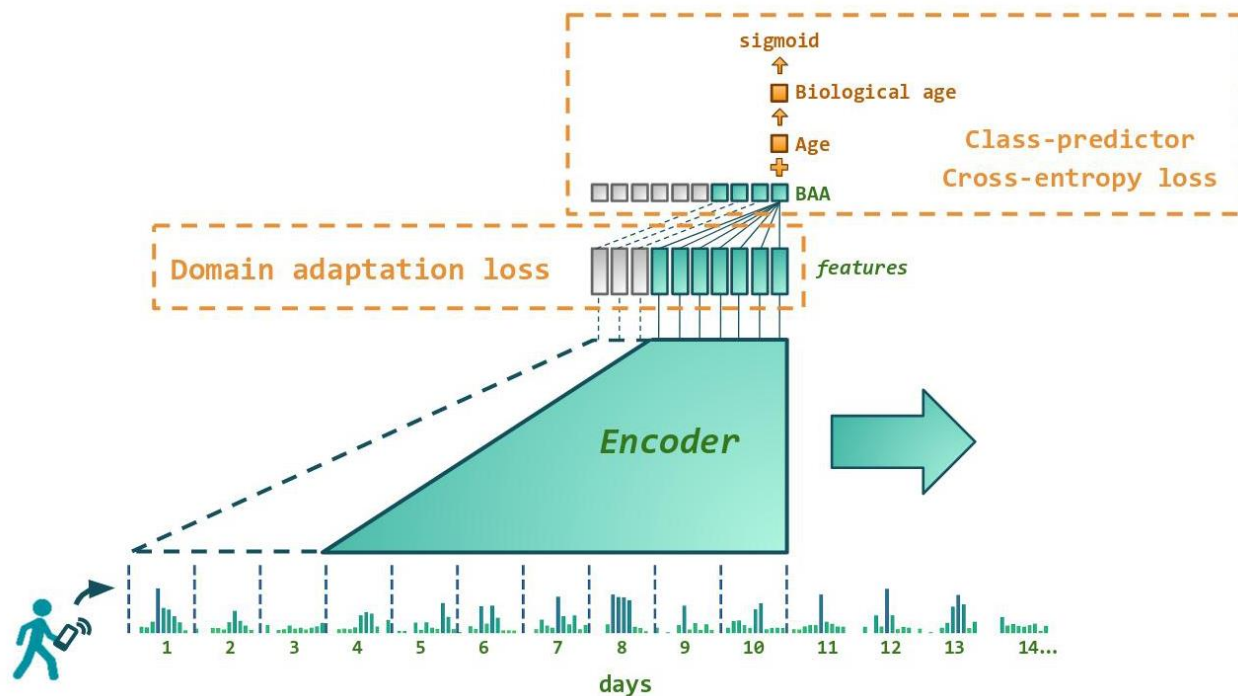


Figure 1. Architecture of the neural network predicting biological age acceleration (BAA). GeroSense model predicts BAA once per day based on step counts recorded by wearable or mobile device sensors using each individual's week-long physical activity tracks. The network components responsible for the feature extraction and BAA output are shown in green. BAA can be predicted for any sample of arbitrary length exceeding one week. For example, BAA on day 10 is predicted using the step counts data coming from day 4 through day 10, and so forth. Shown in red are the network components used only during the training procedure. One is the discriminator responsible for domain adaptation between e.g. smartphones and smartwatches. The other is the class predictor based on the log-odds ratio trained to predict morbidity binary status for UK Biobank and NHANES.

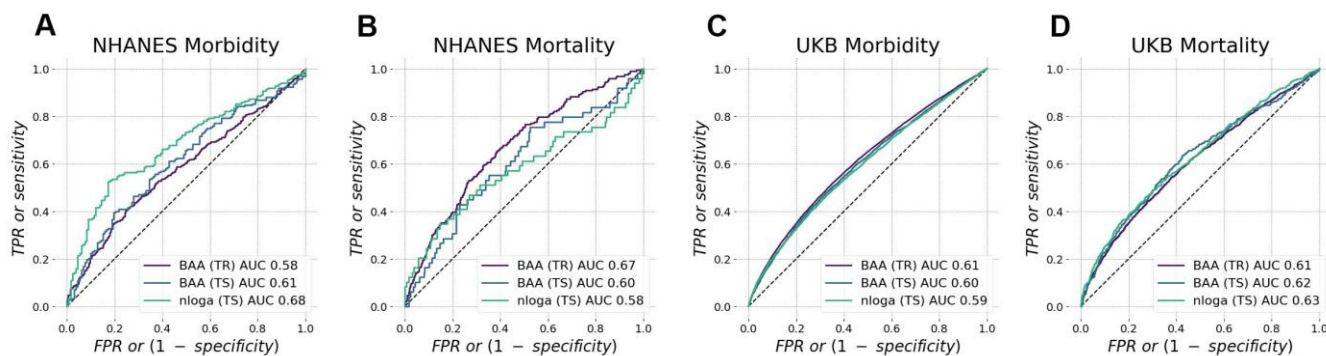


Figure 2. Biological age acceleration (BAA) ranks mortality and morbidity events. BAA estimated from patterns of intraday changes in physical activity level is associated with morbidity and mortality in NHANES (A, B) and UK Biobank (C, D) datasets. The performance was tested in participants aged 45–75 y.o. and was similar in training and test subsets.

physical activity obviously cannot be a good biological age measure. It is strongly affected by social factors and working schedule and therefore has a poor correlation with life expectancy across countries [8] and between groups of different professional occupations (Figure 3A).

Notably, the GeroSense system produced BAA from wearable sensors data, which properly ranked professional occupation groups in NHANES according to both genders' empirical life expectancy (Figure 3B). We did not have access to and hence could not test the association of physical activity and lifespan data across countries. Therefore, GeroSense BAA's ability to score the life-expectancy of populations of different countries remains an open issue.

Cross-platform transferability of BAA and seasonal variations

The embeddings of physical activity tracks depend on the signal source, whether it is a smartphone or a smartwatch. Deep Neural networks are powerful feature-extraction tools and a proper choice to address this issue. We employed the domain adaptation network minimizing the feature-wise Kullback-Leibler divergence loss between samples originating from different devices during the training procedure. The

problem is akin to batch removal. The proposed procedure helped the GeroSense network to learn the most common features between UKB, NHANES, and samples obtained from iPhone and Apple Watch.

Seasonal changes affect blood parameters [21], and physical activity patterns recorded by wearables [12]. The seasonal variations of the activity patterns may be an additional source of unwarranted fluctuations of the biological age estimates. We applied another Kullback-Leibler divergence minimization to penalize pair-wise differences in distributions of features for UK Biobank samples collected in the summer and winter.

The domain adaptation worked well: BAA level distributions were almost indistinguishable between the samples originating from smartphones and smartwatches ($p=2E-5$). In contrast, the levels of negative logarithm of average physical activity were much more different ($p=2.7E-80$). The difference was expected but is still striking since we analyzed the smartphone and smartwatch data from the same users.

The results of the statistical testing (p-values) strongly depend on the sample size. That is why, here and in all the following examples, we report p-values obtained for the same maximum size of 500 in each group. The p-values themselves are calculated using Fisher's

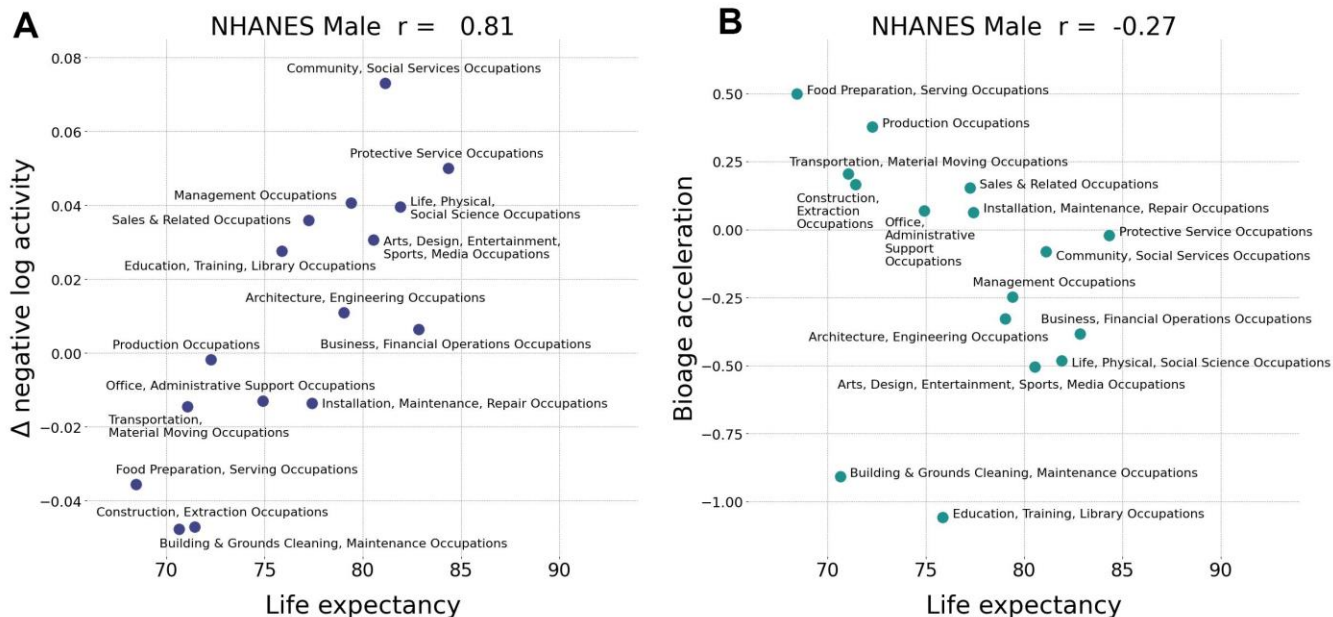


Figure 3. Biological age acceleration (BAA) correctly ranks life expectancy. (A) Assuming the negative logarithm of average daily steps is a proxy for bioage, the observed positive correlation (Pearson's $r = 0.81$ for males) with life expectancy is an incorrect prediction. (B) The negative correlation (Pearson's $r = -0.27$ for males) of GeroSense BAA with life expectancy is correct. Similar results were observed for females with Pearson's $r = 0.19$ and $r = -0.55$, respectively (data not shown). The calculations were performed in NHANES 2005–2006 cohort aged 30–60 y.o.

combined probability test (see details in Materials and Methods section).

Notably, there was a very significant drop in the physical activity levels during the COVID-19 pandemic lockdown in March through May 2020 as compared to the same period in 2019 ($p < 1E-30$ for nloga). This was consistent with what was reported earlier [9]. In contrast, the increase in BAA was much less significant ($p > 1E-10$). This may indicate that BAA responds weaker to the lockdown than the expected decrease in physical activity, see Figure 4F. Moreover, this was in contrast to the improved ability of BAA to predict future risks of COVID-19 incidence and mortality rates in UKB as compared to nloga.

The decreased average level of physical activity (nloga) was associated with the increased COVID-19 risk in UKB [11], although it was not clear if this was not an effect of chronic disease burden (also known for its association with increased BAA). In Figure 4 we report that the excess BAA predicted the increased risk of COVID-19 incidence (for example, $HR=2.4$, $p=4E-2$ for 16 of UKB subjects died from the disease) in the

subset of randomly sampled 500 UK Biobank participants free of chronic diseases at the time of measurements (2013–2015).

Side-by-side comparison of motion data- and blood-based aging clocks

We compared the performance of different BAA models for stratification of cohorts of NHANES participants of various lifestyles and health status. We have already seen in physical activity data [22] that the disease and smoking labels are associated with elevated BAA among individuals without chronic diseases. In our tests, the sensitivity of the BAA derived from blood markers was comparable to that of the self-reported questionnaire. GeroSense BAA performed consistently well in the same set of tests and conditions, see Figures 5, 6.

Estimation of the BAA from wearable sensors has an advantage over blood-based models. It arises from its ability to further improve the signal-to-noise ratio by averaging over sufficiently long motion data streams. We demonstrated this with self-reported morbidity and

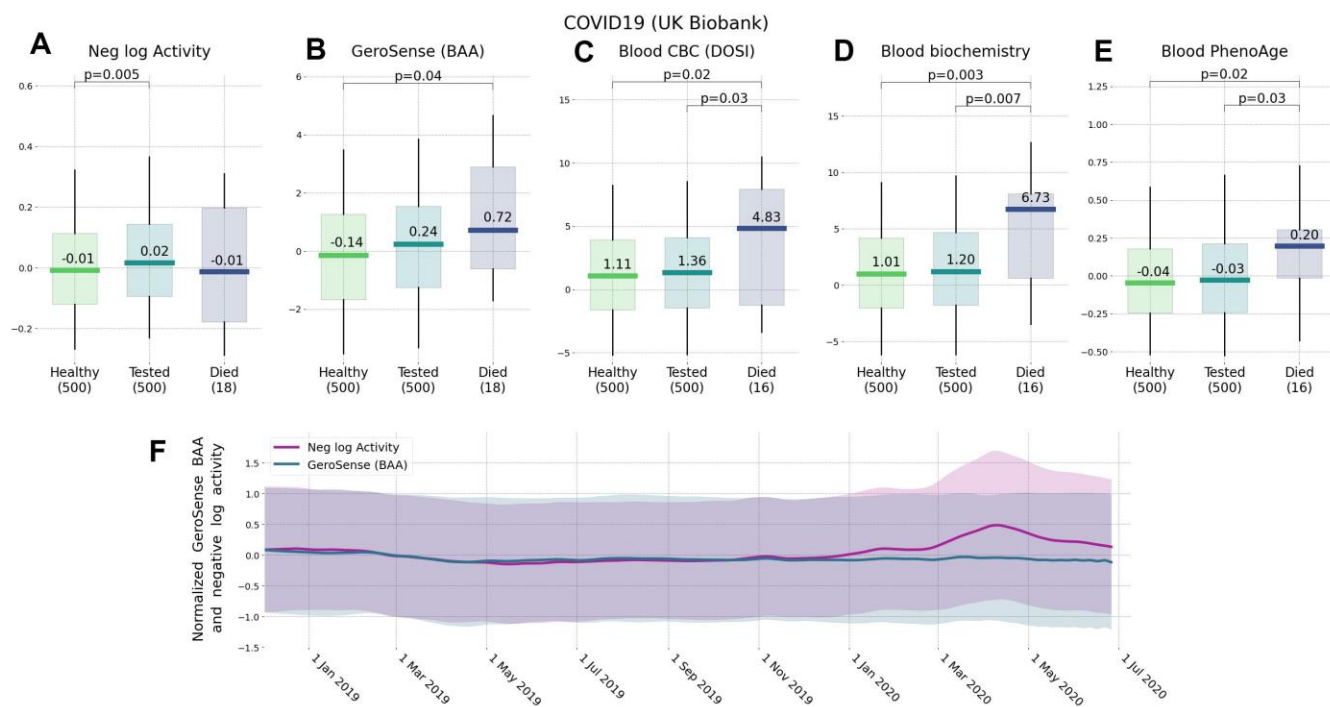


Figure 4. Effect of lockdown and the future risks of COVID-19 ranked by wearable BAA and blood-based bioage models. Association of BAA with the future incidence of COVID-19: (A) BAA in the form of the negative logarithm of daily step counts, (B) GeroSense BAA, (C) CBC-based DOSI [4], (D) CBC and Blood biochemistry hazards model, and (E) Blood-based PhenoAge [2] (all data are given for UK Biobank participants aged 45–75 y.o.). (F) Longitudinal data obtained by smartwatch show that while GeroSense is more sensitive to future risks it is also more selective and does not change immediately (green line) due to merely walking less during March to May 2020 lockdown while the average activity level does (red line). The values in panel (F) were scaled to zero mean and unit variance for comparison; shaded areas show one standard deviation range at each time point.

smoking status provided by a smartphone app and wristband tracker users.

Averaging of GeroSense BAA predictions over a few weeks-long tracks led to a dramatic improvement of association between the BAA and morbidity/smoking status (Figure 7A). As expected, the sensitivity of the model was comparatively lower once we used smartphones instead of wristbands as the source of the data (Figure 7B) but also improved upon averaging over several weeks.

Longitudinal analysis of BAA fluctuations reveals age-dependent loss of resilience

BAA reversibly depends on lifestyles. Hence, BAA is a dynamic variable more characteristic of stress rather than aging and responding to random organism state

perturbations in a stochastic manner. We used longitudinal tracks of step counts from Fitbit devices and calculated the autocorrelation function for every user. The autocorrelation function decayed exponentially. Accordingly, we carried out the exponential fit to infer the autocorrelation time as a measure of recovery rate or resilience. This quantity is a natural quantitative measure of an organism's ability to recover its equilibrium state after stress.

The characteristic decay time was typically in the range of a few weeks and increased with age. Figure 8A shows the dependence of the recovery rate (the inverse auto-correlation time) on chronological age. The graph was produced by averaging over age-stratified cohorts and resembles much what we have previously reported for blood-based marker DOSI [4]. The recovery rate decreased approximately linearly with age, indicating

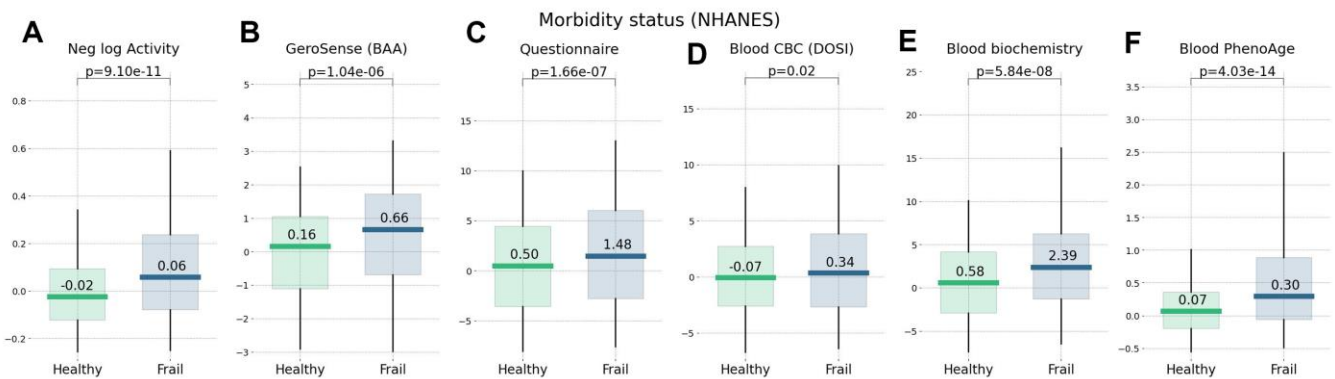


Figure 5. Morbidity status scored by wearable BAA and blood-based bioage models. BAA and chronic diseases: (A) the combined logarithm of daily step counts, (B) GeroSense BAA (C) questionnaire [22], (D) CBC-based DOSI [4], (E) log-mortality risk trained using combined CBC and Blood biochemistry variables, and (F) Blood-based PhenoAge [2]. The plots are produced for NAHNES participants aged 45–75 y.o.

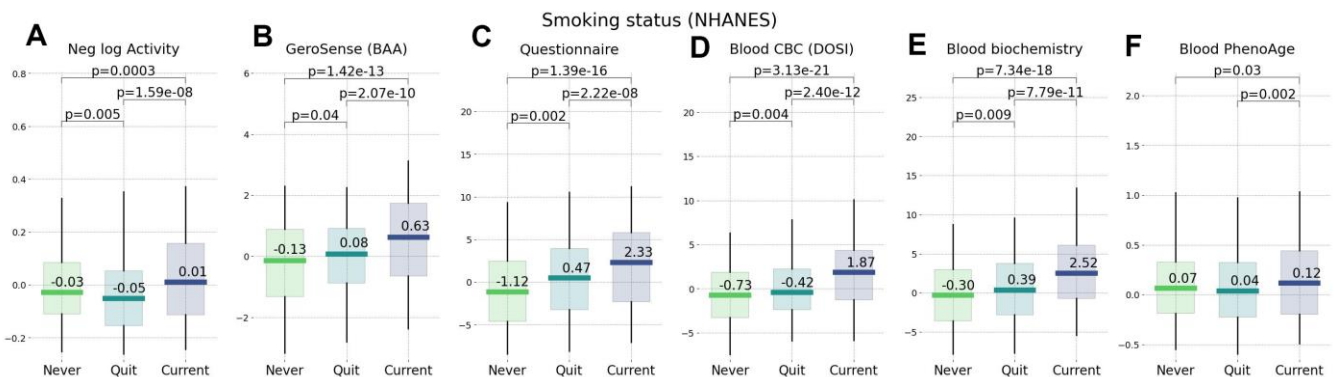


Figure 6. Smoking status representing an unhealthy lifestyle ranked by wearable BAA and blood-based bioage models. BAA and smoking: (A) BAA in the form of the negative logarithm of daily step counts, (B) GeroSense BAA, (C) questionnaire [22], (D) CBC-based DOSI [4], (E) log-mortality risk trained using combined CBC and Blood biochemistry variables (F) PhenoAge [2]. The plots are produced for NAHNES participants aged 45–75 y.o.

the effective loss of resilience at some age exceeding 100 y.o. The same extrapolation would suggest that the recovery time increases approximately hyperbolically and would diverge at the same age, indicating the complete loss of resilience and the dynamic stability of the organism state.

To further investigate the relationship between resilience and aging, we identified individuals, which failed to recover quickly under stress. We established a somewhat arbitrary resilience cutoff corresponding to the recovery time exceeding 3 weeks. The fraction of such “non-resilient” individuals increased exponentially

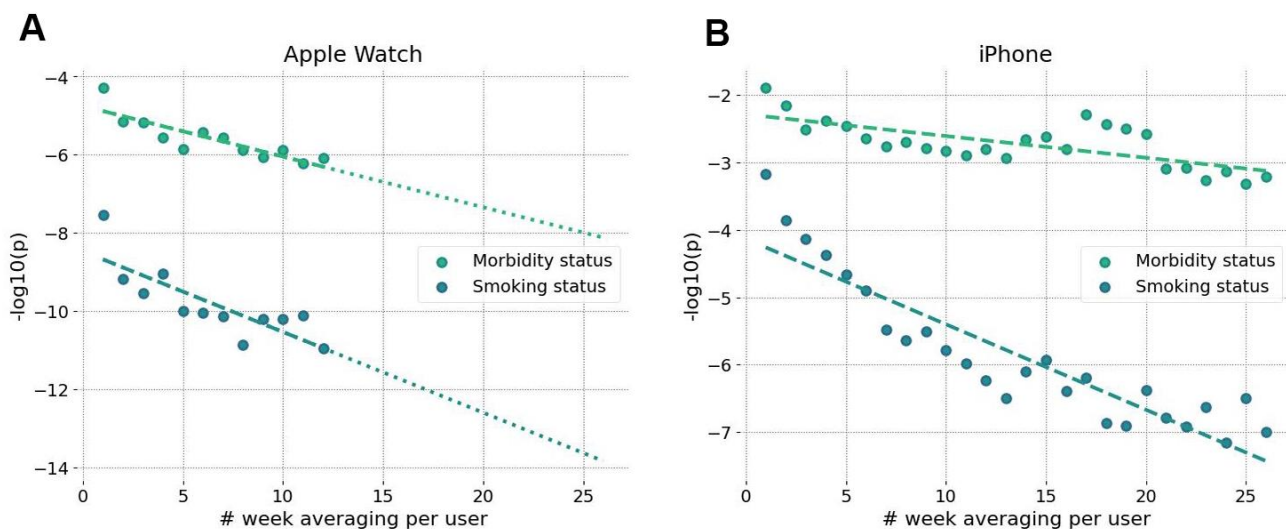


Figure 7. Accuracy of BAA grows with longer data collection intervals. Significance of association of GeroSense BAA with morbidity and current smoking status is improved when BAA is averaged over several weeks of data obtained from sensors of smartwatch (A) and smartphone (B).

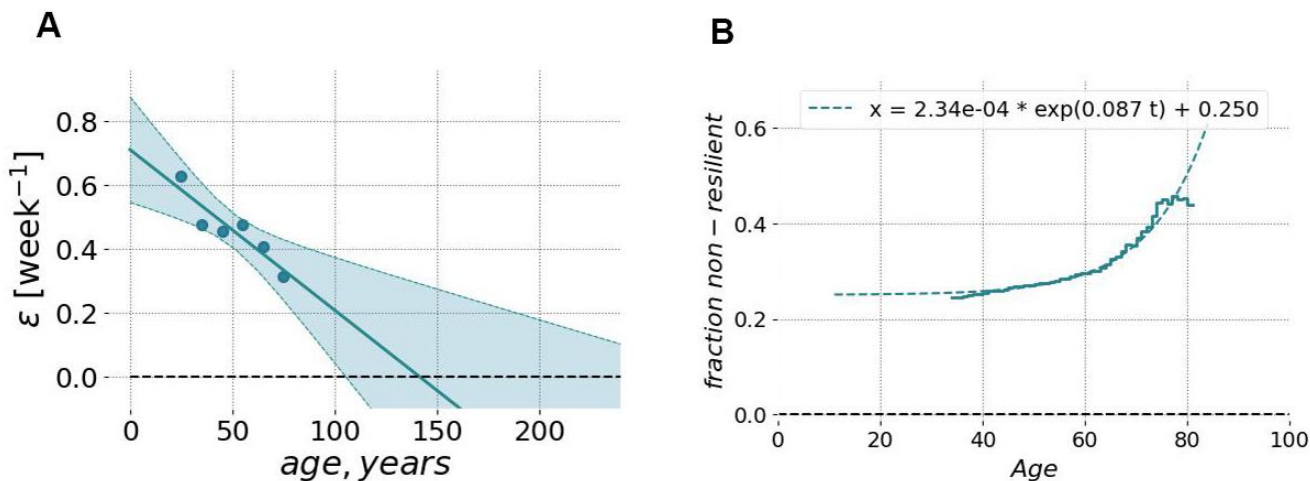


Figure 8. Resilience and its age-related degradation can be measured using longitudinal motion sensor data. (A) The relaxation rate (or the inverse characteristic recovery time) computed for sequential age-matched cohorts of Fitbit users decreased approximately linearly with age. The recovery rate could be extrapolated to zero in the age exceeding ~110 y.o. (at this point, we may expect the complete loss of resilience and, hence, loss of stability of the organism state). The shaded area shows the 95% confidence interval of fit using GeroSense BAA. (B) The fraction of individuals suffering from the lack of resilience (defined as BAA’s autocorrelation time exceeding 3 weeks; the vertical axis) as the function of chronological age (the horizontal axis). The autocorrelation time was computed from longitudinal tracks of GeroSense BAA predicted for Fitbit wristband users.

as a function of age (see Figure 8B). Moreover, this growth demonstrated the characteristic doubling rate of 0.087 per year, which was close to the mortality rate doubling rate according to the Gompertz mortality law.

DISCUSSION

We report the development and characterization of a deep neural network model trained to quantify the state of human health from the analysis of intraday physical activity tracks collected by consumer wearable devices (including mobile phones). The quantity has properties of biological age acceleration (BAA): it is associated with chronic diseases and life-shortening lifestyles, predicts the risks of death and future incidence of chronic diseases in cohorts of individuals free of chronic diseases [4].

Deep neural networks are natural tools for learning non-trivial and highly non-linear representations of the input data. Convolutional and recurrent networks have been used for the analysis of intraday physical activity data streams from wearable devices and predictive modeling of health outcomes [23] including biological age [17, 24]. Often such models demonstrate a moderate improvement in accuracy at a price of a decreased transferability across datasets with different baseline feature levels. This is, of course, is well-known batch effect problem in large-scale studies in biology [25], which is often aggravated by feature-rich deep learning architectures [26, 13].

GeroSense BAA model employs additional neural network components to address this domain shift problem to ensure learning device-independent representations of the input signal. To achieve this goal, we imposed an additional loss in the course of training to penalize model parameters if distributions of learned representations were too far apart for data from different domains (devices). Without such a domain adaptation, the properties of the signal may indeed be very different even in the same biological context. For example, the (log-scaled) average number of daily steps recorded by phone was significantly lower ($p=2.7E-80$) than that by the smartwatch in the data from the same users. GeroSense BAA network successfully resolved this batch effect and yielded essentially indistinguishable BAA distributions for the same population ($p=2E-5$).

The average activity level recorded by the same device in a group of people of the same gender, professional occupation, and country of residence is already an excellent and popular proxy to biological age. The association between the mean activity and health is

robust and hence is the basis for the popular recommendation to take a minimum of 10,000 steps a day [27]. However, the average activity level is highly context-dependent, which is why it is poorly associated with life expectancy across countries [8]. In our study, we demonstrate that the average activity is incorrectly (negatively) associated with the life expectancy across professional occupation groups (Figure 3A).

The device-independent features from intraday physical activity patterns from the GeroSense network are still associated with health but decoupled from the mean activity. The procedure did not undermine the predictive power of GeroSense model, as we could see from the BAA association with mortality events (Figure 2). GeroSense BAA was superior in scoring life expectancy in professional occupation subgroups (Figure 3B). This feature of the model should be useful in applications involving health risk assessment and life insurance applications.

Biological clocks based on mortality risk, including GeroSense BAA, are associated with the prevalence of chronic diseases (Figure 5) and life-shortening lifestyles, such as smoking, in a reversible way (Figure 6). This is totally consistent with earlier observations of the effect of smoking on physical activity [20], blood markers [4, 22], and DNAm PhenoAge [2].

In NHANES cohorts, the GeroSense model produced the association between the BAA and the morbidity and smoking labels at the significance level matching that of the BAA calculated based on self-reported health questionnaire [22], blood test-based bioage including CBC only [4], and blood biochemistry [22], and Phenotypic Age [2].

The longitudinal character of motion data provides a natural way to improve the signal-to-noise ratio by averaging over sufficiently long tracks (see Figure 7A, 7B). This may be critical for mobile phone applications since the step counts recorded by phones suffer from missing data whenever a device is idle and is not recording the user's walks. Our analysis suggests that GeroSense BAA from smartphones can be averaged to a useful level once at least a few months of data are available for an individual. The inferior performance of the biological age model in smartphone data can be compensated by smartphone population coverage compared to that of wristband wearable devices. The smartphone motion data can be used for truly large-scale epidemiological studies involving cohort comparisons. The latter factor might turn important to mitigate the issues of non-representative datasets due to possible income/health status [13, 28] and already observed enrollment biases [29].

We observed, that GeroSense BAA is also associated with the incidence of non-chronic diseases. This is consistent with earlier observations of the association of lower physical activity levels and risks of COVID-19 infection [10, 11], although it was not clear whether this is an effect of chronic diseases, also negatively affecting mobility. GeroSense BAA was better associated with the incidence of COVID-19 than the average physical activity level in UKB among a sub-population of individuals free of chronic health conditions (Figure 4).

The average physical activity dropped worldwide in 2020 in the course of COVID-19 lockdown [9]. We also observed a significant change in (log-scaled) number of daily step counts in our data, but not in GeroSense BAA during March–May 2020 as compared to the same period in 2019. We provided evidence suggesting that GeroSense BAA more efficiently scores those at risk of getting an infection than the physical activity level. The effects of lockdown on morbidity risk may be smaller than one could expect simply by monitoring the drop of the activity. Further studies including direct association with epidemiological data are required to test this hypothesis.

The idea of reducing complex biological signals to as little as one variable, the BAA, in relation to the current or future health arises from the effectively low dimensionality of physiological systems. Typically, physiological and behavioral responses manifest themselves as highly coordinated changes in physiological variables, such as blood tests [4] or daily physical activity patterns [20]. The increasing concordance between the physiological indices is expected to increase late in life, as the range of the fluctuations and the organism state recovery time effectively diverge at advanced ages indicating a maximum attainable lifespan [4]. On the contrary, the number of the relevant variables is expected to increase if we turn to the characterization of the organism state variation at a higher sampling rate. This might be the case for a situation involving response to an acute illness on time scales of days or a few weeks [30], such as increased RHR during fever [31, 32] or change in sleep patterns as potentially a COVID-19 specific signal [33, 34].

The quantitative characterization of the dynamic properties of BAA fluctuations or recovery processes requires a reliable determination of baseline BAA. This task may be hampered by seasonal variation of the physiological state variables, such as blood tests [21, 35], blood pressure [36], resting heart rate [37], and of course physical activity [12]. High-quality research studies acknowledge this problem and adjust for baseline oscillations [12, 28]. Such corrections are

straightforward for relatively short time scales involved in acute respiratory illnesses [30] or post-operative recovery [38].

Unfortunately, proper adjustments are not always possible in practice. Health outcomes associated with BAA may be years apart from the time (and hence the season) of observations [19]. Otherwise, the time of measurements may be available at poor granularity. For example, NHANES provides publicly only the binary labels corresponding to the winter–spring (November–April) or summer–fall (May–October) seasons.

We trained the GeroSense BAA model with an additional loss penalizing the winter-summer distribution difference. In such a way, the model output is decoupled from seasonal variations and yet demonstrated pretty good performance in ranking health outcomes. We expect that this feature of GeroSense BAA will be handy for practical applications.

The longitudinal character of motion data allows the investigation of organism state fluctuations in response to natural stresses and diseases. We computed autocorrelation functions of GeroSense BAA along the individual BAA trajectories. The recovery rate measured as the inverse decay time of the autocorrelation function demonstrated an age-dependent decrease (Figure 8A). Extrapolation to advanced ages shows, that the recovery rate vanishes (and hence the resilience formally diverges) at some age exceeding 100 years, which may be an indication of limiting lifespan [4].

The recovery time among the most healthy individuals was in the range of a few weeks. We used a somewhat arbitrary cutoff corresponding to the recovery rate less than 3 week^{-1} and used it to mark individuals with longer recovery time as those who lack resilience. We observed a progressive exponential increase of the fraction of non-resilient persons in the population with age (Figure 8B). This number grew and doubled every 8 years, which is close to the mortality rate doubling time in Gompertz mortality law for the human population [16].

Long auto-correlation times of state fluctuations are typical for complex systems approaching a tipping point or in the process of disintegration [14] and represent a hallmark of aging [15, 4]. Case fatality rates (CFR) accelerate with age in the case of COVID-19, stroke, and probably other diseases. The characteristic doubling rate in the case of COVID-19 is reported in [39] as 6–9 years. Our estimation from the figure in [40] yielded ≈ 10 years for the doubling rate for one-year survival of stroke patients. The physiology of stroke and infection

diseases is apparently very different. The similarity of patterns of age-dependence of CFR is intriguing and may suggest that the loss of resilience may be a good marker of the approaching loss of dynamic stability of an organism and hence a major and universal contributing factor to the fatality.

The reversible character of the association between mortality risk-based BAA and unhealthy lifestyles (such as smoking) suggests that BAA is not a biomarker of aging but instead is a measure of the overall stress level. BAA's dependence on age in large cross-sectional datasets is a marker of stress imposed by the increasing burden of chronic diseases. The high sampling rate achievable by the motion data lends us a richer set of biomarkers associated with age. Aside from the average BAA level, the continuous data collected by wearable sensors provides a practical opportunity to investigate the autocorrelation and variance properties of BAA fluctuations, which are independent organism state variables, each uniquely informing about the user's health state. We can hardly imagine a large-scale blood test study involving sampling more often than once a month or so for healthy people. Therefore, only the motion data analysis exemplified here is the only technology currently up for the task.

Wearable device motion data have already been used for monitoring acute illnesses including detection of early signs of the outbreak of influenza-like illnesses [28] and COVID-19 [30, 34]. Application of motion data, including the wider deployment of the GeroSense system, described here, should provide means to monitor levels of stress and resilience in response to environmental conditions or interventions on a population level in different countries and socio-economic groups in future studies. We hope that future developments will lead to further applications of AI in geroscience research, public health, and policy decision-making.

MATERIALS AND METHODS

Datasets

UK Biobank

Physical activity for UK Biobank participants aged 40–80 y.o. (54,777 female and 42,543 male) was measured by Axivity AX3 tri-axial accelerometers worn on the wrist for one week. We converted 100Hz raw acceleration measurements to step counts per minute to fit the format of data in other datasets used in this study. The number of steps during each consecutive minute was counted as the number of peaks of the absolute value of acceleration exceeding

1.3g. To ensure the local noise does not affect the result, only one peak (the highest) was counted in each 480ms sliding window with a step of 160ms resulting in at most 3 step counts during each 960ms. Steps closer to each other than 90s were combined into walking bouts and bouts with less than 5 steps in total were discarded.

NHANES

Physical activity for NHANES 2005–2006 participants was used in the form of step counts per minute collected by ActiGraph AM-7164 single-axis accelerometer worn on hip. Data were retrieved from the file “Physical Activity Monitor” of the “Examination data” category. Samples for 3,362 female and 3,148 male participants aged 6–85 y.o. were used.

Healthkit

Physical activity for users of Gero app aged 45–75 y.o. (464 female and 1,412 male users of smartphone, 125 female and 598 male users of smartwatch) was obtained from Healthkit. Raw activity data comprised the number of steps recorded by either smartphone or smartwatch during a time period with start and end timestamps and was resampled to equispaced time series of steps per minute.

Morbidity status

Binary morbidity status for the Healthkit dataset was assigned according to response to the survey question "Have you ever been told you have one of the following: diabetes, hypertension, cancer, coronary heart disease, heart failure, heart attack, or stroke?" Binary morbidity status of NHANES and UK Biobank participants was assigned according to the presence of at least one of those diagnoses. We used NHANES data on health condition and age at diagnosis available in the questionnaire category “Medical Conditions” (MCQ). Data on diabetes and hypertension was retrieved additionally from questionnaire categories “Diabetes” (DIQ) and “Blood Pressure and Cholesterol” (BPQ), respectively. For UK Biobank we aggregated ICD10 (block level) data to match that of NHANES and used the following ICD10 codes to cover the health conditions in UK Biobank: diabetes (E10-E14), hypertension (I10-I15), cancer (C00-C99), coronary heart disease (I20-I25), congestive heart failure (I50), myocardial infarction (I21, I22), and stroke (I60-I64).

Life expectancy

Empirical life expectancy from birth was determined for professional occupation groups using linked death register follow-up data for NHANES 2005–2015

surveys. To do that we fitted parameters of Gompertz likelihood adopted from [41]:

$$-\ln LH = \sum_{all} \frac{M_0}{\Gamma} e^{\Gamma t_n} (1 - e^{-\Gamma \Delta t_n}) - \sum_{dead} (\ln M_0 + \Gamma t_n),$$

where M_0 and Γ are the initial mortality rate and mortality doubling rate of the Gompertz mortality law, t_n is the age of n -th participant at the end of followup, and Δt_n is the follow-up time since enrollment in NHANES survey.

Once M_0 and Γ were obtained by the fit for each professional occupation group, the life expectancy \bar{t} in the group was calculated as:

$$\bar{t} = \frac{1}{\Gamma} \ln \left(\frac{\Gamma}{M_0} \right) - \frac{\gamma}{\Gamma},$$

Where $\gamma \approx 0.58$ is the Euler-Mascheroni constant. The expression for life expectancy is asymptotically correct whenever $M_0 / \Gamma \ll 1$, which is definitely true in human cohorts.

Statistical analysis

Statistical analysis of the association of various Biological Age measures with morbidity/smoking status was performed using two-sided Mann-Whitney test. To ensure the reported p-values are comparable between tests we used the same cutoff of maximum of 500 samples in each test with 100 random samplings followed by combining p-value according to Fisher's combined probability method [42]. All statistical tests were carried out using the python package SciPy (version 1.5.2).

Blood tests-based biological age models

In this work, we used blood tests-based biological age models trained using Cox proportional hazards approach in NHANES mortality follow-up data and reported elsewhere earlier. The Blood CBC (DOSI) model was trained using log-scaled values of hemoglobin, mean corpuscular volume, mean corpuscular hemoglobin concentration, red blood cell distribution width, red blood cell, platelet, neutrophil, lymphocyte, monocyte, and eosinophil counts as well as biological sex label [4]. The Blood Biochemistry model additionally included age, and log-scaled values of C-reactive protein, albumin, alkaline phosphatase, gamma-glutamyl transferase, globulin, and serum glucose [22]. The Blood PhenoAge model was based on age, albumin, creatinine, serum glucose, log-scaled C-reactive protein, lymphocyte percent, mean cell volume,

red cell distribution width, alkaline phosphatase, and white blood cell count [2].

Neural network architecture

Deep neural network architecture is schematically shown in Figure 1. Wearable data is input in the form of a continuous array of steps per minute. The input is immediately converted to a one-hot embedding representation, where each bin corresponds to an increment of 4 steps per minute. Next, the encoded data is processed by a block of 16 1D-convolutional layers, each having 16 filters with a kernel size of 3 and "elu"-activation. One in two convolutional layers is followed by a local max-pooling with stride 2, 3 or 5, and each layer is followed by batch-normalization. The output of the convolutional block was 4 features per every 1440 points in the input array, which corresponds to the number of minutes in one day. Finally, the features were subject to a 7 day-long average pooling and linearly combined with binary biological sex label so that the deep neural network was capable of outputting a prediction once per day based on 7 previous days.

The output of the deep neural network was interpreted as the Biological Age Acceleration (BAA) expressed in years of healthy life expectancy gained or lost. To guarantee this, during the supervised training of class label predictor we obtained the value of the Biological age of each NHANES and UK Biobank user by adding the network output (BAA) to the chronological age. The Biological age was then subject to sigmoid activation and fitted to binary morbidity status label, assuming that such procedure is an approximation to fitting proportional hazards model [18, 19].

The Domain adaptation networks were employed in the form of pairwise Kullback-Leibler divergence loss functions applied to enforce similar feature distributions for samples from UK Biobank on one side and NHANES, HealthKit smartphones, and smartwatches on the other side. Additionally, a domain adaptation was applied to UK Biobank samples collected during summer and winter as well as to samples with up to 3 zero-imputed (missing) days.

The training procedure was run for 2000 iterations, each batch comprising 256 samples. The class predictor was trained on each iteration for UK Biobank samples and only on one in five iterations for NHANES to avoid potential overfitting since the number of NHANES samples was small. All domain adaptation networks were trained on each iteration. Each network was trained using Adam optimizer as implemented in python package tensorflow-gpu (version 2.3.1) with a learning rate of $1E-3$.

Data and code availability

Data and code are available at <https://gerosense.ai> from the corresponding authors upon reasonable request.

AUTHOR CONTRIBUTIONS

All authors designed the study and analyzed the results. T.V. Pyrkov and I.S. Sokolov performed calculations and data analysis. All authors discussed the results, wrote and reviewed the manuscript.

ACKNOWLEDGMENTS

This research has been conducted using data from UK Biobank, a major biomedical database (UK Biobank website: <https://www.ukbiobank.ac.uk>; UK Biobank project ID 21988).

CONFLICTS OF INTEREST

P.O. Fedichev is a shareholder of Gero PTE. T.V. Pyrkov, I.S. Sokolov, P.O. Fedichev are employees of Gero PTE. LTD. The study was funded by Gero PTE. LTD.

REFERENCES

1. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013; 14:R115. <https://doi.org/10.1186/gb-2013-14-10-r115> PMID:24138928
2. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, Reiner AP, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY).* 2018; 10:573–91. <https://doi.org/10.18632/aging.101414> PMID:29676998
3. Putin E, Mamoshina P, Aliper A, Korzinkin M, Moskalev A, Kolosov A, Ostrovskiy A, Cantor C, Vijg J, Zhavoronkov A. Deep biomarkers of human aging: application of deep neural networks to biomarker development. *Aging (Albany NY).* 2016; 8:1021–33. <https://doi.org/10.18632/aging.100968> PMID:27191382
4. Pyrkov TV, Avchaciov K, Tarkhov AE, Menshikov LI, Gudkov AV, Fedichev PO. Longitudinal analysis of blood markers reveals progressive loss of resilience and predicts ultimate limit of human lifespan. *BioRxiv.* 2019. <https://doi.org/10.1101/618876>
5. Fahy GM, Brooke RT, Watson JP, Good Z, Vasanaawala SS, Maecker H, Leipold MD, Lin DT, Kobor MS, Horvath S. Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell.* 2019; 18:e13028. <https://doi.org/10.1111/acer.13028> PMID:31496122
6. Vogels EA. About one-in-five americans use a smart watch or fitness tracker. Washington, DC: Pew Research Centre, 2020.
7. Ang C. Fitness Apps Grew by Nearly 50% During the First Half of 2020, Study Finds. *World Economic Forum.* 2020. <https://www.weforum.org/agenda/2020/09/fitness-apps-gym-health-downloads>
8. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. *Nature.* 2017; 547:336–39. <https://doi.org/10.1038/nature23018> PMID:28693034
9. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, Olgin JE. Worldwide effect of COVID-19 on physical activity: a descriptive study. *Ann Intern Med.* 2020; 173:767–70. <https://doi.org/10.7326/M20-2665> PMID:32598162
10. Zhang X, Li X, Sun Z, He Y, Xu W, Campbell H, Dunlop MG, Timofeeva M, Theodoratou E. Physical activity and COVID-19: an observational and Mendelian randomisation study. *J Glob Health.* 2020; 10:020514. <https://doi.org/10.7189/jogh-10-020514> PMID:33312507
11. Ying K, Zhai R, Pyrkov TV, Mariotti M, Fedichev PO, Shen X, Gladyshev VN. Genetic and phenotypic evidence for the causal relationship between aging and COVID-19. *medRxiv.* 2020. <https://doi.org/10.1101/2020.08.06.20169854>
12. Strain T, Wijndaele K, Dempsey PC, Sharp SJ, Pearce M, Jeon J, Lindsay T, Wareham N, Brage S. Wearable-device-measured physical activity and future health risk. *Nat Med.* 2020; 26:1385–91. <https://doi.org/10.1038/s41591-020-1012-3> PMID:32807930
13. Hicks JL, Althoff T, Sosis R, Kuhar P, Bostjancic B, King AC, Leskovec J, Delp SL. Best practices for analyzing large-scale health data from wearables and smartphone apps. *NPJ Digit Med.* 2019; 2:45. <https://doi.org/10.1038/s41746-019-0121-1> PMID:31304391
14. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, Held H, van Nes EH, Rietkerk M, Sugihara G. Early-warning signals for critical transitions. *Nature.* 2009; 461:53–59. <https://doi.org/10.1038/nature08227> PMID:19727193
15. Podolskiy D, Molodtsov I, Zenin A, Kogan V, Menshikov LI, Gladyshev V, Reis RJS, Fedichev PO. Critical

- dynamics of gene networks is a mechanism behind ageing and Gompertz law. arXiv. 2015. <https://dev.arxiv.org/abs/1502.04307>
16. Gompertz B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London*, 1825; 115:513–583. <https://doi.org/10.1098/rstl.1825.0026>
 17. Pyrkov TV, Slipensky K, Barg M, Kondrashin A, Zhurov B, Zenin A, Pyatnitskiy M, Menshikov L, Markov S, Fedichev PO. Extracting biological age from biomedical data via deep learning: too much of a good thing? *Sci Rep*. 2018; 8:5210. <https://doi.org/10.1038/s41598-018-23534-9> PMID:29581467
 18. Green MS, Symons MJ. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. *J Chronic Dis*. 1983; 36:715–23. [https://doi.org/10.1016/0021-9681\(83\)90165-0](https://doi.org/10.1016/0021-9681(83)90165-0) PMID:6630407
 19. Abbott RD. Logistic regression in survival analysis. *Am J Epidemiol*. 1985; 121:465–71. <https://doi.org/10.1093/oxfordjournals.aje.a114019> PMID:4014135
 20. Pyrkov TV, Getmantsev E, Zhurov B, Avchaciov K, Pyatnitskiy M, Menshikov L, Khodova K, Gudkov AV, Fedichev PO. Quantitative characterization of biological age and frailty based on locomotor activity records. *Aging (Albany NY)*. 2018; 10:2973–90. <https://doi.org/10.18632/aging.101603> PMID:30362959
 21. Liu B, Taioli E. Seasonal variations of complete blood count and inflammatory biomarkers in the US population - analysis of NHANES data. *PLoS One*. 2015; 10:e0142382. <https://doi.org/10.1371/journal.pone.0142382> PMID:26544180
 22. Pyrkov TV, Fedichev PO. Biological age is a universal marker of aging, stress, and frailty. In *Biomarkers of Human Aging*, Springer, 2019. 23–36. https://doi.org/10.1007/978-3-030-24970-0_3
 23. Quisel T, Kale DC, Foschini L. Intra-day activity better predicts chronic conditions. arXiv. 2016. <https://arxiv.org/abs/1612.01200>
 24. Rahman SA, Adjero DA. Deep learning using convolutional LSTM estimates biological age from physical activity. *Sci Rep*. 2019; 9:11425. <https://doi.org/10.1038/s41598-019-46850-0> PMID:31388024
 25. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010; 11:733–39. <https://doi.org/10.1038/nrg2825> PMID:20838408
 26. Cohen AA, Morissette-Thomas V, Ferrucci L, Fried LP. Deep biomarkers of aging are population-dependent. *Aging (Albany NY)*. 2016; 8:2253–55. <https://doi.org/10.18632/aging.101034> PMID:27622833
 27. Choi BC, Pak AW, Choi JC, Choi EC. Daily step goal of 10,000 steps: a literature review. *Clin Invest Med*. 2007; 30:E146–51. <https://doi.org/10.25011/cim.v30i3.1083> PMID:17716553
 28. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. *Lancet Digit Health*. 2020; 2:e85–93. [https://doi.org/10.1016/S2589-7500\(19\)30222-5](https://doi.org/10.1016/S2589-7500(19)30222-5) PMID:33334565
 29. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK biobank participants: a prospective population-based study. *Lancet*. 2015; 386:533–40. [https://doi.org/10.1016/S0140-6736\(15\)60175-1](https://doi.org/10.1016/S0140-6736(15)60175-1) PMID:26049253
 30. Natarajan A, Su HW, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. *NPJ Digit Med*. 2020; 3:156. <https://doi.org/10.1038/s41746-020-00363-7> PMID:33299095
 31. Karjalainen J, Viitasalo M. Fever and cardiac rhythm. *Arch Intern Med*. 1986; 146:1169–71. PMID:2424378
 32. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza Rose SM, Perelman D, Colbert E, Runge R, Rego S, Sonecha R, Datta S, McLaughlin T, Snyder MP. Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS Biol*. 2017; 15:e2001402. <https://doi.org/10.1371/journal.pbio.2001402> PMID:28081144
 33. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, Alavi A, Celli A, Higgs E, Dagan-Rosenfeld O, Fay B, Kirkpatrick S, Kellogg R, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. *Nat Biomed Eng*. 2020; 4:1208–20. <https://doi.org/10.1038/s41551-020-00640-6> PMID:33208926

34. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, Kheterpal V, Topol EJ, Steinhubl SR. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nat Med*. 2021; 27:73–77. <https://doi.org/10.1038/s41591-020-1123-x> PMID:33122860
35. Sailani MR, Metwally AA, Zhou W, Rose SM, Ahadi S, Contrepolis K, Mishra T, Zhang MJ, Kidziński Ł, Chu TJ, Snyder MP. Deep longitudinal multiomics profiling reveals two biological seasonal patterns in California. *Nat Commun*. 2020; 11:4933. <https://doi.org/10.1038/s41467-020-18758-1> PMID:33004787
36. Kim KI, Nikzad N, Quer G, Wineinger NE, Vegreville M, Normand A, Schmidt N, Topol EJ, Steinhubl S. Real world home blood pressure variability in over 56,000 individuals with nearly 17 million measurements. *Am J Hypertens*. 2018; 31:566–73. <https://doi.org/10.1093/ajh/hpx221> PMID:29365036
37. Quer G, Gouda P, Galarnyk M, Topol EJ, Steinhubl SR. Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: retrospective, longitudinal cohort study of 92,457 adults. *PLoS One*. 2020; 15:e0227709. <https://doi.org/10.1371/journal.pone.0227709> PMID:32023264
38. Ramirez E, Marinsek N, Bradshaw B, Kanard R, Foschini L. Continuous digital assessment for weight loss surgery patients. *Digit Biomark*. 2020; 4:13–20. <https://doi.org/10.1159/000506417> PMID:32399512
39. Santesmasses D, Castro JP, Zenin AA, Shindyapina AV, Gerashchenko MV, Zhang B, Kerepesi C, Yim SH, Fedichev PO, Gladyshev VN. COVID-19 is an emergent disease of aging. *Aging Cell*. 2020; 19:e13230. <https://doi.org/10.1111/accel.13230> PMID:33006233
40. Olsen TS, Andersen ZJ, Andersen KK. Age trajectories of stroke case fatality: leveling off at the highest ages. *Epidemiology*. 2011; 22:432–36. <https://doi.org/10.1097/EDE.0b013e3182117b3d> PMID:21346579
41. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models by Ralf Bender, Thomas Augustin and Maria Blettner, *Statistics in Medicine* 2005; 24:1713–1723. *Stat Med*. 2006; 25:1978–9. <https://doi.org/10.1002/sim.2369> PMID:16680804
42. Fisher RA. *Statistical methods for research workers*, (5th Ed). Oliver and Boyd, Edinburgh. 1934.