

Cardiac age detected by machine learning applied to the surface ECG of healthy subjects: Creation of a benchmark

Hein E.C. van der Wall¹, Gerardus J. Hassing¹, Robert-Jan Doll¹, Gerard J.P. van Westen², Adam Cohen¹, Jasper Selder³, Michiel Kemme⁴, Jacobus Burggraaf¹, and Pim Gal¹

¹Centre for Human Drug Research

²Leiden Academic Centre for Drug Research

³Amsterdam UMC Locatie VUmc

⁴Amsterdam UMC - Locatie VUMC

April 16, 2024

Abstract

Objective The aim of the present study was to develop a neural network to characterize the effect of aging on the ECG in healthy volunteers. Moreover, the impact of the various ECG features on aging was evaluated. **Methods & Results** A total of 6228 healthy subjects without structural heart disease were included in this study. A neural network regression model was created to predict age of the subjects based on their ECG; 577 parameters derived from a 12-lead ECG of each subject were used to develop and validate the neural network; A ten fold cross-validation was performed, using 118 subjects for validation each fold. Using SHapley Additive exPlanations values the impact of the individual features on the prediction of age was determined. Of 6228 subjects tested, 1808 (29%) were females and mean age was 34 years, range 18 – 75 years. Physiologic age was estimated as a continuous variable with an average error of 6.9 ± 5.6 years ($R^2 = 0.72 \pm 0.04$). The correlation was slightly stronger for men ($R^2 = 0.74$) than for women ($R^2 = 0.66$). The most important features on the prediction of physiologic age were T wave morphology indices in leads V4 and V5, and P wave amplitude in leads AVR and II. **Conclusion** The application of artificial intelligence to the ECG using a neural network regression model, allows accurate estimation of physiologic cardiac age. This technique could be used to pick up subtle age-related cardiac changes, but also estimate the reversing of these age-associated effects by administered treatments.

Cardiac age detected by machine learning applied to the surface ECG of healthy subjects: Creation of a benchmark

Hein E.C. van der Wall MSc^{a,b}, Gerardus J. Hassing MSc^a, Robert-Jan Doll PhD^a, Gerard J.P. van Westen PhD^b, Adam Cohen MD PhD^{a,b,c}, Jasper Selder MD PhD^d, Michiel Kemme MD PhD^d, Jacobus Burggraaf MD PhD^{a,b,c}, Pim Gal MD PhD^{a,c}

a. Centre for Human Drug Research

b. Leiden Academic Centre for Drug Research

c. Leiden University Medical Center

d. Amsterdam University Medical Center

Corresponding author:

Hein E.C. van der Wall, MSc

Centre for Human Drug Research

Zernikedreef 8, 2333 CL Leiden, The Netherlands

Hvdwall@chdr.nl

Funding: (None)

Disclosures: (None)

Abstract

Objective

The aim of the present study was to develop a neural network to characterize the effect of aging on the ECG in healthy volunteers. Moreover, the impact of the various ECG features on aging was evaluated.

Methods & Results

A total of 6228 healthy subjects without structural heart disease were included in this study. A neural network regression model was created to predict age of the subjects based on their ECG; 577 parameters derived from a 12-lead ECG of each subject were used to develop and validate the neural network; A ten fold cross-validation was performed, using 118 subjects for validation each fold. Using SHapley Additive exPlanations values the impact of the individual features on the prediction of age was determined. Of 6228 subjects tested, 1808 (29%) were females and mean age was 34 years, range 18 – 75 years. Physiologic age was estimated as a continuous variable with an average error of 6.9 ± 5.6 years ($R^2 = 0.72 \pm 0.04$). The correlation was slightly stronger for men ($R^2 = 0.74$) than for women ($R^2 = 0.66$). The most important features on the prediction of physiologic age were T wave morphology indices in leads V4 and V5, and P wave amplitude in leads AVR and II.

Conclusion

The application of machine learning to the ECG using a neural network regression model, allows accurate estimation of physiologic cardiac age. This technique could be used to pick up subtle age-related cardiac changes, but also estimate the reversing of these age-associated effects by administered treatments.

Keywords: Aging, ECG, Machine Learning, Healthy Volunteers, Artificial Intelligence

Introduction

Surface electrocardiograms (ECGs) are used frequently in routine clinical care, but also in investigational studies examining the effects of pharmacological and non-pharmacological treatments on the heart. Readout measures include the RR interval, PR interval, QRS duration and (corrected) QT interval. Typically, the pharmacological treatment effects are mediated by recognized channels on the cardiac surface. [1] However, there are cardiac effects that require a longer period of time to become visible on the surface ECG, such as aging induced cardiac fibrosis, and it is largely unknown if these subtle effects can be visualized on a surface ECG. [2, 3]

There has been a number of recent investigations regarding the prediction of physiological age using medical records, vital signs and laboratory data, or epigenetic changes. [4-6] These investigations indicated the existence of a gap between predicted physiological age and actual chronological age. Exploration of this gap is clinically important as a serious gap difference has been shown to be associated with higher risks of

all-cause mortality, cardiovascular disease, obesity, earlier menopause, and frailty. [5, 7-11] Various previous studies have already shown that the 12-lead ECG can be a reliable tool to estimate physiological aging. [5, 7-16]

Previous studies have applied artificial intelligence to the raw ECG data, allowing estimation of physiologic ECG age, which was found to reflect aging and comorbidities. [17] However, these algorithms were based on large hospital datasets, thus including patients that may have disease-induced abnormalities in their ECGs, which makes the outcome difficult to interpret when applied to a healthy volunteer. Therefore, the aim of the present analysis was to develop a neural network in healthy volunteers to characterize the effect of aging on the ECG.

Methods and Materials

Population

All data were collected at the Centre for Human Drug Research in Leiden, the Netherlands, a clinical research organization specialized in early phase drug development studies. Data collected during the mandatory medical screening to verify study eligibility for enrolment in the early phase drug development studies as a volunteer between 2010 and 2019 were included in the present analysis. Ethical approvals from the Medical Ethical Review Committee for the included studies were acquired and informed consent documents were signed by the volunteers prior to any data collection. The present study was performed in accordance to local regulations. All activities were performed in accordance with applicable standard operating procedures.

The medical screening consisted of a single visit to the clinical unit where a detailed history, a physical examination, vital signs including blood pressure, temperature, weight and height measurement, body mass index (BMI) calculation, and a 12-lead ECG were recorded. Additionally, haematology and chemistry blood panels, urine dipstick, and a urine drug test were analysed.

Data collection for the model

ECG parameters of 6228 subjects with an age between 18 and 75 years were included in the present study. From each subject ECG, 574 features were extracted by the MUSE system. Additionally, gender was used as a feature. The age of the subjects was rounded in whole years. At least ten ECGs were available for each age.

Data pre-processing and selection

As validation set two subjects of each age were kept apart as final test set. The rest of the data was used as the training set.

To create a balanced training set the Synthetic Minority Oversampling Technique (SMOTE) algorithm was applied on the training set to create ‘synthetic’ subjects for the less populated age groups based on the values in the concerning age groups. [18]

Machine learning

A neural network was used as a machine learning model. The keras module v. 2.4.3 in python 3.8.5 was used to build a model. Before training, internal cross validation (three-fold) within the training set was used to optimize the model. The network was optimized for number of layers, number of nodes per layer, activation function per layer for each layer and learning rate. A batch size of 300 was used. The number of epochs (defined as the number of cycles through the full training dataset) for internal validation was determined based on validation performance in the internal validation set. The number of epochs for final validation was based on the median of the optimal number of epochs for the internal cross validations. This process of optimization, training, and validation was repeated 10 times with different training and test sets. The optimal models were evaluated on the test set with the R^2 score and mean absolute error. We also evaluated the model performance with respect to gender.

To investigate how such a model could perform in groups of subjects or patients, the mean absolute error in a group of 1 to 50 randomly selected subjects in the test sets was evaluated. This was repeated 10 times for each fold (i.e. 100 times for each number of subjects).

To gain insight into the impact of the individual features on the predicted age, each fold SHapley Additive exPlanations (SHAP) values were calculated [19] based on the training set. The importance of the features was validated by means of permutation importance (defined as the decrease in a model score when a single feature value is randomly shuffled).[20]

Results

Table 1 shows the clinical characteristics of the 6228 included subjects. The study population was divided into ten chronological age groups of 6 years, starting from the age of 18 years. Each age group contained at least 194 subjects, and younger age groups comprised up to 2282 subjects. A total of 1808 (29 %) volunteers were female.

In Figure 1 ECG examples of a young 18 year old male (1A) and an elderly 74 year old male (1B) are shown. Figure 1C shows an ECG of a young 19 year old female and figure 1D shows an ECG of an elderly 74 year old female subject. Several differences between the young and older healthy subjects were discernable. In elderly persons the heart rate was lower, the T wave had a lower (absolute) amplitude in leads I,II,III,AVR, and AVL and the P-wave duration seemed shorter. However, these ECG differences showed considerable variations in the healthy population.

In supplementary Tables 1 and 2, 54 features present in most leads and other ECG features used for the machine learning model are shown, respectively. In addition, gender of each subject was also included in the model.

The relation between the (predicted) physiologic age and the chronological age was assessed in 10 sets of 116 subjects. In Figure 2a, the relation between predicted physiologic age and chronological age of all 10 test sets is shown. The average relationship of the models showed an R^2 of 0.72 ± 0.04 (mean \pm SD). The mean absolute error of all predictions was 6.9 ± 5.5 years.

On average, the predicted physiologic age was 0.3 years younger than the chronological age of the subjects. The median deviation of all predicted ages was 5.6 years from the actual age, indicating that half of the predictions was within the range of 5.6 years of chronological age.

The average prediction line is presented in figure 2b. The average predicted age of the 20 subjects per chronological age had a mean absolute error of 3.4 ± 3.0 years ($R^2 = 0.93$). For subjects between 30 and 60 years old the mean absolute error of the average predicted age per chronological age was 1.6 ± 1.1 years.

Figure 3 shows how such models could perform in new patient groups. It can be seen that the average absolute prediction error is declining fast when multiple subjects are tested. For example, a cohort of 10 healthy subjects with age ranging from 18 to 75 years would have an average absolute error of 2.7 ± 2.1 years. The mean absolute error of a test group of 30 subjects would be only 1.7 ± 1.2 years.

In order to study gender differences, the predicted physiological ages of the male and female subjects in the test sets were separated and are presented in Figure 4. The predicted ages of the male subjects were more accurate ($R^2 = 0.74$) than the predictions of the female subjects ($R^2 = 0.66$).

Figure 5 shows the SHAP values of the 40 most important ECG features used in the prediction model. So, the impact of each individual feature on the model output and physiologic aging can be seen. Some of the most important features on the prediction of physiologic age were T top abnormalities in leads V4 and V5, P top amplitude in leads AVR and II and atrial rate.

An increase of P peak amplitude in lead II for example, indicates a younger physiological age (a long red bar to the left). A longer PR interval both indicate an older physiologic age (longer red bar to the right). A higher atrial rate indicates a younger physiologic age (large red bar to the left). The impact of gender was

only of minor importance with SHAP values ranging from -1.2 to 0.9. The order of the feature permutation importance is similar to the order of the SHAP values, confirming the impact of the features.

Discussion

In this study we developed machine learning models that allow accurate prediction of physiologic cardiac age of healthy subjects based on 12-lead surface ECG parameters. Using a neural network we were able to estimate the age of a healthy subject with an error of 7 years and to analyze the impact of the ECG features. The created models of the present study may serve as a benchmark for testing the effects of new pharmacological drugs on potential decline or improvement of physiologic health of the heart.

Application of Machine Learning

Attia et al. recently sought to determine whether the application of machine learning algorithms, including convolutional neural networks, to a large ECG patient data set would be capable of predicting age and sex reported by patients, independent of additional clinical data [17]. They further investigated whether discrepancies between ECG age and chronological age might be a marker of physiological health. When the convolutional neural network-predicted age exceeded a patient's actual age by at least 7 years, there was a higher incidence of cardiovascular comorbidities, potentially suggesting that the convolutional neural network-predicted age from 12-lead ECGs may correlate with physiological health. Their findings suggested that physiological age is distinct from chronological age, and may have useful clinical applications. For example, if a patient's biologic age is 60 but their ECG age predicts that they are 70, it may indicate underlying cardiovascular disease and potential risk. A limitation of their study was, as also recognized by the authors, that all individuals included were patients, and thus an ECG was obtained for a certain clinical indication. It was questioned by the authors whether their results are similarly accurate among an ostensibly healthy population is unknown, and revalidation in such a cohort will therefore be critical.

The same holds true for the study by Hirota et al., who studied biological age, physiological age, and all-cause mortality by 12-lead ECG in patients without structural heart disease. [21] Their data showed that the gap between ECG-predicted physiological and biological age allowed estimation of increased risk of all-cause mortality. Although their study subjects were assumed to have no structural heart diseases, it was stated by the authors that it will be necessary to validate the results of their study in populations of healthy subjects. In our study, we only studied healthy individuals, giving the advantage of being a much needed benchmark study, which enables the validation of future studies in patients versus our data.

Performance of the model

The relation between chronological and predicted physiologic age was associated with an R^2 of 0.72. Although with a smaller dataset than used by Attia et al., our predictions have a similar performance, probably because of the healthy population in our study, which we expect reduces the variability of the association. Given the large number of influencing factors that can affect ECG parameters the R^2 of 0.72 of our models seems sufficient to detect a pharmacodynamic effect in a cohort of subjects. Use of the entire dataset with a larger number of subjects may improve future performance of the model.

In the present study, the impact of physiologic aging on the various ECG features was analyzed using SHAP values. Several changes are clearly visible in the ECG figures. Some of these are already well known in clinical practice, such as prolongation of PR and QT interval and deceleration of heart rate.[12] Other changes, however, could only be recognized by using machine learning, while these may be evenly important. Moreover, when multiple features change at the same time, it becomes difficult to judge whether the change in the ECG is good or bad without using machine learning. By means of machine learning techniques a combination of various ECG changes allows a more accurate insight into the physiologic health changes of the heart.

Gender differences

The accuracy of predicting physiologic age was found to be higher in males than in the female subjects. This may be due to the somewhat smaller female study population, but it may also reflect the atypical ECG repolarization patterns which are known to occur frequently in women.[22] The SHAP values show that impact of gender on physiologic age prediction was only of minor importance.

Pharmaceutical drug testing and potential implications

The prediction of the physiologic age for one single person is less relevant in this model. However for larger groups or cohorts of multiple subjects, the prediction is more accurate. For example, for a group of 30 test subjects, the average deviation is only less than two years from average physiologic age. Therefore, our models could be particularly suitable as benchmark for testing new pharmaceutical drugs or other interventions which may have impact on cardiac health in the near future. Differences between physiologic ECG age and chronological age have been shown to predict all-cause and cardiovascular mortality and reflect physiologic age, cardiovascular health and long term outcomes. [23]

The proper use of a model - trained on the entire dataset - in early drug development can provide important information that can be used to make a go/no-go decision regarding further development of new drugs. Similarly, this can be used to guide the decision-making process regarding the dosage range to be used in phase II studies, determining a therapeutic window, and even identifying the target study population [24]. This way novel pharmacological drugs could be tested for effect on cardiac physiologic aging in the early phase of development.

Limitations

Our population consisted of only 29% female subjects. This may have influenced the accuracy of the model, but SHAP value analysis showed that gender only had a minimal impact on the predictions of physiologic age.

ECG changes do not need to have a purely cardiac cause, but they may also be caused by effects of age on the position of the heart in the thorax, the presence of fat layers around the heart, and the shape of the thorax shape. Therefore, the found relationship does not necessarily mean older heart per se, but can also mean an older body.

Conclusion

The application of machine learning to the ECG using a neural network regression model, allows estimation of physiologic cardiac age. This technique could be used to pick up subtle age-related cardiac changes, but also estimate the reversing of these age-associated effects by administered treatments.

References

1. van Dam, P.M., et al., *The relation of 12 lead ECG to the cardiac anatomy: The normal CineECG*. Journal of Electrocardiology, 2021.
2. Biernacka, A. and N.G. Frangogiannis, *Aging and cardiac fibrosis*. Aging and disease, 2011. **2** (2): p. 158.
3. Hayashi, H., et al., *Aging-related increase to inducible atrial fibrillation in the rat model*. Journal of cardiovascular electrophysiology, 2002. **13** (8): p. 801-808.
4. Wang, F., T. Syeda-Mahmood, and D. Beymer. *Information extraction from multimodal ECG documents* . in *2009 10th International Conference on Document Analysis and Recognition* . 2009. IEEE.
5. Roetker, N.S., et al., *Prospective study of epigenetic age acceleration and incidence of cardiovascular disease outcomes in the ARIC study (Atherosclerosis Risk in Communities)*. Circulation: Genomic and Precision Medicine, 2018. **11** (3): p. e001937.

6. Kistler, P.M., et al., *Electrophysiologic and electroanatomic changes in the human atrium associated with age*. Journal of the American College of Cardiology, 2004. **44** (1): p. 109-116.
7. Breitling, L.P., et al., *Frailty is associated with the epigenetic clock but not with telomere length in a German cohort*. Clinical epigenetics, 2016. **8** (1): p. 21.
8. Perna, L., et al., *Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort*. Clinical epigenetics, 2016. **8** (1): p. 64.
9. Wang, Z., et al., *Predicting age by mining electronic medical records with deep learning characterizes differences between chronological and physiological age*. Journal of biomedical informatics, 2017. **76** : p. 59-68.
10. Horvath, S., et al., *Obesity accelerates epigenetic aging of human liver*. Proceedings of the National Academy of Sciences, 2014.**111** (43): p. 15538-15543.
11. Levine, M.E., et al., *Menopause accelerates biological aging*. Proceedings of the National Academy of Sciences, 2016. **113** (33): p. 9327-9332.
12. Rijnbeek, P.R., et al., *Normal values of the electrocardiogram for ages 16–90 years*. Journal of electrocardiology, 2014.**47** (6): p. 914-921.
13. Macfarlane, P., et al., *Effects of age, sex, and race on ECG interval measurements*. Journal of electrocardiology, 1994. **27** : p. 14-19.
14. Mason, J.W., E.W. Hancock, and L.S. Gettes, *Recommendations for the standardization and interpretation of the electrocardiogram: part II: Electrocardiography diagnostic statement list: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology*. Circulation, 2007. **115** (10): p. 1325-1332.
15. Kligfield, P., et al., *Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology*. Journal of the American College of Cardiology, 2007. **49** (10): p. 1109-1127.
16. Khane, R.S., A.D. Surdi, and R.S. Bhatkar, *Changes in ECG pattern with advancing age*. 2011.
17. Attia, Z.I., et al., *Age and sex estimation using artificial intelligence from standard 12-lead ECGs*. Circulation: Arrhythmia and Electrophysiology, 2019. **12** (9): p. e007284.
18. Chawla, N.V., et al., *SMOTE: synthetic minority over-sampling technique*. Journal of artificial intelligence research, 2002.**16** : p. 321-357.
19. Lundberg, S.M. and S.-I. Lee. *A unified approach to interpreting model predictions* . in *Advances in neural information processing systems* . 2017.
20. Altmann, A., et al., *Permutation importance: a corrected feature importance measure*. Bioinformatics, 2010. **26** (10): p. 1340-1347.
21. Hirota, N., et al., *Prediction of biological age and all-cause mortality by 12-lead electrocardiogram in patients without structural heart disease*. BMC Geriatrics, 2020. **21** (460).
22. Okin, P.M., *Electrocardiography in women: taking the initiative* . 2006, Am Heart Assoc.
23. Ladejobi, A., et al., *ECG-DERIVED AGE AND SURVIVAL: VALIDATING THE CONCEPT OF PHYSIOLOGIC AGE DETECTED BY ECG USING ARTIFICIAL INTELLIGENCE*. Journal of the American College of Cardiology, 2020.**75** (11 Supplement 1): p. 3469.

24. Groeneveld, G.J., Hay, J. L., Van Gerven, J. M., *Measuring blood–brain barrier penetration using the NeuroCart, a CNS test battery*. Drug Discovery Today: Technologies, 2016. **20** : p. 27-34.

Table 1. Age characteristics of the 6228 healthy subjects

Subject age (years)	N	% female
18 - 23	2282	29
24 - 29	1563	26
30 - 35	449	20
36 - 40	247	17
41- 46	245	24
46 - 52	339	35
53 - 57	241	38
58- 63	194	42
63 - 69	393	40
69 - 75	275	35

Figure 1. ECG samples of young and elderly male and female healthy subjects. A: ECG of a young 18 year old male subject. B: ECG of an elderly 74 year old male subject. C: ECG of a young 19 year old female subject. D: ECG of an elderly 74 year old female subject

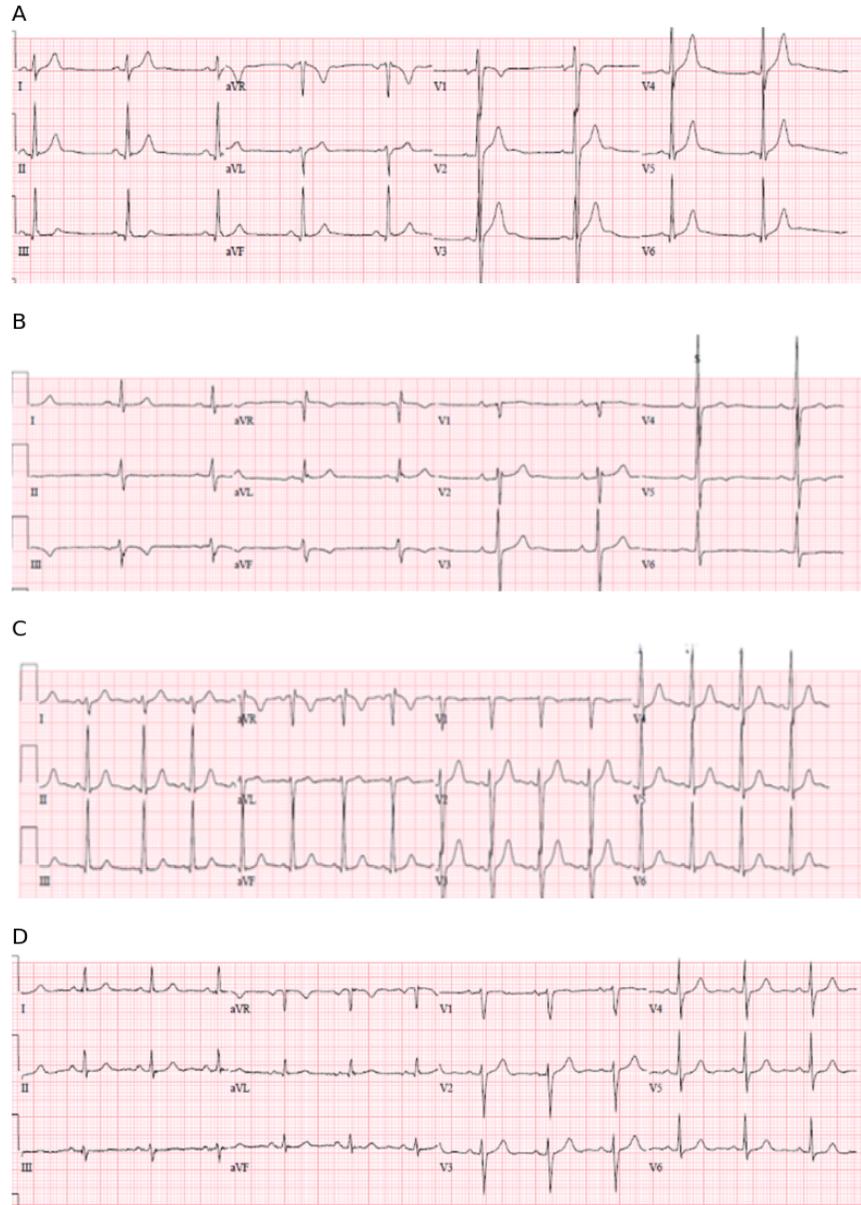
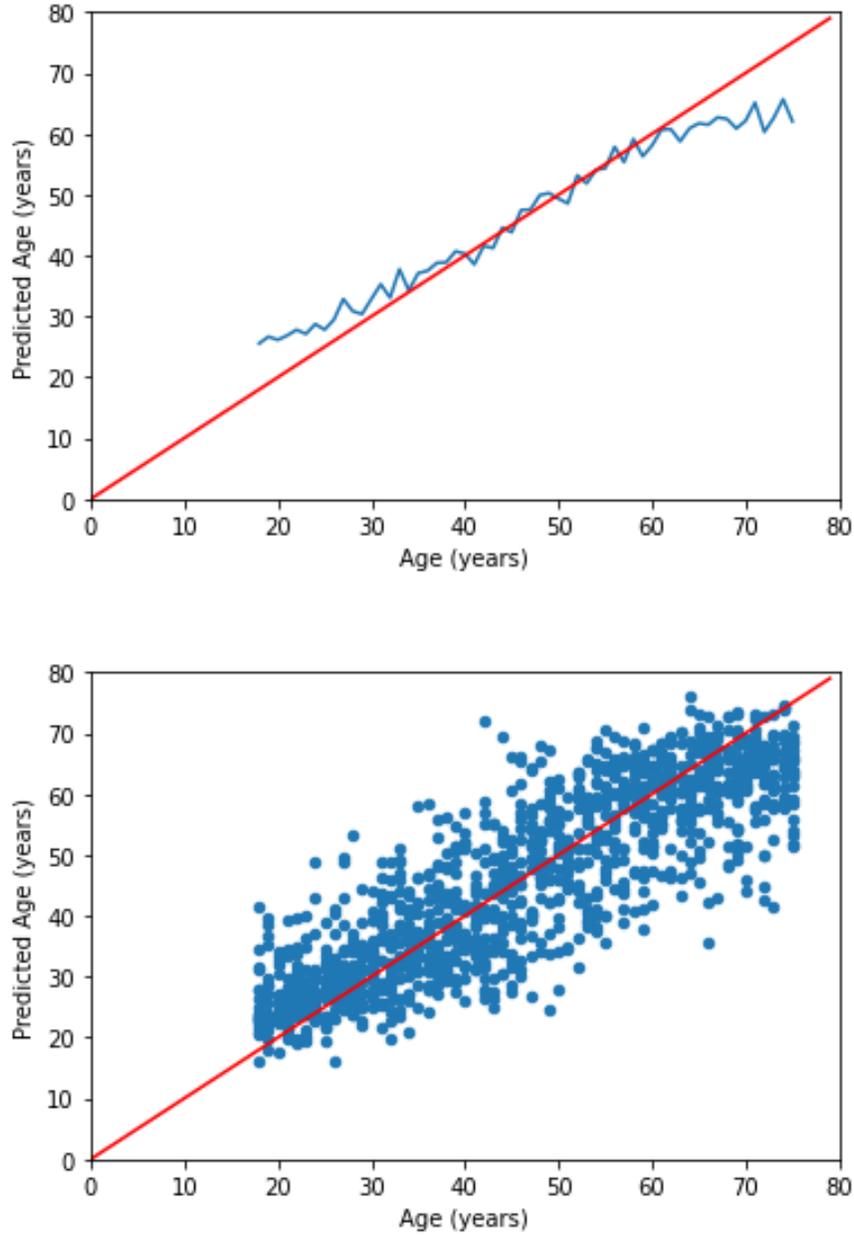
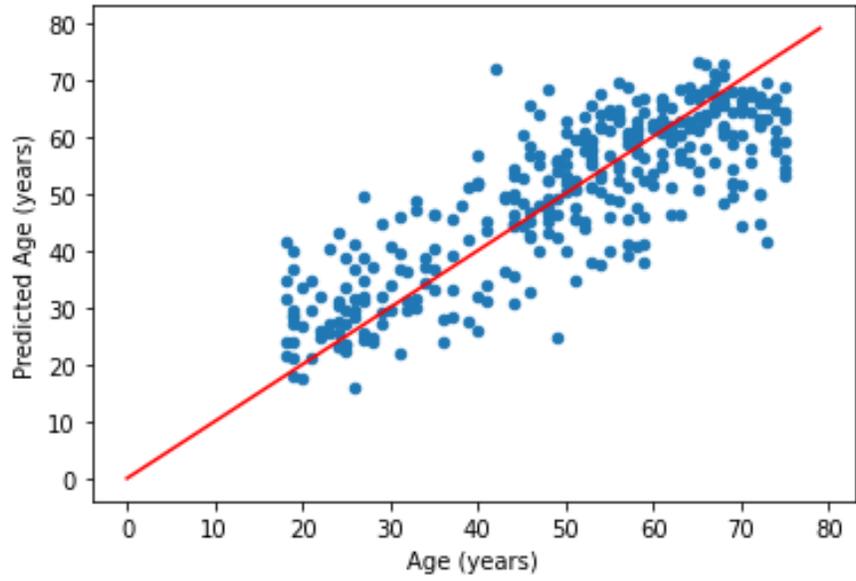
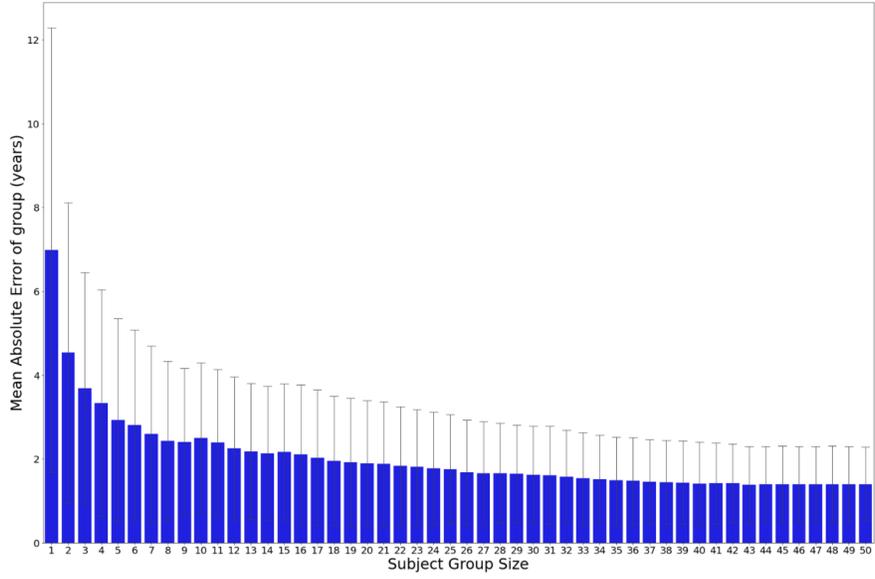


Figure 2. Relationship between predicted) physiologic age and chronological age for 1180 healthy adults (10 test sets, Figure 2a, left). The average prediction line is shown in figure 2b (right).



a b

Figure 3. Average absolute error with standard deviation of subject groups ranging from size 1 to 50 test subjects. On the vertical axis the mean absolute error (in years) is shown, with on the horizontal axis the number of test subjects



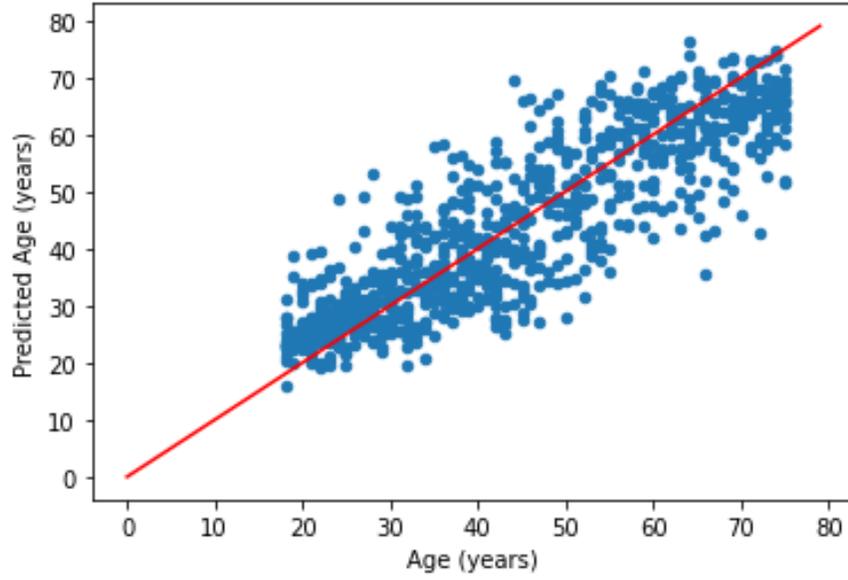
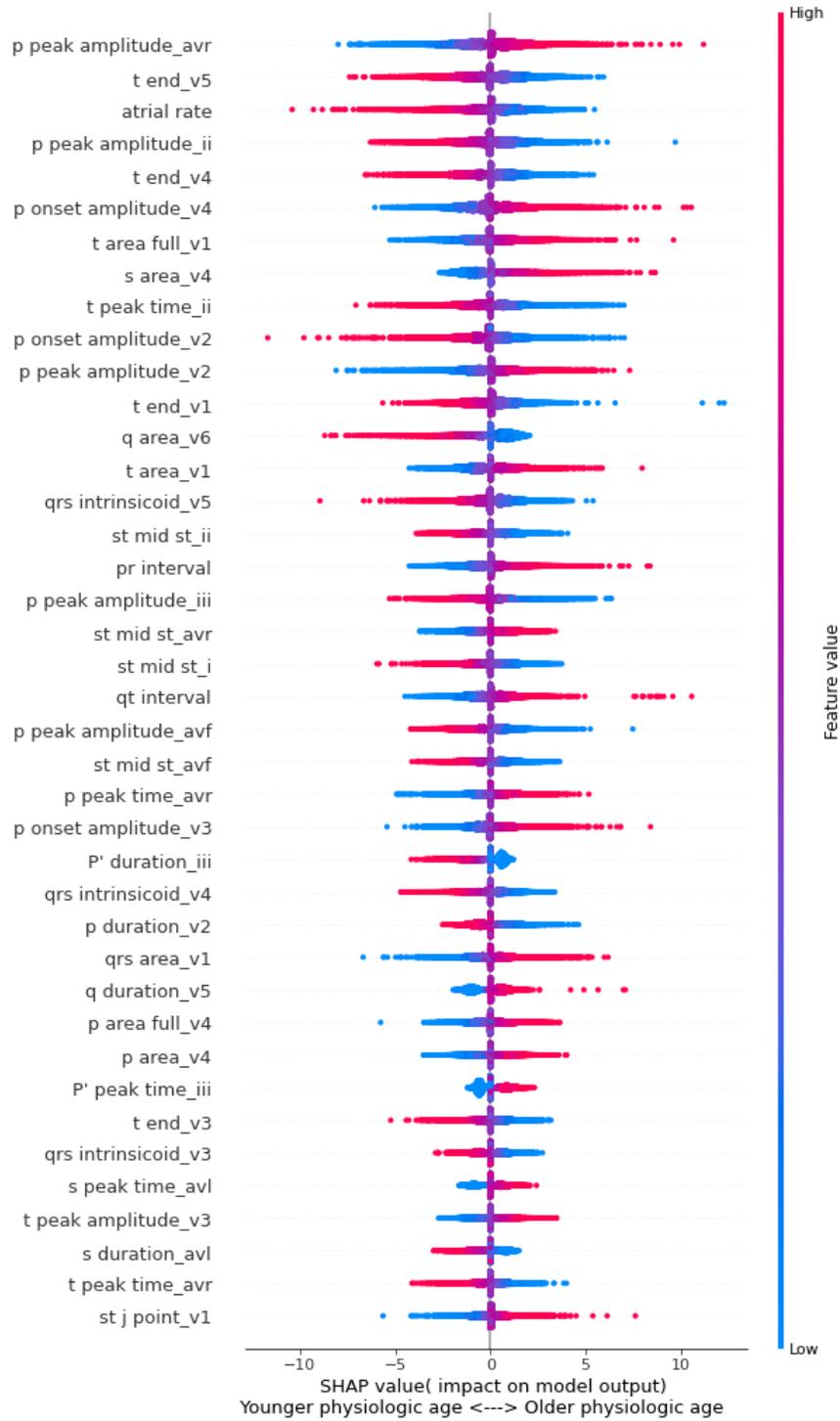


Figure 4. Predictions of 819 male subjects (a, left) and 361 female subjects (b, right).

Figure 5. SHAP values of the 40 most important features for predicting physiologic age. High values of the features are represented in red. Low values are represented in blue. On the x-axis, the predicted physiologic age. Shorter bars mean less impact on physiologic aging .



Supplementary Table 1. ECG features present in most leads extracted by the Muse system included in the model. P', R', S', and T' indicate the second components of P, R, S, and T wave, respectively, which could be positive or negative polarity.

Features per lead

max r amplitude	R PEAK TIME
max s amplitude	R' AREA
maximum st level	R' DURATION
minimum st level	R' PEAK AMPLITUDE
p area	R' PEAK TIME
p area full	S AREA
p duration	S DURATION
p offset	S PEAK AMPLITUDE
p onset	S PEAK TIME
p onset amplitude	SPECIAL T
p peak amplitude	ST END ST
p peak time	ST J POINT
p' area	ST MID ST
p' duration	S' AREA
p' peak amplitude	S' DURATION
p' peak time	S' PEAK AMPLITUDE
q area	S' PEAK TIME
q duration	T AREA
q peak amplitude	T AREA FULL
q peak time	T DURATION
qrs area	T END
qrs balance	T PEAK AMPLITUDE
qrs deflection	T PEAK TIME
qrs intrinsicoid	T' AREA
R AREA	T' DURATION
R DURATION	T' PEAK AMPLITUDE
R PEAK AMPLITUDE	T' PEAK TIME

Supplementary Table 2. Other ECG features extracted by the Muse system included in the model.

Features per ECG

atrial rate P AXIS Q OFFSET QT INTERVAL QTC BAZETT Q ONSET R AXIS Number of qrs complexes T AXIS T O
