DOI:10.19221/202023

Sahan V.,
Ph. D. Student (System Analysis) of the Department of Mathematical Methods of System Analysis, NTUU "Igor Sikorsky KPI", Kyiv, Ukraine, E-mail: vitalii@mawi.band

TOWARDS THE ECG-BASED BIOMARKERS OF HUMAN AGING

Abstract. *The rates of aging may vary substantially among the different individuals and population groups. Multiple attempts have been made to develop the biologically- relevant biomarkers of human aging using molecular and imaging data. Electrocardiogram (ECG) is one of the most accessible clinical tests used by physicians to examine the health state of patients. Wearable devices that can measure ECG on demand (often on the form of wristbands or smartwatches like Apple Watch or Mawi Band) or continuously (chest patches like iRhythm or Qardio) have gained a lot of popularity in recent years offering an accessible and affordable alternative to other biomedical tests. This increased availability of data provides unprecedented opportunities not only for medical diagnosis but also for aging research and longevity biotechnology. This work explores the possibility of using an electrocardiogram (ECG) and derived features as a biomarker of aging. Our contribution is twofold. Firstly, we collected a comprehensive dataset from two different sources: 24-hours Holter cardiograph recordings of around 1000 individuals and 5-minute wearable cardiograph recordings of 500 individuals. Secondly, we identified several potential biomarkers of aging including fiducial features of raw ECG signal (QT segment variability), heart rate variability characteristics (time-domain measures including standard deviation of NN intervals (SDNN) and such frequency-domain measures as powers of the low-frequency (LF) and high-frequency (HF) bands) that show high correlations with chronological age. Further validation has demonstrated that most of these features remain significant both for long-term recordings from medical cardiographs and short-term recordings from wearable devices. These results suggest the emerging potential of combining wearable sensors and machine learning technologies for continuous health risk monitoring with real-time feedback to life and health insurance, healthcare, and wellness providers.*

Key words: *aging, ECG, biomarker*

Introduction. Numerous physiological parameters show significant correlation with an individual's age. Recently research in this direction had been driven by experiments with DNA methylation [1], gene expression [2], plasma proteome [3] and similar data, which is relatively difficult to obtain at the moment (both from the pricing and processual point of view). In this study we aim to develop a set of biomarkers of aging from a more accessible source of data. We have chosen the humans heart as this source, because ECG collecting is a non-invasive and relatively cheap procedure. Furthermore, with growing popularity of pulse trackers and wearable ECG monitors it can be very accessible for the broader masses. The obtained biomarkers can be used for several purposes, in particular, evaluation of aging dynamics in different populations, clinical trials, drug development and testing.

There are several related works showing potential of the use of ECG as a source of biomarker of aging (4], [5]). Most of them show, that heart rate variability (HRV) related parameters correlate with the age at most, namely SDNN, RMSSD, pNN50, LF, HF (low frequency and high frequency) and others. Minority of the works show correlation of fiducial ECG features like QT interval variability [7], ST wave depression[8] and others with chronological

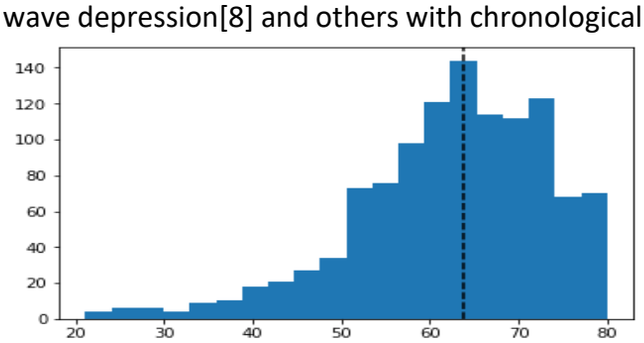

Fig. 1: Ages distribution in the dataset

age. There are also studies with promising approaches based on information theory approaches, in particular, entropy or detrended fluctuation analysis (DFA), see [6], extracted from RR-intervals time series.

Numerous physiological parameters show significant correlation with an individual's age. Recently research in this direction had been driven by experiments with DNA methylation [1], gene expression [2], plasma proteome [3] and similar data, which is relatively difficult to obtain at the moment (both from the pricing and processual point of view). In this study we aim to develop a set of biomarkers of aging from a more accessible source of data. We have chosen the human heart as this source, because ECG collecting is a non- invasive and relatively cheap procedure. Furthermore, with the growing popularity of pulse trackers and wearable ECG monitors it can be very accessible for the broader masses. The obtained biomarkers can be used for several purposes, in particular, evaluation of aging dynamics in different populations, clinical trials, drug development and testing.

There are several related works showing potential of the use of ECG as a source of biomarker of aging (see [4], [5]). Most of them show that heart rate variability (HRV) related parameters correlate with the age at most, namely SDNN, RMSSD, pNN50, LF, HF (low frequency and high frequency) and others. Minority of the works show correlation of fiducial ECG features like QT interval variability [7], ST wave depression [8] and others with chronological age. There are also studies with promising approaches based on information theory approaches, in particular, entropy or detrended fluctuation analysis (DFA) (see [6]), extracted from RR-intervals time series.

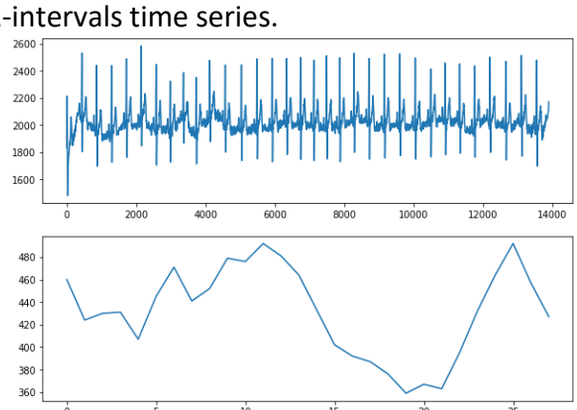

Fig. 2: Example of ECG and corresponding RR intervals lengths

Nevertheless, most of them do not have enough participants in the practical experiments or use the long-term ECG recordings that are not always accessible. On the contrary, in this work we have gathered a large scale dataset of more than 1000 individuals and we show several experiments exploiting both long-term and short-term recordings measured with different frequencies.

This paper is organized as follows: first, we will describe a dataset used for analysis, individuals in it, extracted ECG features and how they are correlated with a chronological age. Next, we will describe the modeling experiment, validation scheme and the numerical results. Then we will discuss the obtained results and, at the end, we will show the room for practical use of this work and further improvements.

Materials and methods. In this section we describe the dataset and methods we used in this study, its origin, statistical properties alongside with the ECG features and their correlations with the age that were extracted from the raw ECG signal.

Our study population initially consists of people who measured their cardiograms based on the doctor's prescription, or for the purposes of preventive diagnostics in one of the medical institutions in Kyiv, Ukraine. The ECG recordings were measured using standard Holter cardiograph during 24 hours, while the patients under study were acting like they're used to.

Since doctors prescrpit Holter monitoring only to the aging groups that are in the potential risk of different heart diseases, the initial dataset consists of the individuals, whose chronological age is biased towards higher values. See Figure 1 for the detailed histogram of the initial dataset.

Also, we needed to filter our dataset from the individuals with heart pathologies that appear with the age. We do this to avoid a situation, when the extracted variables from the ECG signal do not correlate with the actual individual's age itself, but with the diseases. This way our models will estimate not only the biological age of the heart, but the presence or absence of some particular diseases, namely suffered heart strokes, various arrhythmia types, ischemia and others.

The raw ECG signal was obtained from Holter cardio- graphs that have 200Hz sampling rate and

were measured during 24 hours. After that we have extracted the features, described in detail in the table 1 further. First, there is a group of variables, derived on the basis of the time domain analysis of the RR intervals:

1. SDNN-Standard deviation of all normal RR intervals. Considered to reflect the total effect of autonomic regulation of blood circulation.
2. RMMSD-The square root of the sum of differences in a series of cardio intervals. Considered to reflect activity of parasympathetic link of autonomic regulation.
3. Heart rate is the speed of the heartbeat measured by the number of contractions (beats) of the heart per minute (bpm).
4. pNN50 - The proportion of pairs of successive NNs that differ by more than 50 ms divided by the total num- ber of NNs. It is an indicator of the degree of predominance of the parasympathetic link of regulation over the sympathetic.
5. MxDMn - The difference between the maximum and minimum values of cardio intervals. It was considered to represent the maximum amplitude of regulatory influences.
6. SI-Stress-index developed by Professor Roman M. Baevsky. Represents the degree of tension of regulatory systems (the degree of predominance of the activity. of central regulatory mechanisms over autonomous), see [13] for details.

Then, there are features based on the frequency do- main analysis:

1. Total power - The total power of the HRV spectrum. Considered to represent the total absolute level of activity of regulatory systems.
2. HF power - Power spectra in high frequency range (0.15 - 0.40Hz). Shows the relative level of parasympathetic regulation activity.
3. LF power - Power spectra in low frequency range (0.04 - 0.15Hz). Shows the relative level of activity of the sympathetic level of regulation.
4. VLF power - Power spectra in very low frequency range (Less than 0.04Hz).
5. LF/HF - LF to HF ratio. Considered to characterize the Sympathetic to Parasympathetic Autonomic Balance and reflect relative activity of the subcortical sympathetic nerve center.
6. LFnu - Normalized LF value calculated as LF/(HF+LF)
7. HFnu - Normalized HF value calculated as HF/(HF+LF)

Finally, there is a group of other features as well:

1. CC0 - The number of shifts of the autocorrelation function until the value of the correlation coefficient is less than zero. Represents the degree of activity of the central regulation loop.
2. CC1 - The value of the first coefficient of the autocorrelation function. Considered to represent the degree of activity of the autonomous regulation loop.
3. SD1 - RR-interval Poincare plot index - the standard deviation of the distance of each point from the y=x axis. Specifies the RR-interval Poincare plot ellipses width. SD1 measures short-term HRV in ms.
4. SD2 - RR-interval Poincare plot index - the standard deviation of each point from the y=x+(average RR interval). Specifies the RR-interval Poincare plot ellipses length. SD2 measures short- and long-term HRV in ms and correlates with LF power.
5. SD1/SD2 - SD1 to SD2 ratio. Measures the unpredictability of the RR time series, may be used to measure autonomic balance. SD1/SD2 is correlated with the LF/HF ratio.
6. Artefact amount - Amount of artefacts detected in the record.

It’s also important to mention that no specific data was available related to patients' mortality. Of course, additional clinical and biological data will be needed in the future experiments for more accurate heart biomarkers development.

In this section, we show how based on the extracted features from the raw 24-hours ECGs we build experi- ments on the prediction of the individuals’ age in different settings: from the long-term or short-term recordings. Then, we describe algorithms we used for prediction and evaluation schemes.

Since we were extracting the features both from the full 24-hours recording and 5-minutes subsequences, we're interested in three realistic experiments. First one is related to the situation, when we want to predict individuals’ age based on the whole 24 hours of Holter monitoring. In this case, we can use all the extracted features

Table 1

This is my one big table

Feature	24 hours	Mean short terms	Std short terms
CC!	-0.25	-0.38	0.12
HystAmo	0.15	0.10	0.14
HystMo	0.15	0.13	-0.06
MxDMn	0.03	0.14	0.08
PARS	0.06	0.35	0.07
SD1/SD2	0.22	0.03	0.03
Stress Index	0.04	0.08	0.09
Artefact	0.09	0.10	0.03
HF power	0.21	0.22	0.15
HF nu	0.39	-0.09	-0.09
LF power	0.10	0.10	0.08
LF nu	-0.38	-0.44	-0.13
LF / HF	-0.37	-0.40	-0.38
Heart rate	-0.17	-0.13	-0.22
RMMSD	0.23	0.23	0.13
pNN50	0.21	0.20	0.03
SDNN	0.01	0.14	0.05
Total power	0.16	0.17	0.10
VLF HF	-0.23	-0.16	0.02
VLF Power	-0.10	-0.12	-0.12
Hurst Exponent	-	-0.11	-

together concatenated. The second setting is designed to mimic a situation, when a person is making just several short-term recordings during a single day and we want to predict the age based on them. The last setting is in the use of a single short-term recording for the same purposes.

In terms of data processing, at first, we checked the correlations of all the features described in section 2 with the chronological age of the individual, to select potential candidates for biomarkers. They are all shown in table 1 and the top-correlated features (both negatively and positively) are LFnu and LF/HF both from 24-hours and short term, recordings mean (positively correlated) and RMSSD, HFnu (negatively).

Before applying an algorithm that learns from data, input samples have to be prepared properly, i.e. normalized or scaled to satisfy certain range or distribution. In this work we have compared min-max scaling, max-abs scaling, normalization and z-score scaling and have chosen max-abs scaling as the one, that performs the most in terms of further explained metrics. This estimator scales and translates each feature individually such that the maximal absolute value of each feature in the training set will be 1.0. It does not shift/center the data, and thus does not destroy any sparsity. To the output log-transformation was applied and after restored with exponentiation operation.

To validate the performance of the age prediction algorithm, we need to split a dataset into training and testing sets and use the first one for fitting the model and the second one for performance tracking.

We have used a well-known cross-validation scheme with 10 folds and after average the results to see the mean performance and its statistical significance. As the main performance metrics, R2 and mean absolute error (MAE) were chosen. R2 shows general correlation of the prediction with the real age and should be closer to 1.0 and MAE represents the error in the terms of the age prediction itself.

To visualize the dataset and obtain intuition about possible outliers, clusters and dependencies types, we have used principal component analysis (PCA) projection and t-SNE, see [9], projection on the 2D space. If any strong

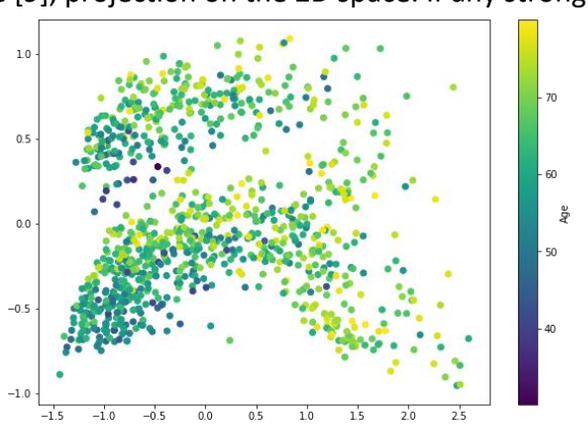

Fig. 3: PCA projection of all the features

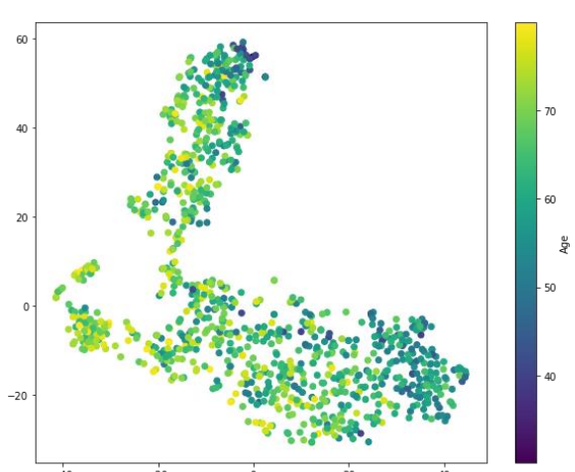

Fig. 4: t-SNE projection of all the features

linear dependencies are present, they will be shown on the PCA plot. Clusters and outliers are usually very well explained by the t-SNE embedding. As it can be seen from the figures X and Y, there is no strong linear dependence, but some clusters (of the mixed ages) are present, so we may hypothesis that strongly regularized linear model or non-linear estimators will be helpful in this case.

In this subsection, we will briefly describe regression models used in this work, motivation of the choices and main characteristics.

In machine learning, k-Nearest Neighbors Regression (kNN), see [10] for more details, is a non-parametric model that can be used both for classification and regression. The age is predicted by local interpolation of the targets associated with the k nearest neighbors in the training set. We can consider our dataset as the set of pairs {(Xi,Yi)}i=1...N, where Xi is the ith individual ECG features and Yi is its age. Given a distance in the space of {Xi} and a member of this space Xj we can predict the age for it as such Yj from {(Xi,Yi)}i=1...N, where the distance between Xi and Yj is the smallest. You can find more details on kNN regression in X.

This algorithm has several drawbacks, for example the need for storing all the dataset, but it is easy to implement and it is non-linear, which is our main assumption based on the visualizations of the data in the figures 3 and 4.

Linear assumption still can hold, as is shown in several works mentioned before. Then, the PCA and t-SNE plots can be explained with the redundant features in the dataset, that should be pruned, or taken into account with a smaller weight. It can be done with regularization, and one of the ways to do it is a ridge regression (see [11]). We rely on the Bayesian version of the algorithm (see [12] for more details) where we should learn a predictive distribution from the data to predict from it.

Ridge regression minimizes squared error while regularizing the norm of the parameters:

$$J(w) = \lambda w^2 + \sum_i (w^T X_i - Y_i)^2, \quad (1)$$

where J is a loss function, in our case mean squared error (MSE), w is the set of parameters of a linear model, X_i is the ith individual ECG features and Y_i is its age, λ is a regularization parameter that, in our case, is inferred from the dataset itself.

Since the main hypothesis about the dependence between features and the age is that it is nonlinear, we need an appropriate model for it. There are several families of models designed to model nonlinear data, namely, polynomial models, neural networks, SVMs with different kernels and gradient boosting machines. We have chosen the latter, because they tend less to overfitting and work well with datasets that consist of numerous features of different types (categorical and numerical).

The idea behind gradient boosting regression, see [14] for more details, is in training an ensemble of so-called ”weak learners”, that is trained stage-wise. On each stage m, 1<m<M, the new member of an ensemble F_m is firstly considered as a weak model, that predicts the mean $Y_i$from the dataset. It is improved by constructing a new model $h(x)$as $F_{m+1} = F_m + h(x)$, where $h(x)$is found with gradient descent while optimizing a loss function$J$, see more details in [14].

Results. This section explains the performance of the trained algorithms within different experimental settings. On the table 3 the results of age prediction based on the features calculated from 24 hours, mean values of 6 short-term 5 minutes recordings and corresponding standard deviations. We see that gradient boosting regression performs reasonably well with MAE error close to 5 and high R^2 .

Table 4 shows how modeling of age based just on the mean values of 6 short-term 5 minutes recordings and corresponding standard deviations. As expected, the average performance dropped slightly, but not significantly. It proves our initial hypothesis, that age estimation can be done with several short-term measurements during the day.

Finally, we provide the results of regressing age based only on the feature set from a single short-term recording. The R^2 is very low and shows that predictions are not far from predicting the mean age value from the training set.

We also researched what particular features are the most important for the regression algorithm to understand top-performing biomarkers of aging. We extracted them from the gradient boosting regression model as the one

Table 2

24-hours features regression results

	Nearest Neighbors	Bayesian Ridge	Gradient Boosting
MAE	5.86 ± 0.64	5.47 ± 0.48	5.16 ± 0.41
R^2	0.31 ± 0.07	0.39 ± 0.04	0.41 ± 0.03

Table 3

Several short-term features regression results

	Nearest Neighbors	Bayesian Ridge	Gradient Boosting
MAE	7.3 ± 0.54	7.21 ± 0.78	6.97 ± 0.68
R^2	0.28 ± 0.03	0.33 ± 0.05	0.35 ± 0.02

that performed the best. We have used the model from the first experiment, which takes into account all the features both from the 24-hour and short-term recordings. The results are presented on fig. 5. As we can see, top-5 performing features are mean of LF/HF of short-term recordings, standard deviation of MxDMn of the short-term recordings, standard deviation of HystAmo of the short-term recordings, mean of HF of the short-term recordings and artefact standard deviation of the short-term recordings. It proves the hypothesis of usefulness of applicability of short-term recordings for the biological age estimation with ECG.

Additionally, we have performed an experiment with changing the objective function from regression to classification into age groups, each in the range of 10 years, to check how the features from our dataset are able to predict drastic changes in the age group. Eventually, it can be stated as 6-class classification, where with a 10-fold cross-validation scheme we could obtain 55 ± 0.71% accuracy.

Table 4

Single short-term features regression results

	Nearest Neighbors	Bayesian Ridge	Gradient Boosting
MAE	11.03 ± 1.2	10.9 ± 0.98	10.3 ± 0.9
R^2	0.05 ± 0.11	0.07 ± 0.2	0.8 ± 0.12

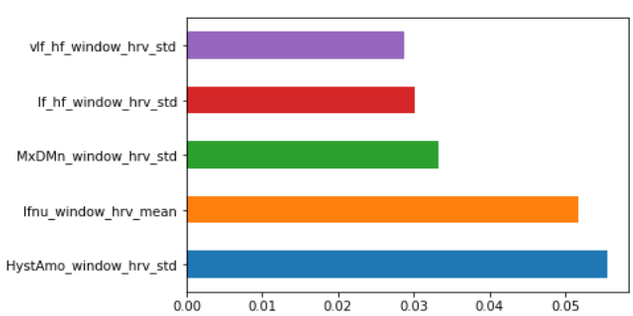

Fig. 5: Feature importance from gradient boosting regression

Conclusions. In this work, we have explored the ECG as the source of biomarkers of aging. With the rise of popularity of wearable devices that are able to measure first-lead ECG, it could become an affordable and comfortable way to track the biological age of the heart and biological age of the person itself. We have collected and preprocessed a dataset of more than 1000 healthy individuals, compiling 3 experiments that aim to simulate a situation, when we predict the age based on the 24-hour continuous recording, several short, 5 minutes long, recordings per day and a single short recording accordingly. We have compared 3 different algorithmic approaches, namely, nearest neighbors similarity-based approach, Bayesian regression, and gradient boosting based regression. It is already a known result that we can predict hearts' biological age from the full 24-hours ECG recording, but we also show that with doing a couple of short-term recordings during a day, we can estimate this value with an acceptable loss of accuracy.

This research shows a room of consequent work both from the practical and theoretical point of view. First, it can be used for drug discovery research and to see how newly developed drugs affect the aging of the heart. Second, further experiments are needed with combining different sources of data from the wearable devices, like accelerometer-based data. Last but not least, from the theoretical point of view better algorithms for regression should be developed, that take into account more sophisticated age distributions and novel features have to be developed as well.

References.

1. *Horvath S., DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.*
2. *Cells and Surveys: Should Biological Measures Be Included in Social Science Research? National Research Council (US) Committee on Population; Finch CE, Vaupel JW, Kinsella K, editors. Washing- ton (DC): National Academies Press (US); 2001.*
3. *Toshiko Tanaka et al., Plasma proteomic signature of age in healthy humans Aging Cell. 2018 Oct; 17(5): e12799. Published online 2018 Jul 11. doi: 10.1111/acel.12799*
4. *Vito Starc, Manja Leban et al., Can Functional Car- diac Age be Predicted from the ECG in a Normal Healthy Population?*
5. *David Blokh, Ilia Stambler, The use of information theory for the evaluation of biomarkers of aging and physiological age Mechanisms of Ageing and Devel- opment Volume 163, April 2017, Pages 23-29*
6. *Natalia Botsva et al., Predictors of aging based on the analysis of heart rate variability PACE, Volume 40, Issue11, November 2017, Pages 1269-1278*
7. *Simon W Rabkin, Xin-Bo Justin Cheng, Detailed analysis of the impact of age on the QT interval J Geriatr Cardiol. 2016 Sep; 13(9): 740748.*
8. *Marcelo Franken, Amit Nussbacher, ST Elevation Myocardial Infarction in the elderly J Geriatr Car- diol. 2012 Jun; 9(2): 108114.*
9. *L.J.P. van der Maaten. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learn- ing Research 15(Oct):3221-3245, 2014.*
10. *Altman, N. S. (1992). "An introduction to ker- nel and nearest-neighbor nonparametric regression". The American Statistician. 46 (3): 175185.*
11. *Section 3.3 in Christopher M. Bishop: Pattern Recognition and Machine Learning, 2006*
12. *David J.C. MacKay, Bayesian Interpolation Sub- mitted to Neural Computation as a review paper, May 18, 1991*
13. *Roman . Baevsky, Anna G. Chernikova, Heart rate variability analysis: physiological founda- tions and main methods DOI: 10.12710/cardiome- try.2017.10.6676*
14. *Friedman,J.H.(February1999)."GreedyFunc tion Approximation: A Gradient Boosting Machine"*
15. *https://medlineplus.gov/ency/article/003 877.htm*
16. *https://mawi.band*
17. *https://www.apple.com/apple-watch- series- 4/health/*
18. *https://www.irhythmtech.com/*
19. *https://www.getqardio.com/*