# ESTIMATION OF THE BIOLOGICAL AGE OF HUMAN HEART USING MACHINE LEARNING

*Volodymyr Slipchenko*

*Department of Digital Technologies in Energy1*

*ORCID: https://orcid.org/0000-0002-3405-0781*

*ddpolytechnic2016@gmail.com*

*Liubov Poliahushko*

*Department of Digital Technologies in Energy1*

*ORCID: https://orcid.org/0000-0003-3287-8523*

*Scopus Author ID 58246840200*

*liubovpoliagushko@gmail.com*

*Vladyslav Shatylo*

*Department of Digital Technologies in Energy1*

*ORCID: https://orcid.org/0000-0001-5395-2097*

*Scopus Author ID 58247671300*

*v.shatylo@kpi.ua*

*Volodymyr Rudyk*

*Department of Digital Technologies in Energy1*

*ORCID: https://orcid.org/0009-0004-4774-6579*

*rudykviv@gmail.com*

*1National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”*

*6 Polytechnichna St., Kyiv, Ukraine, 03056*

## Abstract

This article explores the issue of assessing the biological age of human heart using machine learning methods and neural networks. Biological age is an indicator that demonstrates the degree of aging of the human body, taking into account not only the number of days since birth but also the biological characteristics of the body or individual organs. Accurate determination of this indicator will help medical professionals understand patients' conditions and detect heart diseases early. The study aims to to reveal the idea of estimating biological age based on ECG parameters and the complexity of the heart rate variability time series. The object of study is machine learning methods and neural networks. The subject of study is the use of machine learning and neural networks as a means of assessing the biological age of human heart based on ECG records. QRS duration, R amplitude, T amplitude, Higuchi fractal dimension based on RR time series data were chosen as biomarkers. The correlation of biomarkers with age was verified using Pearson correlation coefficients. The training part of the processed dataset was fed into various machine learning models based on different methods (linear regression, k-nearest neighbors, boosting, ensembles, etc.), followed by accuracy testing on the test set. Additionally, the method for assessing biological age based on two neural networks and the Klemera-Doubal method was tested. As a result of the research, the chosen set of models and machine learning methods for assessing the biological age of human bones based on data on the condition of the skeletal system was tested. The best results were shown by machine learning models based on the boosting method, such as XGBRegressor, LGBMRegressor, and CatBoostRegressor, with MAE ranging from 2.1 to 2.2 and a correlation coefficient from 0.93 to 0.94, which indicates high accuracy given the limited dataset. The scientific novelty of the research is the first use of a method based on two consecutive neural networks for bone data. Although this method showed worse results than machine learning models, the difference is not significant, indicating the versatility of the method. The study also provided important information on determining the BA of bones based on data from Ukrainian citizens, contributing to the development of the biological age field in Ukraine.

**Keywords**: biological age, bone age, BMD, neural networks, machine learning.

Annotation

Autonomic function regulating cardiac rhythm progressively declines with increasing age.

A well-structured article on **biological age assessment using fractal dimension on RR-interval time series** should include the following sections:

**Introduction**

The same individuals may have the same chronological age, but different states of the body's subsystems (estimated by biological age). Biological age is determined based on biomarkers of aging, which reflect the functional state of the body. Few hypothesis on correspondence among BA, CA and BMs are given here [1]. Assessment of biological age is important because each system, each tissue has its own, unique aging characteristics. Biological age, which reflects the degree of aging of the body both at the level of various subsystems and at the cellular level, can differ significantly from chronological age and provide more accurate information about the general state of health and the risk of developing various diseases. Assessment of biological age is important not only for clinical medicine, but also for preventive measures, as well as for monitoring the effectiveness of various anti-aging interventions, the use of geroprotectors. Determination of biological age will make it possible to differentiate geroprotectors based on the sign of slowing down aging at different stages of ontogenesis and select for further study those of them that will most closely correspond to the ideal. Due to the development of information technology and the rapid growth of available data, machine learning methods are gaining significant attention, which can be used to analyze large amounts of biological information and create accurate models for predicting biological age. One of the methods of assessing a person's biological age can be its determination using an ECG (heart age).

The need for reliable methods for assessing biological age is due to the fact that ECG data is a fairly common and easy-to-obtain biomarker of aging, which qualitatively and quantitatively reflects the functional state of the heart. It is known that heart rate variability (HRV) changes with age. Estimation of certain parameters of the HRV time series could allow analyzing age-dependent factors.

One method for analyzing RR interval time series data is to use fractal dimension (FD). FD methods (e.g., Detrended Fluctuation Analysis (DFA), Higuchi’s FD, Box-Counting FD) allow for a scale-invariant and system-wide assessment of heart rhythm complexity. In this work, the Higuchi fractal dimension was used [2]. Fractal Dimension (FD) quantifies the complexity and self-similarity of a signal. It provides a measure of how a pattern changes across different scales, making it useful for analyzing physiological signals, such as RR-interval time series in heart rate variability (HRV). Biological signals, including heart rate dynamics, exhibit fractal properties due to the interplay of multiple regulatory mechanisms (e.g., autonomic nervous system, baroreflex). FD can capture the loss of complexity in aging and disease, which traditional linear methods may miss. Healthy heart rate dynamics exhibit a fractal-like structure, meaning they are neither completely random nor entirely regular. It is known that the series of RR-intervals have a fractal structure, so they can be investigated with the help of multifractal methods [3]. With aging and certain diseases (e.g., cardiovascular disorders, diabetes), HRV tends to become more regular (lower FD), reflecting a loss of adaptability. FD serves as a biomarker for biological aging, distinguishing younger individuals (higher complexity) from older individuals (lower complexity). It can be used in predictive models to estimate biological age based on HRV patterns.

**Objective**

The aim of this study is to develop an algorithm for estimating human biological age based on fractal dimension from time series of ECG RR intervals. Attention will be paid to modern information technologies that allow processing large data sets and automating analysis processes.

**Related works**

Previous research on fractal dimension applied to physiological signals. In work [4] the classification of 7 arrhythmias is given from ECG Using Fractal Dimension and multilayer perceptron. In work [5] Artificial intelligence-estimated biological heart age using a 12-lead electrocardiogram predicts mortality and cardiovascular outcomes. In the article [6], biological age was estimated based on heart rate variability. Higuchi’s fractal dimension (HFD) is used widely to understand the complexity and non-linearities in brain signals. Age classification using HFD was slightly better than classification using spectral features (power and slope). Therefore, HFD could effectively integrate various spectral features as well as some non-linearities not captured using spectral analysis, which could enhance our understanding of brain dynamics underlying healthy aging [7]. A method based on the assessment of the Fractal Dimension (FD) of ECG recordings is suggested for the identification of cardiac diseases [8].

**Data Acquisition**

A dataset to quantify changes of cardiovascular autonomic function during healthy aging published on July 30, 2021.

A study [9] aims to provide a database of high-resolution biological signals to describe the effect of healthy aging on cardiovascular regulation. Electrocardiogram signals were recorded simultaneously at rest in 1,121 healthy volunteers.

Data set have been collected over the last decade in Jena University Hospital. All measurements were recorded at the department of psychosomatic medicine and psychotherapy. The study was approved by the ethics committee of the Medical Faculty of the Friedrich Schiller University Jena. All research was performed in accordance with relevant guidelines and regulations. The informed written consent was obtained from all subjects.

An ECG (lead II) was recorded at 1000 Hz either by an MP150 (ECG100C, BIOPAC systems inc., Golata, CA, USA) or Task Force Monitor system (CNSystems Medizintechnik GmbH, Graz AUT). Pre-gelled Ag/AgCl electrodes (BlueSensor VL, Ambu BmbH, Bad Nauheim, GER) were attached according to an Einthoven triangle.

Measurements were performed in an examination room that was temperature controlled at 22°C. During the recordings it was absolutely quiet and fully shaded. The illumination level was kept constant via an indirect light source.

After the subjects lied down comfortably on the examination tilt table, electrodes and pressure cuffs were placed. For the resting state recording, we instructed participants to avoid movement, yawning or coughing.

The instructor waited a few minutes for the participant to calm down and checked the quality of the acquired signals. In case of insufficient signal quality, electrodes and cuffs were re-arranged. Otherwise, the recording was started. The length of the recording was on average 19 minutes (8 - 45 minutes) and was supervised by the instructor.

The data files are provided in open WFDB standard format and named in consecutive numbers after random ordering. Additional patient information is stored in the file *subject-info.csv*.

Age groups are defined as follows: 1 (18-19 years), 2 (20-24 years), 3 (25-29 years), 4 (30-34 years), 5 (35-39 years), 6 (40-44 years), 7 (45-49 years), 8 (50-54 years), 9 (55-59 years), 10 (60-64 years), 11 (65-69 years), 12 (70-74 years), 13 (75-79 years), 14 (80-84 years), 15 (85-92 years). Gender is coded 0 (male) or 1 (female). Recording device is either 0 (TFM, CNSystems) or 1 (CNAP 500, CNSystems; MP150, BIOPAC Systems).

**Methodology**

**Preprocessing steps.** Recordings with artifacts (missing values) were removed from the ECG set (figure 1).

Зображення, що містить текст, знімок екрана, Графік, ряд

Вміст, створений ШІ, може бути неправильним.

Figure 1 – Extraction of ECG with artifacts (missing time series values)

A part of the ECG signal is shown in Figure 2.

Зображення, що містить текст, ряд, Графік, Шрифт

Вміст, створений ШІ, може бути неправильним.

Figure 2 – A part of the ECG signal

Preprocessing of ECG recordings includes removing baseline wander and removing high-frequency noise (figure 3, figure 4.a).

Зображення, що містить текст, знімок екрана, Шрифт, ряд

Вміст, створений ШІ, може бути неправильним.

Figure 3 – Removing baseline wander and high-frequency noise

**R-peaks detection and extracting HRV time series.** R-peak detection is performed using the Pan-Tompkins algorithm (figure 4.b).

Зображення, що містить текст, знімок екрана, Шрифт, ряд

Вміст на основі ШІ може бути неправильним.

Figure 4 – Output and filtered signals (a), R-peak detection (b)

The time series of RR-intervals is shown in the figure 5.

A graph with blue lines and dots

AI-generated content may be incorrect.

Figure 5 - Rhythmogram

**Fractal Dimension Calculation**

To calculate the fractal dimension of a time series, the Higuchi fractal dimension is used [2].

Let’s consider a finite set of HRV time series observations taken at a regular interval (1).

|  |  |  |
| --- | --- | --- |
|  | X(1), X(2), X(3) *.... , X ( N ).* | (1) |

From given time series, we first construct a new time series, , defined as (2):

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

where [ ] denotes the integer with the fractional part discarded and both k and m are integers, m and k indicate the initial time and the interval time, respectively. For a time interval equal to k, we get k sets of new time series.

We define the length of the curve, , as (3):

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

We define the length of the curve for the time interval k, <L ( k )> , as the average value over k sets of . If <L(k)> ∝ then the curve is fractal with the dimension D. Then, if *<L(k)>* is plotted against k on a doubly logarithmic scale, the data should fall on a straight line with a slope - D. Идея сделать значений k не 50 на интервале, а с шагом 1.

**Machine Learning Model**

To obtain the p-value, the value of t-statistics is first calculated. It quantifies how much the observed correlation deviates from zero in terms of standard error [10]. Larger absolute values of t-statistics indicate stronger evidence of a significant correlation. It is calculated as (4).

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

where 𝑟 is the correlation coefficient (Pearson); and *n* is the sample size. After obtaining this value, the p-value is derived from the t-distribution, considering that the degrees of freedom two less than sample size.

For each patient, the expected age was determined by averaging the values ​​of the age range.

Then a regression model with one factor was built for male (figure 6) and female (figure 7) respectively.

A screen shot of a graph

AI-generated content may be incorrect.

Figure 6 - Linear regression model for male

A screen shot of a graph

AI-generated content may be incorrect.

Figure 7 - Linear regression model for female

**Results**

As can be seen from the figures, the fractal dimension decreases with age. This means that the complexity of the HRV time series decreases with age. The larger the fractal dimension, the smaller is the biological age.

Higuchi fractal dimension (HFD) was calculated for 368 men and had Pearson correlation with age r = -0,23918 (p-value = 3,48121∙). Also HFD was calculated for 593 women and had Pearson correlation with age r = -0,2769 (p-value = 7,29297∙).

For male LR model had Mean Average Error (MAE): 16.62 years and Mean Squared Error (MSE): 526.21 years, for female had Mean Average Error (MAE): 15.76 years and Mean Squared Error (MSE): 511.05 years.

**Comparison with Other Methods**. A study conducted at the Institute of Gerontology of the Academy of Sciences of Ukraine by O.V. Korkushko and co-authors [11] showed that HRV changes naturally in the process of ontogenesis: first it increases, in accordance with the development of the organism, and then, after 40 years, it progressively decreases. The conclusion is made about the development of vegetative regulation of the cardiovascular system in the first period of ontogenesis and its gradual involution during aging. The linear regression model developed by A. V. Pisaruk gave the average absolute value of the error in calculating HRV for both sexes of 5.67 years.

Therefore, we can conclude that the linear regression model based only on the Higuchi fractal dimension gives an average absolute error of 10–11 years more compared to the model of the Institute of Gerontology of the Academy of Sciences of Ukraine, which includes three biomarkers.

**5. Discussion**

* **Interpretation of Results:**
  + Explain the physiological implications of fractal dimension changes with aging.
* **Limitations:**
  + Small sample size, noise in RR intervals, individual variability.
* **Future Directions:**
  + Improvements in feature extraction, deep learning applications, larger datasets.

**6. Conclusion**

* Summary of key findings.
* Practical implications for health monitoring and aging research.
* Potential applications in wearable health devices.

**References**

1. Klemera, P. and Doubal, S. (2006) ‘A new approach to the concept and computation of biological age’, *Mechanisms of Ageing and Development*, 127(3), pp. 240–248. doi:10.1016/j.mad.2005.10.004.
2. HIGUCHI T. APPROACH TO AN IRREGULAR TIME SERIES ON THE BASIS OF THE FRACTAL THEORY. *Physica D:* [*Nonlinear Phenomena*](https://www.sciencedirect.com/journal/physica-d-nonlinear-phenomena)*.* 1988. Vol. 31, № 2. P. 277-283.
3. Кириченко Л., Радивилова Т. ФРАКТАЛЬНЫЙ АНАЛИЗ РЕАЛЬНЫХ ДАННЫХ. *International Journal "Information Content and Processing"*. 2018. Т. 5, вип. 2. С. 142–199.
4. Kiani, K. and Maghsoudi, F. (2019) ‘Classification of 7 arrhythmias from ECG using fractal dimensions’, *Journal of Bioinformatics and Systems Biology*, 02(03). doi:10.26502/jbsb.5107008.
5. Artificial intelligence-estimated biological heart age using a 12-lead electrocardiogram predicts mortality and cardiovascular outcomes / Y.-S. Baek та ін. *Frontiers in Cardiovascular Medicine*. URL: <https://10.3389/fcvm.2023.1137892>.
6. Biological age estimation based on heart rate variability: А pilot study / O. Bashkirtsev та ін. *Праці НТШ Медичні науки*. 2021. Т. 65, вип. 2. С. 218–239. ISSN 2708‐8642.
7. Aggarwal, Srishty, і Supratim Ray. *Age-Related Changes in Higuchi’s Fractal Dimension in Healthy Human EEG Are Anti-Correlated with Changes in Oscillatory Power and 1/f Slope*. 16, Червень 2024. *Neuroscience*, <https://doi.org/10.1101/2024.06.15.599168>.
8. Ben Ali S., Aguili T. Application of Fractal Dimension for Cardiac Arrhythmias Classification. *Computational Biology and Bioinformatics*. Т. 12, вип. 1. С. 12–17. URL: <https://10.11648/j.cbb.20241201.12>.
9. Schumann, A., & Bär, K. (2021). Autonomic Aging: A dataset to quantify changes of cardiovascular autonomic function during healthy aging (version 1.0.0). *PhysioNet*. <https://doi.org/10.13026/2hsy-t491>.
10. E. Johannesson, J. Ohlson, and S. Zhai, “The Explanatory Power of Explanatory Variables,” Rev Account Stud, vol. 29, pp. 3053–3083, Jul. 2023, doi:https://doi.org/10.1007/s11142-023-09781-w
11. Писарук А.В., Шатило В.Б., Антонюк-Щеглова І.А., Кошель Н.М. Біологічний вік і профіль старіння людини (регресійні та нейромережні моделі). К.: «Фенікс», 2023. – 220 с.

**Изолиния**

Между волной Р и началом комплекса QRS. Изолиния на ЭКГ — это базовая линия, относительно которой происходят все электрические колебания сердца. Она представляет собой состояние покоя сердца между сокращениями, когда нет деполяризации или реполяризации.

**3. Проверь сегмент PQ (или PR)**

Сегмент PQ (или PR) — это наиболее стабильный участок, который часто принимается за изолинию. Он идёт после зубца P и до начала комплекса QRS.

**Как определить изолинию при искажённой TP:**

1. **Смотреть на сегмент PR (или PQ)** — он часто более устойчив, особенно у медленного или нормального ритма.

Изолиния — это идеализированная базовая линия, которая может не совпадать с TP. Она определяется по максимально плоским, спокойным участкам ЭКГ, и лучше всего — по PQ или между окончаниями T и началом P, если сигнал чист.

Вот пример кода на Python с использованием библиотеки **NumPy** и **SciPy**, который находит участок изолинии на ЭКГ автоматически, ориентируясь на **PQ-сегмент**:

**🧠 Идея:**

* Найти все **P-концы** и **Q-начала** (у тебя они уже размечены).
* В каждом интервале между ними взять сигнал.
* Посчитать **среднее значение амплитуды** — это и будет уровень изолинии.

**Піки**

Піки нормально визначає.

Начало зубца P — это первая точка, где сигнал уходит от изолинии, а не от пика и не от нижнего значения.

То есть нужно отсчитывать P сегмент от начала выпуклости вниз от изолинии?

**Програма ChatGPT сказала:**

Да, именно так — ты схватываешь суть очень точно.

**✅ Краткий и точный ответ:**

**Да, зубец P начинается с момента, когда сигнал начинает выпуклость вниз или вверх от изолинии, то есть — от первого устойчивого отклонения от изолинии, даже если это вниз перед основным подъёмом.**

**🔍 Подробное объяснение:**

**🔹 Зубец P — это вся электрическая активность предсердий.**

* Она **не обязательно начинается сразу подъёмом вверх**.
* Может быть **предварительное отклонение вниз** (небольшое «плечо» или волна).

**🔹 Физиологически правильно:**

* Начало P — **первая точка, где кривая отрывается от изолинии**;
* Это может быть как **вниз**, так и **вверх**;
* Даже если сигнал уходит сначала немного вниз, потом вверх — **вся эта форма входит в зубец P**.

**🔎 На твоём графике:**

* Перед главным подъёмом зубца P (пик в ~1470 мс) есть **плавная выпуклость вниз начиная примерно с ~1420–1425 мс**.
* Именно эта точка и есть **начало зубца P**.

**📐 Как распознать это программно:**

1. Взять небольшой отрезок до P-пика (100–120 мс).
2. Найти изолинию (например, по PQ-сегменту).
3. Пройти от пика назад:
   * найти **точку, где сигнал стабильно уходит от изолинии**;
   * либо найти **точку изгиба**, где вторая производная сигнала ≠ 0.

**💬 Вывод:**

**Зубец P начинается от момента, где сигнал начинает "отклоняться" от изолинии — вне зависимости от направления (вверх или вниз). Главное — это устойчивое, не шумовое, отклонение.**

Хочешь, я помогу тебе построить алгоритм, который находит **точку изгиба или устойчивого отклонения от изолинии** для точного определения начала зубца P?