# 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:

1. **Introduction**

**Одні і ті самі індивіди можуть мати один і той самий хронологічний вік, проте різний стан підсистем організму (оцінєються біологічним віком). Біологічний вік визначається на основі біомаркерів старіння, які відображають функціональний стан організму. Few hypothesis on correspondence among BA, CA and BMs are given here [1]. Оцінка біологічного віку є важливою, оскільки** кожна система, кожна тканина має свої, тільки їй притмаманні особливості старіння. Біологічний вік, який відображає ступінь старіння організму як на рівні різноманітних підсистем, так і на клітинному рівні, може суттєво відрізнятися від календарного віку та надавати більш точну інформацію про загальний стан здоров'я та ризик розвитку різноманітних захворювань. Оцінка біологічного віку є важливою не тільки для клінічної медицини, але й для профілактичних заходів, а також для моніторингу ефективності різних антивікових втручань, використання геропротекторів. Визначення біологічного віку дасть змогу диференціювати геропротектори за ознакою уповільнення старіння на різних етапах онтогенезу і відбирати для подальшого вивчення ті з них, котрі в найбільшій ступені будуть відповідати ідеалу. У зв'язку з розвитком інформаційних технологій та швидким зростанням обсягів доступних даних, значної уваги набувають методи машинного навчання, які можуть бути використані для аналізу великих обсягів біологічної інформації та створення точних моделей прогнозування біологічного віку. Одним з методів оцінки біологічного віку людини може бути його визначення за допомогою ЕКГ.

Потреба в надійних методах оцінки біологічного віку зумовлена тим, що дані ЕКГ є досить поширеним та простим для отримання біомаркером старіння, який якісно та кількісно відображає функціональний стан серця. Відомо, що з віком варіабельність серцевого ритму (ВСР) змінюється. Оцінка певних параметрів часового ряду ВСР могла б дати змогу проаналізувати вік-залежні фактори.

Одним з методів для аналізу даних часових рядів RR-інтервалів є використання фрактальної розмірності (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. В даній роботі використовувалася фрактальна розмірність Хігучі. 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. With aging and certain diseases (e.g., cardiovascular disorders, diabetes), HRV tends to become more **regular (lower FD) or erratic (higher 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.

* Метою даного дослідження є розробка алгоритму оцінювання біологічного віку людини на основі фрактальної розмірності за даними часового ряду RR-інтервалів ЕКГ. Увага буде приділена сучасним інформаційним технологіям, які дозволяють обробляти великі масиви даних, автоматизувати процеси аналізу.

1. **Related Work**

**В работе [3] приведена классификация 7 аритмий from ECG Using Fractal Dimensions. В роботі [4] наведена класифікація аритмій Using Fractal Dimensions and Neural Networks. В роботі [5] Artificial intelligence-estimated**

**biological heart age using a**

**12-lead electrocardiogram**

**predicts mortality and**

**cardiovascular outcomes**

* **Biological Age Models:** Overview of existing biological age assessment techniques (e.g., HRV-based, metabolic models). В статті [5] оцінювався біологічний вік на основі варіабельності серцевого ритму
* **Fractal Analysis in Medicine:** Previous research on fractal dimension applied to physiological signals.
* A method based on the assessment of the Fractal Dimension (FD) of ECG recordings is suggested for the identification of cardiac diseases [7].

**Significance Statement** 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 [6].

**Fractal Properties of RR Intervals:** Prior studies on fractal complexity of heart rate variability (HRV) and aging. **Известно, что ряды RR-интервалов имеют фрактальную структуру, поэтому их можно исследовать с помощью мультифрактальных методов. [2].**

**3. Methodology**

* **Data Acquisition:**

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

A study [8] 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).

**Preprocessing steps.**

З набору ЕКГ були вилучені записи с артефактами (пропусками значень) (рисунок 1). Зображення, що містить текст, знімок екрана, Графік, ряд

Вміст, створений ШІ, може бути неправильним.

Рисунок 1 – Вилучення ЕКГ з артефактами (пропущені значення часового ряду)

(artifact removal, filtering, normalization).

Попередня обробка записів ЕКГ включає в себе видалення блукання базової лінії, видалення високочастотних завад (рисунок 2).

Частина вихідного сигналу ЕКГ наведена на рисунку 1.

Зображення, що містить текст, ряд, Графік, Шрифт

Вміст, створений ШІ, може бути неправильним.

Рисунок 1 – Частина вихідного сигналу ЕКГ

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

Вміст, створений ШІ, може бути неправильним.

Рисунок 2 – Видалення блукання базової лінії, видалення високочастотного шуму

Детектування R-піків виконується за допомогою алгоритму Пана-Томпкінса.

A screenshot of a computer screen

AI-generated content may be incorrect.A graph with blue lines and dots

AI-generated content may be incorrect.

Рисунок 5 - Ритмограма

A screenshot of a computer screen

AI-generated content may be incorrect.

* **RR-Interval Extraction:**
  + Explain how RR-intervals are derived from ECG signals.
* **Fractal Dimension Calculation:**
  + Methods used (e.g., **DFA (Detrended Fluctuation Analysis), Higuchi’s FD, Katz’s FD, Box-Counting Method**).
* **Feature Selection:**
  + Discuss key fractal parameters that may correlate with biological age.
* **Machine Learning Model (if applicable):**
* A screen shot of a graph

  AI-generated content may be incorrect.

Чи треба шукати середню фрактальну розмірність по вікам?

Нейронна мережа. На вході фрактальна розмірність. На виході мітка з віком (регресія чи класифікація?)

A screen shot of a graph

AI-generated content may be incorrect.

* + Explain how fractal features are mapped to biological age.
  + Algorithms used (e.g., regression models, neural networks).

**4. Results**

* **Fractal Dimension vs. Age:**
  + Present findings on how FD changes with age.
* **Statistical Analysis:**
  + Correlation coefficients, significance tests (e.g., Pearson/Spearman correlation).
* **Performance Evaluation:**
  + If ML models are used, report RMSE, R², MAE, or classification accuracy.
* **Comparison with Other Methods:**
  + Compare FD-based age estimation with conventional methods.

**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.

**7. References**

* Cite relevant research papers, books, and resources used in the study.

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. Кириченко Л., Радивилова Т. ФРАКТАЛЬНЫЙ АНАЛИЗ РЕАЛЬНЫХ ДАННЫХ. *International Journal "Information Content and Processing"*. 2018. Т. 5, вип. 2. С. 142–199.
3. 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.
4. 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>.
5. 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.
6. 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>.
7. 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>.
8. 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>.

**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?