Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

AI-ECG–Derived Biological Age Estimation from 12-Lead ECG

Master's Thesis

2027
Filip Zrubák

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

AI-ECG–Derived Biological Age Estimation from 12-Lead ECG

Master's Thesis

Study Programme:	Computer Science
Field of Study:	Computer Science
Department:	Department of Computer Science
Supervisor:	Ing.Adam Rafajdus

Bratislava, 2027
Filip Zrubák

ZADANIE ZÁVEREČNEJ PRÁCE

[download this from AIS]

THESIS ASSIGNMENT

[download this from AIS]

Acknowledgments: You can thank anyone who helped you with the thesis here (e.g. your supervisor).

Abstrakt

Slovenský abstrakt v rozsahu 100–500 slov, jeden odstavec. Abstrakt stručne sumarizuje výsledky práce. Mal by byť pochopiteľný pre bežného informatika. Nemal by teda využívať skratky, termíny alebo označenie zavedené v práci, okrem tých, ktoré sú všeobecne známe.

Kľúčové slová: Slovak, keywords, here

Abstract

Abstract in the English language (translation of the abstract in the Slovak language).

Keywords: English, keywords, here

Contents

Introduction

Put your introduction here.

Chapter 1

Background

This chapter introduces the clinical and methodological foundations required for ECG-based age estimation with deep learning. We first summarize the physiological meaning of a standard 12-lead ECG and typical waveform components. We then describe how ageing and cardiovascular pathology influence ECG morphology. Finally, we introduce deep learning approaches for raw ECG waveforms and the basic concepts of survival analysis that are commonly used to evaluate risk associations.

1.1 Electrocardiogram basics

A standard 12-lead electrocardiogram (ECG) is a non-invasive measurement of the heart's electrical activity. It is typically recorded over a short time window (commonly 10 seconds) and represented as multiple synchronous one-dimensional time series, each corresponding to a lead with a different projection of the cardiac electrical vector.

An ECG heartbeat is commonly described by the **P–QRS–T** pattern. The P-wave reflects atrial depolarization, the QRS complex corresponds to ventricular depolarization, and the T-wave represents ventricular repolarization. In addition to morphology, clinically used summary parameters include intervals (e.g., PR and QT/QTc) and durations (e.g., P-wave duration, QRS duration), which can be derived from fiducial points on the waveform.

1.2 Ageing effects observable in ECG

Ageing is associated with gradual changes in cardiac structure and conduction. Two important long-term processes are cardiac remodelling (e.g., hypertrophy) and fibrosis, which can manifest as subtle alterations in waveform morphology and timing. Because these changes are often small at the level of individual ECG parameters, they may be difficult to capture with rule-based analysis, but can still be present in the raw signal

in a distributed way across leads and waveform segments.

Recent work studying ECG-age models has shown that groups with higher predicted ECG ageing effects tend to exhibit longer durations and intervals such as P duration, PR interval, QRS duration and QTc, even though distributions overlap substantially between groups [1]. This motivates the use of data-driven models that can combine many weak ECG cues.

1.3 Deep learning for raw ECG waveforms

Modern ECG models typically process raw waveforms using one-dimensional convolutional neural networks (1D-CNNs) or residual networks (ResNets). Compared to hand-crafted features, such networks can learn hierarchical representations directly from the signal, potentially capturing subtle, non-obvious patterns linked to physiology, ageing, and disease.

In the ECG-age setting, a model is trained to predict chronological age from the ECG waveform. The resulting predicted age can be interpreted as an ECG-derived age estimate. A commonly used derived quantity is the age gap (also called Δ-age or predicted age deviation), defined as the difference between predicted ECG age and chronological age. This age gap is often evaluated both continuously and by categorizing individuals into groups such as underestimation, correct prediction, and overestimation using a fixed threshold (e.g., ± 8 years) [1, 2].

1.4 Regression bias and age-gap confounding

When predicting age with regression, models often show *regression-to-the-mean*: younger individuals tend to have their age overestimated and older individuals underestimated. This produces a systematic negative correlation between chronological age and the raw age gap, which can distort downstream analyses if the age gap is used as a biomarker.

A practical correction strategy is to remove the dependence of the age gap on chronological age by regressing the age deviation on age and using the residual. One concrete formulation defines predicted age PA, chronological age CA, and predicted age deviation $PAD = PA - CA$. A linear correction can be derived from a fitted model

$$PA = \alpha + \beta \cdot CA + \varepsilon, \tag{1.1}$$

leading to

$$PAD_c = PA - (\alpha + \beta \cdot CA). \tag{1.2}$$

Residual non-linear dependence across ages can be further reduced by subtracting the mean PAD_c within each integer age (or age band for sparse ages), yielding a

bias-corrected PAD_{bc} [3]. Such corrections are especially relevant when associating age gaps with outcomes that are themselves age-dependent.

1.5 Survival analysis essentials

Clinical outcomes are frequently studied with time-to-event analysis. A key challenge is that not all individuals experience the event during observation; for those individuals the event time is unknown beyond their last follow-up. This is handled via **right-censoring**, where an individual contributes follow-up time until a censoring time (e.g., end of study), but the event is not observed.

Two common tools are:

- **Kaplan–Meier estimation**, which provides a non-parametric estimate of the survival function and enables group comparisons via log-rank tests.
- **Cox proportional hazards models**, which estimate relative risk (hazard ratios) associated with predictors, optionally adjusting for covariates.

In ECG-age research, survival analysis is typically used to quantify whether larger positive age gaps are associated with increased mortality risk and whether negative age gaps are protective, both in population-based cohorts and high-risk clinical cohorts [1, 2, 3].

Chapter 2

State of the Art and Related Work

This chapter summarizes current research on AI-based ECG age estimation and the interpretation of the ECG age gap as a biomarker. We focus on three themes that directly inform this thesis: (i) evidence that the ECG age gap is associated with risk and outcomes, (ii) the role of longitudinal (serial) ECGs, and (iii) methodological pitfalls such as regression bias and confounding, including practical bias-correction strategies. We also review explainability findings that suggest plausible physiological mechanisms.

2.1 ECG age and age gap as a biomarker

Deep learning models can estimate chronological age from raw 12-lead ECG waveforms. The resulting *ECG age* is often used to define an *age gap* Δ-age (or predicted age deviation, PAD), which is hypothesized to reflect biological rather than purely chronological ageing.

A common evaluation approach categorizes individuals by whether predicted ECG age deviates from chronological age beyond a predefined threshold. For example, in a longitudinal population study, δ-age was categorized into three groups: *overestimation* (δ-age > 8 years), *correct prediction* ($|\delta\text{-age}| \leq 8$ years), and *underestimation* (δ-age < -8 years) [1]. Such groupings are attractive in clinical narratives (biologically older vs. younger), but require careful statistical handling, especially across age ranges and settings.

2.2 Model architectures and training pipelines for ECG-age estimation

From a machine learning perspective, ECG-age models are typically trained as supervised regressors that map raw multi-lead ECG waveforms to chronological age. Most approaches follow an end-to-end paradigm: the network receives minimally processed

signals and learns its own feature representations without explicit hand-crafted ECG measurements.

Residual networks trained on large-scale ECG corpora. A widely used open-source ECG-age model is based on a residual network (ResNet) architecture and was trained on ECG exams from 1,558,415 patients from a Brazilian telehealth network. The model processes 12-lead ECGs end-to-end, without manual filtering or feature extraction stages, and outputs a numerical age estimate [1]. In a hospital-based validation study, the same originally published ResNet-based model was applied without recalibration; importantly, the algorithm used only pre-processed raw ECG signals and no patient metadata (e.g., sex, clinical history), supporting a clean assessment of signal-derived ageing information [2].

1D-CNN baselines with temporal feature extraction and lead aggregation. An alternative (and conceptually simpler) family of models uses 1D convolutional blocks to extract temporal features from each lead, followed by an explicit aggregation step across leads. For example, one recent implementation consists of multiple sequential blocks of convolution, batch normalization, and max pooling, followed by a spatial aggregation block across leads and final fully connected layers producing the age estimate [3]. In that setup, the input is a raw 12×5000 matrix corresponding to a 10-second recording at 500 Hz, and the model is trained by minimizing mean squared error (MSE) [3].

Preprocessing and data splitting choices. Across studies, preprocessing is intentionally kept lightweight to preserve morphological information. Reported pipelines include resampling ECGs to a common sampling rate and standardizing amplitude units, with limited or no additional filtering [3, 2]. To avoid leakage, splits are performed at the patient level (all ECGs from a patient in one split), and analyses often restrict model training to the first ECG per patient to prevent over-representation of frequent visitors [3].

Evaluation. Model accuracy is commonly summarized using MAE and correlation between predicted ECG age and chronological age, sometimes complemented by regression slope/intercept analyses [3]. Downstream biomarker analyses then study the age gap (or its bias-corrected variants) in relation to outcomes such as mortality, typically using Cox proportional hazards models with adjustment for chronological age and other confounders [3, 2, 1].

2.3 Longitudinal ECGs and risk association

Most early ECG-age evaluations used only one ECG per individual. However, serial ECGs can capture progression and stability of ageing effects. In a cohort with long follow-up, incorporating two ECGs recorded 5–6 years apart strengthened the association between pronounced ageing effects and mortality: the hazard ratio for the overestimation group increased when using serial ECGs compared to a single follow-up ECG [1]. This supports the hypothesis that persistent or consistently elevated ECG-age signals a higher-risk phenotype than a single measurement alone.

Longitudinal analyses also motivate methodological design choices in new studies: aligning time-to-event from a follow-up ECG to avoid label leakage, assessing the stability of age-gap classifications, and quantifying whether consistent predictions over time improve risk stratification [1].

2.4 ECG age in high-risk clinical cohorts

Beyond population screening, recent evidence suggests that ECG age retains prognostic value in high-risk clinical settings. A study of patients with cardiovascular disease or acute medical conditions used a validated open-source ECG-age model and reported that a positive Δ-age (e.g., $\geq +8$ years) was associated with higher long-term mortality, while negative Δ-age (e.g., ≤ -8 years) was associated with lower risk [2]. Importantly, an exploratory analysis found an optimal cutoff close to the commonly used ± 8 years threshold [2].

Such results broaden the potential utility of ECG age from preventive care toward real-world cardiology and acute care, but also highlight that comorbidity burden may interact with the correlation between ECG age and chronological age, suggesting that the age gap can reflect both ageing and disease processes [2].

2.5 Regression bias and bias-adjusted age deviation

A key methodological challenge is regression bias, which creates a systematic dependence of the raw age deviation on chronological age. This can invert or obscure associations when PAD is correlated with outcomes that themselves vary strongly with age.

A recent large-scale study explicitly quantified this effect. The authors observed a substantial negative correlation between chronological age and PAD, consistent with regression-to-the-mean, and proposed bias-corrected variants: a linear correction using a fitted regression of predicted age on chronological age, and an additional age-level correction to remove remaining non-linear dependencies, yielding PAD_{bc} [3]. Using PAD_{bc} changed the direction of associations with multiple cardiovascular risk factors

and reversed the qualitative interpretation of Kaplan–Meier survival curves compared to uncorrected PAD, while multivariable Cox models that adjust for chronological age produced more consistent results [3].

For new studies aiming to interpret ECG-age as a biomarker, this implies that:

- analyses should test and report the dependence between the age gap and chronological age;
- categorical thresholds derived from prediction error (e.g., MAE-based cutoffs) should be interpreted with caution across ages;
- bias-corrected age deviation (or explicit adjustment for age in outcome models) is often necessary to avoid misleading conclusions [3].

2.6 Explainability and mechanistic interpretation

To build trust and generate hypotheses, ECG-age studies often use post-hoc explainability methods such as saliency maps or integrated gradients. Across multiple cohorts, explainability analyses repeatedly highlight the importance of atrial activity, particularly the P-wave and adjacent segments, for age prediction [1, 2, 3]. Lead-wise relevance analyses also suggest that multiple leads contribute, with notable importance of precordial leads (e.g., V1 and V4) that provide complementary information about atrial and ventricular activity [1].

While these observations are biologically plausible, it is important to remember that saliency indicates *sensitivity* of the model output to input regions, not causal mechanisms. Therefore, explainability should be combined with robust validation, subgroup analyses (age, sex, rhythm), and sanity checks to mitigate over-interpretation.

2.7 Summary and open challenges

The current state of research supports ECG age and the age gap as promising signals associated with cardiovascular risk and mortality across settings, including longitudinal population cohorts and high-risk clinical cohorts [1, 2]. At the same time, methodological issues such as regression bias can materially affect downstream associations, motivating bias correction and careful study design [3]. Finally, emerging explainability results suggest that age-related information is distributed across leads and waveform segments, with repeated emphasis on atrial conduction markers, providing a direction for mechanistic follow-up [1, 2].

Chapter 3

Data and Cohort Construction

This chapter describes the data sources and the cohort definition used in this thesis. The goal is to build a reproducible pipeline that links ECG waveforms to patient-level clinical information and constructs analysis cohorts suitable for training ECG-age models and downstream association analyses.

3.1 Data sources

We use two linked critical-care resources:

- **MIMIC-IV-ECG**: a waveform database containing standard 12-lead ECG recordings.
- **MIMIC-IV (hospital module)**: structured electronic health record (EHR) data including demographics, diagnoses, and outcomes.

ECGs are linked to EHR information primarily via `subject_id` (patient identifier). When available, `hadm_id` enables admission-level linkage to hospitalization-specific events. This design allows enrichment of each ECG with patient-level variables such as age at recording, sex, comorbidities derived from diagnosis codes, and mortality.

3.2 Label definition: chronological age at ECG

For each ECG, the primary label is chronological age at the time of recording. In de-identified hospital datasets, date shifting and anchor-year schemes may require special handling. We compute age using the dataset provided demographic fields and the ECG acquisition time, ensuring consistency across all records and preventing information leakage from future events.

Cohort / filtering step	**#ECGs**	**#Patients**
All linked ECG records (matched dataset)	800,035	161,352
No-CVD adults ($\geq$ 18; exclude CVD by ICD)	124,680	62,893
One ECG per patient (index ECG)	62,893	62,893
Age-restricted (18–65 years; one ECG per patient)	54,947	54,947

Table 3.1: Cohort overview and filtering steps. Counts are reported as ECG recordings and unique patients.

3.3 Inclusion and exclusion criteria

We define an analysis cohort with the following general criteria:

- adult patients ($\geq$ 18 years at ECG),
- ECGs passing basic signal quality checks (e.g., excluding corrupted waveforms or acquisition failures),
- one ECG per patient for model training and primary analyses, to avoid over-weighting frequent visitors.

Depending on the experiment, we further define a *near-healthy* cohort by excluding patients with evidence of cardiovascular disease (CVD) based on diagnosis codes. Concretely, we filter out subjects with ICD codes indicating CVD (e.g., ICD-10 codes starting with `I` and ICD-9 codes in the range 390–459), using diagnosis tables indexed by `subject_id` and `hadm_id`. This yields a cohort enriched for subjects without known CVD at the time of ECG.

Dataset scale. Starting from the full matched ECG dataset, the linked MIMIC-IV-ECG and MIMIC-IV resources contain **800,035** ECG recordings from **161,352** unique patients. After excluding patients with cardiovascular disease diagnoses (ICD-10 `I*` and ICD-9 390–459) and restricting to adults ($\geq$ 18 years at ECG time), the resulting *near-healthy* (no-CVD adult) subset contains **124,680** ECGs from **62,893** patients. To prevent patient-level leakage and over-representation of frequent visitors, we further reduce the cohort to one ECG per patient, resulting in **62,893** ECGs. For the initial baseline experiments, we additionally restrict the age range to 18–65 years, yielding **54,947** ECGs.

3.4 Train/validation/test splitting

To prevent patient-level leakage, data are split by `subject_id`, ensuring that all ECGs from a given patient appear in exactly one split. We use three-way splits

(train/validation/test) and, when performing risk factor or survival analyses, we keep a sufficiently large test split to support statistical power.

Because age distributions can be highly imbalanced, weconstruct age-balanced evaluation subsets by capping the number of patients per integer age (e.g., a maximum count per age-year) and by stratifying splits by sex and age bands. This reduces the impact of regression to the mean artefacts and improves interpretability of subgroup performance.

3.5 Signal preprocessing

ECG waveforms are converted into a consistent tensor format:

- lead ordering is standardized across all records,
- signals are resampled if required to a common sampling frequency,
- amplitude units are normalized to a consistent scale,
- optional per-lead normalization (zero mean, unit variance) is applied based on training data statistics.

Quality control includes rejecting ECGs with missing leads, extreme noise, or inconsistent lengths. All preprocessing steps are implemented deterministically to ensure reproducibility.

3.6 Outcome and follow-up definitions

For association analyses, we consider outcomes available in the EHR such as all cause mortality. Time-to-event is defined from the ECG acquisition time to the event time (death) or censoring time (end of data availability), applying right-censoring for patients without an observed event. For diagnosis based outcomes, we define incident events by requiring that relevant diagnosis codes occur after the index ECG (with sensitivity analyses to address imperfect timing in retrospective labels).

3.7 Reproducibility

All cohort extraction, linkage, and preprocessing steps are implemented as a modular pipeline. Each step logs counts and exclusion reasons (e.g., number of patients removed due to CVD codes, age constraints, or waveform issues), enabling traceability and straightforward auditing of cohort construction decisions.

Conclusion

Put your conclusion here.

Bibliography

[1] Philip Hempel, Antônio H. Ribeiro, Marcus Vollmer, Theresa Bender, Marcus Dörr, Dagmar Krefting, and Nicolai Spicher. Explainable ai associates ecg aging effects with increased cardiovascular risk in a longitudinal population study. *npj Digital Medicine*, 8:25, 2025.

[2] Daniel Pavluk, Fabian Theurl, Samuel Proell, Michael Schreinlecher, Florian Hofer, Patrick Rockenschaub, Angus Nicolson, Mercedes Gauthier, Sebastian Reinstadler, Clemens Dlaska, and Axel Bauer. Ai-ecg-derived biological age as a predictor of mortality in cardiovascular and acute care patients. *European Heart Journal - Digital Health*, 2025.

[3] Myrte Barthels, Elisa Verhofstadt, Inigo Bermejo Delgado, Henri Gruwez, Laurent Pison, Noëlla Pierlet, and Pieter Vandervoort. Artificial intelligence-predicted ecg age gap as a biomarker: bias-adjusted correlation with mortality and cardiovascular risk factors. *European Heart Journal - Digital Health*, 2025.