Philip Hempel*, Tabea F. A. Steinbrinker, Lennart Graf, Henning Dathe, Nicolai Spicher, and Dagmar Krefting

Predicting ECG Age in 24-hour Holter Recordings of Heart Failure Patients

https://doi.org/10.1515/cdbme-2025-0240

Abstract: Recent advances in deep learning enable estimation of a patient's "ECG age" from standard short-term 12-lead electrocardiogram (ECG). This study introduces sequential ECG age predictions in continuous 24-hour Holter ECG recordings of chronic heart failure (CHF) patients. Using publicly available data from the MUSIC study, we analyzed data from 869 CHF patients, assessing both the differences (predicted ECG age vs actual chronological age) and dynamic —including variability and entropy-based complexity—to characterize temporal fluctuations over 24h. After automated removal of segments based on signal-to-noise ratio (SNR), 222 deceased patients were matched 1:1 with 222 surviving patients by sex, NYHA class, and age. While the predicted ECG age is similar across CHF patients, our findings indicate lower variability but increased complexity (approximate and sample entropy) in deceased compared to surviving patients, suggesting more irregular predicted ECG age dynamics among those who experienced adverse outcomes. Our results suggest that longitudinal evaluation of predictions from an end-to-end deep learning model can uncover subtle temporal dynamics potentially valuable for risk stratification.

Keywords: electrocardiogram; heart failure; ECG age; Holter monitoring; entropy; complexity; deep learning; biomarker

***Corresponding author: Philip Hempel,** Department of Medical Informatics, University Medical Center Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany, e-mail: philip.hempel@med.uni-goettingen.de
Tabea F. A. Steinbrinker, Department of Medical Informatics, University Medical Center Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany
Lennart Graf, Department of Medical Informatics, University Medical Center Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany
Henning Dathe, Department of Medical Informatics, University Medical Center Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany
Nicolai Spicher, Department of Medical Informatics, University Medical Center Göttingen, 37075 Göttingen, Germany and with the Department of Health Technology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
Dagmar Krefting, Department of Medical Informatics, University Medical Center Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany

1 Introduction

Chronological aging is a recognized risk factor for cardiovascular diseases [1], yet the processes underlying biological aging can diverge considerably among individuals. In prior work [2], we used a deep learning approach to estimate a patient's "heart" age from standard 12-lead electrocardiograms (ECGs), demonstrating correlations with the incidence and prevalence of chronic heart failure (CHF) [2]. Although these findings indicate promise for ECG-based age estimation as a novel bioindicator, its stability over extended monitoring periods and under ambulatory conditions remains unclear.

Holter ECGs are routinely employed in CHF management for prolonged arrhythmic surveillance, offering the unique opportunity to monitor ECG-derived age continuously for 24 hours or more [3, 4]. However, ambulatory recordings are frequently affected by patient movement, speech, and environmental factors that introduce substantial noise. Recent evidence suggests that deep neural networks can effectively manage such variability, with initial layers often learning robust features resilient to common artifacts [2, 5]. Building on these insights, our study investigates the stability and variability of ECG-derived age predictions across a full day of Holter recordings in a CHF cohort. Specifically, we quantify both the overall differences (predicted ECG age - chronological age) and the temporal fluctuations by non-linear dynamics metrics like entropy, hypothesizing that pronounced variability or elevated complexity might reflect latent pathological processes (see Figure 1).

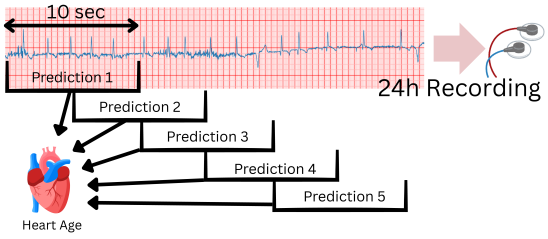

Fig. 1: Graphical abstract: Differences between chronological and predicted ages are computed across the recording period, and time-series analyses (including entropy-based complexity metrics) are applied to capture dynamic variations and underlying trends.

2 Materials and Methods

2.1 Dataset

We used the publicly available *MUSIC* (Muerte Súbita en Insuficiencia Cardíaca) database [6] from Spain, a prospective, multi-center registry designed to identify risk factors in patients with symptomatic CHF. The cohort predominantly comprises individuals in NYHA class II–III (indicating mild to moderate limitations in physical activity due to heart failure) who each underwent a 24-hour Holter ECG recording. From the full dataset, 869 patients were identified with Holter ECG and complete follow-up, which lasted a median of 44 months with outpatient visits every six months.

From these patients, we selected 444 individuals (222 who later died and 222 who remained alive) and matched them 1:1 by sex, NYHA class, and age (within 2 years). This matched sub-cohort had a mean age of 63.4 years (SD: 10.7), with 75.2% being male. By NYHA classification, 152 were in NYHA II and 70 were in NYHA III.

To prepare each Holter recording, the first 30 seconds were discarded and then segmented into 10-second windows with a 5-second overlap. The three recorded orthogonal Frank leads (X, Y, Z) sampled at 200 Hz were transformed into a 12-lead ECG configuration via the Uijen matrix [7, 8] and resampled to 400 Hz. On average, each 24-hour Holter ECG yielded roughly 16,584 ten-second windows (SD: 884).

Automated filtering was performed to remove segments with poor signal quality due to motion artefacts (SNR, threshold: 2 dB), resulting in approximately 15.64% of segments being excluded (SD:11.89%) [9].

For subsequent analyses, patients were categorized into two primary outcome groups:

- **Alive:** Patients who remained alive at the end of the follow-up period.
- **Deceased:** Patients who died from cardiac-related or other complications during the follow-up period.

2.2 Analysis

We applied a previously validated 1DResNet architecture [2, 5, 10] to each filtered 10-second ECG segment, thereby generating window-level predicted ECG age estimates. These estimates were subsequently aggregated per patient, yielding both moment statistics (mean, median, minimum, maximum, and standard deviation (SD)) and complexity measures (approximate entropy and sample entropy). While the SD explicitly captures variability among the segment-level predictions, the minimum and maximum additionally reflect the overall range. Approximate entropy quantifies the likelihood that patterns of length m in the prediction time series remain similar upon extension by one observation; the distance between two patterns x and y is defined by the Chebyshev distance $D(x, y) = \max_i(|x_i - y_i|)$ for $i = 1, \ldots, m$, with the similarity threshold r typically set to $r = 0.2$ SD and $m = 2$. Higher approximate entropy values indicate greater unpredictability and irregularity [11]. Sample entropy likewise characterizes complexity by evaluating the regularity of the time series but is less sensitive to data length; it uses the same pattern length, distance, and threshold definitions as approximate entropy yet differs in its calculation. Higher sample entropy values signify increased complexity and reduced predictability [11].

We summarized the differences between predicted ECG age and chronological age to assess predictive accuracy. Nonlinear dynamics were examined through these entropy metrics to elucidate temporal complexity and irregularity patterns. The resulting distributions were visualized via violin plots, which combine kernel density estimates and boxplot. All statistical analyses and visualizations were implemented in Python using the libraries "matplotlib" and "seaborn", and the corresponding code is openly available for reproducibility.

3 Results

Figure 2 displays the distribution of ECG-derived age differences (predicted ECG age minus chronological age) for the matched *Alive* and *Deceased* groups. Both groups exhibit similar distributions centered near zero, without statistically significant differences across mean, median, minimum, or maximum differences (Mann–Whitney U test, 0.7090, 0.6574, 0.4220, and 0.1923, respectively). Most CHF patients exhibit positive mean and median differences, indicating a general overestimation of their predicted ECG age. Furthermore, the considerable spread between minimum and maximum differences in both groups highlights notable within-patient variability in ECG age throughout the 24-hour recording period.

Figure 3 illustrates the distributions of SD, Approximate Entropy, and Sample Entropy of ECG-derived age predictions over 24 hours for the same matched groups. Deceased patients demonstrate significantly lower SD values, indicating reduced variability, yet exhibit significantly higher entropy values, suggesting increased complexity or instability in their predicted ECG age dynamics. These findings indicate that death in CHF patients is associated with less variable but more irregular predicted ECG age fluctuations compared to surviving matched counterparts.

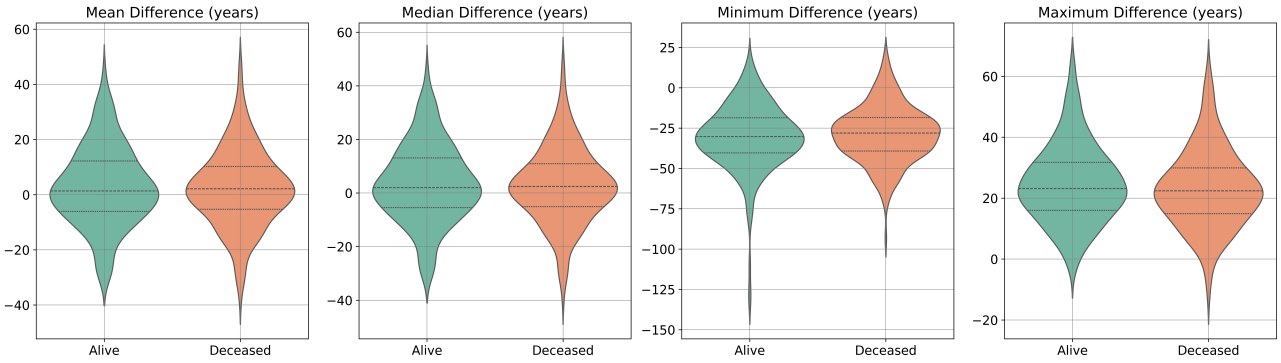

Fig. 2: Violin plots displaying difference between predicted ECG age and chronological age for matched *Alive* and *Deceased* groups. Patients were matched 1:1 by sex, NYHA class, and age within 2 years (222 Deceased matched with 222 Alive). Each metric was compared between Deceased and Alive using the Mann–Whitney U test, yielding no significant differences. The respective *p* values were 0.7090, 0.6574, 0.4220, and 0.1923, indicating no significant differences. The majority of CHF patients are overestimated in their predicted ECG age, indicated by positive mean and median difference.

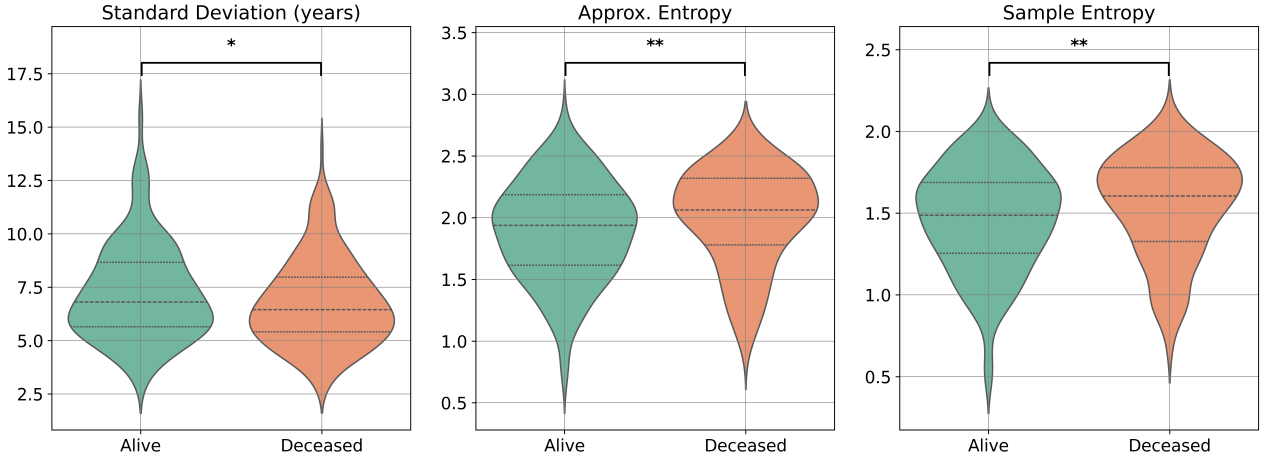

Fig. 3: Violin plots displaying the distributions of SD, Approximate Entropy, and Sample Entropy from predicted ECG age over 24 hours for matched *Alive* and *Deceased* groups. Patients were matched 1:1 by sex, NYHA class, and age within 2 years (222 deceased matched with 222 alive). Each metric was compared between Deceased and Alive with the Mann–Whitney U test. Statistical significance is indicated with * for p <0.05 and ** for p < 0.01. Deceased patients exhibited significantly lower variability (SD, p = 0.0343) and higher entropy (Approximate Entropy, p = 0.0032; Sample Entropy, p = 0.0046), suggesting increased instability in their predicted ECG age dynamics.

4 Discussion

Our findings indicate that variability and non-linear dynamics of the predicted ECG age provide subtle distinctions between alive and deceased CHF patient groups. Utilizing Frank leads suited for 24-hour recordings and excluding noise artifacts via SNR thresholds, we observed ECG age fluctuations across the day—likely attributable to diurnal variability and patient activities. However, these fluctuations were not time-synchronous between recordings, so the exact timing (e.g., morning vs. evening) and nature of patient activity remain unknown. Despite this limitation, distinct patterns emerged based on clinical outcome: deceased patients displayed significantly reduced variability (lower SD) accompanied by increased complexity metrics compared to matched survivors. These findings suggest a heightened degree of disorganization in cardiac electrical activity or autonomic dysregulation, potentially reflecting fibrotic or hypertrophic changes often reported in heart failure [2, 12].

While variations in predicted ECG age may stem partly from unfiltered artifacts, the consistent differences among the deceased cohort indicate that these fluctuations could reflect pathophysiological processes rather than artifacts alone. Repeated ECG age assessments across multiple time windows may help overcome inherent ambulatory recording fluctua-

tions, boosting reliability for longitudinal patient evaluation. Indeed, previous research underscores the prognostic value of measuring variability in heart failure [12].

Although Frank leads facilitate continuous monitoring, single-time-point ECG age estimates from standard 12-lead ECGs may suffice in more controlled clinical settings [2, 5]. Future studies should validate these observations in larger prospective cohorts to elucidate whether heightened complexity in predicted ECG age directly signifies pathological remodeling or elevated cardiovascular risk in CHF management. Moreover, integrating population-based data[2, 10] can further distinguish physiological variability from disease-specific, potentially advancing ECG age as a robust non-invasive prognostic biomarker in diverse clinical contexts.

5 Conclusion

This study demonstrates the application of ECG age prediction in 24-hour Holter recordings as a dynamic biomarker for CHF patients. Our analysis reveals that metrics derived from continuous ECG age monitoring, particularly entropy-based complexity measures, could provide meaningful insights into patient outcomes. Specifically, decreased variability and increased complexity among deceased patients highlight the potential clinical value. Further validation in larger, prospective cohorts is needed to fully explore its predictive power and practical clinical utility.

Code availability:

All code is open source available here.

Acknowledgments

This work was partially funded by the DZHK (Germany center for cardiovascular diseases) within the “Postdoc Start-up Grant on advancing digital aspects” guideline (81X3300116) as well as by the Lower Saxony “Vorab” of the Volkswagen Foundation and the Ministry for Science and Culture of Lower Saxony, Germany under Grant 76211-12-1/21.

The authors gratefully acknowledge the computing time granted by the Resource Allocation Board and provided on the supercomputer Emmy/Grete at NHR-Nord@Göttingen as part of the NHR infrastructure. The calculations for this research were conducted with computing resources under the project nib00042.

References

[1] D. Zhao, Y. Wang, N. D. Wong, and J. Wang, “Impact of aging on cardiovascular diseases: from chronological observation to biological insights: Jacc family series,” *JACC: Asia*, vol. 4, no. 5, pp. 345–358, 2024.

[2] P. Hempel, A. H. Ribeiro, M. Vollmer, T. Bender, M. Dörr, D. Krefting, and N. Spicher, “Explainable ai associates ecg aging effects with increased cardiovascular risk in a longitudinal population study,” *npj Digital Medicine*, vol. 8, no. 1, p. 25, 2024.

[3] F. E. Camazzola, P. V. Schwartzmann, M. Sabedotti, R. Massuti, T. Zortea, V. Chen, A. C. G. Maggi, F. F. d. Souza, A. d. S. Cardoso, and L. d. S. Selistre, “Comparative analysis of ecg and holter monitoring in the assessment of heart rate in heart failure with reduced ejection fraction and sinus rhythm,” *Arquivos Brasileiros de Cardiologia*, vol. 121, p. e20230771, 2024.

[4] A. Chetran, A. D. Costache, C. I. Ciongradi, S. T. Duca, O. Mitu, V. Sorodoc, C. M. Cianga, C. Tuchilus, I. Mitu, R. D. Mitea *et al.*, “Ecg and biomarker profile in patients with acute heart failure: a pilot study,” *Diagnostics*, vol. 12, no. 12, p. 3037, 2022.

[5] E. M. Lima, A. H. Ribeiro, G. M. Paixão, M. H. Ribeiro, M. M. Pinto-Filho, P. R. Gomes, D. M. Oliveira, E. C. Sabino, B. B. Duncan, L. Giatti *et al.*, “Deep neural network-estimated electrocardiographic age as a mortality predictor,” *Nature communications*, vol. 12, no. 1, p. 5117, 2021.

[6] A. Martín-Yebra, A. Bayés de Luna, R. Vázquez, P. Caminal, P. Laguna, and J. P. Martínez, “The MUSIC Database: Sudden Cardiac Death in Heart Failure Patients,” in *Computing in Cardiology*, vol. 51, 2024.

[7] G. Uijen, A. Van Oosterom, and R. Van Dam, “The relationship between the 12-lead standard ecg and the xyz vector leads,” in *Proceedings of the X14th International Congress on Electrocardiology*, 1988, pp. 301–307.

[8] P. W. Macfarlane, A. Van Oosterom, O. Pahlm, P. Kligfield, M. Janse, and J. Camm, *Chapter 11.5: Derived 12-Lead ECG Systems; book: Comprehensive electrocardiology*, 2nd ed. Springer Science & Business Media, 2010.

[9] T. Bender, P. Gemke, E. Idrobo-Avila, H. Dathe, D. Krefting, and N. Spicher, “Benchmarking the impact of noise on deep learning-based classification of atrial fibrillation in 12-lead ecg,” in *Caring is Sharing–Exploiting the Value in Data for Health and Innovation*. IOS Press, 2023, pp. 977–981.

[10] L. C. C. Brant, A. H. Ribeiro, M. M. Pinto-Filho, J. Kornej, S. R. Preis *et al.*, “Association between electrocardiographic age and cardiovascular events in community settings: The framingham heart study,” *Circulation: Cardiovascular Quality and Outcomes*, vol. 16, no. 7, p. e009821, 2023.

[11] J. S. Richman, D. E. Lake, and J. R. Moorman, “Sample entropy,” in *Methods in enzymology*. Elsevier, 2004, vol. 384, pp. 172–184.

[12] A. Cacciotti, C. Pappalettera, F. Miraglia, L. Valeriani, E. Judica, P. M. Rossini, and F. Vecchio, “Complexity analysis from eeg data in congestive heart failure: A study via approximate entropy,” *Acta Physiologica*, vol. 238, no. 2, p. e13979, 2023.