

Utilizing longitudinal data in assessing all-cause mortality in patients hospitalized with heart failure

Robert Herman^{1,2,3*} , Marc Vanderheyden⁴, Boris Vavrik¹, Monika Beles⁴, Timotej Palus¹, Olivier Nelis⁴, Marc Goethals⁴, Sofie Verstreken⁴, Riet Dierckx⁴, Martin Penicka⁴, Ward Heggermont⁴ and Jozef Bartunek^{4*}

¹Powerful Medical, Bratislava, Slovak Republic; ²Sigmund Freud University, Vienna, Austria; ³Department of Advanced Biomedical Sciences, University of Naples Federico II, Naples, Italy; and ⁴Cardiovascular Center, OLV Hospital, Aalst, Belgium

Abstract

Aims Risk stratification in patients with a new onset or worsened heart failure (HF) is essential for clinical decision making. We have utilized a novel approach to enrich patient level prognostication using longitudinally gathered data to develop ML-based algorithms predicting all-cause 30, 90, 180, 360, and 720 day mortality.

Methods and results In a cohort of 2449 HF patients hospitalized between 1 January 2011 and 31 December 2017, we utilized 422 parameters derived from 151 451 patient exams. They included clinical phenotyping, ECG, laboratory, echocardiography, catheterization data or percutaneous and surgical interventions reflecting the standard of care as captured in individual electronic records. The development of predictive models consisted of 101 iterations of repeated random subsampling splits into balanced training and validation sets. ML models yielded area under the receiver operating characteristic curve (AUC-ROC) performance ranging from 0.83 to 0.89 on the outcome-balanced validation set in predicting all-cause mortality at aforementioned time-limits. The 1 year mortality prediction model recorded an AUC of 0.85. We observed stable model performance across all HF phenotypes: HFpEF 0.83 AUC, HFmrEF 0.85 AUC, and HFfrEF 0.86 AUC, respectively. Model performance improved when utilizing data from more hospital contacts compared with only data collected at baseline.

Conclusions Our findings present a novel, patient-level, comprehensive ML-based algorithm for predicting all-cause mortality in new or worsened heart failure. Its robust performance across phenotypes throughout the longitudinal patient follow-up suggests its potential in point-of-care clinical risk stratification.

Keywords Heart failure; Machine learning; Mortality prediction; Risk stratification; Big data analysis; Precision medicine

Received: 27 November 2021; Revised: 14 April 2022; Accepted: 31 May 2022

*Correspondence to: Robert Herman, Powerful Medical, Bratislavská 81/37, Šamorín, Slovak Republic. Phone: +43 676 374 6517. Email: robert@powerfulmedical.com; Jozef Bartunek, Cardiovascular Center, OLV Hospital, Moorselbaan 164, 9300 Aalst, Belgium. Phone: +32 53 72 4439. Email: jozef.bartunek@olvz-aalst.be

Introduction

Heart failure (HF) is a heterogeneous syndrome with a complex pathophysiology and broad range of phenotypes associated with poor clinical outcomes.¹ Several biomarkers and clinical risk scores have been introduced to aid in prognostic stratification.² However, they often fail to provide optimal patient-level precision.³ They utilize data from a single observational timepoint and do not capture the entire care pathway with variations in individual patient management. This reductionist approach is also dismissive of ongoing adaptations in patient care with dynamic changes in patient phenotype along heart failure progression or—in some cases—partial regression. To yield in-depth predictive insights within these

cycles, it is appealing to utilize a continuum of all captured electronic health record parameters comprehensively reflecting the entire patient journey during care delivery. Such multivariate time-series dataset contains sizeable and necessary granularity for patient-level prognostication, being of exceptional relevance in the setting of new onset of worsened HF. However, it is usually aggregated into a static-time dataset due to multi-dimensionality, heterogeneity and irregularity making it challenging to preprocess and interpret.

Machine learning (ML), artificial intelligence (AI) and deep learning⁴ enable an efficient use of the ‘big data’ from electronic health records. Various AI-based segmentations, image reconstructions, and automated detection algorithms⁵ have been shown to improve diagnostic workflows. Likewise, the use of

ML showcased improvement of the empirical risk models in patient selection for cardiac resynchronization⁶ or acute myocardial injury.⁷

We sought to develop a dynamic and comprehensive ML-based predictive algorithm for all-cause mortality in a cohort of patients hospitalized with new onset or worsened heart failure. Our novel findings are instrumental in showcasing the potential of ML predicting time-dependent mortality within a point-of-care Powerful Medical (PM) - Aalst HF system, reflecting the multidimensional care pathway at individual patient level.

Methods

Study population

We have selected 2449 patients admitted with the primary diagnosis of a new onset or worsened heart failure between years 2011–2017 at the Cardiovascular Center, OLV Hospital Aalst in Belgium. From each patient a total of 902 clinically relevant parameters, routinely collected per standards of

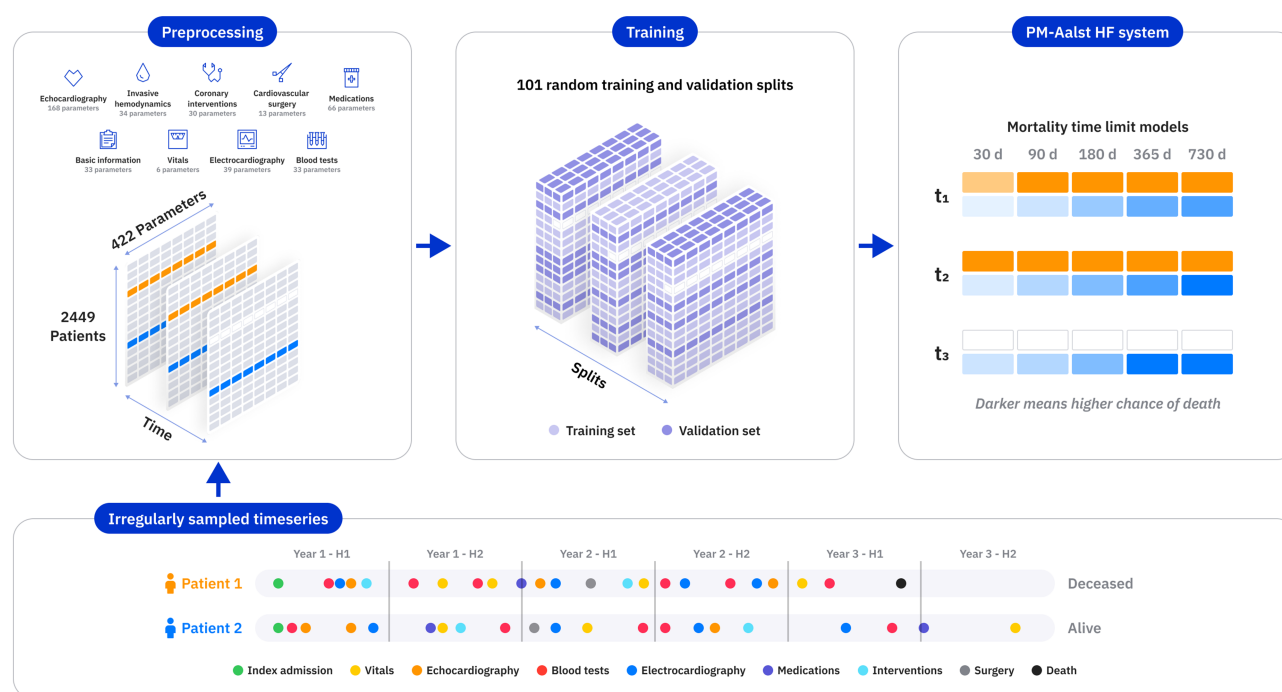
care, were included into the analysis and development of the predictive prognostic models. Besides demographics, they included vital parameters, laboratory tests, ECG recordings, medications, echocardiography parameters, findings from invasive haemodynamics, coronary or electrophysiology exams, surgical and percutaneous cardiac and coronary interventions. A graphic representation of different types of input data for the preprocessing stages is shown in *Figure 1*. Parameters were gathered during the index hospitalization and continuously updated throughout the clinical follow-up at outpatient visits or re-admissions. A time series diagram visualizing collection of parameters including prognostic outcomes of two sample patient's pathways of care can be observed in *Figure 1*. This retrospective study was approved by the local ethics committee for human research and complied with the Declaration of Helsinki.

Data pre-processing

Raw anonymized data from 13 tables were merged to create a large time-series dataset of all ambulatory and in-hospital contacts for each patient. The dataset was aggregated to reflect a

Figure 1 ML methodology in utilizing electronic records over time reflecting longitudinal clinical care. Bottom panel shows an example of two irregular time series of exams and endpoints for two different patients. Left panel shows the dimensionality of the preprocessed dataset created from different sources of data gathered in the electronic health records of both patients. In the centre, model training with 101 cross validation splits of training and validation sets. On the right, recalibrated predictions of the PM-Aalst HF system are shown at three different follow-up patient contacts for both example patients. See text for more details.

Machine Learning for Mortality Prediction throughout Standard of Care



daily granularity. Parameters often missing and occurring in <0.15% of patients were deleted. Missing values between observations were imputed using the last observation carried forward (LOCF) method.⁸ Where no prior results were available, missing values were filled with the parameter's median.

The final pre-processing stage included the encoding of all nominal categorical variables, such as the HF phenotypes, into binary columns.⁹ LV and RV functional assessment, as well as presence or absence of LV hypertrophy were discretized using categorical values according to the recommendations for cardiac chamber quantification by echocardiography in adults.¹⁰ This option was chosen to avoid confounding effect of interindividual variations when using continuous values for these parameters. Furthermore, all continuous parameters extracted from the ECG recordings, such as cardiac axis, PR interval, or QRS duration have been standardized to have a mean of 0 and standard deviation of 1. For other parameters, we have used the original unscaled values.

Upon pre-processing and deletion of parameters with high missingness (occurring in <0.15% of patients), the final dataset contained 422 clinically relevant input parameters and the primary endpoint, all-cause mortality, represented by a continuous variable 'days until death'. For all patients who remained alive, exams recorded close to the data collection date occurring within the respective mortality time limit were deleted ensuring that only exams for which a definite outcome has been established are used in the analysis. The last pre-processing task was a sigmoid transformation (Supporting Information, Figure S1) of the output variable days until death based on the mortality time limit (30, 90, 180, 365, or 730 days) reflected by the following function:

$$\text{output_transformation(days)} = \frac{1}{1 + e^{-\left(\frac{\text{days}}{\text{time_limit}} - 1\right)}}$$

All ML models for mortality, were trained to predict the transformed continuous variable, transformed *days until death* resulting from the equation above.

Development of models

In total, the PM-Aalst HF prognosis prediction system consists of five multivariate linear regression models each predicting mortality at a different time limit (30 day, 90 day, 180 day, 1 year, and 2 year). For each of these predictive models, the same methodology has been applied as follows. Model development began with a pruning phase, which ranked 422 input features remaining after the last data pre-processing stage based on their significance. The dataset was then randomly split into a balanced train and validation set with a ratio of

3:1 respectively. In the first stage of training, the process was repeated 10 times and for each iteration, feature importance coefficients were saved for each input feature. Finally, each mortality time limit model was trained with features, which were significant (P -value <0.05) in at least 20% of random cross validation runs.

The validation set was balanced by a down sampling of exams based on the number of exams in the minority class, upon a 3:1 train (75% patients; 75% exams) and validation (25% patients; 25% exams) split. The training process consisted of 101 iterations of repeated random subsampling splits into training and validation sets, conform to the Monte Carlo cross-validation approach.¹¹ Exams from the same patient only appeared in either the training or the validation set.

Models' evaluation and statistical analyses

Area under the curve (AUC) was chosen as the main metric for the evaluation of the performance of all developed models. For every ML-model we report the median AUC achieved on the balanced validation set. This median AUC value out of 101 cross-validation iterations corresponds to the performance of a specific model. Model performance was also assessed across different patient segments based on risk factors, patient characteristics and data collection time-intervals. HF phenotypes were defined according to the 2016 ESC Guidelines.¹² In addition to the AUC, during the model cross-validations, feature significance and influence towards the mortality endpoint was determined for all predictors remaining after the pruning process. P values <0.05 were considered statistically significant.

Results

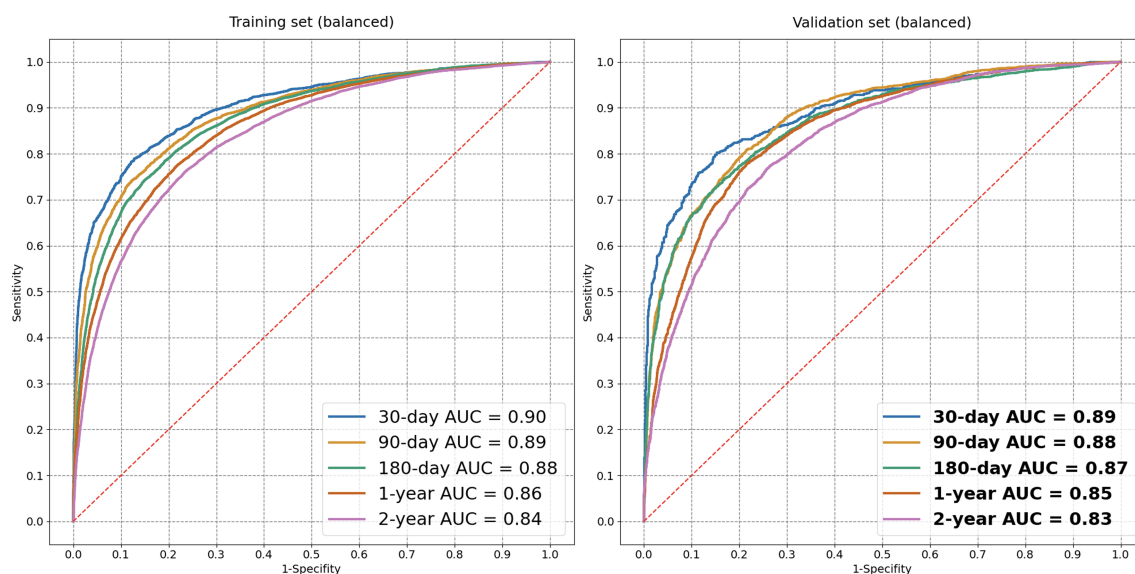
Baseline characteristics

A total of 2449 patients [73.8 ± 12 years, 1479 (60.4%) males] and 151 451 patient exams retained after the preprocessing phase have been included into the development of the mortality prediction models. The average patient follow-up reflecting either time to death or last recorded follow-up was 2.95 years (2 days to 9.5 years). A total of 857 (35%) patients died during the follow-up period. Baseline characteristics of balanced training and validation patient sets (3:1 split) were similar (Table 1). The average rate of missing data across parameters was 31%, a detailed overview is available in Supporting Information, Table S3. Baseline characteristics including distribution of HF phenotypes, risk factors and medication were comparable.

Table 1 Baseline characteristics at index hospital admission, reflecting a balanced patient population in both training and validation sets

Parameter	Category	Overall	Training set	Validation set
Patients, <i>n</i> (%)		2449	1838 (75.0)	611 (25.0)
Exams, <i>n</i> (%)		151 451	116,229 (76.7)	35,222 (23.3)
Age, mean (SD)		73.8 (12.0)	73.6 (12.0)	74.3 (11.9)
Gender, <i>n</i> (%)	M	1,479 (60.4)	1,113 (60.6)	366 (59.9)
	W	970 (39.6)	725 (39.4)	245 (40.1)
Weight, mean (SD)		77.7 (18.6)	77.7 (18.9)	77.5 (17.8)
BMI, mean (SD)		27.1 (5.5)	27.2 (5.7)	27.1 (5.1)
BSA, mean (SD)		1.9 (0.3)	1.9 (0.3)	1.9 (0.2)
SBP, mean (SD)		128.1 (24.1)	128.1 (24.0)	127.8 (24.2)
DBP, mean (SD)		73.8 (13.8)	73.6 (13.8)	74.4 (13.9)
Heart rate, mean (SD)		77.7 (21.4)	77.0 (20.8)	79.5 (23.1)
HF class, <i>n</i> (%)	HFmrEF	443 (18.1)	321 (17.5)	122 (20.0)
	HFpEF	842 (34.4)	646 (35.1)	196 (32.1)
	HFrEF	918 (37.5)	686 (37.3)	232 (38.0)
MR grade, <i>n</i> (%)	1.0	719 (29.4)	543 (29.5)	176 (28.8)
	2.0	353 (14.4)	262 (14.3)	91 (14.9)
	3.0	139 (5.7)	100 (5.4)	39 (6.4)
Diseased vessels, <i>n</i> (%)	1.0	475 (19.4)	356 (19.4)	119 (19.5)
	2.0	22 (0.9)	13 (0.7)	9 (1.5)
	3.0	30 (1.2)	19 (1.0)	11 (1.8)
Ischaemic aetiology, <i>n</i> (%)	1.0	1550 (63.3)	1167 (63.5)	383 (62.7)
Smoking, <i>n</i> (%)	Current	294 (12.0)	220 (12.0)	74 (12.1)
	None	1116 (45.6)	826 (44.9)	290 (47.5)
	Previous	448 (18.3)	345 (18.8)	103 (16.9)
	Unknown	591 (24.1)	447 (24.3)	144 (23.6)
Hypertension, <i>n</i> (%)	1.0	1105 (45.1)	833 (45.3)	272 (44.5)
Dyslipidaemia, <i>n</i> (%)	1.0	912 (37.2)	668 (36.3)	244 (39.9)
Diabetes, <i>n</i> (%)	1.0	790 (32.3)	586 (31.9)	204 (33.4)
COPD, <i>n</i> (%)	1.0	757 (30.9)	579 (31.5)	178 (29.1)
Atrial fibrillation, <i>n</i> (%)	1.0	550 (22.5)	405 (22.0)	145 (23.7)
LBBB, <i>n</i> (%)	1.0	225 (9.2)	172 (9.4)	53 (8.7)
Beta-blockers, <i>n</i> (%)	1.0	1929 (78.8)	1454 (79.2)	475 (77.6)
Diuretics, <i>n</i> (%)	1.0	1966 (80.3)	1484 (80.8)	482 (78.8)
Potassium sparing diuretics, <i>n</i> (%)	1.0	1932 (78.9)	1453 (79.1)	479 (78.3)
RAS agents, <i>n</i> (%)	1.0	1508 (61.6)	1124 (61.2)	384 (62.7)

BMI, body mass index; BSA, body surface area; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; HF, heart failure; HFpEF, HF with preserved ejection fraction; HFmrEF, HF with mid-range ejection fraction; HFrEF, HF with reduced ejection fraction; LBBB, left bundle branch block; MR, mitral regurgitation; RAS, renin-angiotensin system; SBP, systolic blood pressure.

Figure 2 Receiver operating characteristic curves (ROC) for individual ML models predicting 30 day, 90 day, 180 day, 1 year, and 2 year all-cause mortality. Left panel: training set. Right panel validation set. AUC, area under curve.

All-cause mortality prediction

Figure 2 shows receiver operating curves (ROC) for all-cause mortality at 30 day, 90 day, 180 day, 1 year, and 2 year. Training and validation models predicting all-cause mortality at the respective time limits were comparable, no model overfitting was observed. The validation set performance ranged from 0.83 to 0.89 as shown in Figure 2B. The 1 year mortality prediction model, predicting the primary endpoint has recorded an AUC score of 0.85.

Performance of the primary model was tested across HF phenotypes and patient segments based on baseline characteristics. As shown in Figure 3, we observed stable model per-

formance across all HF phenotypes: HFpEF 0.83 AUC, HFmrEF 0.85 AUC and HFrEF 0.86 AUC, respectively. The model yielded comparable predictive power regardless of the risk factors or disease modifiers including diabetes, hypertension, presence of secondary mitral regurgitation >2+, renal insufficiency and atrial fibrillation. Performance was more favourable in younger patients under 65 years versus older patients over 85 years with 0.89 AUC and 0.75 AUC, respectively.

We have also tested the performance of the ML model predicting 1 year mortality at different time points of data collection after index admission. Shown in Figure 4, model performance predicting 1 year mortality improved with data accumulation from patient contacts over time. Predictive

Figure 3 Model performance in predicting one-year all-cause mortality in different patient subsets according to risk factors, comorbidities or heart failure phenotype. Abbreviations as in Table 1 and Figure 2.

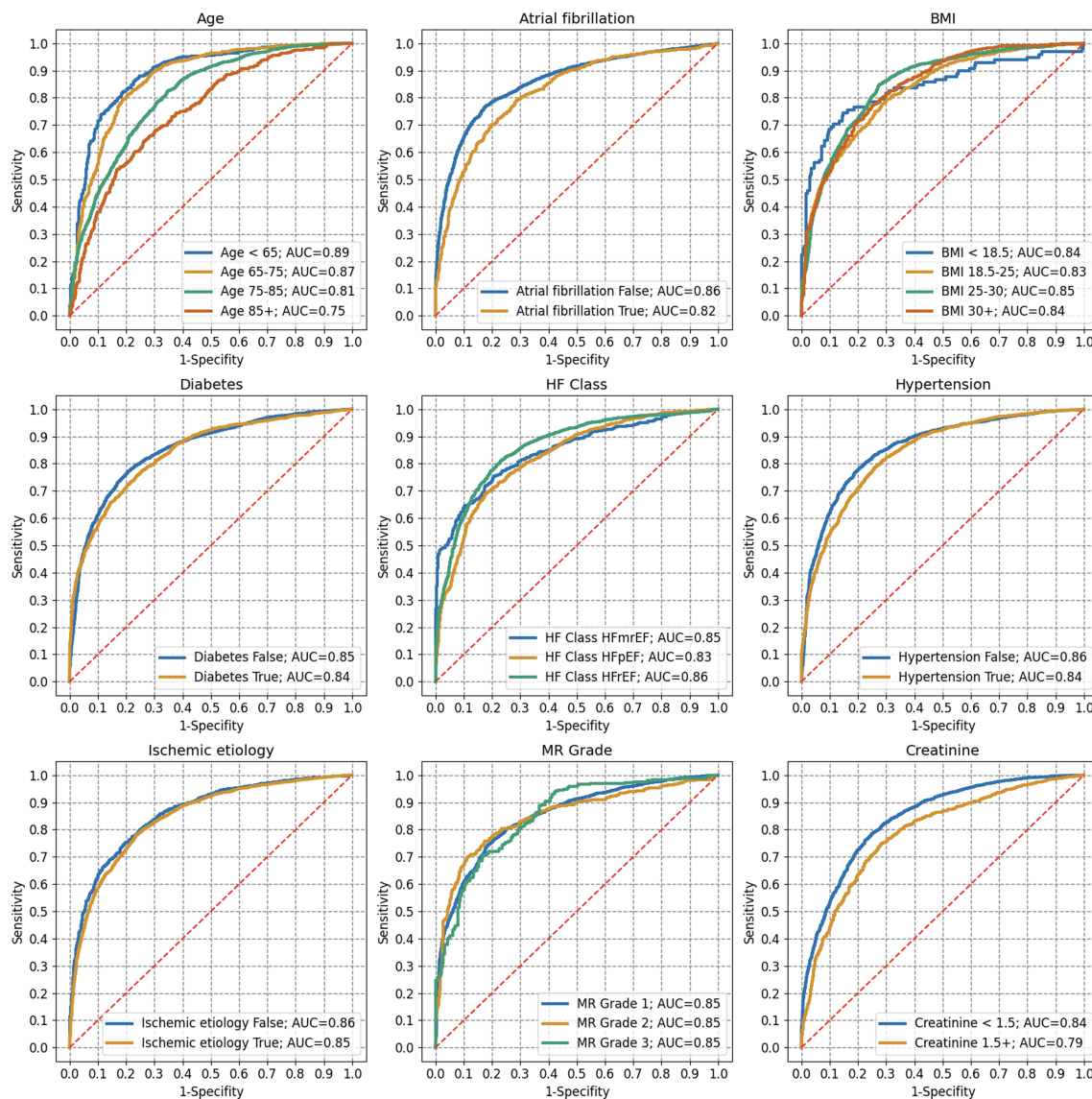
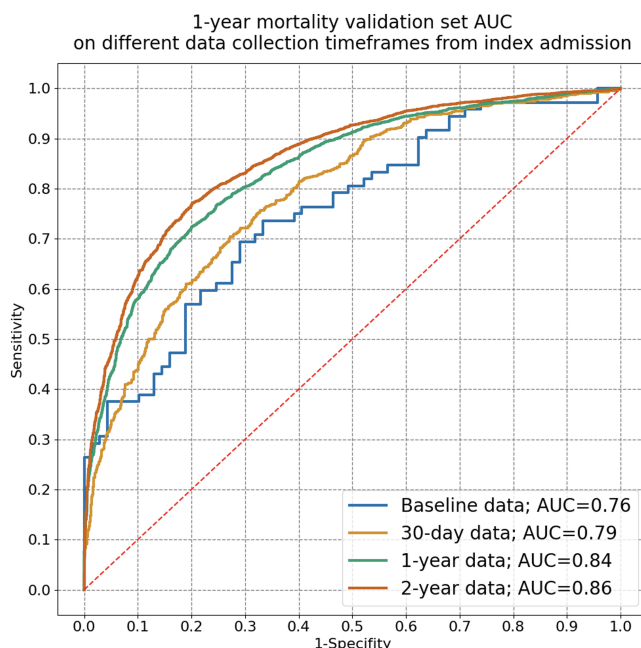


Figure 4 One year mortality model AUC on validation set showcasing performance on different data collection timeframes from index hospital admission.



performance on baseline data, collected within the index hospitalization, yielded an AUC of 0.76, compared with 0.86 when predicting on all longitudinal data gathered over follow-up of 2 years after index admission.

As shown in *Figure 5*, the 1 year mortality prediction ML model yielded higher performance compared with either Seattle Heart Failure or MAGGIC risk score in the entire validation set or as compared with Semmelweis CRT score on the subset of CRT patients.

Significant features predicting mortality

Upon analysing the final ML models, 40 features were identified as most significant for all-cause mortality at individual time limits. The full list of features, including their coefficients for reproducibility of the models is available in the Supporting Information, *Tables S1* and *S2*. *Figure 6* shows a heatmap diagram of these features with their weighted influence on the time-related model represented by the colour intensity. High recorded values for features coloured in blue (e.g. high haemoglobin or absence of hypochloraemia) correspond to a lower probability of death, while an increase in red features (e.g. higher Age or longer hospitalizations) represent a higher chance of death. The darker the colour, the greater feature impact on mortality for given time limit. Individual models for each time limit included predictive features

derived from patient demographics, vitals, laboratory tests reflecting control of anaemia, liver and kidney function as well as minerals. ECG parameters included features reflecting atrio-ventricular activation and repolarization. From Doppler echocardiography data, features reflected concomitant aortic and mitral valve disease, LV filling pattern and function. For the primary model predicting 1 year mortality, 10 most significant ($P < 0.001$) predictors included age, haemoglobin, any hospitalization multiplied by length of stay, severely reduced systolic LV function, NT-proBNP, hyperlipidaemia, urea, chloride, WBC count and maximum transaortic valve pressure gradient. A PM-Aalst HF mortality calculator is available for external use on independent datasets in the supporting information.

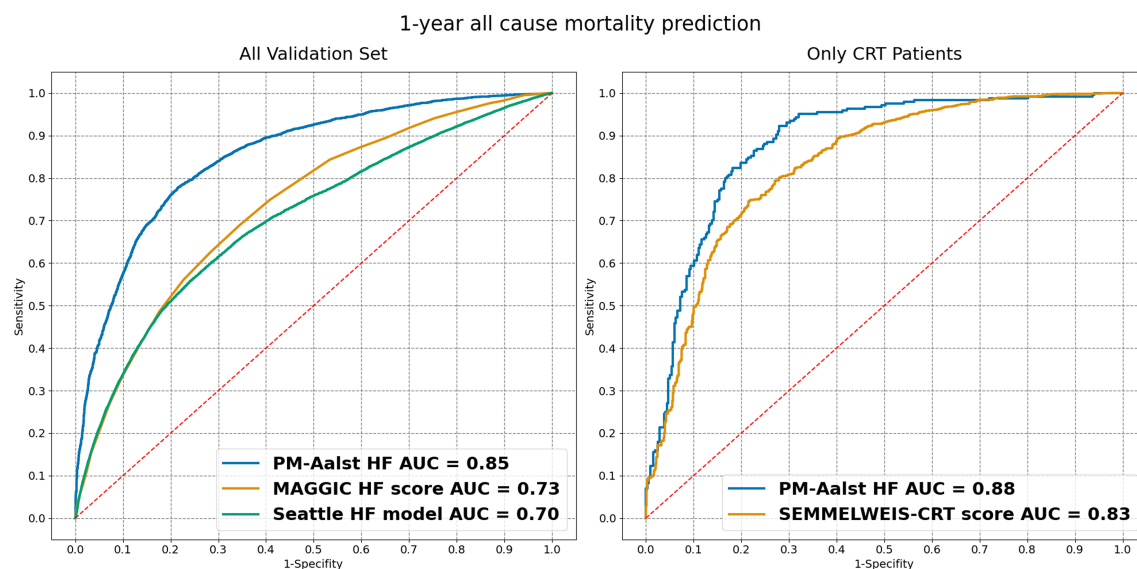
Discussion

The present study investigates the potential of machine learning in predicting mortality using a continuum of digitally recorded parameters reflecting real-world care pathways in patients initially hospitalized with a new onset or worsened HF. A novel PM-Aalst HF system consisting of multiple ML models yielded robust performance in predicting all-cause 30 day, 90 day, 180 day, 1 year, and 2 year mortality. Performance was consistent across heart failure phenotypes and disease modifiers. We have identified networks of significant features within each model that are pointing at congruent pathophysiological mechanisms impacting the mortality risk at individual time limits. The model appears to show enhanced performance when predicting on data accumulated over time capturing trends and patterns in medical records during the follow-up.

Circulating biomarkers and clinical risk scores are often suboptimal for patient-level risk stratification. Besides the heterogeneity of heart failure phenotypes, this is likely related to insufficient multimodality of individual risk scores and their static nature, relying mainly on conventional statistical analyses predicting outcome at baseline or single observation timepoints. Machine learning and artificial intelligence may efficiently dissect the big data from electronic health records and identify traits linking predictive factors to disease state and its outcome in a comprehensive way.¹³ Their use has been recently showcased by the ML-based Semmelweis CRT score in the HF subset of patients undergoing cardiac resynchronization.

Our approach is innovative as we seized the entirety of digital health care records reflecting individual patient management from the initial contact at index hospitalization longitudinally throughout the care pathway. This approach yielded a large dataset capturing disease evolution and modifications in the clinical management over time. We selected patients hospitalized with a new onset or worsened heart failure in a tertiary centre with a dedicated heart failure clinic. Utilizing

Figure 5 Head-to-head comparison of PM-Aalst HF 1 year mortality prediction model as compared to other risk scores. Left panel shows comparison on full validation set. Right panel shows model performance only on the subset of CRT implanted patients to ensure fair comparison with the SEMMELWEISS-CRT score.



more than 400 clinically relevant parameters over a follow-up till 9.5 years, the presented PM-Aalst HF system identified a set of 40 most significant predictors of 1 year mortality with a robust performance across heart failure phenotypes, risk factors and disease modifiers such as gender, obesity, diabetes, renal insufficiency or ischemic aetiology. Predictive performance was higher in patients under 65 years as compared with elderly patients (>85 years). Likewise, longitudinal data accumulation over time enhanced performance of 1 year mortality model compared to initial evaluation at initial index admission. The system appears to provide higher predictive performance as compared to traditional risk scores^{14–16} or similar ML-based approaches in subset of HF patients undergoing CRT.^{6,17} Monitoring over 400 parameters over time and re-calibrating mortality prediction at any patient contact, it can serve as a point-of-care approach in personalized care pathways and risk stratification.

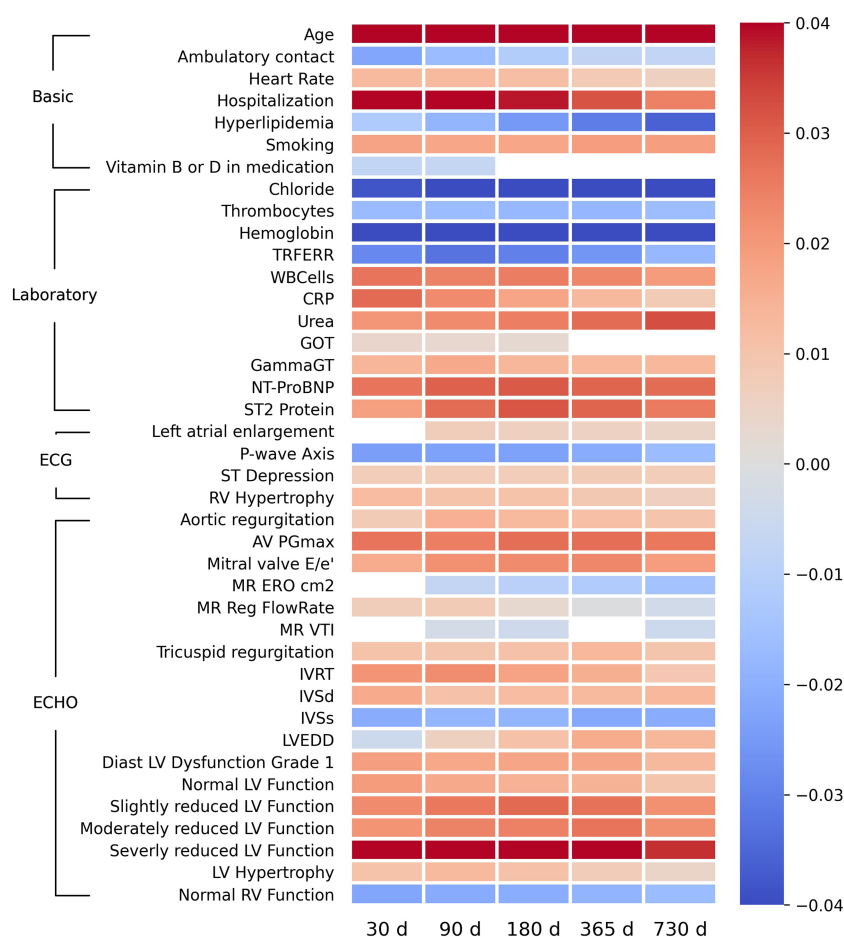
Our findings also showcase the potential of non-selective machine learning methodology to identify a network of features acting congruently in one system to predict mortality at individual time limits. Besides reinforcing the relevance of expected features such as age, hospitalization and its length or established biomarkers such as NT-proBNP and ST2 protein, it points at the relevance of proactive ambulatory patient management within the heart failure clinic in mitigating the patient's risk. Others provide insights into potential interplay of parameters reflecting inflammation, liver and renal function in affecting the outcome. Here, higher chloride levels emerged as strongly correlating with reduced mortality risk pointing at its relevance in patient's surveillance.¹⁸ Likewise,

hyperlipidaemia¹⁹ as a risk factor was associated with a protective effect on mortality. ML models also confirm the relevance of abnormal ECG repolarization, abnormal Doppler filling pattern or right sided function as well as aortic or mitral valve disease. Features reflecting LV structure and regional and global function complemented the overall traits underlying the model performance. It is of note that all patients were under optimal medical regimen in given time period of standard of care and yet use of vitamin B and D appeared to contribute by lowering the mortality risk in short term. These notions suggest that machine learning approaches may be instrumental in providing comprehensive insights into pathophysiological traits underlying diseases states and potentially reveal new insights into the mechanisms of their progression.

Following methodological aspects should be noted. In the developmental phase, we tested multiple preprocessing strategies and their effect on the performance metric as reflected by area under the curve. In this approach, days until death appeared superior to the use of categorical variables for outcome classification. As shown in Supporting Information, *Figure S1*, transforming this outcome variable using a sigmoid function has forced the models to detect differences between data points adjacent to the observed mortality time limit with higher distinction.

Several limitations are to be acknowledged. Our study cohort represents a single center experience from a tertiary centre with a dedicated heart failure unit and findings needs to be validated in an external patient cohort. The generation of a baseline parameter balanced third split, representing a testing data set to provide an unbiased evaluation was not

Figure 6 Heatmap of 40 most significant ($p < 0.05$) features predicting all-cause mortality for respective time limits with their weighted influence. Blue color relates to the lower probability of death, the red color indicates higher death probability. Color intensity corresponds to the weighted influence on mortality. ECG, Electrocardiography parameters; ECHO, Echocardiography parameters; TRFERR, Transferrin; WBC, White blood cells; CRP, C-Reactive protein; GOT, Aspartate Aminotransferase; RV, Right Ventricle; AV, aortic valve; PG, pressure gradient; MR, Mitral regurgitation; ERO, Effective Regurgitant Orifice; VTI, Velocity time integral; IVRT, Isovolumic relaxation time; IVS, Interventricular septum thickness; LVEDD, Left Ventricular Dimension in diastole; LV, Left Ventricle.



possible due to the limited size of the included patient population. In this regard, model overfitting was mitigated by subsequent random cross-validation. Our model is also based on electronic data recordings and model performance needs to be addressed in other settings of care and in all-around patients with heart failure. Our comprehensive model is scalable and consists primarily of routine parameters gathered in the current standard of care. However, some features, such as the ST2 biomarker may not be available across different geographies of the healthcare systems. It should be noted that only NT-proBNP and ST2 biomarkers were used and potential contribution of other novel biomarkers in improving the model performance needs to be investigated.²⁰ This is of potential relevance with regard to the observed consistency of the model performance in all heart failure phenotypes. Current features identified by the model point at their synergistic in-

volvement in pathophysiologic traits underlying overall heart failure syndrome. Nevertheless, it is possible that the inclusion of other parameters, in particular biomarkers or novel strain-derived echocardiographic parameters, could provide more granularity in the refinement of models with regard to distinctive feature panels for either HF phenotype.²¹

Clinical implications and conclusions

The current study presents a novel, reproducible, ML-based risk prediction system for all-cause mortality in heart failure patients that combines a multimodality of electronic data recordings in the continuum of standard of care. Such prediction of mortality is important for either short or long-term patient-tailored decision making to optimize the care and

overall management. Deploying the in-depth granular approach in a cohort of patients hospitalized with a new onset or worsened heart failure, the models demonstrated robust predictive performance across heart failure phenotypes and background risk factors. Deeper analysis revealed the relevance of features reflecting multiple pathophysiological traits fundamental for the prognosis of a heart failure patient. The presented model also allows for external validation on independent patient datasets (see supporting information). The system's ability to recalibrate prediction at every follow-up patient contact, suggests the potential of ML-based predictive models as a point-of-care approach embedded into electronic hospital information systems. This may guide clinical risk stratification and efficiently modify personalized care pathways. The findings warrant further validation in broader patient populations including all-comers with heart failure regardless of their index presentation or other cardiovascular disease subsets. Future studies should also explore the systems potential in personalized risk prediction of other relevant outcomes including unplanned hospital readmissions or responsiveness to therapeutic interventions.

Acknowledgements

We appreciate support of the heart failure nursing team and study team in supporting the execution of the data collection.

Conflict of interest

RH is the Co-founder and Chief Medical Officer of Powerful Medical; BV, ZK and TP are employees of Powerful Medical. Other authors report no conflict of interest.

Funding

The study has been supported by VZW Cardiovascular Research Center, Aalst, Belgium.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Effect caused by sigmoid transformation of the output variable time until death showing smaller steps at time limits further away from the mortality prediction threshold. For details see methods and discussion section.

Table S1. Prediction models input feature pre-processing.

Table S2. PM-Aalst HF models feature coefficients at different time limits.

Table S3. Parameter missingness.

File S4. PM-Aalst HF mortality calculator.

References

1. Senthong V, Kirsop JL, Tang WHW. Clinical phenotyping of heart failure with biomarkers: Current and future perspectives. *Curr Heart Fail Rep*. 2017; **14**: 106–116.
2. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J, Woodward M, Patel A, McMurray J, MacMahon S. Risk prediction in patients with heart failure: A systematic review and analysis. *JACC. Heart Failure*. 2014; **2**: 440–446.
3. Nadar SK, Shaikh MM. Biomarkers in routine heart failure clinical care. *Card Fail Rev*. 2019; **5**: 50–56.
4. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep learning*. MIT press Cambridge; 2016.
5. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, Kapa S, Friedman PA. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction. *The Lancet*. 2019; **394**: 861–867.
6. Tokodi M, Schwertner WR, Kovács A, Tóser Z, Staub L, Sárkány A, Lakatos BK, Behon A, Boros AM, Perge P, Kutya V, Széplaki G, Geller L, Merkely B, Kosztin A. Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: The SEMMELWEIS-CRT score. *Eur Heart J*. 2020; **41**: 1747–1756.
7. Than MP, Pickering JW, Sandoval Y, Shah AS, Tsanas A, Apple FS, et al. Machine learning to predict the likelihood of acute myocardial infarction. *Circulation*. 2019; **140**: 899–909.
8. Kenward M. The handling of missing data in clinical trials. *Clin Invest*. 2013; **3**: 241–250.
9. Draper NR, Smith H. *Applied regression analysis*. John Wiley & Sons; 1998.
10. Lang RM, Badano LP, Mor-Avi V, Afzal J, Armstrong A, Ernande L, Flachskampf FA, Foster E, Goldstein SA, Kuznetsova T, Lancellotti P, Muraru D, Picard MH, Rietzschel ER, Rudski L, Spencer KT, Tsang W, Voigt JU. Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *European Heart Journal-Cardiovascular Imaging*. 2015; **16**: 233–271.
11. Dubitzky W, Granzow M, Berrar DP. *Fundamentals of data mining in genomics and proteomics*. Springer Science & Business Media; 2007.
12. Oeing C, Tschöpe C, Pieske B. The new ESC guidelines for acute and chronic heart failure 2016. *Herz*. 2016; **41**: 655–663.
13. Ayala Solares JR, Diletta Raimondi FE, Zhu Y, Rahimian F, Canoy D, Tran J, Pinho Gomes AC, Payberah AH, Zottoli M, Nazarzadeh M, Conrad N, Rahimi K, Salimi-Khorshidi G. Deep learning for electronic health records: A comparative

- review of multiple deep neural architectures. *J Biomed Inform.* 2020; **101**: 103337.
14. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, Anand I, Maggioni A, Burton P, Sullivan MD, Pitt B, Poole-Wilson PA, Mann DL, Packer M. The Seattle heart failure model. *Circulation.* 2006; **113**: 1424–1433.
 15. Pocock SJ, Ariti CA, McMurray JJ, Maggioni A, Køber L, Squire IB, Swedberg K, Dobson J, Poppe KK, Whalley GA, Doughty RN, Meta-Analysis Global Group in Chronic Heart Failure. Predicting survival in heart failure: A risk score based on 39 372 patients from 30 studies. *Eur Heart J.* 2013; **34**: 1404–1413.
 16. Chen Y, Voors AA, Jaarsma T, Lang CC, Sama IE, Akkerhuis KM, Boersma E, Hillege HL, Postmus D. A heart failure phenotype stratified model for predicting 1-year mortality in patients admitted with acute heart failure: Results from an individual participant data meta-analysis of four prospective European cohorts. *BMC Med.* 2021; **19**: 1–11.
 17. Tokodi M, Behon A, Merkel ED, Kovács A, Tóser Z, Sárkány A, Csákvári M, Lakatos BK, Schwertner WR, Kosztin A, Merkely B. Sex-specific patterns of mortality predictors among patients undergoing cardiac resynchronization therapy: A machine learning approach. *Front Cardiovasc Med.* 2021; **8**: 87.
 18. Li Z, Xing C, Li T, Du L, Wang N. Hypochloremia is associated with increased risk of all-cause mortality in patients in the coronary care unit: A cohort study. *J Int Med Res.* 2020; **48**: 0300060520911500.
 19. Horwich TB, Hamilton MA, MacLellan WR, Fonarow GC. Low serum total cholesterol is associated with marked increase in mortality in advanced heart failure. *J Card Fail.* 2002; **8**: 216–224.
 20. Woolley RJ, Ceelen D, Ouwerkerk W, Tromp J, Figarska SM, Anker SD, Dickstein K, Filippatos G, Zannad F, Metra M, Ng L, Samani N, Veldhuisen DJ, Lang C, Lam CS, Voors AA. Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction. *Eur J Heart Fail.* 2021; **23**: 983–991.
 21. Swat SA, Njoroge J, Sanchez CL, Shah SJ. Proteomic differences between heart failure with preserved ejection fraction patients and comorbidity controls. *Circulation.* 2018; **138**: A16584-A.