

Machine learning-based prediction of 1-year all-cause mortality in patients undergoing CRT implantation: validation of the SEMMELWEIS-CRT score in the European CRT Survey I dataset

Márton Tokodi ^{1,2,*}, Annamária Kosztin ¹, Attila Kovács ^{1,2},
László Gellér ¹, Walter Richard Schwertner ¹, Boglárka Veres ¹,
Anett Behon ¹, Christiane Lober³, Nigussie Bogale ⁴, Cecilia Linde ⁵,
Camilla Normand ^{6,7}, Kenneth Dickstein ^{6,8}, and Béla Merkely ^{1,*}

¹Heart and Vascular Centre, Semmelweis University, 68 Városmajor Street, 1122 Budapest, Hungary; ²Department of Surgical Research and Techniques, Semmelweis University, 68 Városmajor Street, 1122 Budapest, Hungary; ³Institut für Herzinfarktforschung, Ludwigshafen, Germany; ⁴Department of Heart Disease, Haukeland University Hospital, Bergen, Norway; ⁵Division of Cardiology, Department of Medicine, Karolinska Institutet, Stockholm, Sweden; ⁶Cardiology Division, Stavanger University Hospital, Stavanger, Norway; ⁷Department of Quality and Health Technology, University of Stavanger, Stavanger, Norway; and ⁸Institute of Internal Medicine, University of Bergen, Bergen, Norway

Received 26 February 2024; revised 23 April 2024; accepted 1 July 2024; online publish-ahead-of-print 12 July 2024

Aims

We aimed to externally validate the SEMMELWEIS-CRT score for predicting 1-year all-cause mortality in the European Cardiac Resynchronization Therapy (CRT) Survey I dataset—a large multi-centre cohort of patients undergoing CRT implantation.

Methods and results

The SEMMELWEIS-CRT score is a machine learning-based tool trained for predicting all-cause mortality in patients undergoing CRT implantation. This tool demonstrated impressive performance during internal validation but has not yet been validated externally. To this end, we applied it to the data of 1367 patients from the European CRT Survey I dataset. The SEMMELWEIS-CRT predicted 1-year mortality with an area under the receiver operating characteristic curve (AUC) of 0.729 (0.682–0.776), which concurred with the performance measured during internal validation [AUC: 0.768 (0.674–0.861), $P = 0.466$]. Moreover, the SEMMELWEIS-CRT score outperformed multiple conventional statistics-based risk scores, and we demonstrated that a higher predicted probability is not only associated with a higher risk of death [odds ratio (OR): 1.081 (1.061–1.101), $P < 0.001$] but also with an increased risk of hospitalizations for any cause [OR: 1.013 (1.002–1.025), $P = 0.020$] or for heart failure [OR: 1.033 (1.015–1.052), $P < 0.001$], a less than 5% improvement in left ventricular ejection fraction [OR: 1.033 (1.021–1.047), $P < 0.001$], and lack of improvement in New York Heart Association functional class compared with baseline [OR: 1.018 (1.006–1.029), $P = 0.003$].

Conclusion

In the European CRT Survey I dataset, the SEMMELWEIS-CRT score predicted 1-year all-cause mortality with good discriminatory power, which confirms the generalizability and demonstrates the potential clinical utility of this machine learning-based risk stratification tool.

* Corresponding authors. Tel: +361 458 68 10, Fax: +361 458 68 17, Email: merkely.study@gmail.com (B.M.); Tel: +361 458 68 10, Fax: +361 458 68 42, Email: tokmarton@gmail.com (M.T.)

© The Author(s) 2024. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Structured Graphical Abstract

Key Question

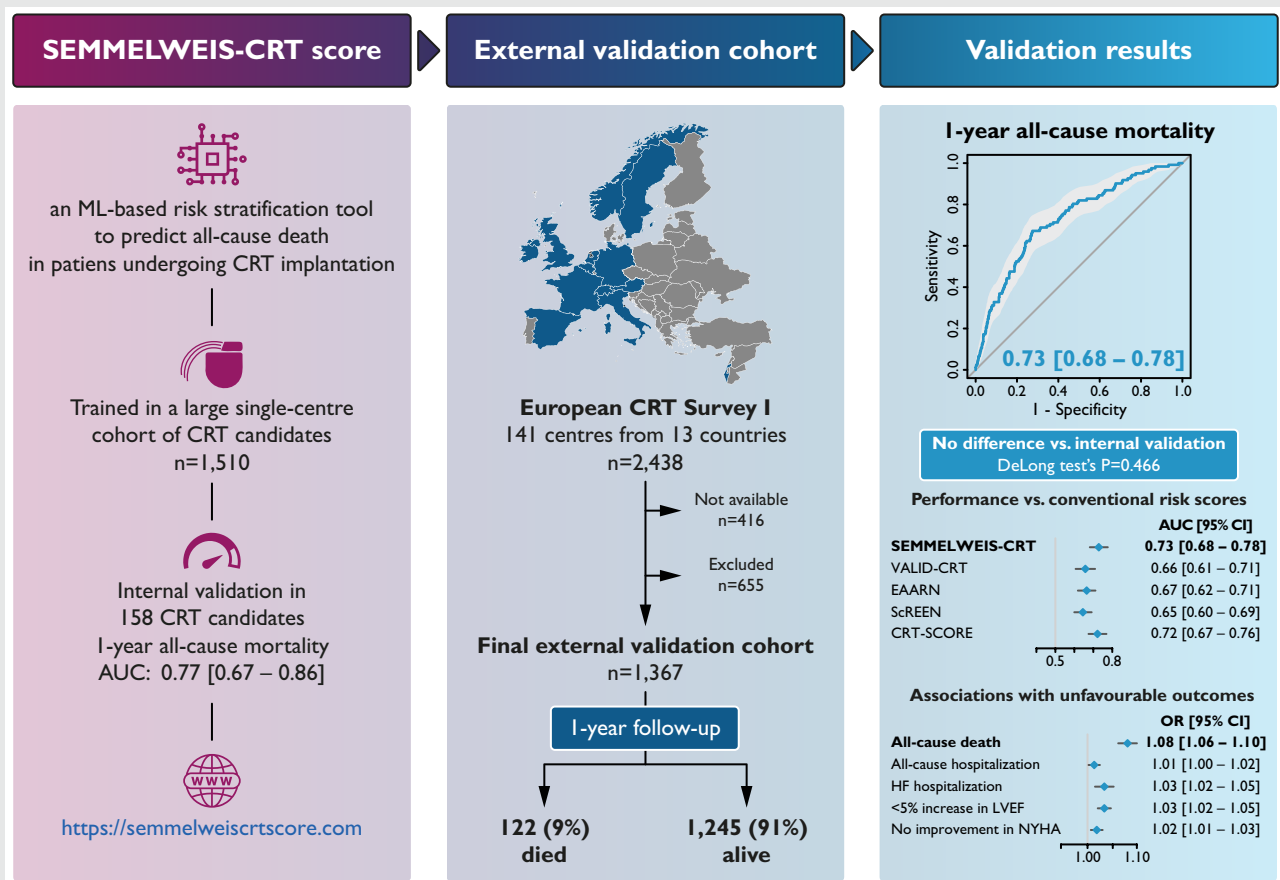
Can the machine learning (ML)-based SEMMELWEIS-CRT score, developed in a large single-centre cohort of cardiac resynchronization therapy (CRT) candidates, efficiently predict 1-year all-cause mortality in the European CRT Survey I dataset?

Key Finding

In the European CRT Survey I dataset, the SEMMELWEIS-CRT score predicted 1-year all-cause mortality with good discriminatory power, which was non-inferior to what was observed during internal validation, and it also outperformed multiple conventional statistics-based risk scores.

Take Home Message

Besides proving the generalizability of the SEMMELWEIS-CRT score, this external validation study also underscored the potential clinical utility of ML in the risk stratification of patients undergoing CRT implantation.



Keywords

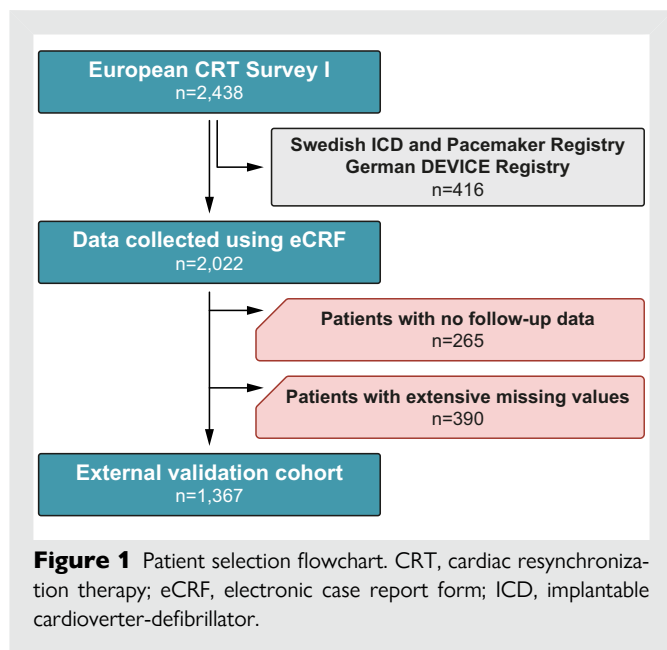
Heart failure • Cardiac resynchronization therapy • Machine learning • Risk stratification • All-cause death

Introduction

Since the first large clinical trial [MUSTIC (Multisite Stimulation in Cardiomyopathies)] showed that cardiac resynchronization therapy (CRT) improves symptoms and quality of life in heart failure patients with reduced ejection fraction and electrical dyssynchrony,¹ CRT has become an integral component of heart failure management.² Various aspects of CRT have evolved over the past decades,^{3,4} leading to a notable improvement in the prognosis of patients following CRT implantation.⁵ Nevertheless, significant variation in clinical outcomes can still be observed among patients undergoing CRT implantation,

and more than one-third of them do not experience clinical benefit despite fulfilling the guideline-defined criteria for implantation.⁶ These inconsistencies have prompted the development of machine learning (ML) models for predicting echocardiographic response and other outcomes in this patient population using data from landmark trials or retrospective observational single- or multi-centre datasets.⁷⁻¹²

In a recent paper, we presented the SEMMELWEIS-CRT score, an ML-based risk stratification tool developed using a large retrospective single-centre dataset to predict all-cause mortality in patients undergoing CRT implantation.⁹ Although it showed impressive performance during internal validation, the SEMMELWEIS-CRT score has not yet been validated in patients



from other centres. Nevertheless, external validation is pivotal to assessing the reproducibility and generalizability of ML-based prognostic models; hence, it is imperative before clinical adoption.¹³

Accordingly, we aimed to validate the SEMMELWEIS-CRT score for predicting 1-year all-cause mortality in the European CRT Survey I dataset¹⁴—a large multi-centre cohort of patients undergoing CRT implantation.

Methods

The European CRT Survey I

The European CRT Survey I was a joint initiative of the Heart Failure Association and the European Heart Rhythm Association of the European Society of Cardiology that evaluated the European implantation practice of CRT between 1 November 2008 and 30 June 2009.¹⁴ The rationale, design, and results of the Survey have been published previously.^{14–16} Briefly, the Survey enrolled 2438 patients who underwent successful CRT implantation from 141 centres in 13 European countries and collected baseline demographic, clinical, electrocardiographic, echocardiographic, laboratory, and procedure-related data.¹⁴ 9–15 months after CRT implantation, follow-up data, including information regarding outcomes (e.g. death or hospitalizations), were also acquired from 2111 (87%) patients.¹⁶

Although data were collected using electronic case report forms (eCRFs) for the majority of patients, two device registries (i.e. the Swedish Implantable Cardioverter-Defibrillator and Pacemaker Registry and the German DEVICE Registry), including CRT patients and capturing most of the information contained in the eCRF, were also merged into the European CRT Survey I dataset. Nevertheless, we were only granted access to the data collected using eCRFs but not the data from the other two registries (Figure 1).

The SEMMELWEIS-CRT score

The SEMMELWEIS-CRT score is an ML-based risk stratification tool designed to predict 1-, 2-, 3-, 4-, and 5-year all-cause mortality in patients undergoing CRT implantation.⁹ It is a random forest classifier that takes 33 numeric features as input (see [Supplementary material online, Table S1](#)) and returns probability values denoting the risk of all-cause death 1–5 years after implantation. It was trained using the data of 1510 patients who underwent successful CRT implantation at the Heart and Vascular Centre of Semmelweis

Table 1 Clinical characteristics of the entire European CRT Survey I cohort and the subset of patients used for the external validation of the SEMMELWEIS-CRT score

	The entire European CRT Survey I cohort ¹⁴ n = 2438	External validation cohort n = 1367
Age, years	70 (62–76)	71 (63–77)
Male	73%	77%
Height, cm	Not reported	170 (165–177)
Weight, kg	Not reported	78 (69–90)
BMI, kg/m ²	26 (24–29)	26 (24–30)
SBP, mmHg	Not reported	120 (110–135)
NYHA functional class		
I	2%	2%
II	20%	20%
III	70%	69%
IV	8%	9%
Ischemic HF aetiology	51%	55%
LVEF, %	27 ± 8	27 ± 7
CRT-D	73%	75%
Previously implanted PPM or ICD	26%	26%
Hypertension	Not reported	24%
Diabetes	30%	30%
Atrial fibrillation	23%	25%
QRS duration, ms	157 ± 32	156 ± 32
LBBB	68%	73%
Haemoglobin, g/dL	13.0 (11.7–14.3)	13.1 (11.8–14.3)
Serum sodium, mmol/L	139 (137–141)	139 (137–141)
Creatinine, µmol/L	103 (81–133)	106 (85–136)
GFR, mL/min/1.73m ²	Not reported	56 (42–74)
NT-proBNP, pg/mL	1740 (655–3542)	2113 (876–4455)
ACE-I	67%	65%
ARB	24%	25%
Beta-blockers	84%	84%
Diuretics	88%	87%
MRA	46%	47%
Statin	56%	56%
Antiarrhythmic medication	24%	27%

Continuous variables are presented as mean ± standard deviation or median (interquartile range), as appropriate. Categorical variables are reported as percentages to remain consistent with the reporting format of the paper describing the European CRT Survey I dataset.¹⁴

ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; BMI, body mass index; CRT, cardiac resynchronization therapy; CRT-D, cardiac resynchronization therapy defibrillator; GFR, glomerular filtration rate; HF, heart failure; ICD, implantable cardioverter-defibrillator; LBBB, left bundle branch block; LVEF, left ventricular ejection fraction; MRA, mineralocorticoid receptor antagonist; NT-proBNP, N-terminal pro-brain natriuretic peptide; NYHA, New York Heart Association; PPM, permanent pacemaker; SBP, systolic blood pressure.

University between 1 September 2000 and 31 December 2017 and was validated internally on an independent cohort of 158 patients. With areas under the receiver operating characteristic curves (AUCs) over 0.750,

Table 2 Clinical characteristics of the training and the internal and external validation cohorts

	Training cohort n = 1510	Internal validation cohort n = 158	External validation cohort n = 1367
Age, years	67 (60–74)	68 (61–74)	71 (63–77)***
Male	1141 (76)	127 (80)	1058 (77)
Height, cm	172 (165–176)	172 (167–176)	170 (165–177)
Weight, kg	80 (70–91)	80 (72–92)	78 (69–90)***
BMI, kg/m ²	27 (24–31)	27 (24–30)	26 (24–30)*
SBP, mmHg	121 (110–134)	118 (110–134)	120 (110–135)
NYHA functional class			
I	7 (1)	1 (1)	21 (2)
II	393 (33)	27 (17)	280 (20)
III	523 (44)	115 (73)	937 (69)
IV	274 (23)	15 (9)	129 (9)
Ischemic HF aetiology	767 (51)	95 (60)	714 (55)*
LVEF, %	28 (23–33)	28 (24–33)	25 (20–30)***
CRT-D	688 (46)	32 (20)	1026 (75)***
Hypertension	1055 (70)	97 (61)	319 (24)***
Diabetes	560 (37)	64 (41)	404 (30)***
Atrial fibrillation	584 (39)	48 (30)	337 (25)*
QRS duration, ms	160 (141–180)	160 (150–180)	160 (134–180)***
LBBB	1054 (70)	128 (81)	997 (73)**
Haemoglobin, g/dL	13.6 (12.3–14.8)	14.0 (13.0–15.0)	13.1 (11.8–14.3)***
Serum sodium, mmol/L	138 (136–141)	138 (136–140)	139 (137–141)***
Creatinine, µmol/L	103 (84–133)	110 (80–136)	106 (85–136)
GFR, mL/min/1.73m ²	59 (43–74)	62 (46–85)	56 (42–74)**
NT-proBNP, pg/mL	1950 (841–4468)	2608 (1377–5087)	2113 (876–4455)
ACE-I/ARB	1197 (91)	146 (95)	1187 (87)***
Beta-blockers	1165 (88)	143 (91)	1144 (84)***
Diuretics	1144 (76)	129 (82)	1187 (87)*
MRA	845 (64)	110 (70)	636 (47)***
Statin	772 (59)	97 (61)	761 (56)
Antiarrhythmic medication	394 (30)	47 (30)	363 (27)
1-year all-cause mortality	230 (15)	39 (25)	122 (9)

Continuous variables are presented as mean ± standard deviation or median (interquartile range), whereas categorical variables are reported as frequencies (percentages). The characteristics of the external validation cohort were compared with those of the training and internal validation cohorts using unpaired Student's t-test or Mann–Whitney U test for continuous variables and Chi-squared or Fisher's exact test for categorical variables, as appropriate.

Abbreviations as in [Table 1](#).

*P < 0.05 vs. the training cohort.

**P < 0.05 vs. the internal validation cohort.

the SEMMELWEIS-CRT score outperformed several conventional statistics-based risk scores, including the Seattle Heart Failure Model,¹⁷ the VALID-CRT,¹⁸ the EAARN,¹⁹ the ScREEN,²⁰ and the CRT-SCORE.²¹ The SEMMELWEIS-CRT score is publicly available at <https://semmelweisrtscore.com>.

Data pre-processing and external validation

As the European CRT Survey I reported outcomes approximately 1 year after implantation, the primary objective of the current analysis was to externally validate the SEMMELWEIS-CRT score for predicting 1-year all-cause mortality. Although the follow-up might have occurred anytime in the 9- to 15-month window following CRT implantation, it was considered a 1-year follow-up, and the survival status reported on the follow-up form was used as the label (i.e. ground truth) in our analysis. Patients with no follow-up data were excluded from the present study ([Figure 1](#)). Given that 8 of the 33 input features of the

SEMMELWEIS-CRT score (namely lymphocyte percentage, total cholesterol, uric acid, urea, other loop diuretics, thiazide diuretics, digitalis, and allopurinol) were not collected in the European CRT Survey I or could not be deduced from the collected ones (see [Supplementary material online, Table S1](#)), we had to limit our analysis to patients who did not have an excessive amount of missing value in the rest of the features. Thus, patients with missing values in any of the most important input features [i.e. age, sex, height, weight, New York Heart Association (NYHA) functional class, left ventricular ejection fraction (LVEF), left bundle branch block, serum sodium, and haemoglobin concentration] or missing values in more than 5 features besides those 8 features that were not collected in the survey were also excluded ([Figure 1](#)). The proportion of missing values in the remaining subset of patients is presented in [Supplementary material online, Table S2](#).

After applying the above-mentioned exclusion criteria, we uploaded the final external validation set to the online platform of the SEMMELWEIS-CRT score (<https://semmelweisrtscore.com>) and used the batch prediction

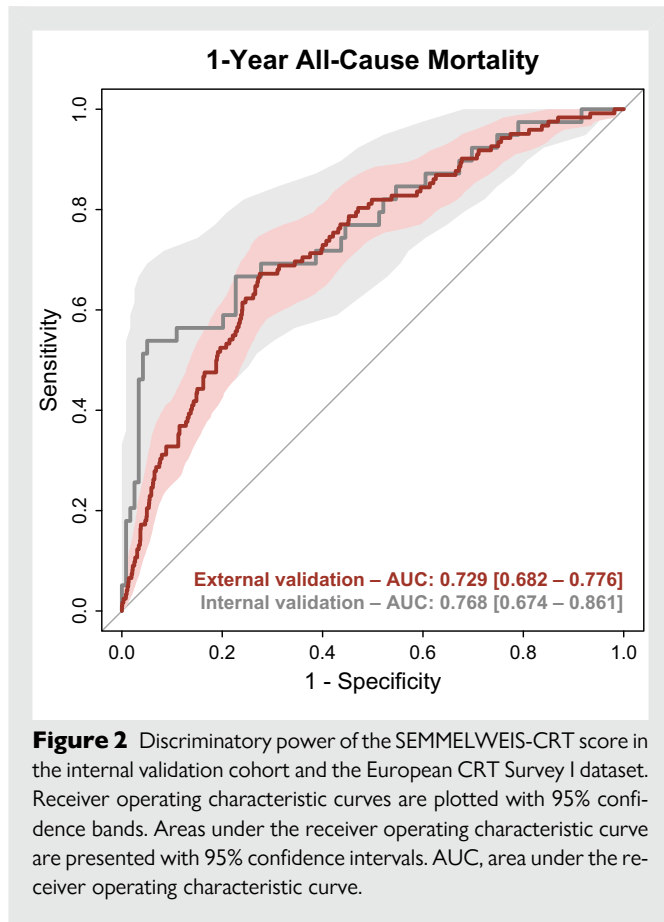


Figure 2 Discriminatory power of the SEMMELWEIS-CRT score in the internal validation cohort and the European CRT Survey I dataset. Receiver operating characteristic curves are plotted with 95% confidence bands. Areas under the receiver operating characteristic curve are presented with 95% confidence intervals. AUC, area under the receiver operating characteristic curve.

mode to evaluate the entire set at once. Importantly, we did not make any modifications to the original version of the SEMMELWEIS-CRT score; we just applied it to the external validation set as it is.⁹ Mean imputation and Z-score transformation were performed automatically on the platform using the same imputer and transformer that were previously applied to the training and the internal validation set. The performance (i.e. AUC) of the SEMMELWEIS-CRT score was compared with the discriminatory power of multiple conventional statistics-based risk scores, namely the VALID-CRT,¹⁸ the EAARN,¹⁹ the ScREEN,²⁰ and the CRT-SCORE.²¹ In addition, we assessed the associations between the SEMMELWEIS-CRT-predicted probabilities and other outcomes, including hospitalizations for any cause or heart failure and improvement in LVEF and NYHA functional class compared with baseline.

To identify the input features contributing the most to the SEMMELWEIS-CRT score's performance in predicting 1-year all-cause mortality in the analysed cohort, we computed permutation feature importances. This technique measures the importance of each input feature by calculating the decrease in the model's performance (AUC in our case) after randomly shuffling the feature's values (while keeping the other features the same as before). A feature is considered important if shuffling its values (i.e. breaking the relationship between the feature and the true outcome) results in a substantial decrease in the model's performance. In the current study, we performed permutation 20 times for each feature, and correlated features (a Pearson's correlation coefficient of >0.5 or <-0.5) were permuted jointly. The final order of importance was determined based on the median decrease in AUC.

Statistical analysis

Continuous variables are expressed as mean standard \pm deviation or median (interquartile range), whereas categorical variables are reported as frequencies (percentages), unless indicated otherwise. The performance of the SEMMELWEIS-CRT and other risk scores was quantified using AUCs,

which were then compared using DeLong tests. Brier score was also calculated to measure the accuracy of probabilistic predictions of the SEMMELWEIS-CRT. Risk ratios (RRs) were calculated using unconditional maximum likelihood estimation, and the corresponding 95% confidence intervals (CIs) were determined by normal approximation. The associations between the SEMMELWEIS-CRT-predicted probabilities and outcomes were assessed using univariable logistic regression. Odds ratios (ORs) with 95% CIs are calculated for a 0.01 increase in the predicted probability. A P -value of <0.05 was considered statistically significant. Statistical analysis was performed in R (version 4.1.2, R Foundation for Statistical Computing, Vienna, Austria).

Ethical approval

The current study was conducted in accordance with the principles of the Declaration of Helsinki, and its protocol was approved by Semmelweis University's Regional and Institutional Committee of Science and Research Ethics (approval no. 161-0/2019).

Results

Of the 2438 patients in the CRT Survey I dataset, we were granted access to the data of 2022 (83%) patients. We excluded 265 patients with no follow-up data and 390 with extensive missing values. Accordingly, the final cohort used for external validation comprised 1367 patients (Figure 1). The baseline clinical characteristics of the entire European CRT Survey I cohort and its subset analysed in the current study (i.e. the external validation cohort) are presented in Table 1. In the external validation cohort, patients were older, exhibited a lower prevalence of hypertension and diabetes, and had lower LVEF and haemoglobin concentration values and slightly higher serum sodium compared with both the training and internal validation cohorts (Table 2). Moreover, a higher proportion of patients received a CRT defibrillator (CRT-D) device, but a lower proportion used angiotensin-converting enzyme inhibitors or angiotensin receptor blockers, beta-blockers, and mineralocorticoid receptor antagonists in this cohort than in the other two (Table 2).

During the 1-year follow-up period, 122 (9%) patients died, whereas 363 (27%) were hospitalized for any cause and 111 (8%) for heart failure in the external validation cohort. Among those with available LVEF at baseline and follow-up ($n = 941$), a positive response to CRT, defined as an absolute increase of 5% or more in LVEF compared with baseline, was reported in 529 (56%) patients. Of the 1116 patients for whom NYHA functional class was reported both at baseline and follow-up, improvement in functional class was observed in 754 (65%).

In the European CRT Survey I dataset, the SEMMELWEIS-CRT predicted 1-year mortality with an AUC of 0.729 (95% CI: 0.682–0.776) (Brier score: 0.150) (Figure 2), which did not differ significantly from the AUC measured during internal validation [AUC: 0.768 (95% CI: 0.674–0.861), $P = 0.466$].⁹ Moreover, the SEMMELWEIS-CRT exhibited similarly good performance in CRT-D and CRT pacemaker (CRT-P) patients [AUC in CRT-D vs. CRT-P patients: 0.737 (95% CI: 0.682–0.792) vs. 0.705 (95% CI: 0.612–0.798), $P = 0.558$]. In addition, as shown in Figure 3A, it outperformed the VALID-CRT [AUC: 0.658 (95% CI: 0.607–0.709), $P = 0.022$], the EAARN [AUC: 0.665 (95% CI: 0.619–0.711), $P = 0.016$], the ScREEN [AUC: 0.645 (95% CI: 0.600–0.691), $P = 0.003$], but not the CRT-SCORE [AUC: 0.715 (0.668–0.762), $P = 0.541$], which was in line with our observations in the internal validation cohort.⁹

By dividing the training cohort of the SEMMELWEIS-CRT score into tertiles based on the predicted probabilities of 1-, 2-, 3-, 4-, and 5-year mortality, we have previously determined probability thresholds, allowing the stratification of CRT candidates into low-risk, moderate-risk, and high-risk groups. From these thresholds, we used those (0.283 and 0.451) determined based on the probability of 1-year mortality to stratify patients of the European CRT Survey I into low-risk (<0.283), moderate-risk (≥ 0.283 and <0.451), and high-risk groups (≥ 0.451). Patients stratified into the moderate-risk [$n = 706$ (52%)]

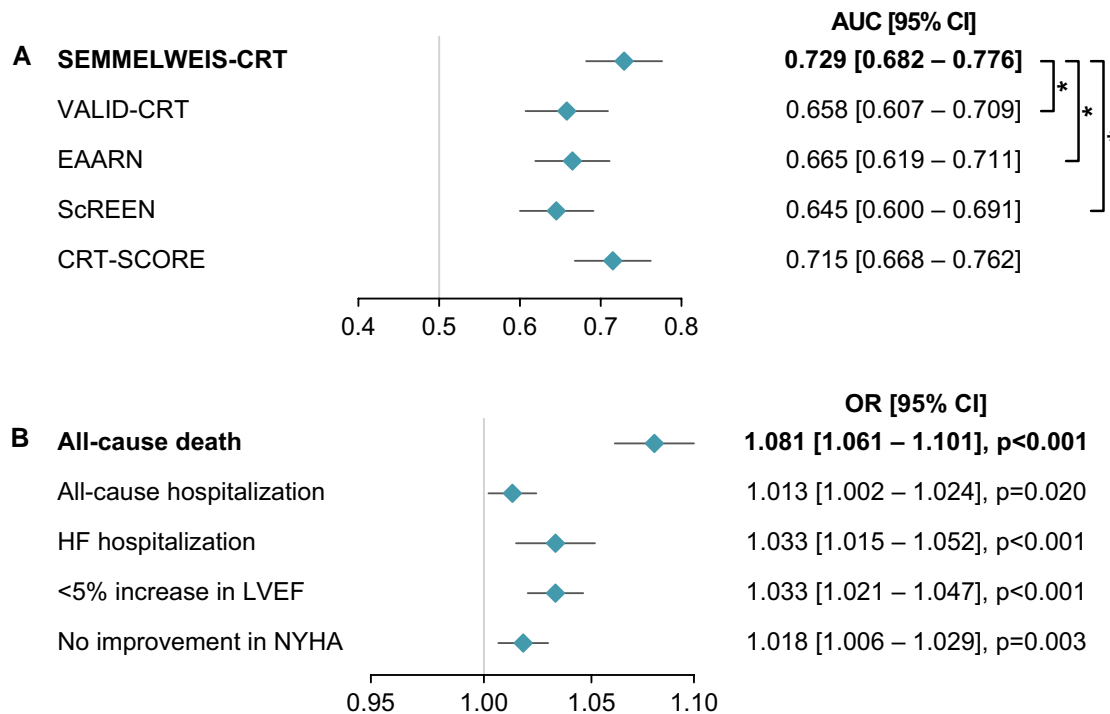


Figure 3 Performance of the SEMMELWEIS-CRT score vs. other conventional statistics-based risk scores (A) and the associations between the SEMMELWEIS-CRT-predicted probability and unfavourable outcomes (B). * $P < 0.05$ vs. SEMMELWEIS-CRT, DeLong test. Odds ratios with 95% confidence intervals are calculated for a 0.01 increase in the predicted probability. AUC, area under the receiver operating characteristic curve; CI, confidence interval; HF, heart failure; NYHA, New York Heart Association; LVEF, left ventricular ejection fraction; OR, odds ratio.

and high-risk groups [$n = 281$ (21%)] exhibited an approximately 2.3-fold [RR: 2.332 (95% CI: 1.261–4.315), $P = 0.005$] and 6.5-fold higher risk of all-cause death [RR: 6.536 (95% CI: 3.579–11.935), $P < 0.001$] during the first year following CRT implantation than those in the low-risk group [$n = 380$ (28%)]. Patients classified into the high-risk group had a 2.8-fold higher risk of all-cause death than those in the moderate-risk group [RR: 2.802 (95% CI: 1.979–3.968), $P < 0.001$].

Although the accumulation of patients who died in the first year following CRT implantation could be observed among those in whom the SEMMELWEIS-CRT predicted a higher probability of death (see [Supplementary material online, Figure S1A](#)), the calibration was suboptimal in the European CRT Survey I dataset (see [Supplementary material online, Figure S1B](#)).

In univariable logistic regression models, a higher SEMMELWEIS-CRT-predicted probability was significantly associated with an increased risk of all-cause death [OR: 1.081 (95% CI: 1.061–1.101), $P < 0.001$] and hospitalization for any cause [OR: 1.013 (95% CI: 1.002–1.025), $P = 0.020$] or heart failure [OR: 1.033 (95% CI: 1.015–1.052), $P < 0.001$] ([Figure 3B](#)). Moreover, a greater predicted probability value was also associated with a less than 5% absolute improvement in LVEF [OR: 1.033 (95% CI: 1.021–1.047), $P < 0.001$] and no improvement or worsening in the NYHA functional class compared with baseline [OR: 1.018 (95% CI: 1.006–1.029), $P = 0.003$] ([Figure 3B](#)).

In the subset of the European CRT Survey I dataset analysed in the current study, the 10 most important features, listed in descending order of importance, were creatinine, haemoglobin concentration, serum sodium, glomerular filtration rate, LVEF, age, QRS morphology, N-terminal pro-brain natriuretic peptide, sex, and systolic blood pressure ([Figure 4; Supplementary material online, Table S3](#)).

Discussion

In the European CRT Survey I dataset, the SEMMELWEIS-CRT score predicted the 1-year all-cause mortality of patients undergoing CRT implantation with good performance. Importantly, its discriminatory power (i.e. AUC) in this external dataset was non-inferior to what we previously observed during internal validation, and the SEMMELWEIS-CRT score outperformed multiple conventional statistics-based risk scores in this cohort as well. In addition, we showed that a higher predicted probability was not only associated with a higher risk of all-cause death but also with other unfavourable outcomes, such as the increased risk of all-cause or heart failure hospitalizations and the lack of improvement in LVEF and NYHA functional class.

Machine learning-based prognostic models—such as the SEMMELWEIS-CRT score—have been postulated to enhance the care of CRT patients in several ways, e.g. through optimizing patient selection and resource allocation, aiding shared decision-making, streamlining and prioritizing the referral of high-risk patients for advanced therapies, and ultimately translating into better patient outcomes.^{22,23} Therefore, developing novel and potentially better ML models might be tempting, even though most will remain research prototypes and never mature into products ready for clinical adoption.^{24,25} External validation is pivotal to combat this research waste and help bridge the gap between the development and clinical implementation of ML models.¹³ Successful external validation is also essential before prospective randomized comparative impact studies can be conducted to investigate whether applying the proposed ML models improves patient outcomes. Considering the results of the present analysis, assessing clinical effectivity by an impact study would be one of the next logical steps for the SEMMELWEIS-CRT score.

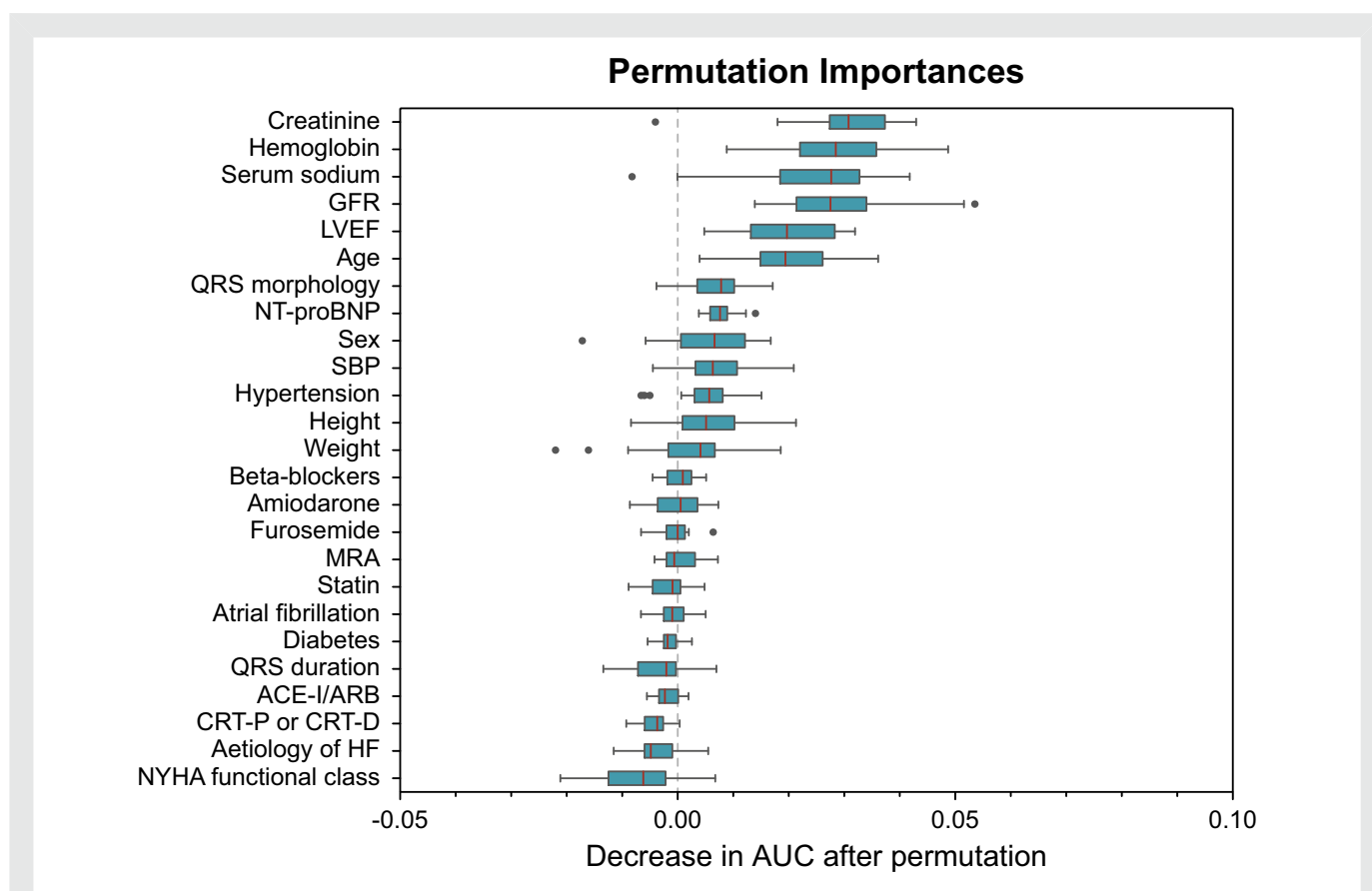


Figure 4 Contribution of each input feature to the performance of the SEMMELWEIS-CRT score in predicting 1-year mortality in the European CRT Survey I dataset. To identify the input features contributing the most to the SEMMELWEIS-CRT score's performance in the analysed subset of the European CRT Survey I dataset, we computed permutation feature importance. This technique measures the importance of each input feature by calculating the decrease in the model's performance (i.e. area under the receiver operating characteristic curve) after randomly shuffling the feature's values (while keeping the other features the same as before). A feature is considered important if shuffling its values results in a substantial decrease in the model's performance. In the current study, we performed permutation 20 times for each feature. The final order of importance was determined based on the median decrease in area under the receiver operating characteristic curve (marked by a vertical line in each box). ACE-I, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; AUC, area under the receiver operating characteristic curve; CRT-D, cardiac resynchronization therapy defibrillator; CRT-P, cardiac resynchronization therapy pacemaker; GFR, glomerular filtration rate; HF, heart failure; LVEF, left ventricular ejection fraction; MRA, mineralocorticoid receptor antagonist; NT-proBNP, N-terminal pro-brain natriuretic peptide; NYHA, New York Heart Association; SBP, systolic blood pressure.

According to a recently published systematic review, the AUCs of most ML models predicting echocardiographic response or other clinical outcomes in CRT candidates fall between 0.70 and 0.80.²⁶ The fact that both the internal and external validation AUCs of the SEMMELWEIS-CRT score for predicting 1-year all-cause mortality is within this range underscores its robustness. Nevertheless, direct comparisons with the AUCs of other ML models targeting the risk stratification of CRT candidates should be performed and interpreted with caution, as there are notable differences between models in terms of the prediction target, and the vast majority of the other models have undergone internal validation only, which may substantially overestimate their true performance.²⁶ Importantly, this rarity of external validation also increases the uniqueness and value of our current study.

Although the SEMMELWEIS-CRT score was not specifically designed to aid clinicians in choosing between the implantation of a CRT-D or a CRT-P, its predictions might still provide valuable information to support this clinical decision. Given that life expectancy is one of

the factors that should be considered during device selection,² the predicted probability of 1-year all-cause mortality—an indicator of life expectancy—can be useful in pinpointing CRT candidates in whom a CRT-D might be preferred over a CRT-P. Moreover, since the type of device is included among the input features of the SEMMELWEIS-CRT score, it is also possible to model whether implanting a CRT-D instead of a CRT-P in a given CRT candidate would result in a lower predicted probability of death by changing only the type of device while leaving all other input features unchanged. However, it is important to note that the SEMMELWEIS-CRT score was developed using data from CRT candidates in whom a CRT-D or a CRT-P was implanted based on the physicians' clinical judgment and guideline recommendations, not in a randomized fashion. Therefore, the extent of bias resulting from the non-randomized nature of the training data, as well as the impact of this bias on our model's capabilities in aiding device selection, must be evaluated in patients randomized to receive either a CRT-D or a CRT-P. Nevertheless, to eliminate this bias entirely, we would need to train a new ML model specifically for device selection using data

from prospective trials in which patients were randomized to a CRT-D or a CRT-P.

Besides the shortcomings of the European CRT Survey I discussed in previous publications,^{14,16} additional limitations related to our analysis should also be acknowledged. First, many of the input features of the SEMMELWEIS-CRT score were not collected in the European CRT Survey I, and even the collected features contained missing values. However, the fact that this missingness did not result in diminished performance compared with internal validation confirmed the robustness of the SEMMELWEIS-CRT score. Nevertheless, simplifying the model by reducing the number of input features would be desirable, which we will consider while developing the next version of our tool. Second, although a higher predicted probability was associated with an increased risk of death and other adverse outcomes, we observed suboptimal calibration in the European CRT Survey I dataset, which might be attributable to the 8 missing features. Third, although the SEMMELWEIS-CRT score was designed to predict 1- to 5-year all-cause mortality, the European CRT Survey I dataset was suitable to validate its performance only for predicting 1-year mortality. Thus, future studies should be conducted to validate the model in patients with longer follow-ups. Last, since the European CRT Survey I, which was conducted more than a decade ago, considerable advancements have occurred in the pharmacological management and device therapy of heart failure, leading to improved outcomes.⁵ Therefore, we will endeavour to validate the SEMMELWEIS-CRT score using more recent retrospectively collected datasets and through prospective studies as well.

Conclusions

In conclusion, the SEMMELWEIS-CRT score can predict 1-year all-cause mortality in patients undergoing CRT implantation with good discriminatory power. Besides proving the generalizability of our tool, the presented findings also underscore the potential clinical utility of ML in the risk stratification of these patients.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Funding

Project numbers RRF-2.3.1-21-2022-00003 and RRF-2.3.1-21-2022-00004 (MILAB) have been implemented with the support provided by the European Union. TKP2021-EGA-23 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-EGA funding scheme. TKP2021-NVA-12 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development, and Innovation Fund, financed under the TKP2021-NVA funding scheme. M.T. was supported by the New National Excellence Program (ÚNKP-23-4-II-SE-39) of the Ministry of Culture and Innovation in Hungary from the National Research, Development, and Innovation Fund. A.Kosz. was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Conflict of interest: M.T. was a former employee of Argus Cognitive. M.T. has also received consulting fees from CardioSight, outside the submitted work. A.Kosz. has received consulting fees from Medtronic and Biotronik and personal fees from Biotronik, Boehringer Ingelheim, Boston Scientific, AstraZeneca, Bayer, and Novartis, outside the submitted work. A.Kov. has received personal fees from Argus

Cognitive and CardioSight, outside the submitted work. L.G. has received lecture fees from Medtronic, Biotronik, Johnson & Johnson Medical, and Abbott, outside the submitted work. C.L. has received research support from the Swedish Heart-Lung Foundation, Swedish Royal Society of Science, Stockholm County Council, consulting fees from AstraZeneca, Roche Diagnostics, speaker honoraria from Novartis, Astra, Bayer, Vifor Pharma, Medtronic, and Impulse Dynamics and has served on advisory boards for AstraZeneca. B.M. has received personal fees from Biotronik, Boehringer Ingelheim, Abbott, AstraZeneca, and Novartis, as well as grants from Medtronic, outside the submitted work. Other authors declare that they have no conflicts of interest regarding this manuscript.

Data availability

The European CRT Survey I dataset was provided by the Institut für Herzinfarktforschung (Ludwigshafen, Germany). Access to this dataset may be requested from the Institut für Herzinfarktforschung (Ludwigshafen, Germany). The SEMMELWEIS-CRT score is publicly available at <https://semmelweisrtscore.com>. The data generated during this external validation study are available in the article and its online [Supplementary material](#) or will be shared upon reasonable request by the corresponding authors.

References

- Cazeau S, Leclercq C, Lavergne T, Walker S, Varma C, Linde C, et al. Effects of multisite biventricular pacing in patients with heart failure and intraventricular conduction delay. *N Engl J Med* 2001;**344**:873–880.
- Glikson M, Nielsen JC, Kronborg MB, Michowitz Y, Auricchio A, Barbash IM, et al. 2021 ESC guidelines on cardiac pacing and cardiac resynchronization therapy. *Eur Heart J* 2021;**42**:3427–3520.
- Ellenbogen KA, Auricchio A, Burri H, Gold MR, Leclercq C, Leyva F, et al. The evolving state of cardiac resynchronization therapy and conduction system pacing: 25 years of research at *EP Europace* journal. *Europace* 2023;**25**:euaad168.
- Leyva F, Zegard A, Patel P, Stegemann B, Marshall H, Ludman P, et al. Timing of cardiac resynchronization therapy implantation. *Europace* 2023;**25**:euaad059.
- Leyva F, Zegard A, Patel P, Stegemann B, Marshall H, Ludman P, et al. Improved prognosis after cardiac resynchronization therapy over a decade. *Europace* 2023;**25**:euaad141.
- Chatterjee NA, Singh JP. Cardiac resynchronization therapy: past, present, and future. *Heart Fail Clin* 2015;**11**:287–303.
- Kalscheur MM, Kipp RT, Tattersall MC, Mei C, Buhr KA, DeMets DL, et al. Machine learning algorithm predicts cardiac resynchronization therapy outcomes. *Circ Arrhythm Electrophysiol* 2018;**11**:e005499.
- Feeny AK, Rickard J, Patel D, Toro S, Trulock KM, Park CJ, et al. Machine learning prediction of response to cardiac resynchronization therapy: improvement versus current guidelines. *Circ Arrhythm Electrophysiol* 2019;**12**:e007316.
- Tokodi M, Schwertner WR, Kovács A, Tóser Z, Staub L, Sárkány A, et al. Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: the SEMMELWEIS-CRT score. *Eur Heart J* 2020;**41**:1747–1756.
- Howell SJ, Stivland T, Stein K, Ellenbogen KA, Tereshchenko LG. Using machine-learning for prediction of the response to cardiac resynchronization therapy: the SMART-AV study. *JACC Clin Electrophysiol* 2021;**7**:1505–1515.
- Tokodi M, Behon A, Merkel ED, Kovács A, Tóser Z, Sárkány A, et al. Sex-specific patterns of mortality predictors among patients undergoing cardiac resynchronization therapy: a machine learning approach. *Front Cardiovasc Med* 2021;**8**:611055.
- Schwertner WR, Tokodi M, Veres B, Behon A, Merkel ED, Masszi R, et al. Phenotyping and risk stratification of patients undergoing cardiac resynchronization therapy upgrade using topological data analysis. *Sci Rep* 2023;**13**:20594.
- Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2020;**14**:49–58.
- Dickstein K, Bogale N, Priori S, Auricchio A, Cleland JG, Gitt A, et al. The European cardiac resynchronization therapy survey. *Eur Heart J* 2009;**30**:2450–2460.
- CRT Survey Scientific Committee. European cardiac resynchronization therapy survey: rationale and design. *Eur J Heart Fail* 2009;**11**:326–330.
- Bogale N, Priori S, Cleland JG, Brugada J, Linde C, Auricchio A, et al. The European CRT Survey: 1 year (9–15 months) follow-up results. *Eur J Heart Fail* 2012;**14**:61–73.
- Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation* 2006;**113**:1424–1433.

18. Gasparini M, Klersy C, Leclercq C, Lunati M, Landolina M, Auricchio A, et al. Validation of a simple risk stratification tool for patients implanted with cardiac resynchronization therapy: the VALID-CRT risk score. *Eur J Heart Fail* 2015;**17**:717–724.
19. Khatib M, Tolosana JM, Trucco E, Borràs R, Castel A, Berruezo A, et al. EAARN score, a predictive score for mortality in patients receiving cardiac resynchronization therapy based on pre-implantation risk factors. *Eur J Heart Fail* 2014;**16**:802–809.
20. Providencia R, Marijon E, Barra S, Reitan C, Breitenstein A, Defaye P, et al. Usefulness of a clinical risk score to predict the response to cardiac resynchronization therapy. *Int J Cardiol* 2018;**260**:82–87.
21. Höke U, Mertens B, Khidir MJH, Schlij MJ, Bax JJ, Delgado V, et al. Usefulness of the CRT-SCORE for shared decision making in cardiac resynchronization therapy in patients with a left ventricular ejection fraction of ≤ 35 . *Am J Cardiol* 2017;**120**:2008–2016.
22. Gautam N, Ghanta SN, Clausen A, Saluja P, Sivakumar K, Dhar G, et al. Contemporary applications of machine learning for device therapy in heart failure. *JACC Heart Fail* 2022;**10**:603–622.
23. Eckhardt LL, Kalscheur MM. Machine learning in CRT outcomes: implementing the right tool for the right outcome. *JACC Clin Electrophysiol* 2021;**7**:1516–1518.
24. Seneviratne MG, Shah NH, Chu L. Bridging the implementation gap of machine learning in healthcare. *BMJ Innov* 2020;**6**:45–47.
25. Zhang A, Xing L, Zou J, Wu JC. Shifting machine learning for healthcare from development to deployment and from models to data. *Nat Biomed Eng* 2022;**6**:1330–1345.
26. Nazar W, Szymanowicz S, Nazar K, Kaufmann D, Wabich E, Braun-Dullaes R, et al. Artificial intelligence models in prediction of response to cardiac resynchronization therapy: a systematic review. *Heart Fail Rev* 2024;**29**:133–150.