

SUPPLEMENTARY DATA

Table 1 of the supplementary data. Baseline demographic, clinical and echocardiographic characteristics of patients with and without cardiac events and in the derivation cohort

	All patients (n = 558)	Endpoint (-) (n = 343)	Endpoint (+) (n = 215)	P-
<i>Age, y</i>	74 ± 13	75 ± 13	73 ± 14	.183
<i>Male sex</i>	253 (45)	163 (48)	90 (42)	.191
<i>Hypertension</i>	300 (54)	193 (56)	107 (50)	.134
<i>Diabetes mellitus</i>	101 (18)	59 (17)	42 (20)	.486
<i>Chronic obstructive pulmonary disease</i>	70 (13)	36 (11)	34 (16)	.065
<i>Previous myocardial infarction</i>	105 (19)	48 (14)	57 (27)	< .001
<i>Previous valve surgery</i>	58 (10)	36 (11)	22 (10)	.921
<i>Severe chronic kidney disease</i>	170 (31)	80 (23)	90 (42)	< .001
<i>NYHA</i>				< .001
I	126 (23)	97 (28)	29 (13)	
II	264 (47)	170 (50)	94 (44)	
III	153 (27)	71 (21)	82 (38)	
IV	15 (3)	5 (1)	10 (5)	
<i>Atrial fibrillation</i>	302 (54.1)	183 (53.4)	119 (55.3)	.645
<i>Mitral regurgitation ≥ 2+</i>	157 (28.1)	97 (28.3)	60 (27.9)	.924
<i>Right atrial longitudinal strain, %</i>	13 [8-22.2]	15 [9-25]	12 [7-19]	< .001
<i>Right atrial volume, mL/m²</i>	48 [35-66]	46 [34-62]	53 [36-76]	.002
<i>STR etiology</i>				< .001
Ventricular	299 (54)	155 (45)	144 (67)	
Atrial	259 (46)	188 (55)	71 (33)	
<i>TAPSE, mm</i>	17 ± 5	18 ± 5	16 ± 4	< .001
<i>RV end-diastolic volume, mL/m²</i>	87 ± 31	81 ± 26	96 ± 35	< .001
<i>RV ESV, mL/m²</i>	44 ± 24	39 ± 19	53 ± 28	< .001
<i>RV SV, mL/m²</i>	42 ± 13	42 ± 12	43 ± 14	.128
<i>Left ventricular ejection fraction, %</i>	49 ± 15	50 ± 14	48 ± 16	.055
<i>RV ejection fraction, %</i>	51 ± 11	53 ± 10	47 ± 12	< .001
<i>Effective RV ejection fraction, %</i>	23 ± 12	26 ± 12	19 ± 11	< .001
<i>RV FWS, %</i>	18.9 ± 7.0	20.1 ± 7.0	17.1 ± 6.7	< .001
<i>TAPSE/PASP, mm/mmHg</i>	0.45 ± 0.22	0.48 ± 0.22	0.39 ± 0.19	< .001
<i>RV FWS/PAP, %/mmHg</i>	0.50 ± 0.27	0.55 ± 0.27	0.42 ± 0.24	< .001
<i>RV forward SV/ESV</i>	0.56 ± 0.38	0.64 ± 0.38	0.44 ± 0.35	< .001
<i>PASP, mmHg</i>	43 ± 16	41 ± 14	48 ± 18	< .001
<i>Regurgitant volume, mL</i>	41 ± 21	37 ± 18	47 ± 23	< .001
<i>Effective regurgitant orifice area, cm²</i>	0.40 [0.3-0.6]	0.40 [0.3-0.5]	0.50 [0.3-0.7]	< .001
<i>Regurgitant fraction, %</i>	55 ± 21	51 ± 20	61 ± 20	< .001

ESV, end-systolic volume; RV, right ventricular; NYHA, New York Heart Association; PASP, pulmonary artery systolic pressure; STR, secondary tricuspid regurgitation; SV, stroke volume; RVFWS, right ventricular free-wall longitudinal strain; TAPSE, tricuspid annular plane systolic excursion.

Data are expressed as No. (%), mean ± standard deviation or median [interquartile range].

Table 2 of the supplementary data. Comparison of baseline demographic, clinical and echocardiographic characteristics between patients included in the derivation and validation cohorts

	Derivation cohort (n = 558)	Validation cohort (n = 200)	P
Age, y	74 ± 13	73 ± 12	.621
Male sex	242 (43)	81 (40)	.506
Hypertension	300 (54)	198 (99)	< .001
Diabetes mellitus	101 (18)	92 (46)	< .001
Chronic obstructive pulmonary disease	70 (13)	5 (2)	< .001
Previous myocardial infarction	105 (19)	16 (8)	< .001
Previous valve surgery	58 (10)	40 (20)	< .001
Severe chronic kidney disease	170 (31)	48 (24)	.085
NYHA			< .001
I	126 (23)	6 (3)	
II	264 (47)	119 (59)	
III	153 (27)	56 (28)	
IV	15 (3)	19 (10)	
Atrial fibrillation	302 (54)	175 (87)	< .001
Mitral regurgitation ≥ 2+	157 (28)	43 (21)	.076
Right atrial longitudinal strain, %	13 [8-22]	6 [4-9]	< .001
Right atrial volume, mL/m ²	48 [35-66]	78 [55-115]	< .001
STR etiology			< .001
Ventricular	417 (75)	115 (58)	
Atrial	141 (25)	85 (42)	
TAPSE, mm	17 ± 5	18 ± 6	.092
RV end-diastolic volume, mL/m ²	87 ± 31	116 ± 49	< .001
RV ESV, mL/m ²	44 ± 24	61 ± 31	< .001
RV SV, mL/m ²	42 ± 13	54 ± 21	< .001
Left ventricular ejection fraction, %	49 ± 15	59 ± 14	< .001
RV ejection fraction, %	51 ± 11	48 ± 8	< .001
Effective RV ejection fraction, %	23 ± 12	27 ± 8	< .001
RV FWS, %	18.9 ± 7.0	17.7 ± 5.6	.013
TAPSE/PASP, mm/mmHg	0.45 ± 0.22	0.39 ± 0.16	< .001
RV FWS/PAP, %/mmHg	0.50 ± 0.27	0.39 ± 0.20	< .001
RV forward SV/ESV	0.56 ± 0.38	0.97 ± 0.29	< .001
PASP, mmHg	43 ± 16	50 ± 18	< .001
Regurgitant volume, mL	41 ± 21	39 ± 24	.227
Effective regurgitant orifice area, cm ²	0.4 [0.3-0.6]	0.4 [0.3-0.6]	.587
Regurgitant fraction, %	55 ± 21	43 ± 13	< .001

ESV, end-systolic volume; RV, right ventricular; NYHA, New York Heart Association; PASP, pulmonary artery systolic pressure; STR, secondary tricuspid regurgitation; SV, stroke volume; RVFWS, right ventricular free-wall longitudinal strain; TAPSE, tricuspid annular plane systolic excursion.

Data are expressed as No. (%), mean ± standard deviation or median [interquartile range].

APPENDIX A of the supplementary data. The FUTURE 3D Study methodology

The inclusion criteria were individuals over 18 years of age with moderate or severe STR. Exclusion criteria included primary TR, cardiac implantable electronic device-related TR, previous surgical or transcatheter tricuspid valve (TV) interventions, pregnancy, poor apical acoustic windows with inadequate echocardiographic images, and lack of follow-up data. This retrospective analysis was approved by the Ethics Committee of the Istituto Auxologico Italiano, IRCCS (record #2021_05_18_13, approved on May 18, 2021). The investigation conforms to the principles outlined in the Declaration of Helsinki.

Clinical and echocardiographic variables

Clinical, demographic, and laboratory data were recorded at baseline. Clinical comorbidities included in the analysis were hypertension, diabetes, coronary artery disease, previous cardiac surgery, chronic lung disease (obstructive or restrictive), atrial fibrillation, and chronic kidney disease with creatinine > 2 mg/dL. These conditions were considered present if documented as diagnoses in the patient's medical record and were identified by International Classification of Diseases, Ninth or Tenth Revision diagnosis codes prior to the echocardiogram.

Patients underwent standard Doppler, 2-dimensional and 3DE imaging using Vivid E9/E95 systems (GE Healthcare, Chicago, IL), equipped with M5S and 4V/4Vc probes. The images were analyzed offline utilizing EchoPAC 204 (GE Healthcare, Chicago, IL) by a single experienced researcher who was blinded to the patients' medical history and follow-up data. Left ventricular (LV) volumes, LV ejection fraction, left atrial volumes, and pulmonary artery systolic pressure (PASP) were assessed according to the most recent recommendations.¹ Conventional echocardiographic parameters of RA and RV size and function were measured from the RV-focused apical view.²⁻⁴ The effective regurgitant orifice area (EROA) and regurgitant volume (RegVol) of the TV were calculated using the proximal isovelocity surface area (PISA) method corrected by the angle of tethering of the TV leaflets and the velocity of the regurgitant flow.^{5,6}

The regurgitant fraction (RegFr) was calculated as the ratio between the RegVol and the total RV stroke volume (SV).⁷

The severity of mitral regurgitation and TR was determined according to the latest guidelines.⁸ 3DE acquisitions of the RV were obtained from the RV-focused apical view using electrocardiogram gating over 4 to 6 consecutive cardiac cycles during a single breath-hold.⁹ The 4D AutoRVQ software package was used to measure RV end-diastolic (EDV) and end-systolic (ESV) volumes, and the RV ejection fraction (RVEF).⁹ The effective RVEF (eRVEF) was calculated as the ratio between the net pulmonary flow and RV end-diastolic volume, where net pulmonary flow is obtained by subtracting the STR RegVol from the total RV SV.

RVFWLS and RA longitudinal strain measurements were obtained following current recommendations.¹⁰ RV-PA coupling was estimated using the following parameters: TAPSE/PASP (mm/mmHg), RVFWS/PASP (%/mmHg), and RV forward SV/ESV.¹¹ Patients with STR were classified into atrial or ventricular phenotypes according to recent recommendations.^{12,13}

Follow-up and study endpoint

The primary study endpoint was the occurrence of death from any cause or hospitalization due to heart failure (HHF). Information about survival and hospitalization was gathered through regular telephone interviews with patients or their families, direct communication with their physicians, and reviews of their electronic medical records. HHF events were included if they lasted more than 24 hours, were confirmed by diagnostic tests, and required specific treatments. Mortality status was verified using the Social Security Death Index and death certificates as independent sources. Follow-up was completed at the time of the first event or the last check-in for patients who did not experience any events. Clinical events were adjudicated by physicians who were blinded to the patients' echocardiographic and clinical characteristics.

Appendix B of the supplementary data. Introduction to eXtreme Gradient Boosting (XGBoost) algorithm

XGBoost is an integrated machine learning (ML) algorithm based on boosted decision trees, designed to predict the value of a dependent variable. Tree-based algorithms are among the most effective ML techniques. Using an ensemble strategy, a "strong" predictive model is constructed by combining simple trees. Compared to other tree-based ensemble algorithms, XGBoost can handle missing values.

During the training phase, the model builds an additive expansion of the objective function by minimizing a loss function. Specifically, the objective function consists of two components: a loss function and a regularization term. The XGBoost algorithm expands the loss function to the second-order Taylor series and incorporates the regularization term into the optimization process. This approach helps prevent overfitting and enhances the model's generalization ability, as follows:

$$obj = \sum_i L(y_i, \hat{y}_i) + \sum_j \Omega(f_j)$$

with $i \in 1:N$, and $j \in 1:m$, where N denotes the number of samples, $L(y_i, \hat{y}_i)$ denotes the training loss function, and Ω denotes the regularization term:

$$\Omega(f_j) = \gamma T + \frac{1}{2} \lambda \omega_j^2$$

where T denotes the number of leaves, and ω_j denotes the leaf node output in each subdecision tree model. The variable γ is the L_1 penalty parameters that controls the continued splitting of leaf nodes, and λ is the penalty parameter of L_2 , which prevents the leaf nodes from being overweighted. Both parameters are constants that control the degree of regularization. Then, second order Taylor approximation can be used for optimization, so the specific objective at step j becomes:

$$obj^{(j)} = \sum_i \left[g_i f_j(x_i) + \frac{1}{2} h_i f_j^2(x_i) \right] + \Omega(f_j)$$

where g_i and h_i are defined as follows:

$$g_i = \frac{\partial L(y_i, f_{j-1}(x_i))}{\partial f_{j-1}(x_i)} \text{ and } h_i = \frac{\partial^2 L(y_i, f_{j-1}(x_i))}{\partial f_{j-1}^2(x_i)}$$

Appendix C of the supplementary data. Hyperparameter tuning

The study cohort was randomly split into a training cohort (70%), which was used to train the machine learning model and tune its parameters, and a testing cohort (30%), which was used to evaluate the developed model on unseen data. To determine the optimal hyperparameters during the training process, the model was tuned using the grid search algorithm and optimized by minimizing cross-entropy loss in a nested 4-fold cross-validation. This approach ensured an unbiased and robust performance evaluation. The best-performing configuration during tuning was selected for the final model. XGBoost hyperparameters included:

max_depth: maximum depth of a tree.

gamma: minimum reduction in loss required to partition a leaf node in the tree

colsample_bytree: subsample ratio of training data to be used when growing each tree in the ensemble

learning rate: step size shrinkage used in updates to prevent overfitting.

The grid search space for “*max_depth*”, “*gamma*”, “*colsample_bytree*” and “*learning rate*” was set to [0, 0.5, 1], [0.8, 1.0], [4, 5, 6] and [0.1, 0.01, 0.05], respectively.

Appendix D of the supplementary data. Feature selection

To ensure the effectiveness of the model, we calculated the linear (Pearson) correlation coefficient between each feature and eliminated the features that were too similar (correlation coefficient > 0.7). This approach balances the inclusion of potentially informative variables with the robustness of tree-based algorithms, such as XGBoost, which effectively handle multicollinearity through internal regularization mechanisms.

Out of a total of 16 parameters, after exclusion of correlated features, 12 features were selected, including age, tricuspid annulus plane systolic excursion (TAPSE), right ventricular (RV) end-diastolic volume index, RVEF, RV free-wall strain, PASP, EROA, regurgitant fraction, right atrial longitudinal strain, right atrial volume index, TAPSE/PASP and forward RV stroke volume/ end-systolic volume.

Appendix E of the supplementary data. Definition of statistical parameters for evaluating the XGBoost model.

The statistical parameters were as follows:

$$\text{Balanced accuracy} = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right)$$

$$\text{Macro-averaged precision} = \frac{1}{n} \sum_{i=1}^n \left(\frac{TP_i}{TP_i+FP_i} \right)$$

$$\text{Macro-averaged recall} = \frac{1}{n} \sum_{i=1}^n \left(\frac{TP_i}{TP_i+FN_i} \right)$$

$$\text{Macro-averaged F1-score} = \frac{1}{n} \sum_{i=1}^n \left(2 \frac{\frac{TP_i}{TP_i+FP_i} \frac{TP_i}{TP_i+FN_i}}{\frac{TP_i}{TP_i+FP_i} + \frac{TP_i}{TP_i+FN_i}} \right)$$

where TP are true positives, TN true negatives, FN false negatives, FP false positives and n is the number of classes.

Appendix F of the supplementary data. Explainable AI with SHapley Additive exPlanation (SHAP).

The SHAP technique is based on the theoretically optimal Shapley values from game theory and provides an interpretability method that aligns more closely with human intuition.¹ The concept of the SHAP value was introduced to explain model predictions, identify the most influential features impacting the outcome, and understand how these features contribute to the prediction. The SHAP technique calculates each variable's contribution to the model by using a predicted model for every combination of variables $S \subseteq F$, where F is the set of all possible variables. The SHAP value of a variable corresponds to the average measure of its additive feature attribution (marginal contribution), computed for all subsets $S \subseteq F \setminus \{i\}$.

The formula for the SHAP value is defined as follows:

$$\phi_i(f) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)]$$

where F is the total number of features. $f(S)$ is the prediction given the subset S and $f(S \cup \{i\})$ is the prediction given S including feature i . $f(S \cup \{i\}) - f(S)$ is the marginal contribution.

The SHAP technique offers 2 approaches for model explainability: global explainability, which provides insights into the overall structure of how a model makes decisions, and local explainability, which focuses on understanding how the model arrived at a specific prediction.¹

One advantage of the SHAP technique over earlier perturbation methods is that SHAP values also account for interactions between different features.

- 1- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in neural information processing systems. 2017;30.
- 2- Welcome to the shap documentation. <https://shap.readthedocs.io/en/latest/>

Appendix G of the supplementary data. Software packages

Statistical and survival analyses were performed in R (version 4.3.2) using the following packages: readxl, corrplot, survival, ggplot2, pROC, PredictABEL, FactorMineR, factoextra, dplyr, tibble, logistf, lme4, rms, and survminer.

The remaining analyses were conducted in Python (version 3.10.12) using the following libraries: pandas, numpy, sklearn, XGBoost, matplotlib and shap packages.

Figure 1 of the supplementary data. Patient selection flow-chart

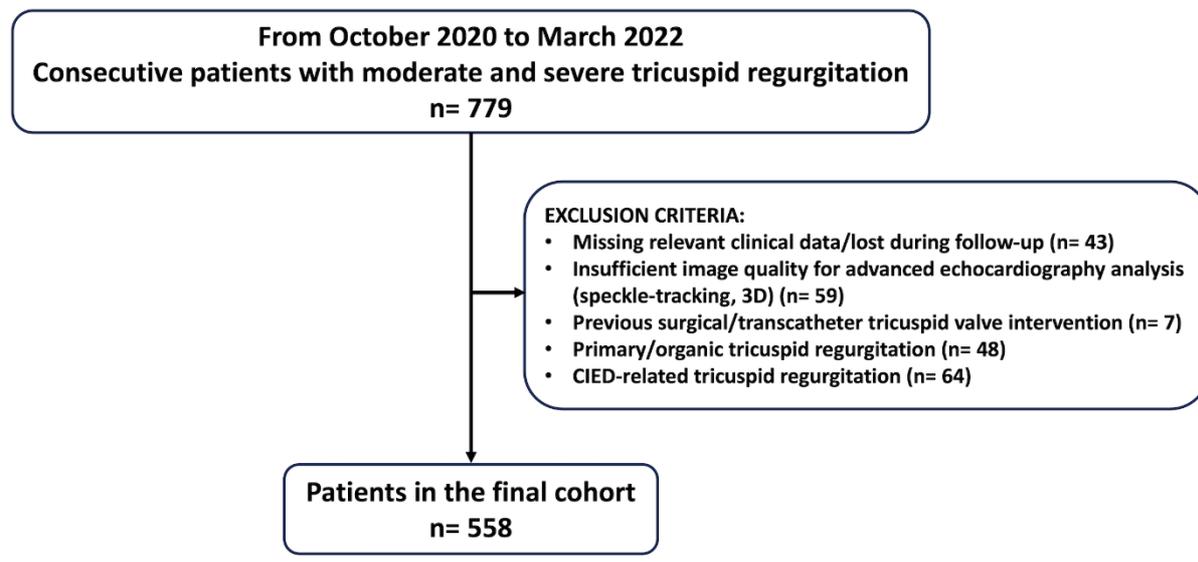
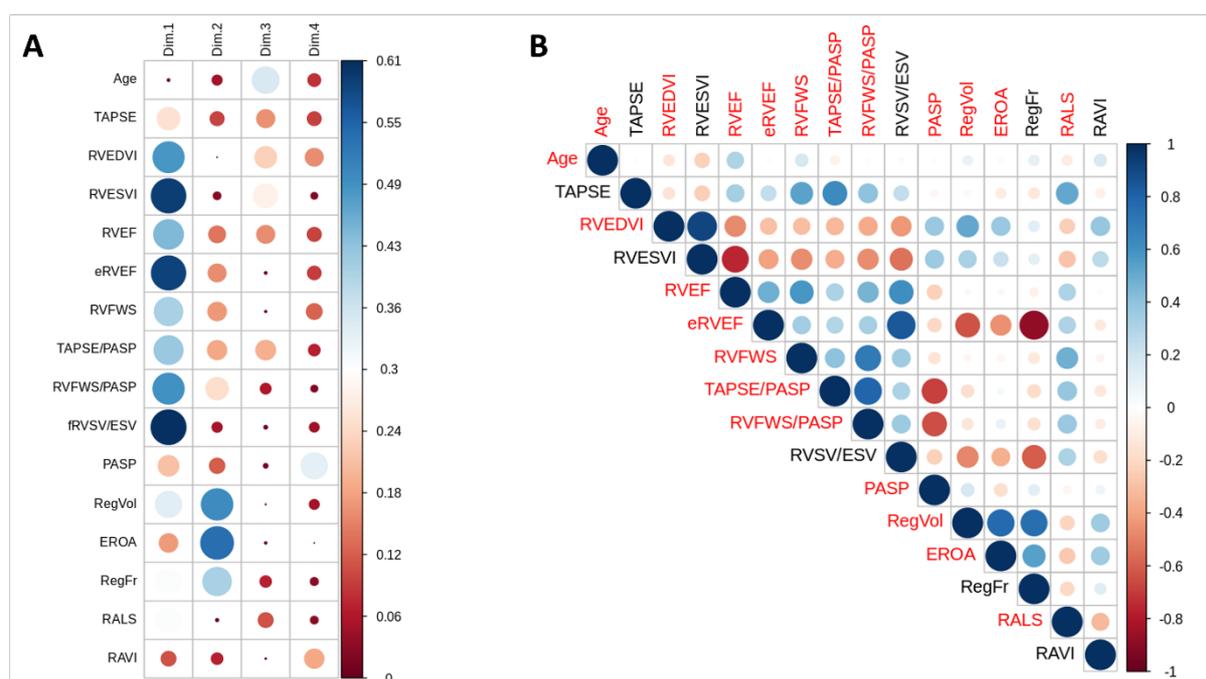


Figure 2 of the supplementary data. Variable selection and resulting cluster analysis.

Representation of the candidate variables expressed as squared cosine into 4 main principal component analysis dimensions. A high squared cosine indicates a good representation of the variable on the principal component. B. correlation matrix with Pearson correlation coefficient of candidate variables. The color intensity and circle size are proportional to the correlation coefficients (if $P > .05$ they are left blank). Variables selected for clustering are shown in red.



eRVEF, effective right ventricular ejection fraction; EROA, effective regurgitant orifice area; RALS, right atrial longitudinal strain; RAVI, right atrial volume index; RegF, regurgitant fraction; RegV, regurgitant volume; RVFWS, right ventricular free-wall strain; PASP, pulmonary artery systolic pressure; RVEDVI, right ventricular end-diastolic volume index; RVEF, right ventricular ejection fraction; RVESVI, right ventricular end-systolic volume index; fRVSV, forward right ventricular stroke volume; TAPSE, tricuspid annulus plane systolic excursion.

Figure 3 of the supplementary data. Survival plots for the derivation (left panel) and validation (right panel) cohorts.

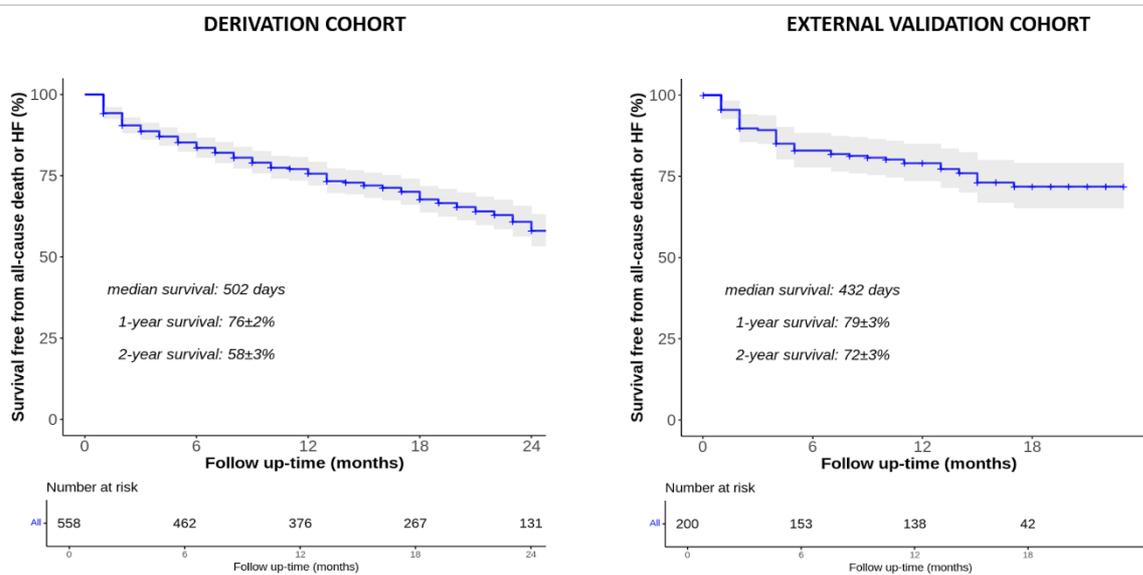


Figure 4 of the supplementary data. Survival plots for the derivation (left panel) and validation (right panel) cohorts.

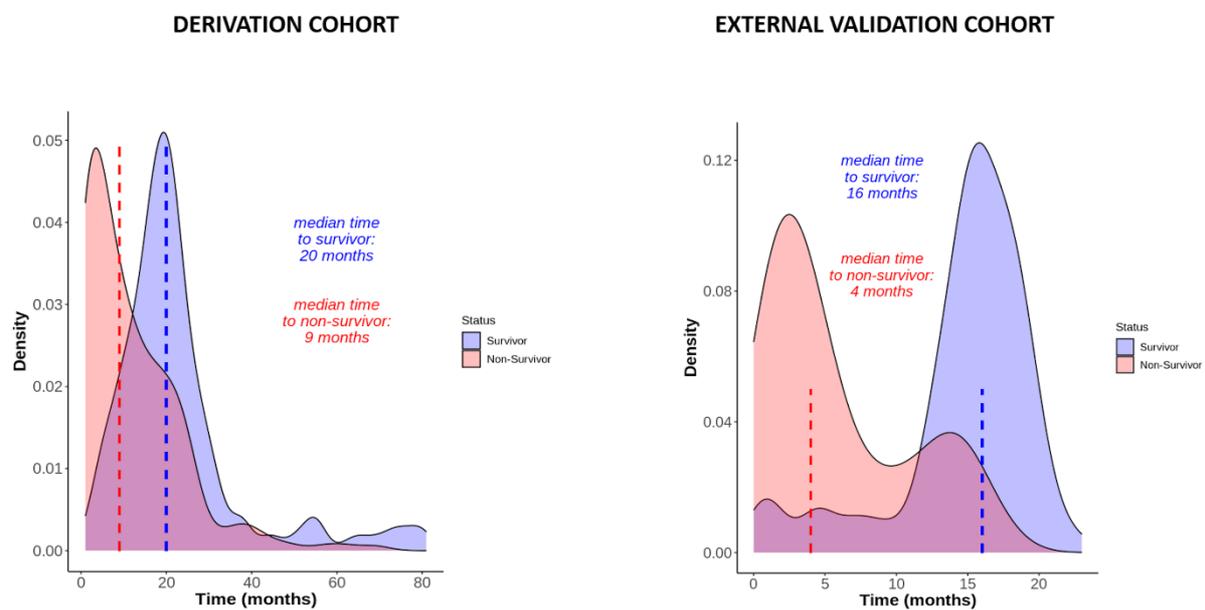


Figure 5 of the supplementary data. Density plots for the derivation (left panel) and validation (right panel) cohorts.

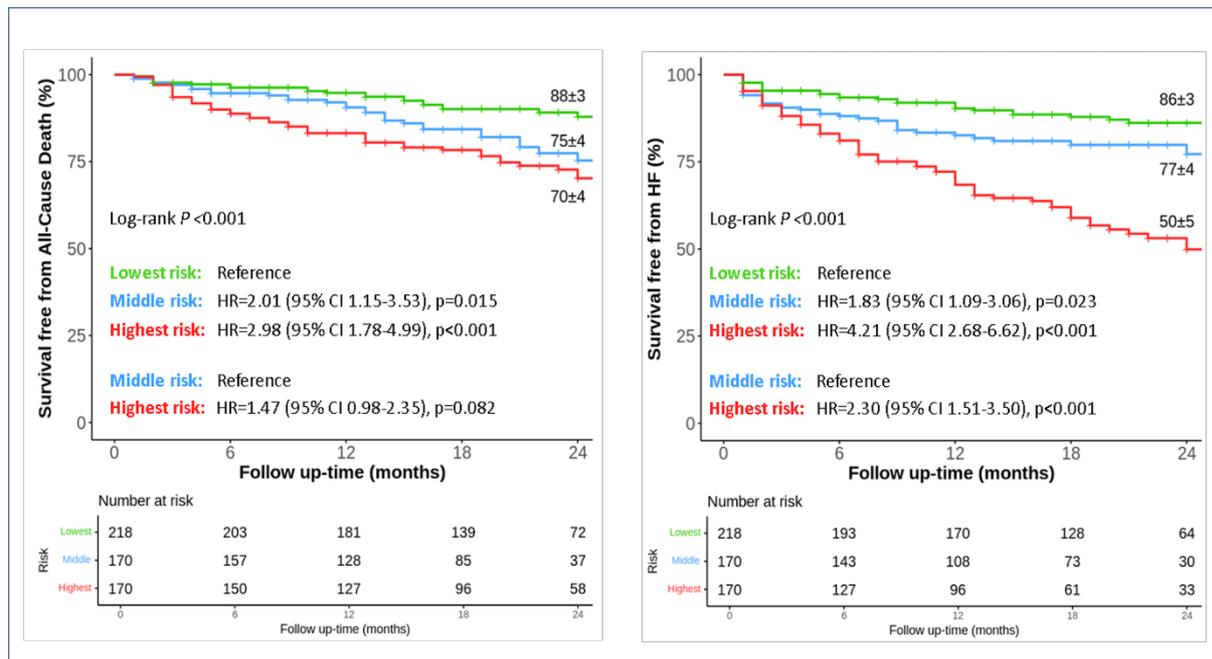


Figure 6 of the supplementary data. Kaplan-Meier plots for the analysis of the event-free survival from all-cause death (left panel) and hospitalization for heart failure of the patients of the derivation cohort included into the lowest-, intermediate, and highest-risk clusters.

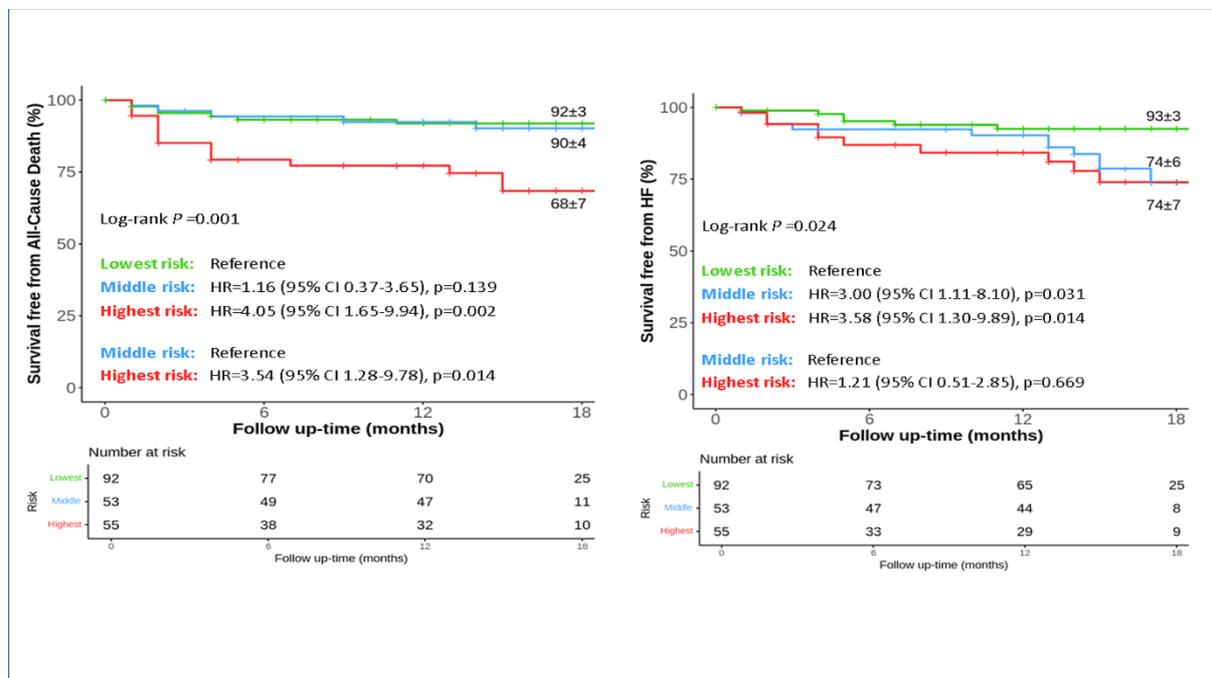


Figure 7 of the supplementary data. True table of the derivation and validation cohorts.

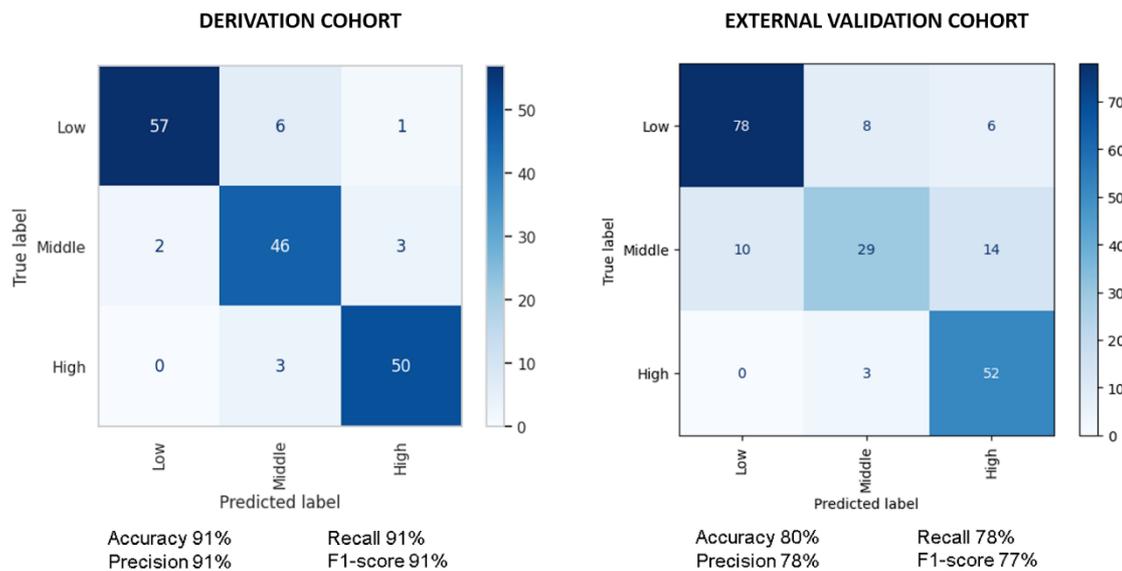


Figure 8 of the supplementary data. Results of 100-fold Monte-Carlo cross-validation for accuracy, precision (top panels), recall and F1-score (bottom panels). Green line = mean value; red lines = 95% confidence interval.

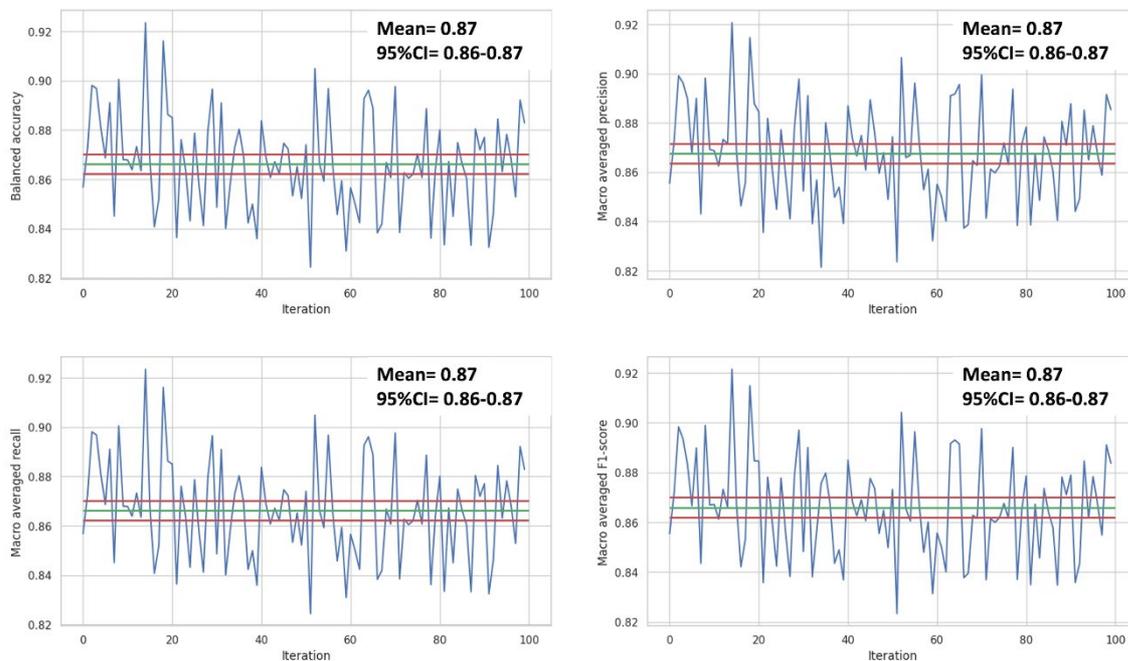
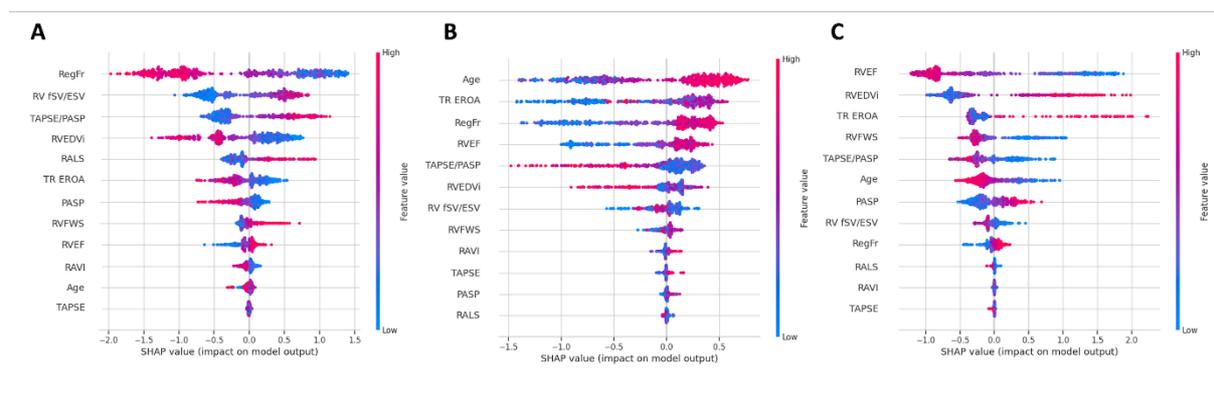


Figure 9 of the supplementary data. Summary plots combining feature importance with feature effects to predict the patient-specific phenogroup. SHAP feature importance assessment for A) low-risk, B) middle-risk, C) high-risk phenogroups. Each point on the summary plot is a SHAP value for a feature and an instance. The importance of the feature, based on the magnitude of its influence on the model output, determines its position on the y-axis, while its SHAP value is shown on the x-axis. The color represents the value of the feature value from low (blue) to high (red).



REFERENCES OF THE SUPPLEMENTARY DATA

1. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging*. 2015;16:233-270.
2. Ciampi P, Badano LP, Florescu DR, et al. Comparison of RA Volumes Obtained Using the Standard Apical 4-Chamber and the RV-Focused Views. *JACC Cardiovascular imaging*. 2023;16:248-250.
3. Genovese D, Mor-Avi V, Palermo C, et al. Comparison Between Four-Chamber and Right Ventricular-Focused Views for the Quantitative Evaluation of Right Ventricular Size and Function. *J Am Soc Echocardiogr*. 2019;32:484-494.
4. Gavazzoni M, Badano LP, Pugliesi GM, et al. Assessing Right Atrial Size in Patients with Tricuspid Regurgitation: Importance of the Right Ventricular-Focused View. *Eur Heart J Cardiovasc Imaging*. 2024;25:1743-1750.
5. Tomaselli M, Badano LP, Mene R, et al. Impact of correcting the 2D PISA method on the quantification of functional tricuspid regurgitation severity. *Eur Heart J Cardiovasc Imaging*. 2022;23:1459-1470.
6. Clement A, Tomaselli M, Badano L, et al. Association with outcome of the regurgitant-volume adjusted right ventricular ejection fraction in secondary tricuspid regurgitation. *J Am Soc Echocardiogr*. 2025. <https://doi.org/10.1016/j.echo.2025.01.008>.
7. Muraru D, Previtero M, Ochoa-Jimenez RC, et al. Prognostic validation of partition values for quantitative parameters to grade functional tricuspid regurgitation severity by conventional echocardiography. *Eur Heart J Cardiovasc Imaging*. 2021;22:155-165.
8. Lancellotti P, Pibarot P, Chambers J, et al. Multi-modality imaging assessment of native valvular regurgitation: an EACVI and ESC council of valvular heart disease position paper. *Eur Heart J Cardiovasc Imaging*. 2022;23:e171-e232.
9. Muraru D, Spadotto V, Cecchetto A, et al. New speckle-tracking algorithm for right ventricular volume analysis from three-dimensional echocardiographic data sets: validation with cardiac magnetic

resonance and comparison with the previous analysis tool. *Eur Heart J Cardiovasc Imaging*. 2016;17:1279-1289.

10. Badano LP, Kolas TJ, Muraru D, et al. Standardization of left atrial, right ventricular, and right atrial deformation imaging using two-dimensional speckle tracking echocardiography: a consensus document of the EACVI/ASE/Industry Task Force to standardize deformation imaging. *Eur Heart J Cardiovasc Imaging*. 2018;19:591-600.

11. Gavazzoni M, Badano LP, Cascella A, et al. Clinical Value of a Novel Three-Dimensional Echocardiography-Derived Index of Right Ventricle-Pulmonary Artery Coupling in Tricuspid Regurgitation. *J Am Soc Echocardiogr*. 2023;36:1154-1166.e1153.

12. Hahn RT, Lawlor MK, Davidson CJ, et al. Tricuspid Valve Academic Research Consortium Definitions for Tricuspid Regurgitation and Trial Endpoints. *Eur Heart J*. 2023;44:4508-4532.

13. Muraru D, Badano LP, Hahn RT, et al. Atrial secondary tricuspid regurgitation: pathophysiology, definition, diagnosis, and treatment. *Eur Heart J*. 2024;45:895-911.