External Validation of Machine Learning Models Predicting Global Longitudinal Strain from Conventional Echocardiography in Cancer Patients

地方独立行政法人
東京都立病院機構

安西　耕[1,2,3],平田健司[3,4,5]

[1] 東京都立多摩総合医療センター　循環器内科
[2] 静岡県立静岡がんセンター　腫瘍循環器科
[3]北海道大学大学院医学研究院　医療AI開発者養成プログラム
[4]北海道大学大学院医学研究院　画像診断学教室
[5]北海道大学病院　医療AI研究開発センター

北海道大学病院
医療AI研究開発センター

開示すべき利益相反状態はありません

Introduction

Global longitudinal strain (GLS) is a well-established prognostic marker for the early detection of heart failure and cancer therapy–related cardiac dysfunction (CTRCD) [1]. Despite the widespread availability of GLS measurement across healthcare institutions, its routine implementation in daily clinical practice remains limited, largely due to practical constraints such as time limitations and the need for specialized training [2]. We previously developed machine learning (ML) models that predicted reduced GLS (Low-GLS, defined as GLS <16%) from conventional echocardiographic parameters in cancer patients, which were internally validated [3]. We aim to externally validate the generalizability of such models across institutions.

Methods

This multicenter study included patients from Tokyo Metropolitan Tama Medical Center (TMC) (n = 1,531) and Shizuoka Cancer Center (SCC) (n=230) who underwent echocardiography with GLS measurement before or after anticancer chemotherapy (**Fig.1**).

Patients with left ventricular ejection fraction <50% were excluded. The primary dataset from TMC, used for machine learning model development was randomly divided into a training and an internal validation dataset at a ratio of 4:1. Twenty-two conventional echocardiographic parameters were used as predictors of Low-GLS. Using an automated machine learning (AutoML) framework, five ML models, including bagged tree-based algorithms (Random Forest, Extra Trees, LightGBM, and CatBoost) and a weighted ensemble were trained and compared in the internal validation dataset. All models, were subsequently evaluated in an independent external validation dataset from SCC.

Model performance was assessed using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 score.

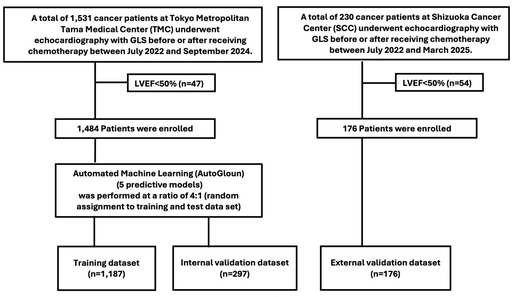


Fig. 1 Study Design

Results

Table 1. Patients Characteristics

	Primary dataset (n=1,484)	External validation dataset (n=176)	p value
Age (y/o)	63.7 ± 13.3	56.4 ± 15.1	<.001
Female, n(%)	1022 (69.0)	115 (65.3)	0.93
BMI (kg/m²)	22.1 ± 3.8	22.3 ± 3.9	0.89
EF (%)	66.0 ± 6.0	60.1 ± 7.1	<.001
GLS(%)	17.7 ± 3.2	17.4 ± 2.4	0.19
AAD (mm)	20.5 ± 2.2	29.5 ± 3.6	<.001
LAD (mm)	31.9 ± 5.5	29.0 ± 6.3	<.001
LVDd (mm)	43.0 ± 4.7	44.4 ± 5.4	<.001
LVDs (mm)	27.3 ± 3.7	30.2 ± 4.7	<.001
IVST (mm)	8.5 ± 1.4	8.3 ± 1.4	0.07
PWT (mm)	8.5 ± 1.3	8.4 ± 1.2	0.65
E (cm/s)	70.2 ± 18.2	60.4 ± 14.1	<.001
A (cm/s)	75.6 ± 20.1	68.0 ± 21.0	<.001
DCT (ms)	225.5 ± 61.2	229.5 ± 66.4	0.49
Septal e` (cm/s)	7.0 ± 2.3	7.4 ± 2.5	0.03
Septal a` (cm/s)	9.2 ± 1.9	9.0 ± 2.3	0.16
Lateral e`(cm/s)	9.2 ± 2.8	9.3 ± 2.9	0.64
Lateral a`(cm/s)	9.9 ± 2.5	9.5 ± 2.7	0.03
E/A	1.0 ± 0.4	1.0 ± 0.4	0.25
E/e'	11.0 ± 4.0	9.2 ± 3.4	<.001
LVOT-Vmax (m/s)	1.0 ± 0.2	0.9 ± 0.2	<.001
AR, n (%)	222 (15.0)	13 (7.4)	0.84
MR, n (%)	435 (29.3)	9 (5.1)	0.43
TR, n (%)	545 (36.7)	0 (0)	<.001
PR, n (%)	113 (7.6)	0 (0)	<.001

Values are presented as numbers, mean ± standard deviation. Continuous variables are analyzed using t-test or the Mann–Whitney U test if the data are not normally distributed. Categorical variables are analyzed using the chi-squared test or Fisher's exact test. Significant differences are indicated in bold. AAD, Aortic Root Diameter; AR, Aortic Regurgitation (>=mild); A, Atrial Contraction Wave; DCT, Deceleration Time of E wave; EF, Ejection Fraction; E/A, Transmitral Early Filling Velocity to Mitral Annular Early Diastolic Velocity Ratio; E, Transmitral Early Filling Velocity; IVST, Interventricular Septal Thickness; LAD, Left Atrium; LVDd, Left Ventricular Diastolic Diameter; LVDs, Left Ventricular Systolic Diameter; LVOT-Vmax, Left Ventricular Outflow Tract Velocity Maximum; MR, Mitral Regurgitation (>=mild); PWT, Left Ventricular Posterior Wall Thickness; TR, Tricuspid Regurgitation (>=mild); e`, Mitral Annular Early Diastolic Velocity; a`, Mitral Annular Atrial Systolic Velocity.

Table 2. Performance comparison of machine learning models

	AUC	ACC	Sensitivity (Recall)	Specificity	PPV (Precision)	NPV	F-1
Internal validation dataset							
Bagged Random Forest	0.8006	0.7811	0.2963	0.9630	0.7500	0.7849	0.4248
Bagged Extra Trees	0.8005	0.7710	0.2346	0.9722	0.7600	0.7721	0.3585
Weighted ensemble (AutoGluon)	0.7926	0.7845	0.2963	0.9676	0.7742	0.7857	0.4286
Bagged CatBoost	0.7892	0.7811	0.3086	0.9583	0.7353	0.7871	0.4348
Bagged LightGBM	0.7865	0.7778	0.2346	0.9815	0.8261	0.7737	0.3654
External validation dataset							
Bagged Random Forest	0.7618	0.6989	0.7200	0.6905	0.4800	0.8614	0.5760
Bagged Extra Trees	0.8202	0.7670	0.6000	0.8333	0.5882	0.8400	0.5941
Weighted Ensemble (AutoGluon)	0.7805	0.7216	0.7600	0.7063	0.5067	0.8812	0.6080
Bagged CatBoost	0.7760	0.6761	0.7800	0.6349	0.4588	0.8791	0.5778
Bagged LightGBM	0.7622	0.7557	0.6600	0.7937	0.5593	0.8547	0.6055

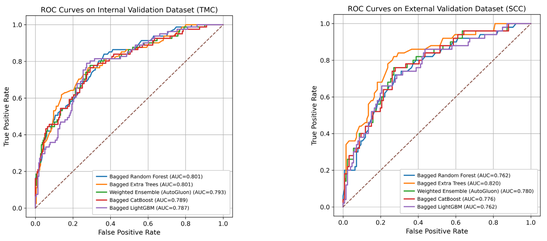
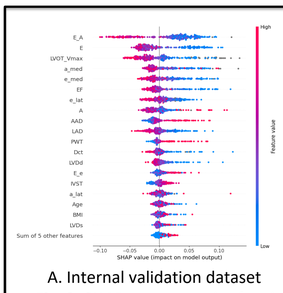
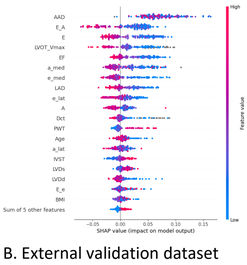


Fig. 2 Receiver operating characteristic (ROC) curves

Patients in the primary dataset were significantly older and demonstrated higher values of LVEF than those in the external validation dataset (**Table 1**).
In the internal validation dataset, the bagged Random Forest model achieved the highest discriminative performance in terms of AUC (0.801). Comparable performance was observed for the other models, including bagged Extra Trees, the weighted ensemble, bagged CatBoost, and bagged LightGBM, with AUC values ranging from 0.787 to 0.801 and F1 scores between 0.359 and 0.435 (**Table 2 and Fig. 2**).
In the external validation dataset, model performance was generally preserved across all algorithms, with the bagged Extra Trees model demonstrating the highest discriminative ability, achieving an AUC of 0.820. Similar levels of performance were achieved by other models, with AUCs ranging from 0.762 to 0.781 (**Table 2 and Fig. 2**).



A. Internal validation dataset



B. External validation dataset

Fig. 3. SHAP summary plot for the Bagged Random Forest model.

The SHAP (Shapley Additive exPlanations) method was used to interpret the model predictions. Each dot represents an individual patient. The color indicates the feature value (red: higher, blue: lower). The x-axis represents the SHAP value, indicating the magnitude and direction of each feature's contribution to the prediction. Features are ranked by their importance from top to bottom.

In the internal validation dataset, lower values of E/A, E, LVOT-Vmax, septal e', septal a', and EF were associated with a greater likelihood of Low-GLS (**Fig. 3A**), whereas in the external validation dataset, lower values of E/A, E, LVOT-Vmax, EF, septal e', and septal a' were associated with a greater likelihood of Low-GLS (**Fig. 3B**).

These findings indicate that similar echocardiographic features contributed to the prediction of Low-GLS in both datasets.

Discussion

This model may enable identification of patients with reduced GLS using conventional echocardiographic measurements when direct GLS assessment is unavailable or unreliable, thereby supporting early risk stratification for CTRCD. This strategy may complement emerging deep learning-based automated GLS measurement techniques. However, the retrospective design, single-country population, and limited vendor diversity warrant further prospective validation in larger and more heterogeneous cohorts.

Conclusion

This study externally validated machine learning models predicting reduced GLS from conventional echocardiographic parameters in patients with cancer.

Reference

[1] Oikonomou EK, et al. Assessment of Prognostic Value of Left Ventricular Global Longitudinal Strain for Early Prediction of Chemotherapy-Induced Cardiotoxicity: A Systematic Review and Meta-analysis. JAMA Cardiol. 2019;4(10):1007
[2] Sade LE, et al. Current clinical use of speckle-tracking strain imaging: insights from a worldwide survey from the European Association of Cardiovascular Imaging (EACVI). Eur Heart J Cardiovasc Imaging 2023, 24(12):1583-1592.
[3] Anzai T, et al. Machine learning for cardio-oncology: predicting global longitudinal strain from conventional echocardiographic measurements in cancer patients. Cardiooncology 2025, 11(1):49.