POSTER MSR79

Early Detection of Cancer Therapy–Related Cardiac Dysfunction in Lung Cancer Patients Using Machine Learning

Wen-Li Kuan, MD[1]; Hsiu-Ting Chien, PhD[1]; Chih-Fan Yeh, PhD[2]; Sandy Hsu, BA[1]; Wan Tseng Hsu, PhD[1]; Fang-Ju Lin, RPh, PhD[3]

[1] Graduate Institute of Clinical Pharmacy, College of Medicine, National Taiwan University, Taipei, Taiwan
[2] Division of Cardiology, Department of Internal Medicine and Cardiovascular Center, National Taiwan University Hospital, Taipei, Taiwan
[3] Department of Pharmacy, National Taiwan University Cancer Center, Taipei, Taiwan

BACKGROUND & OBJECTIVES

- Advances in lung cancer (LC) therapy have significantly improved patient survival but have also increased the risk of cancer therapy–related cardiac dysfunction (CTRCD).[1]
- Machine learning (ML) offers the potential for early and accurate detection of CTRCD by integrating complex clinical and treatment-related data.[2]
- Although electrocardiograms (ECGs) provide rich information on cardiac function, their unstructured data format has limited their use in previous CTRCD prediction models.[2]
- Study aims: (1) To develop ML-based models for early detection of CTRCD in patients with LC, and (2) To evaluate whether the addition of unstructured ECG data improves model performance.

METHODS

➢ Data Source

NTUH-iMD, the electronic health record database at National Taiwan University Hospital (NTUH)

➢ Study Design

Retrospective case-control study

➢ Patient Population

Inclusion criteria:

- Patients with newly diagnosed primary LC who initiated first lung cancer treatment at NTUH

Exclusion criteria:

- Age <18 or missing age data
- Missing medication or radiotherapy records
- Pre-existing cardiac dysfunction or secondary malignancy
- Development of secondary malignancy during the follow-up period

➢ Case & Control Definition (**Table 1**)

- A two-step case–control screening was performed: algorithmic pre-screening followed by cardiologist adjudication of all cases and controls.

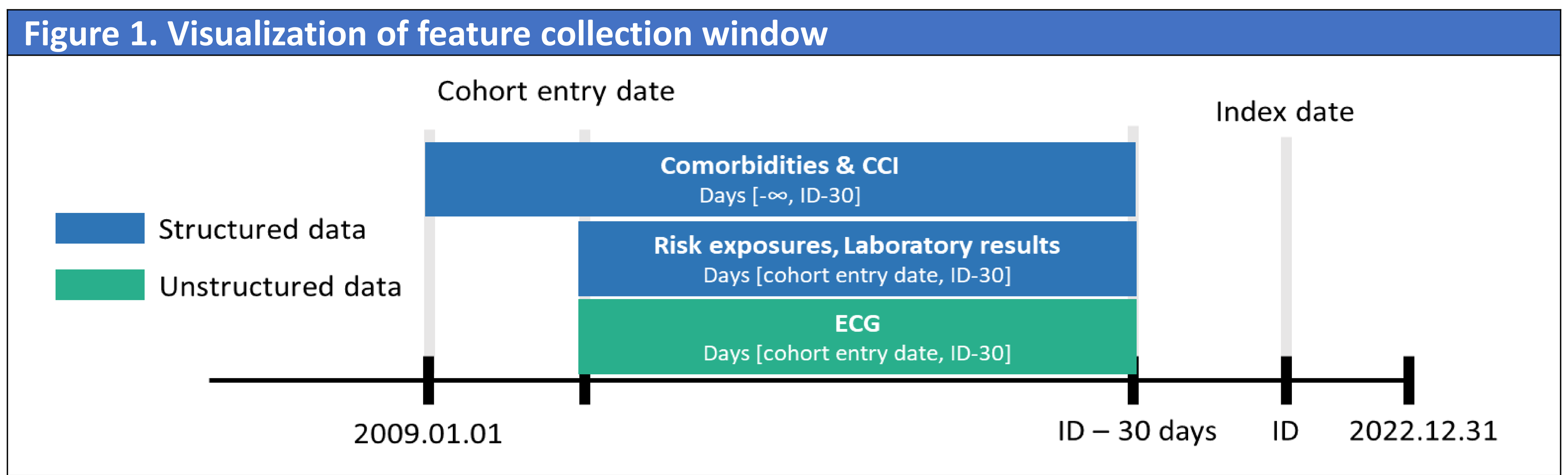
Table 1. Definition of cases and controls

	Cases	Controls
Definition	• With a confirmed decline in LVEF • Heart failure diagnosis meeting eligibility criteria (elevated NT-proBNP or augmented HF medication ±7 days).	• With echocardiographic follow-up showing preserved LVEF • No HF symptoms.
Cohort entry date	Date of first therapy for lung cancer	
Index date	Date of cardiac dysfunction onset	Date of the random echocardiogram performed after the first therapy for LC

➢ Feature Collection (**Figure 1**)

- Multi-dimensional features (n=172) were extracted from the period between cohort entry and 30 days before the index date.
- A **rule-based natural language processing (NLP) method** was developed to extract cardiac features from unstructured ECG report texts.

Figure 1. Visualization of feature collection window



➢ Model development

Two independent models were developed — one with ECG features and one without — each processed and trained separately.

- ***Data Pre-processing***
  - Removed features with >70% missingness
  - Imputed missing values using missForest; applied one-hot encoding for categorical variables
  - Selected non-redundant features via Pearson correlation analysis
- ***Model Building & Evaluation***
  - ML models: LASSO, Random Forest (RF), XGBoost, Naïve Bayes (NB)
  - Validation: 10-fold cross-validation
  - Addressing class imbalance: oversampling, undersampling, SMOTE, weighting
  - Performance metrics: AUPRC, AUROC, accuracy, PPV, recall, specificity, F1 score
- ***Model Interpretation***
  - Applied SHAP to identify top 20 influential features, and define clinical thresholds

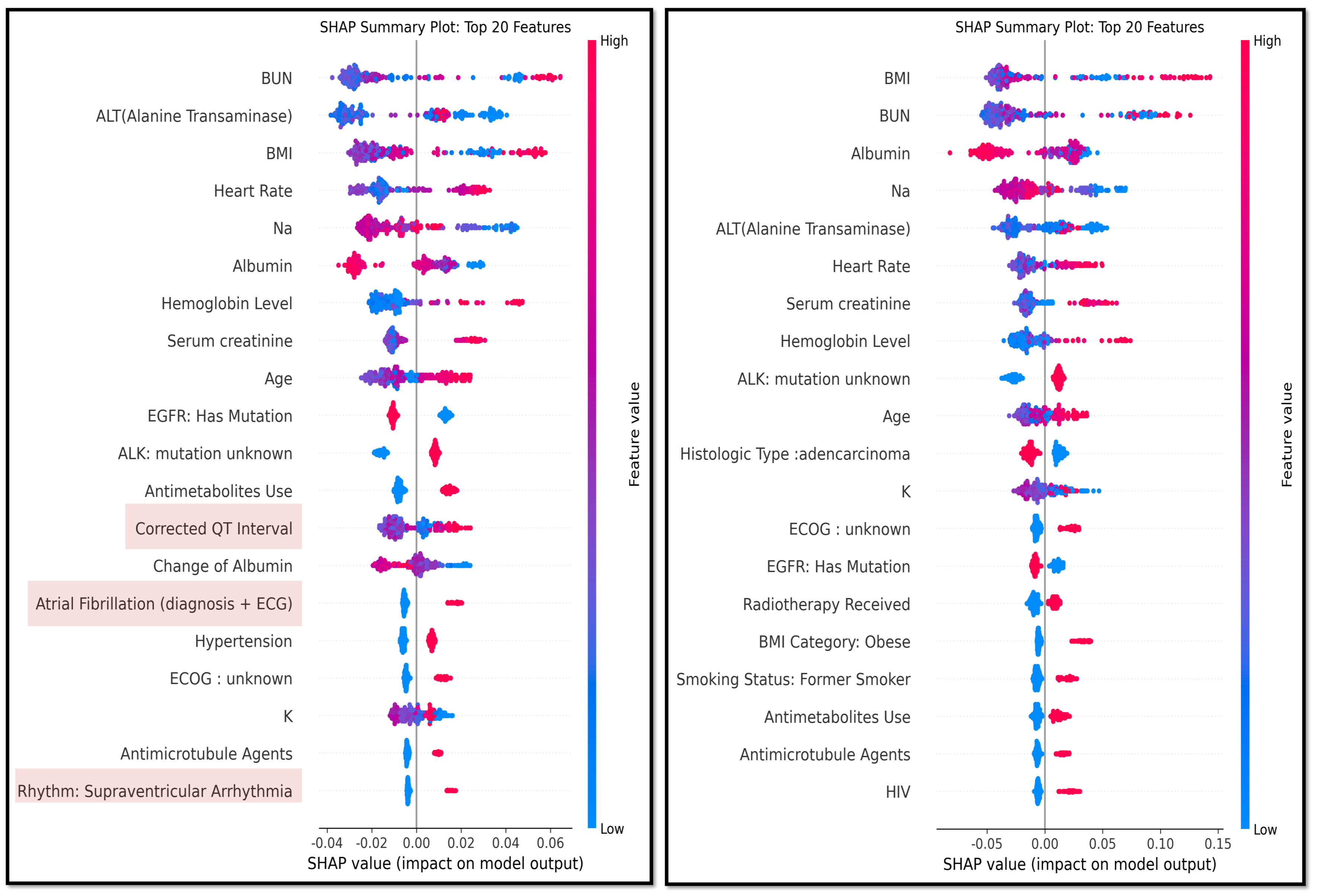
RESULTS

- From 6,032 newly diagnosed LC patients**, 52 CTRCD cases and 341 controls** were identified for model development.
- **The RF model with undersampling performed best** in both ECG-inclusive (93 features) and non-ECG (67 features) models, showing similar performance (**AUPRC = 0.6782 vs. 0.6987**) and indicating limited added value from ECG data (**Table 2**).
- SHAP analysis identified lower ALT, sodium, hemoglobin, and albumin; higher age, BUN, and creatinine; and BMI (U-shaped) as key predictors of CTRCD risk (**Figure 2**).
- In the ECG-inclusive model, additional predictors included QTc prolongation, atrial fibrillation, and supraventricular arrhythmia (**Figure 2**).
- Clinical thresholds at age >70.5 years, heart rate >87 bpm, BMI <18 or >25 (U-shaped), and albumin <4.19 g/dL, indicating physiologic tipping points for increased risk (**Figure 3**).

Table 2. Performance of four models with/without ECG features (using undersampling)

Model (including ECG)	AUROC	AUPRC	Accuracy	Precision (PPV)	Sensitivity (Recall)	Specificity	F1
LASSO	0.8348	0.5081	0.7089	0.2903	0.9000	0.6812	0.4390
XGBoost	0.8797	0.5313	0.7342	0.2963	0.8000	0.7246	0.4324
RF	0.9304	**0.6782**	0.7468	0.3214	**0.9000**	0.7246	0.4737
NB	0.8696	0.3571	0.7722	0.3571	1.0000	0.7391	0.5263
Model (not including ECG)	**AUROC**	**AUPRC**	**Accuracy**	**Precision (PPV)**	**Sensitivity (Recall)**	**Specificity**	**F1**
LASSO	0.8435	0.5648	0.7089	0.2759	0.8000	0.6957	0.4103
XGBoost	0.9377	0.6974	0.8608	0.4737	0.9000	0.8551	0.6207
RF	0.9145	**0.6987**	0.7848	0.3600	**0.9000**	0.7681	0.5143
NB	0.8891	0.4039	0.8101	0.3913	0.9000	0.7971	0.5455

Figure 2. Interpretation of SHAP Plot – model with (Left) / without (Right) ECG features



: Extracted from ECG text report through NLP

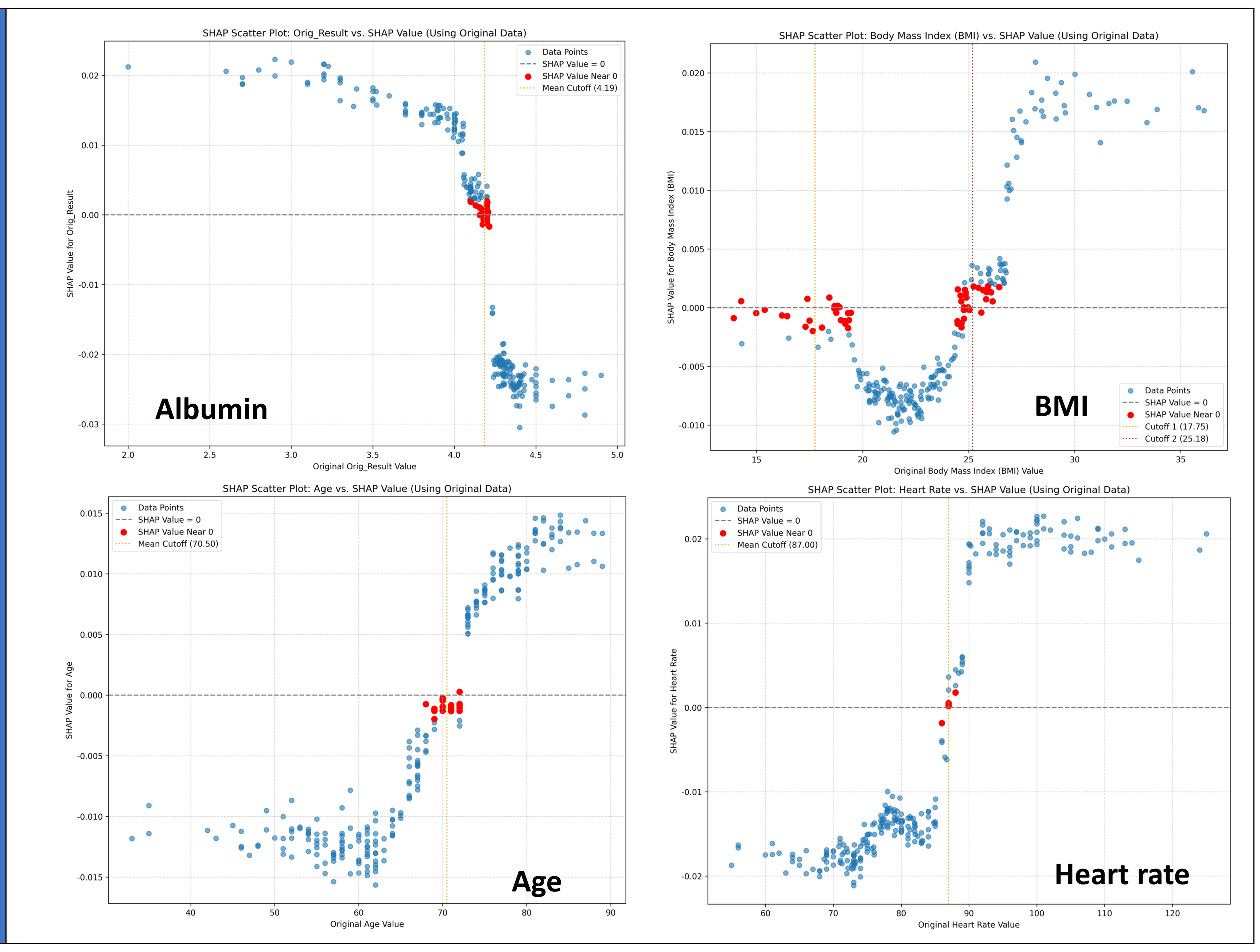


Figure 3. Risk contribution inflection point

CONCULSION

- Both models showed good performance in early CTRCD detection among LC patients.
- ECG features offered modest incremental value, primarily enhancing interpretability through heart rate–related patterns rather than predictive power.
- Further studies with multicenter data and larger sample sizes are needed to validate the findings.

REFERENCES

1. Lyon AR, Lopez-Fernandez T, Couch LS, et al. *Eur Heart J Cardiovasc Imaging*. Sep 10 2022;23(10):e333–e465. doi:10.1093/ehjci/jeac106
2. Yagi R, Goto S, Himeno Y, et al. *Nat Commun*. Mar 21 2024;15(1):2536. doi:10.1038/s41467-024-45733-x

CONTACT INFORMATION: kuanlily0201@gmail.com