中文題目：深度學習預測低收縮率心臟衰竭患者之死亡風險

英文題目：Deep Learning-Based Prediction of Mortality in Patients with Heart Failure with Reduced Ejection Fraction

作　　者：賴政謙[1,2]，黃少嵩[1,2,4]，唐德成[1,2,3,5]，曾偉誠[1,2,3,5*]

服務單位：[1]台北榮民總醫院內科部，[2]國立陽明交通大學醫學系，[3]國立陽明交通大學智慧型藥物與智能生物裝置研究中心，[4]台北榮民總醫院內科部心臟內科，[5]台北榮民總醫院內科部腎臟科

Background: Despite the ever-improving treatment modalities, patients suffering from heart failure with reduced ejection fraction (HFrEF) are still at extremely high risk of mortality. Risk stratification models utilizing machine learning algorithms to predict the survival in heart failure patients have been proposed. However, current machine learning models seldom include comprehensive echocardiographic parameters, and the model predictability on long-term mortality is unclear. Herein, we aimed to determine whether advanced deep learning algorithms improve risk prediction of long-term mortality among HFrEF patients.

Method: A longitudinal cohort of 12,012 patients undergoing transthoracic echocardiography between 2011 and 2018 for heart failure from a tertiary medical center in northern Taiwan were screened. Thereafter, 1197 patients (male, 74.6%; mean [SD] age, 69.6 [14.8] years) diagnosed with HFrEF (left ventricular ejection fraction of ≤40%) were enrolled for model construction, and randomly split into the training dataset (70%, 837 patients) as well as the hold-out testing dataset (30%, 360 patients). All patients were followed until death or Dec. 31, 2021. Deep learning models including extreme gradient boosting (XGBoost), random forest (RF), support vector machine, artificial neural network, logistic regression, and stacked generalization ensemble learning (Ensemble) were constructed to predict the 3-year all-cause mortality by including 71 clinical and 17 echocardiographic parameters. The performances of models were analyzed with the area under the receiver operating characteristic curve (AUROC) and Brier score. Top-ranked features were discovered by Shapley additive explanations (SHAP) analyses. Model usefulness was compared against a current standard risk score (Meta-Analysis Global Group in Chronic Heart Failure [MAGGIC] score) by net reclassification index (NRI).

Results: During the 3-year follow-up period, 166 (13.9%) patients passed away. The Ensemble, RF, and XGBoost models resulted in superior performance with AUROC of 0.8632 (95% confidence interval [CI] 0.814-0.913), 0.8631 (95% CI, 0.815-0.912), and 0.859 (95% CI, 0.814-0.904), respectively. The Brier scores of Ensemble, RF, and XGBoost models were 0.0898, 0.0975, and 0.0976, respectively, reflecting outstanding calibration. These three models all significantly out-performed the traditional MAGGIC score (AUROC 0.741; 95% CI, 0.673-0.810). Notably, NRI analyses showed that Ensemble, XGBoost, and RF models more accurately reclassified 30.3% (95% CI 15.2%-45.3%, $p = 0.001$), 25.2% (95% CI 8.0%-42.3%, $p = 0.011$), and 21.2% (95% CI 4.5%-37.5%, $p = 0.018$) of HFrEF patients for long-term mortality as compared to the MAGGIC score. SHAP analyses indicated that the mitral medial e’ velocity and E/e’ ratio were top-ranked echocardiographic prognostic features.

Conclusion: In this study, we demonstrated that Ensemble, XGBoost and RF algorithms are capable of improving prediction of long-term mortality in HFrEF patients after incorporating echocardiographic parameters as compared to traditional MAGGIC score. With these advanced deep learning models, we can not only early identify and timely treat HFrEF patients with grave prognosis, but also provide insights into future exploration of the complicated interaction between clinical features and undesirable outcomes.