Purpose: Early prediction of radiation pneumonitis (RP) could help physicians provide locally advanced non-small cell lung cancer (NSCLC) patients at high risk of RP with targeted radiotherapy. To date, several RP predictive models have been developed using statistical methods and machine learning techniques. We designed an improved risk model for predicting RP using a recently introduced machine learning method.

Methods: We retrospectively analyzed a single NSCLC dataset collected at Washington University School of Medicine, which consisted of 123 NSCLC patients who received 3D conformal radiation therapy with a median prescription dose of 66 Gy and had a median follow-up of 17 months. In this analysis, 59 patients with grade ≥2 were considered to have RP. Using clinical variables and lung and heart dose-volume variables extracted from this dataset, we built a novel RP predictive model based on multivariate logistic regression and supervised principal component analysis (SPCA). The performance of our proposed method was compared with that of multivariate logistic regression alone.

Results: Using lung-related dosimetric variables, SPCA-based logistic regression model had a Spearman correlation coefficient (Rs) of 0.33 (p=0.0001), whereas logistic regression model had Rs=0.20 (p=0.014). Using heart-related dosimetric variables, Rs=0.42 (p < 0.0001) and Rs=0.40 (p < 0.0001) were obtained in the SPCA-based logistic regression and logistic regression models, respectively. Interestingly, when clinical variables alone were used, logistic model showed better performance with Rs=0.34 (p < 0.0001) than SPCA-based logistic regression model with Rs=0.32 (p < 0.0001). Integrating all variables, SPCA-based logistic regression model obtained the best performance with Rs=0.44 (p< 0.0001) while logistic model had Rs=0.37 (p< 0.0001).

Conclusions: Incorporating multivariate logistic regression with a new machine learning method allowed us to produce a better predictive RP model that should be further tested for potential clinical use.