understand underlying mechanisms. PFOs had no significant association with survival, although this should be evaluated in a larger study.

Reference

130

Machine-learning approach to estimating percentage predicted forced expiratory volume in 1 second in cystic fibrosis using computed tomography–derived lung parenchymal biomarkers

G. Mylavarapu1, 2, A. Pillai1, L. He3, 4, R. Fleck4, R. Amin1, 1Division of Pulmonary Medicine, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH; 2Department of Pediatrics, University of Cincinnati Medical Center, Cincinnati, OH; 3Imaging Research Center, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH; 4Department of Radiology, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH

Background: Computed tomography (CT) attenuation patterns in the lung become more complex with disease progression in people with cystic fibrosis (PwCF) because of chronic inflammation of lung parenchyma [1]. Although pulmonary function tests (PFTs) are the gold standard for monitoring lung disease in CF, some studies have demonstrated that CT-based radiology scores can detect changes earlier than PFT, suggesting CT sensitivity in CF [2], but obtaining radiology scores is time consuming and subjective. Lung parenchymal biomarkers derived from CT image processing are quick and reproducible and have been demonstrated to quantify parenchymal abnormalities [3]. Our central hypothesis was that imaging biomarkers would have translational applications in CF. To test this hypothesis, we developed a machine learning (ML) model to estimate percentage predicted forced expiratory volume in 1 second (FEV1,pp) in PwCF using CT-derived lung parenchymal measures alone.

Methods: Lung parenchymal biomarkers describing different components of CT attenuation distribution were derived using an in-house-developed automated program and pyradiomics [4]. Individuals with CT and PFTs within 3 months from our CF center were included. We trained ML regression models using decision trees, random forest, and extreme gradient boosting (XGBoost) to estimate FEV1,pp measured using PFT using CT-derived parenchymal measures. The dataset was split 70:30 between training and testing tasks of ML using stratified sampling. Ten-fold cross-validation was used to find the best hyperparameters for each model. Root mean squared error (RMSE) was used to assess ML model performance.

Results: We included 153 PwCF (age 14.3 ± 3.6; 59 male/94 female; FEV1pp as measured by PFT: 82.8 ± 22.8%). ML regression models were trained on 23 of 121 lung parenchymal measures derived from CT scans, after feature selection. The XGBoost regression model performed best in estimating FEV1,pp (RMSE 6.4%). The ML model error in FEV1,pp is on the same order of repeatability errors in FEV1,pp in spirometry reported in the 2012 NIOSH Spirometry Quality Assurance Study.

Conclusions: This study demonstrates that ML models have translational applications in predicting FEV1,pp in CF. Future studies will further refine this ML model by including clinical data and demographic characteristics and will eventually deploy the model to inform clinical practice.

Acknowledgements: Supported by a Cystic Fibrosis Foundation RDP grant.

References