Early View

Original research article

Chronic lung allograft dysfunction phenotype and prognosis by machine learning CT analysis

Micheal C. McInnis, Jin Ma, Gauri Rani Karur, Christian Houbois, Liran Levy, Jan Havlin, Eyal Fuchs, Jussi Tikkanen, Chung-Wai Chow, Ella Huszti, Tereza Martinu

CHRONIC LUNG ALLOGRAFT DYSFUNCTION PHENOTYPE AND PROGNOSIS BY MACHINE LEARNING CT ANALYSIS

Micheal C. McInnis[1], Jin Ma[2], Gauri Rani Karur[1], Christian Houbois[1,3], Liran Levy[4], Jan Havlin[4], Eyal Fuchs[4], Jussi Tikkanen[4], Chung-Wai Chow[4,5], Ella Huszti[2], Tereza Martinu[4,5]

1. Department of Medical Imaging, University Health Network, University of Toronto, Toronto, ON, Canada

2. Biostatistics Research Unit, University Health Network, University of Toronto, Toronto, ON, Canada

3. Department of Diagnostic and Interventional Radiology, University of Cologne, Cologne, Germany

4. Toronto Lung Transplant Program, Ajmera Transplant Centre, University Health Network, Toronto, ON, Canada

5. Division of Respirology, Department of Medicine, University of Toronto, ON, Canada

Corresponding author:

Micheal C. McInnis MD

Toronto General Hospital

585 University Avenue, 1PMB-298

Toronto, Ontario, Canada, M5G 2N2

Email: micheal.mcinnis@uhn.ca

Phone: +1-416-340-4800 ext. 3709

Fax: +1-416-340-3900

Author Contributions: Conception and design: M.C.M. and T.M. Acquisition of data: M.C.M., G.R.K., C.H., L.L., J.H., E.F., J.T., and T.M. Analysis and interpretation of data: M.C.M., T.M., J.M., and E.H. Drafting the manuscript and intellectual content: M.C.M., J.T., C.W.C. and T.M.

Take Home Message: Machine learning CT lung texture analysis and radiologist analysis both predict prognosis in chronic lung allograft dysfunction (CLAD), independent of CLAD phenotype. Machine learning can discriminate between CLAD phenotypes without expiratory CT.

Abstract

Background: Chronic lung allograft dysfunction (CLAD) is the principal cause of graft failure in lung transplant recipients and prognosis depends on CLAD phenotype. We used machine learning computed tomography (CT) lung texture analysis tool at CLAD diagnosis for phenotyping and prognostication compared to radiologists' scoring.

Methods: This retrospective study included all adult first double-lung transplant patients (01/2010-12/2015) with CLAD (censored 12/2019) and inspiratory CT near CLAD diagnosis. The machine learning tool quantified ground-glass opacity, reticulation, hyperlucent lung, and pulmonary vessel volume (PVV). Two radiologists scored for ground-glass opacity, reticulation, consolidation, pleural effusion, air trapping and bronchiectasis. Receiver operating characteristic curve analysis was used to evaluate the diagnostic performance of machine learning and radiologist for CLAD phenotype. Multivariable Cox proportional-hazards regression analysis for allograft survival controlled for age, sex, native lung disease, cytomegalovirus serostatus, and CLAD phenotype (bronchiolitis obliterans syndrome [BOS] and restrictive allograft syndrome [RAS]/mixed).

Results: 88 patients were included (57 BOS, 20 RAS/mixed, and 11 unclassified/undefined) with CT a median 9.5 days from CLAD onset. Radiologist and machine learning parameters phenotyped RAS/mixed with PVV as the strongest indicator (AUC 0.85). Machine learning hyperlucent lung phenotyped BOS using only inspiratory CT (AUC=0.76). Radiologist and machine learning parameters predicted graft failure in the multivariable analysis, best with PVV (HR=1.23, 95%CI 1.05-1.44, p=0.01).

Conclusions: Machine learning discriminated between CLAD phenotypes on CT. Both radiologist and machine learning scoring were associated with graft failure, independent of CLAD phenotype. PVV, unique to machine learning, was the strongest in phenotyping and prognostication.

Key words: Lung Transplantation, Lung Allograft Rejection, Quantitative CT

Introduction

Chronic lung allograft dysfunction (CLAD) is the principal cause of graft failure in lung

transplant recipients. Bronchiolitis obliterans syndrome (BOS) is the most frequent CLAD

phenotype and is characterized by progressive obstructive lung disease [1]. A restrictive

phenotype, restrictive allograft syndrome (RAS), is now well recognized and is associated with

parenchymal fibrosis and a worse prognosis but is less frequent than BOS [2]. In addition to

these two phenotypes, the International Society of Heart and Lung Transplantation (ISHLT)

2019 consensus statement added a mixed phenotype and an undefined phenotype [1].

Furthermore, a small but significant proportion of patients may be unclassifiable using the

existing framework [3].

Imaging of CLAD is best documented for BOS [4]. On computed tomography (CT), BOS

is characterized by geographic regions of air trapping that manifest as hyperlucent lung,

accentuated by exhalation. Bronchial wall thickening occurs with bronchiectasis developing over

time, often in association with mucus plugging and clustered nodules [5]. The typical appearance

of RAS is consolidation and ground-glass opacity evolving into reticulation and traction

bronchiectasis with pleural and parenchymal fibrosis, more commonly in the upper lung zones

[5]. CT features have been associated with allograft survival in double lung transplant patients

using a radiologist-scored approach [3, 6-8] and when using quantitative measures of lung

density, volume, parametric response mapping, and airway measures [9-15].

The computer-aided lung informatics for pathology evaluation and rating (CALIPER)

tool is a validated machine learning (ML) algorithm that performs automated, quantitative

analysis of lung texture on CT [16]. It segments the lung from the chest and classifies each

portion of the lung as either normal, hyperlucent, reticular, ground-glass, or honeycomb texture. It has been principally used in the evaluation of idiopathic pulmonary fibrosis where it is superior to visual analysis in quantifying extent of disease and predicting survival [17-19]. The strongest predictor of survival is pulmonary vessel volume (PVV), a biomarker derived from the ML analysis with no direct visual correlate [20].

Although hyperlucent, ground-glass and reticular textures are typical features of diffuse lung disease in non-transplant patients, they are also reported features of CLAD. For this reason, we hypothesized that an existing ML analysis tool may be useful in the phenotyping of CLAD and in predicting graft survival. Indeed, quantitative assessment of chest CT has been identified as a key future direction in CLAD research [1] and we sought to address this need.

Materials and Methods

Patient Selection

This single-center retrospective study has been approved by the institutional research ethics board (REB# 15-9531-AE). All adult first double lung transplant patients transplanted between 2010 and 2015 were selected. Exclusion criteria included single-lung transplants, heart-lung transplants, and re-transplants. We then excluded those who died without CLAD, who were CLAD-free up to December 31, 2019, and those who had non-CLAD pulmonary function decline. Patients were required to have had a thin section CT within 100 days after CLAD onset or, if unavailable, within 28 days before onset. Clinical characteristics were extracted from the medical record and date of CLAD onset was derived. CLAD phenotype and date of onset was

assigned according to previously described methodology and considered the gold standard for sensitivity analyses below [3]. Graft failure was defined as death or re-transplantation.

We further collected virtual crossmatch status at the time of transplant (i.e. donor specific antibodies [DSA] at transplant), the presence of DSA at CLAD onset based on the last test before onset or not more than three weeks after onset, as well as development of *de novo* DSA between transplant and CLAD onset. The pre-CLAD infection score (number of pre-CLAD bronchoalveolar lavage with significant pathogens divided by the number of all pre-CLAD bronchoalveolar lavage) and pre-CLAD acute cellular rejection score (sum of histologic A-grades divided by the number of all evaluable biopsies) were also calculated.

Radiologic follow-up

Our routine lung transplant CT protocol is a low dose scan acquired at end-inspiration and followed by a minimal dose scan at end expiration. Generally, a low dose scan is performed at around one third the dose of a regular dose scan, and the minimal dose scan is performed at one tenth the dose of a regular scan. Surveillance CT scans are performed at three-month intervals for the first-year post-transplant and then at 18 and 24 months with CT also being performed as needed when CLAD is suspected. All patients were scanned on one of three multi-slice CT units (Aquilion, Aquilion ONE, or Aquilion PRIME, Canon, Japan) at Toronto General Hospital. For inclusion in the study, the CT scan had to be non-contrast and have a transverse image series in a mediastinal reconstruction algorithm with contiguous slices of a thickness ≤ 1 mm. Each scan was reviewed by a fellowship-trained thoracic radiologist (M.M., five years of experience) to assess for factors that may interfere with texture analysis including respiratory motion artifact and patient intubation.

Machine learning analysis

Qualifying CT scans were analyzed using the Lung Texture Analysis tool (Imbio LLC, Minneapolis, Minnesota) based on CALIPER with the development and training of the machine learning algorithm having been previously described [16]. This is a tool that performs a texture analysis in interstitial lung disease but was not trained on our lung transplant cohort. In brief, the lung is segmented from the chest wall and the large airways and central vasculature are removed. It then assigns a texture (normal lung [NL_{ML}], hyperlucent lung [HL_{ML}], ground-glass opacity [GGO_{ML}], reticular [RET_{ML}], or honeycombing) to each voxel and a color overlay output is made available for review on the picture archiving and communications system (Coral, Toronto, Ontario, Canada) (**figure 1**). Quantitative data in the form of total lung capacity (CT_{TLC}), volumes for each assigned texture, and pulmonary vessel volume (PVV_{ML}) are provided in a comma-separated values file output from a locally hosted server. A thoracic radiologist (M.M.) reviewed all CT scans with the output overlay present to exclude studies where segmentation failed in that extra-pulmonary structures comprised an estimated >1% of lung volume.

Radiologist analysis

CT scans were independently evaluated by two fellowship-trained cardiothoracic radiologists (RAD) each with five years of experience (G.R.K. and C.H.) and who were blinded to all clinical information, outcome, and ML results. A semi-quantitative scoring system was used as previously described by Suhling et al [6]. In brief, a CT image of the upper, mid, and lower lung was selected at a predetermined interval (25th, 50th, and 75th percentile image) and presented in lung windows (width 1500 Hounsfield units (HU), level -600 HU) and mediastinal windows (width 350 HU, level 40 HU). Each lung was evaluated for consolidation (CON_{RAD}), ground-glass opacity (GGO_{RAD}), reticulation (RET_{RAD}), traction bronchiectasis, and pleural

effusion using established definitions [21]. Each lung was also evaluated for air trapping (AT_{RAD}) using the same selected images paired with an expiratory image at the same level, when expiratory was performed. Each finding in each lung was scored on a grading of 0-2 where "0" meant the abnormality was not present, "1" it involved < 10% of the lung, and "2" it involved >10% of the lung. The radiologists were initially trained on a set of 18 images derived from six representative CLAD patients who were not included in this study.

Statistical analysis

Continuous variables were described using median and inter quartile range [IQR], or mean and standard deviation, and categorical variables using numbers and percentages. Comparisons between two groups were made by independent samples *t*-test or Mann-Whitney *U* test for continuous variables and Fisher's exact test or Chi square test for categorical variables. Variables analyzed by a t-test were evaluated for normality using histograms or a Shapiro-Wilks test, when appropriate.

For radiologist scoring, agreement was assessed with intraclass correlation coefficients (ICC) for GGO_{RAD}, RET_{RAD}, CON_{RAD} and AT_{RAD} and with weighted kappa for pleural effusion and bronchiectasis. The mean of the two radiologist's scores was used in analysis. For ML scoring, all variables were divided by CT_{TLC} to derive the proportion of total lung involved and are expressed as a percentage value. The relationship between AT_{RAD} and HL_{ML} was evaluated using the Pearson correlation coefficient. The relationship between CT_{TLC} and TLC from pulmonary function testing at the time of CLAD diagnosis was evaluated using the Pearson correlation coefficient.

The radiologist and ML scores were compared between CLAD phenotype groups using logistic regression models. RAS and mixed CLAD phenotypes were grouped to account for small numbers as were unclassified and undefined phenotypes. Receiver operating characteristic (ROC) curve analysis was used to determine the diagnostic performance for CLAD phenotyping.

Cox proportional-hazards (PH) regression analysis was used to assess the association of radiologist and ML scoring with graft failure and PH assumption validity was tested. Univariable and multivariable Cox PH regression analyses were performed for each score adjusted for the following covariates selected a priori based on their known association with post-lung transplant outcomes: age (per five-year intervals), male sex, native lung disease (COPD, cystic fibrosis, or interstitial lung disease), CMV recipient-donor serostatus matching, and CLAD phenotype (BOS, RAS/Mixed, or unclassified/undefined) [22]. For illustration purposes, Kaplan Meier curves were generated using the tertiles of each quantitative and semi-quantitative variable assessed. A p value of <0.05 was considered statistically significant. Analysis was conducted in R.4.0 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Patient cohort and CT acquisition

There were 211 patients with a diagnosis of CLAD, of whom 95 had an eligible CT (**figure 2**). ML segmented the lung in 93% of cases. The final CLAD cohort therefore consisted of 88 patients: 57 BOS, 12 RAS, 8 mixed, 8 undefined and 3 unclassified. The only significant difference across CLAD phenotypes in clinical characteristics was a longer time from CLAD

diagnosis to graft failure in patients with BOS (p=0.012) (**table 1**). Analysis of pre-CLAD

immunologic variables (infection score, rejection score, and de novo DSA) as well DSA status at

CLAD onset and virtual crossmatch status at transplant revealed no significant difference across

CLAD phenotypes (**supplementary table E1**).

CT was performed on the day of CLAD onset in 15 patients, in the first 100 days after

CLAD onset in 64 patients, and in the 28 days preceding CLAD onset in nine patients with an

overall median interval of 9.5 days after CLAD diagnosis. The CT_{TLC} was strongly correlated

with TLC (r=0.897, p<0.001) (**supplementary figure E1**) and technical parameters of the CT

examinations can be found in **supplementary table E2**.

Machine learning analysis

The proportion of HL_{ML} was highest in the BOS group (p<0.001). GGO_{ML}, RET_{ML} and

PVV_{ML} were highest in the RAS/Mixed group (p<0.001) (**table 2**). A visual representation of all

ML output in the form of glyphs is presented in **supplementary figures E2** and **E3** within the

online data supplement. No significant volume of honeycombing was identified (median 0.022%

[IQR 0.009%, 0.06%]) and it was therefore excluded from analyses.

ROC analysis demonstrated that GGO_{ML}, RET_{ML} and PVV_{ML} had strong diagnostic

capability in identifying the RAS/mixed phenotype (AUC=0.84-0.85) and this was strongest with

PVV_{ML}. Using a cut-point of 3%, PVV_{ML} had a sensitivity and specificity of 90.0% and 70.1%,

respectively (OR=2.08, p<0.001) (**table 3**). The ROC curves are provided in **supplementary

figure E4**. Phenotyping of BOS with HL_{ML} was achieved using inspiratory CT (AUC=0.76) and

did not utilize the expiratory CT to identify air trapping (**table 3**).

Radiologist scoring

Interrater reliability was good for GGO_{RAD}, RET_{RAD}, and CON_{RAD} (ICC=0.89, 0.81 and 0.84, respectively). There was moderate interrater reliability for pleural effusion (weighted kappa=0.60) and poor interrater reliability for bronchiectasis (weighted kappa=0.42) likely related to the infrequency of these findings. All scored features were highest in the RAS/mixed group (p<0.001) except bronchiectasis (**table 2**). ROC analysis demonstrated strong diagnostic capability for GGO_{RAD}, RET_{RAD}, and CON_{RAD} in identifying the RAS/mixed phenotype (**table 3**).

A subset of 50 patients had expiratory imaging available (34 BOS, 11 RAS/Mixed and 5 Undefined). Interrater reliability was moderate for AT_{RAD} (ICC=0.52) and among the lowest of the radiologic findings. AT_{RAD} was significantly higher in the BOS group (median 6, IQR 3.5-8) than in RAS/Mixed (median 2, IQR 1-5) and undefined (median 2, IQR 0.75-3, p=0.003) but only weakly correlated with HLML (r=0.189, p=0.188).

Allograft survival analysis

Univariable analysis of the baseline clinical variables showed that age (HR=0.89, p=0.003) and RAS/mixed phenotype (HR 2.24, p=0.008) were associated with time to graft failure after CLAD diagnosis (**table E3**). BOS phenotype was not associated with graft failure (HR 0.83, p=0.674) and was not included in the multivariable analysis. Univariable analysis of the radiologist and ML scores showed an association with graft failure for all variables assessed except NL_{ML} and HL_{ML} (**table 4**).

In multivariable analyses, radiologist scoring of GGO_{RAD}, RET_{RAD} and CON_{RAD} were independently associated with graft failure (**table 4**). ML scoring of RET_{ML} and PVV_{ML} were also independently associated with graft failure and GGO_{ML} approached statistical significance

(HR=1.36, 95%CI 1.00-1.86, p=0.050) (**table 4**). NL_{ML} and HL_{ML} were not associated with graft failure.

Although the proportion of NL_{ML} and HL_{ML} were not associated with graft failure, in their multivariable analyses there was an independent association of RAS/Mixed phenotype with graft failure (HR=2.73, 95%CI 1.42-5.28, p=0.00 and HR=2.43, 95%CI 1.21-4.85, p=0.01, respectively), and it was only in these two multivariable analyses that any independent association of CLAD phenotype with graft failure was identified. Details of the ML and radiologist multivariable analyses are provided in **supplementary table E4, E5 and E6**, within the online data supplement.

The proportion of each radiologic finding was stratified by tertiles and Kaplan-Meier curves were generated to facilitate visualization of lung allograft survival after CLAD diagnosis (**figure 3**). The greatest separation of all tertiles was seen with RET_{RAD} with 50% allograft survival at 439 days in the tertile with the most reticulation compared to 1134 days and 1700 days in the middle and lowest tertile, respectively (p=0.002).

Because of the heterogeneity seen in **supplementary figure E2 and E3** where some cases of BOS were seen to have GGO and reticulation, whereas other cases in the undefined/unclassified grouping were seen to contain hyperlucent lung, we empirically regrouped cases into three basic imaging patters: Group 1 – inflammatory presentation (GGO_{ML} and RET_{ML} >1% of CT_{TLC}) ii) Group 2 – hyperlucent presentation (HL_{ML} >10% of CT_{TLC}) and iii) Group 3 – indeterminate presentation (fulfilling neither Group 1 nor Group 2 criteria). The revised glyph grouping is presented in **supplementary figure E5** and Kaplan-Meier curves are presented in **supplementary figure E6.** Using this classification, the inflammatory group would be comprised of 16 RAS/Mixed, five undefined/unclassified, and 15 BOS. The indeterminate

group would be 20 BOS, four RAS/Mixed, and four undefined/unclassified. The hyperlucent group would be 22 BOS and two undefined/unclassified.

Discussion

We found an excellent diagnostic performance of the ML tool in identifying the RAS and mixed CLAD phenotype. Furthermore, ML identified hyperlucent lung in patients with BOS, without necessitating expiratory CT. Both ML and radiologist scoring at the time of CLAD onset were of prognostic importance, independent of CLAD phenotype in our multivariable model, emphasizing the importance of CT early in the evaluation of CLAD. PVV_{ML}, a biomarker unique to ML, had the strongest diagnostic and prognostic performance amongst all variables assessed.

We investigated PVV because it was found to be a strong predictor of mortality in diffuse lung disease [20] [23]. As a marker of total vessel volume in the lung, possible explanations for the significance of PVV in pulmonary fibrosis include redistribution of blood flow from abnormal lung to normal lung, the presence of arterial shunts, or a dilatation of blood vessels related to increased negative pressure at inspiration in the context of lung stiffness [20]. It is also possible that perivascular inflammatory opacity is being classified as vessel volume by ML. Although PVV increases as CT_{TLC} decreases, a reduction in lung volumes alone does not fully account for the significance of PVV as we controlled for CT_{TLC} in our analyses. PVV was expressed as a proportion of CT_{TLC}, and we also added CT_{TLC} into the PVV multivariable model.

In a previous study, histopathologic examination of explanted lungs with end-stage restrictive CLAD revealed focal areas of capillary obliteration with a resultant decreased microvascular density, a finding that would suggest a lower PVV in advanced CLAD [24]. However, the same authors also describe areas of reactive microvascular proliferation amongst fibrosis and ectasia of small lymphatic vessels that, in our estimation, might account for an increased PVV. In the absence of a definite histologic correlate at the onset of CLAD, the prognostic significance of increased PVV in CLAD could relate to increased negative intrathoracic pressure at end-inspiration, particularly in those patients with restriction and lung stiffness, and possibly the inclusion of perivascular inflammation in vessel segmentation, given that the analysis is not capable of reliably distinguishing between a true vessel versus a perivascular opacity. However, a further quantitative analysis of the small vessels in CLAD, both at onset and at the end stage, is warranted given the intriguing findings of microvascular damage in lungs with restrictive CLAD or RAS [24].

ML-identified hyperlucent lung is a feature associated with BOS on fully inspiratory exams and this likely corresponds to air-trapping in this cohort. Conventionally, radiologists assess for air-trapping using paired inspiratory and expiratory imaging and, using this approach to diagnose BOS, Bankier et al describe a threshold of 32% air-trapping as being 83% sensitive, 89% specific, and 88% accurate [25]. Indeed, radiologist assessed air-trapping may be insensitive particularly if there is poor timing of the CT scan and limited expiratory effort (**supplementary figure E7**). More recently, Verleden et al used parametric response mapping alongside a matched stable cohort and found a sensitivity of 62.5% and specificity of 93.8%. With ML, and amongst a cohort of only CLAD patients, we found a cut-off of just 0.81% hyperlucent lung to be 79% sensitive, 71% specific, and 76% accurate. The lower specificity in

our study is likely due to including patients with all forms of CLAD. Interestingly, patients with CLAD of a mixed and RAS phenotype may demonstrate lesions of obliterative bronchiolitis and indeed we found hyperlucent lung in all CLAD groups [26, 27].

In the multivariable analysis, we found that NL_{ML} and HL_{ML} were not independently associated with graft failure whereas CLAD phenotype defined clinically was. Despite this limitation of HL_{ML} as a marker of CLAD prognosis, the ability of ML to identify this important feature of CLAD on inspiratory CT warrants further investigation. Foregoing expiratory CT would offer a radiation dose reduction and improved efficiency in performing and reporting the CT studies.

Although the ML analysis presents unique insights in PVV and HL_{ML} quantification, the radiologist scoring also performed well in our study. The semi-quantitative radiologist scoring system in this study was used in a similarly sized lung transplant cohort by Suhling et al in 2016 who found that, by combining CT findings of consolidation and reticulation with TLC, they could identify a subgroup of patients with restrictive CLAD and poorer survival [6]. Their cohort differed substantially from ours in that one third had severe traction bronchiectasis, a marker of advanced fibrosis, and CT was performed a median of 387 days after CLAD onset. In contrast, CT was performed a median of 9.5 days after CLAD onset in the current study and only one case had severe bronchiectasis. Furthermore, Suhling et al found that GGO was not associated with graft failure whereas we found that it was and this may also relate to CT timing if GGO is attributable to acute inflammation near the onset of CLAD. Indeed, Philippot et al also applied the same scoring system to single lung transplant patients and found it useful in the early detection of RAS [28]. Neither study controlled for CLAD phenotype in their multivariable

analysis and our finding of opacities as independent predictors of graft failure suggest that a quantitative assessment of the lung opacities, achieved through ML, is of value.

Although we demonstrate a strong ability of ML to phenotype CLAD and its determination of graft survival, there are several other potential uses. The quantitative data provided by ML analysis may be of use in future clinical drug trials and treatment monitoring where there is a need for reproducible biomarkers of disease. With validation, the use of an automated ML tool in lung transplant studies would allow improved comparison of cohorts across centers. It should also be noted that ML is sensitive to fractional volumes of lung abnormality, the optimal cut-off for HL_{ML} in the current study being 0.81%. This finding raises the question of the ability for ML to detect lung texture abnormalities in CLAD at its earliest stage, before becoming clinically evident.

Limitations of this study include the single center and retrospective design, although a CALIPER based ML analysis is compatible with a wide range of CT protocols. Furthermore, we were restricted in our ability to include a larger number of patients due to the absence of archived CT thin slices prior to 2013. Despite this, our cohort is among the largest with a comprehensive CT analysis in CLAD. Finally, this ML tool was not trained on lung transplant cases and does not capture all features of particular interest in lung transplant patients, such as pleural thickening or complications at the bronchial anastomoses. A future ML tool trained on and tailored for lung transplant patients might be of even greater value. Of course, the clinician and radiologist remain critically important in evaluating the CT and considering the findings in the appropriate clinical context.

In conclusion, ML strongly discriminated between CLAD phenotypes using automated analysis. Both radiologist and machine learning scoring were associated with graft failure,

independent of the CLAD phenotype, and without using expiratory CT. Integration of ML into clinical use may facilitate automated CT analysis for phenotyping and prognostication in a reliable and reproducible manner. PVV, a biomarker unique to ML, was best in CLAD phenotyping and prognostication and warrants future investigation in this population.

References

1.	Verleden GM, Glanville AR, Lease ED, Fisher AJ, Calabrese F, Corris PA, Ensor CR, Gottlieb J, Hachem RR, Lama V, Martinu T, Neil DAH, Singer LG, Snell G, Vos R. Chronic lung allograft dysfunction: Definition, diagnostic criteria, and approaches to treatment-A consensus report from the Pulmonary Council of the ISHLT. *J Heart Lung Transplant* 2019: 38(5): 493-503.

2.	Sato M, Waddell TK, Wagnetz U, Roberts HC, Hwang DM, Haroon A, Wagnetz D, Chaparro C, Singer LG, Hutcheon MA, Keshavjee S. Restrictive allograft syndrome (RAS): a novel form of chronic lung allograft dysfunction. *J Heart Lung Transplant* 2011: 30(7): 735-742.

3.	Levy L, Huszti E, Renaud-Picard B, Berra G, Kawashima M, Takahagi A, Fuchs E, Ghany R, Moshkelgosha S, Keshavjee S, Singer LG, Tikkanen J, Martinu T. Risk assessment of chronic lung allograft dysfunction phenotypes: Validation and proposed refinement of the 2019 International Society for Heart and Lung Transplantation classification system. *J Heart Lung Transplant* 2020.

4.	Worthy SA, Park CS, Kim JS, Müller NL. Bronchiolitis obliterans after lung transplantation: high-resolution CT findings in 15 patients. *AJR Am J Roentgenol* 1997: 169(3): 673-677.

5.	Hota P, Dass C, Kumaran M, Simpson S. High-Resolution CT Findings of Obstructive and Restrictive Phenotypes of Chronic Lung Allograft Dysfunction: More Than Just Bronchiolitis Obliterans Syndrome. *AJR Am J Roentgenol* 2018: 211(1): W13-W21.

6. Suhling H, Dettmer S, Greer M, Fuehner T, Avsar M, Haverich A, Welte T, Gottlieb J. Phenotyping Chronic Lung Allograft Dysfunction Using Body Plethysmography and Computed Tomography. *Am J Transplant* 2016: 16(11): 3163-3170.

7. Dettmer S, Shin HO, Vogel-Claussen J, Westphal M, Haverich A, Warnecke G, Welte T, Wacker F, Gottlieb J, Suhling H. CT at onset of chronic lung allograft dysfunction in lung transplant patients predicts development of the restrictive phenotype and survival. *Eur J Radiol* 2017: 94: 78-84.

8. Dubbeldam A, Barthels C, Coolen J, Verschakelen JA, Verleden SE, Vos R, Verleden GM, De Wever W. Restrictive allograft syndrome after lung transplantation: new radiological insights. *Eur Radiol* 2017: 27(7): 2810-2817.

9. Horie M, Levy L, Houbois C, Salazar P, Saito T, Pakkal M, O'Brien C, Sajja S, Brock K, Yasufuku K, Keshavjee S, Paul N, Martinu T. Lung Density Analysis Using Quantitative Chest CT for Early Prediction of Chronic Lung Allograft Dysfunction. *Transplantation* 2019: 103(12): 2645-2653.

10. Saito T, Horie M, Sato M, Nakajima D, Shoushtarizadeh H, Binnie M, Azad S, Hwang DM, Machuca TN, Waddell TK, Singer LG, Cypel M, Liu M, Paul NS, Keshavjee S. Low-dose computed tomography volumetry for subtyping chronic lung allograft dysfunction. *J Heart Lung Transplant* 2016: 35(1): 59-66.

11. Horie M, Salazar P, Saito T, Binnie M, Brock K, Yasufuku K, Azad S, Keshavjee S, Martinu T, Paul N. Quantitative chest CT for subtyping chronic lung allograft dysfunction and its association with survival. *Clin Transplant* 2018: 32(5): e13233.

12.   Belloli EA, Degtiar I, Wang X, Yanik GA, Stuckey LJ, Verleden SE, Kazerooni EA, Ross BD, Murray S, Galbán CJ, Lama VN. Parametric Response Mapping as an Imaging Biomarker in Lung Transplant Recipients. *Am J Respir Crit Care Med* 2017: 195(7): 942-952.

13.   Verleden SE, Vos R, Vandermeulen E, Ruttens D, Bellon H, Heigl T, Van Raemdonck DE, Verleden GM, Lama V, Ross BD, Galbán CJ, Vanaudenaerde BM. Parametric Response Mapping of Bronchiolitis Obliterans Syndrome Progression After Lung Transplantation. *Am J Transplant* 2016: 16(11): 3262-3269.

14.   Nascimento DZ, Watte G, Torres FS, Schio SM, Sanchez L, de Sousa JLM, Perin FA, Verma N, Mohammed TH, Hochhegger B. Utilization of Quantitative Computed Tomography Assessment to Identify Bronchiolitis Obliterans Syndrome After Single Lung Transplantation. *Lung* 2021: 199(1): 29-35.

15.   Barbosa EJM, Jr., Lanclus M, Vos W, Van Holsbeke C, De Backer W, De Backer J, Lee J. Machine Learning Algorithms Utilizing Quantitative CT Features May Predict Eventual Onset of Bronchiolitis Obliterans Syndrome After Lung Transplantation. *Acad Radiol* 2018: 25(9): 1201-1212.

16.   Bartholmai BJ, Raghunath S, Karwoski RA, Moua T, Rajagopalan S, Maldonado F, Decker PA, Robb RA. Quantitative computed tomography imaging of interstitial lung diseases. *J Thorac Imaging* 2013: 28(5): 298-307.

17.   Jacob J, Bartholmai BJ, Rajagopalan S, Kokosi M, Nair A, Karwoski R, Raghunath SM, Walsh SL, Wells AU, Hansell DM. Automated Quantitative Computed Tomography Versus Visual Computed Tomography Scoring in Idiopathic Pulmonary Fibrosis: Validation Against Pulmonary Function. *J Thorac Imaging* 2016: 31(5): 304-311.

18.   Maldonado F, Moua T, Rajagopalan S, Karwoski RA, Raghunath S, Decker PA, Hartman TE, Bartholmai BJ, Robb RA, Ryu JH. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J* 2014: 43(1): 204-212.

19.   Lee SM, Seo JB, Oh SY, Kim TH, Song JW, Kim N. Prediction of survival by texture-based automated quantitative assessment of regional disease patterns on CT in idiopathic pulmonary fibrosis. *Eur Radiol* 2018: 28(3): 1293-1300.

20.   Jacob J, Bartholmai BJ, Rajagopalan S, Kokosi M, Nair A, Karwoski R, Walsh SL, Wells AU, Hansell DM. Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures. *Eur Respir J* 2017: 49(1).

21.   Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008: 246(3): 697-722.

22.   Chambers DC, Cherikh WS, Harhay MO, Hayes D, Jr., Hsich E, Khush KK, Meiser B, Potena L, Rossano JW, Toll AE, Singh TP, Sadavarte A, Zuckermann A, Stehlik J, International Society for H, Lung T. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-sixth adult lung and heart-lung transplantation Report-2019; Focus theme: Donor and recipient size match. *J Heart Lung Transplant* 2019: 38(10): 1042-1055.

23.   Jacob J, Hirani N, van Moorsel CHM, Rajagopalan S, Murchison JT, van Es HW, Bartholmai BJ, van Beek FT, Struik MHL, Stewart GA, Kokosi M, Egashira R, Brun AL, Cross G, Barnett J, Devaraj A, Margaritopoulos G, Karwoski R, Renzoni E, Maher TM, Wells AU. Predicting outcomes in rheumatoid arthritis related interstitial lung disease. *Eur Respir J* 2019: 53(1).

24.   von der Thusen JH, Vandermeulen E, Vos R, Weynand B, Verbeken EK, Verleden SE. The histomorphological spectrum of restrictive chronic lung allograft dysfunction and implications for prognosis. *Mod Pathol* 2018: 31(5): 780-790.

25.   Bankier AA, Van Muylem A, Knoop C, Estenne M, Gevenois PA. Bronchiolitis obliterans syndrome in heart-lung transplant recipients: diagnosis with expiratory CT. *Radiology* 2001: 218(2): 533-539.

26.   Ofek E, Sato M, Saito T, Wagnetz U, Roberts HC, Chaparro C, Waddell TK, Singer LG, Hutcheon MA, Keshavjee S, Hwang DM. Restrictive allograft syndrome post lung transplantation is characterized by pleuroparenchymal fibroelastosis. *Mod Pathol* 2013: 26(3): 350-356.

27.   Verleden SE, Vasilescu DM, McDonough JE, Ruttens D, Vos R, Vandermeulen E, Bellon H, Geenens R, Verbeken EK, Verschakelen J, Van Raemdonck DE, Wuyts WA, Sokolow Y, Knoop C, Cooper JD, Hogg JC, Verleden GM, Vanaudenaerde BM. Linking clinical phenotypes of chronic lung allograft dysfunction to changes in lung structure. *Eur Respir J* 2015: 46(5): 1430-1439.

28.   Philippot Q, Debray MP, Bun R, Frija-Masson J, Bunel V, Morer L, Roux A, Picard C, Jebrak G, Dauriat G, Castier Y, Cazes A, Mal H, Taupin JL, Couffignal C, Brugiere O. Use of CT-SCAN score and volume measures to early identify restrictive allograft syndrome in single lung transplant recipients. *J Heart Lung Transplant* 2020: 39(2): 125-133.

Figure Legends
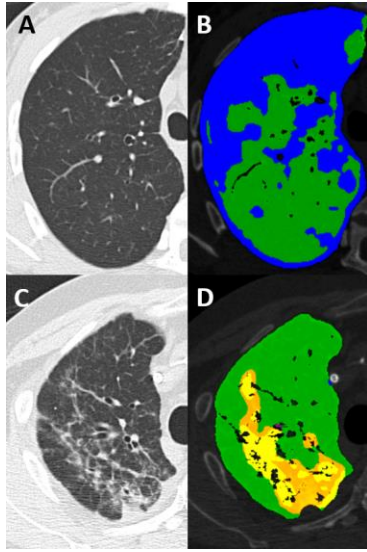
Figure 1. Inspiratory computed tomography (CT) and machine learning output in chronic lung allograft dysfunction. In a 36-year-old woman with bronchiolitis obliterans syndrome, (A) conventional CT analysis demonstrates normal appearing lung which, on texture analysis (B), is a mix of normal (green) and hyperlucent (blue). In a 40-year-old man with a mixed phenotype, (C) conventional CT analysis demonstrates ground-glass opacity and reticulation that, on texture analysis (D) is also classified as a mixture of ground-glass opacity (yellow) and reticulation (orange).
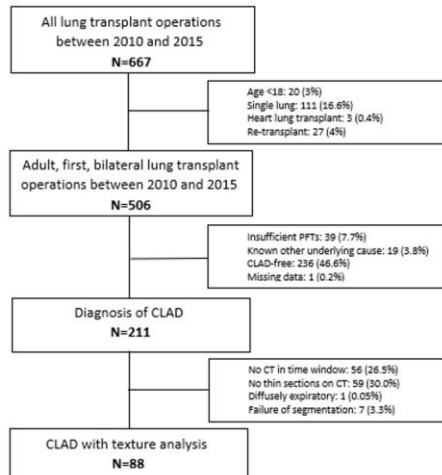
Figure 2. Flow chart of patient inclusion criteria. Failure of segmentation related to classifying trachea (four cases), chest wall (two cases) or abdominal fat (one case) as lung. CLAD – chronic lung allograft dysfunction; CT – computed tomography. PFTs – pulmonary function tests.
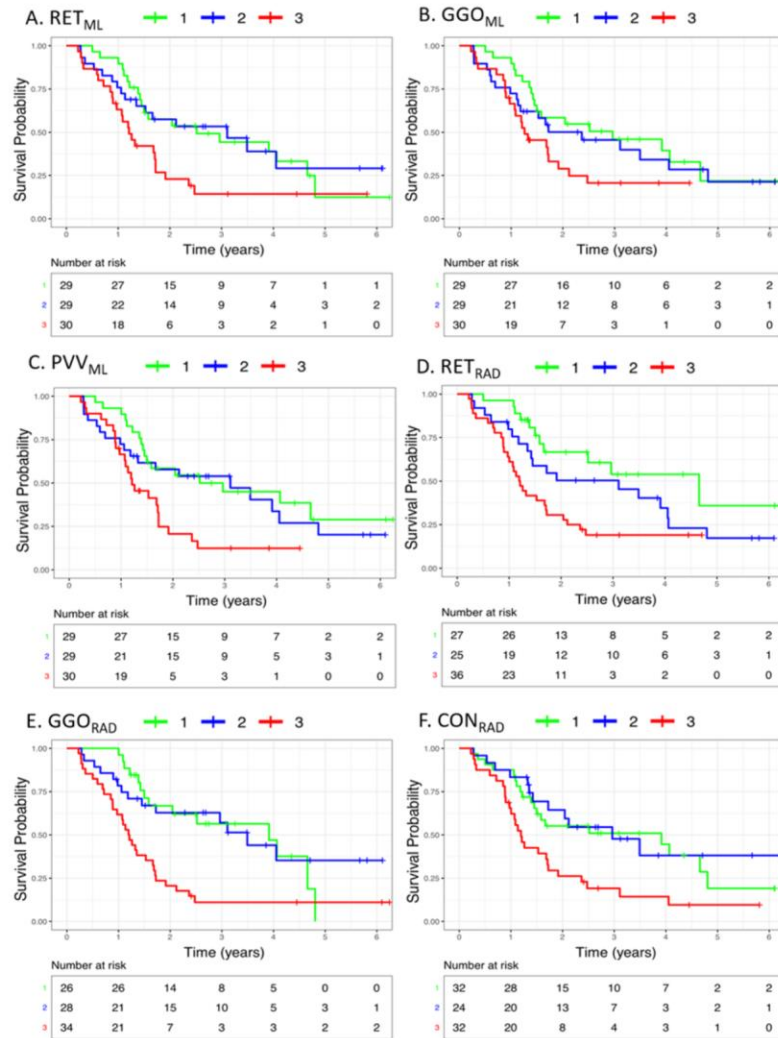
Figure 3. Kaplan-Meier curves demonstrating the allograft survival probability over time for (A) RET$_{ML}$, (B) GGO$_{ML}$, (C) PVV$_{ML}$, (D) GGO$_{RAD}$, (E) RET$_{RAD}$, and (F) CON$_{RAD}$. The red line corresponds to the highest tertile of radiologic abnormality, the blue line is the middle tertile, and the green is the tertile with the least amount of radiologic abnormality. CON – consolidation; GGO – ground-glass opacity; ML – machine learning; PVV – pulmonary vessel volume; RAD – radiologist; RET – reticulation.

Tables

Characteristic	BOS (n=57)	RAS/Mixed (n=12/8)	Unclassified/ Undefined (n=11)	P-value
Age at transplant, years +/- SD	47.81 +/- 15.41	45.45 +/- 17.39	48.09 +/- 17.02	0.840
Sex, n (%)				0.526
Female	31 (54.4)	8 (40.0)	6 (54.5)	
Male	26 (45.6)	12 (60.0)	5 (45.5)	
Native lung disease				0.450
ILD	20 (35.1)	10 (50.0)	1 (9.1)	
Cystic fibrosis	16 (28.1)	4 (20.0)	4 (36.4)	
COPD	14 (24.6)	3 (15.0)	4 (36.4)	
Other	7 (12.3)	3 (15.0)	2 (18.2)	
Time to CLAD onset (median [IQR], days)	700 [367, 1171]	624 [425, 1125]	804 [445, 1370]	0.765
Graft failure				0.803
Death	27 (77.1)	13 (76.5)	6 (66.7)	
Re-transplant	8 (22.9)	4 (23.5)	3 (33.3)	
CMV				0.638
D-R-	8(14.0)	1 (5.0)	1 (9.1)	
D+R-	15 (26.3)	4 (20.0)	4 (36.4)	
R+	34 (59.6)	15 (75.0)	6 (54.5)	
Time from CLAD onset to graft failure (median [IQR], days)	746 [445, 1135]	370 [250, 574]	520 [429, 694]	**0.012**

TABLE 1. Patient characteristics by chronic lung allograft dysfunction phenotype. CLAD – chronic lung allograft dysfunction; BOS – bronchiolitis obliterans syndrome; CMV – cytomegalovirus; RAS – restrictive allograft syndrome; COPD – chronic obstructive pulmonary disease; ILD – interstitial lung disease.

	BOS (n=57)	RAS/Mixed (n=12/8)	Undefined/Unclassified (n=11)	P-value
Radiologist scoring (median score [IQR])				
Consolidation	0.50 [0.00, 1.00]	3.00 [1.50, 5.00]	0.50 [0.00, 1.25]	**<0.001**
Ground-glass	4.00 [2.00, 6.50]	9.00 [7.38, 10.50]	4.00 [1.50, 7.25]	**<0.001**
Reticulation	2.50 [0.50, 6.00]	8.00 [6.50, 9.62]	3.00 [0.75, 6.50]	**<0.001**
Bronchiectasis	0.00 [0.00, 0.50]	0.00 [0.00, 0.62]	0.00 [0.00, 0.00]	0.298
Pleural effusion	0.00 [0.00, 0.00]	1.00 [0.50, 1.88]	0.00 [0.00, 0.50]	**<0.001**
Machine learning (median % [IQR])				
Normal	88.74 [77.14, 94.51]	89.50 [81.59, 93.24]	94.45 [79.59, 96.21]	0.534
Hyperlucent	5.75 [1.00, 19.88]	0.12 [0.05, 1.51]	0.68 [0.23, 5.81]	**<0.001**
Ground-glass	0.45 [0.17, 1.70]	6.43 [1.67, 13.08]	0.85 [0.22, 2.67]	**<0.001**
Reticular	0.53 [0.23, 1.46]	2.35 [1.43, 3.63]	1.15 [0.22, 2.01]	**<0.001**
PVV	2.27 [1.95, 3.14]	4.06 [3.28, 5.25]	2.29 [1.92, 3.31]	**<0.001**

TABLE 2. Radiologist and machine learning analysis of chest CT scans tabulated by chronic lung allograft dysfunction phenotypes. BOS – bronchiolitis obliterans syndrome; PVV – pulmonary vessel volume ; RAS – restrictive allograft syndrome.

	Receiver Operating Characteristic Curve Analysis					Univariable logistic regression	
Texture	Optimal Cut-point	Sensitivity	Specificity	Accuracy	AUC	Odds Ratio	P-value
HL$_{ML}$	0.81%	0.79	0.71	0.76	0.763	1.11	**0.006**
GGO$_{ML}$	0.79%	1.00	0.56	0.66	0.845	1.12	**0.004**
RET$_{ML}$	0.93%	1.00	0.65	0.73	0.835	1.59	**0.003**
PVV$_{ML}$	3.02%	0.90	0.71	0.75	0.851	2.08	**<0.001**
GGO$_{RAD}$	7.00	0.80	0.79	0.80	0.829	1.44	**0.001**
RET$_{RAD}$	6.00	0.90	0.74	0.77	0.838	1.42	**0.001**
CON$_{RAD}$	1.50	0.80	0.76	0.77	0.820	1.88	**0.001**

TABLE 3. Univariable logistic regression and receiver operating curve analysis for the proportion of machine learning and radiologist variables in bronchiolitis obliterans syndrome phenotype for hyperlucent lung (HL$_{ML}$, first row) and restrictive allograft dysfunction/mixed phenotype for all other variables. AUC – Area under the curve; CLAD – chronic lung allograft dysfunction; CON – consolidation; GGO – ground-glass opacity; HL – hyperlucent lung; ML – machine learning; NL – normal lung; PVV – pulmonary vessel volume; RAD – radiologist; RET – reticulation.

	Univariable analysis			Multivariable analysis		
Radiologic variable	Hazard ratio	95% CI	P-value	Hazard ratio	95% CI	P-value
NL_{ML}	0.93	0.78-1.10	0.37	0.96	0.80-1.15	0.65
HL_{ML}	0.92	0.76-1.12	0.40	0.90	0.72-1.13	0.37
GGO_{ML}	1.45	1.12-1.88	**<0.001**	1.36	1.00-1.86	0.05
RET_{ML}	1.25	1.11-1.39	**<0.001**	1.20	1.05-1.37	**0.01**
PVV_{ML}	1.30	1.14-1.48	**<0.001**	1.23	1.05-1.44	**0.01**
GGO_{RAD}	1.16	1.08-1.25	**<0.001**	1.14	1.05-1.23	**<0.001**
RET_{RAD}	1.18	1.10-1.26	**<0.001**	1.17	1.08-1.27	**<0.001**
CON_{RAD}	1.26	1.12-1.41	**<0.001**	1.16	1.01-1.34	**0.04**

TABLE 4. Univariable and multivariable analysis of automated and semi-quantitative scoring for the highest tertile of each radiologic abnormality, for the diagnosis of graft failure (death or re-transplantation). The multivariable model was performed for each radiological variable separately and is adjusted for sex, CLAD phenotype, age, native lung disease and CMV serostatus matching. CLAD – chronic lung allograft dysfunction; CON – consolidation; GGO – ground-glass opacity; HL – hyperlucent lung; ML – machine learning; NL – normal lung; PVV – pulmonary vessel volume; RAD – radiologist; RET – reticulation.

CHRONIC LUNG ALLOGRAFT DYSFUNCTION PHENOTYPE AND PROGNOSIS BY MACHINE LEARNING CT ANALYSIS

Online Data Supplements

Characteristic	BOS (n=57)	RAS/Mixed (n=12/8)	Undefined/ Unclassified (n=11)	P-value
Pre-CLAD infection score, mean±SD	0.27±0.23	0.32±0.25	0.22±0.29	0.545
Pre-CLAD acute cellular rejection score, mean±SD*	0.30±0.33	0.31±0.26	0.40±0.38	0.620
Positive DSA at CLAD onset, n (%)	22 (38.6)	10 (50)	6 (54.5)	0.485
Positive pre-CLAD dnDSA, n (%)	33 (57.9)	11 (55)	6 (54.5)	0.962
Positive virtual crossmatch at transplant, n (%)**	9 (15.8)	1 (5.3)	4 (40)	0.054

TABLE E1. Pre-CLAD immunologic events by CLAD phenotype. Infection score was calculated as the sum of bronchoalveolar lavage samples with at least one significant pathogen on culture, divided by the total number of samples. Acute cellular rejection score was calculated as the sum of histologic acute cellular rejection A-grades, divided by the number of evaluable biopsies. CLAD – chronic lung allograft dysfunction; BOS – bronchiolitis obliterans syndrome; DSA – donor-specific antibodies; dnDSA – de novo donor-specific antibodies; RAS – restrictive allograft syndrome; Tx – transplant. *Data absent in two patients. **Unknown in one patient with RAS and one patient with undefined CLAD.

Computed Tomography Parameter	
Scanner model, n (%)	
Aquilion	57 (64.8)
Aquilion ONE	22 (25)
Aquilion PRIME	9 (10.2)
Reconstruction kernel, n (%)	
FC03	56 (63.6)
FC04	31 (35.2)
FC11	1 (1.1)
Slice thickness 0.5 mm, n (%)	88 (100)
Tube current (mean ± SD, mAs)	49.8 ± 29.9
Tube voltage	
120 kVp	7 (8)
135 kVp	81 (92)
Effective dose (median [IQR], mSv)	2.0 [1.6, 2.5]*

TABLE E2. Computed tomography parameters for all cases included in the final cohort. *The effective dose includes both the inspiratory and expiratory CT. The expiratory CT accounts for approximately 25% of the total dose and was not required for machine learning analysis.

Characteristic	Hazard ratio	95% Confidence Interval	P-value
Age (per 5 years)	0.89	0.82-0.96	**0.003**
Sex (male)	1.10	0.66-1.83	0.724
Native lung disease – COPD	0.96	0.38-2.45	0.936
Native lung disease – CF	1.89	0.80-4.46	0.148
Native lung disease – ILD	0.99	0.41-2.36	0.976
CMV matching D+R-	0.89	0.37-2.15	0.791
CMV matching R+	0.80	0.36-1.82	0.600
CLAD - RAS/Mixed	2.24	1.24-4.08	**0.008**
CLAD – Unclassified/Undefined	1.41	0.65-3.05	0.383
CLAD – BOS	0.83	0.34-2.00	0.674

TABLE E3. Univariable Cox proportional hazards model assessing the association of each confounder variable with time to graft failure (death or re-transplantation). CLAD – chronic lung allograft dysfunction; CF – cystic fibrosis; COPD – chronic obstructive pulmonary disease; CMV – cytomegalovirus; D+R- – donor positive and recipient negative; R+ – recipient positive; ILD – interstitial lung disease; RAS – restrictive allograft syndrome.

|	Normal lung		HL$_{ML}$		GGO$_{ML}$		RET$_{ML}$		PVV$_{ML}$	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
ML variable	0.96 (0.80-1.15)	0.65	0.90 (0.72-1.13)	0.37	1.36 (1.00-1.86)	0.05	1.20 (1.05-1.37)	**0.01**	1.23 (1.05-1.44)	**0.01**
CLAD-RAS/Mixed	2.73 (1.42-5.28)	**0.00**	2.43 (1.21-4.85)	**0.01**	2.02 (0.96-4.23)	0.06	1.69 (0.78-3.68)	0.18	1.63 (0.74-3.59)	0.23
CLAD-UNC/UND	1.44 (0.62-3.35)	0.39	1.26 (0.54-2.95)	0.59	1.27 (0.55-2.95)	0.57	1.36 (0.59-3.12)	0.47	1.33 (0.58-3.06)	0.51
Male Sex	1.15 (0.66-2.00)	0.63	1.12 (0.65-1.95)	0.69	1.26 (0.72-2.22)	0.42	1.18 (0.67-2.08)	0.56	1.18 (0.67-2.08)	0.56
Age	0.87 (0.73-1.02)	0.08	0.87 (0.74-1.02)	0.09	0.86 (0.73-1.01)	0.07	0.85 (0.72-1.00)	0.05	0.87 (0.73-1.02)	0.09
NLD-COPD	1.37 (0.46-4.07)	0.57	1.36 (0.46-4.00)	0.58	1.36 (0.45-4.07)	0.58	1.32 (0.44-3.92)	0.62	1.36 (0.46-4.08)	0.58
NLD-CF	1.10 (0.37-3.23)	0.86	1.26 (0.43-3.73)	0.67	1.09 (0.37-3.18)	0.88	1.02 (0.35-3.02)	0.97	1.17 (0.40-3.42)	0.77
NLD-ILD	1.06 (0.40-2.77)	0.91	1.08 (0.41-2.83)	0.88	1.07 (0.41-2.80)	0.89	1.25 (0.48-3.26)	0.64	1.23 (0.47-3.21)	0.68
CMV D+R-	0.72 (0.28-1.83)	0.49	0.73 (0.29-1.88)	0.52	0.79 (0.31-2.03)	0.62	0.88 (0.34-2.30)	0.80	0.92 (0.35-2.43)	0.87
CMV R+	0.70 (0.30-1.62)	0.40	0.70 (0.30-1.63)	0.41	0.73 (0.31-1.72)	0.47	0.80 (0.34-1.90)	0.62	0.83 (0.35-1.97)	0.68

TABLE E4. Details of multivariable analysis results of the machine learning variables using a Cox proportional hazards regression model for graft failure (re-transplantation or death). Each column represents a separate multivariable analysis for the machine learning variable at the column header. A p-value of <0.05 was considered significant. CLAD – chronic lung allograft dysfunction; CF – cystic fibrosis; COPD – chronic obstructive pulmonary disease; CMV – cytomegalovirus; D+R- – donor positive and recipient negative; R+ – recipient positive; GGO – ground-glass opacity texture; HL$_{ML}$ – hyperlucent lung texture; ILD – interstitial lung disease; ML – machine learning; NLD – native lung disease; PVV – pulmonary vessel volume; RET – reticular texture; RAS – restrictive allograft syndrome; UNC – unclassified; UND – undefined.

|	GGO$_{RAD}$		RET$_{RAD}$		CON$_{RAD}$	
	HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
Radiologist variable	1.14 (1.05-1.23)	**0.00**	1.17 (1.08-1.27)	**0.00**	1.16 (1.01-1.34)	**0.04**
CLAD-RAS/Mixed	1.69 (0.81-3.51)	0.16	1.34 (0.63-2.86)	0.45	1.85 (0.82-4.14)	0.14
CLAD-UNC/UND	1.20 (0.50-2.91)	0.69	1.14 (0.48-2.72)	0.77	1.52 (0.66-3.50)	0.32
Male Sex	1.06 (0.59-1.91)	0.85	0.99 (0.55-1.80)	0.98	1.17 (0.66-2.06)	0.60
Age	0.85 (0.72-1.00)	0.05	0.82 (0.69-0.97)	**0.02**	0.85 (0.72-1.01)	0.06
NLD-COPD	1.65 (0.53-5.11)	0.38	1.76 (0.57-5.50)	0.33	1.45 (0.48-4.39)	0.51
NLD-CF	1.15 (0.40-3.33)	0.80	0.96 (0.33-2.78)	0.94	1.00 (0.34-2.92)	0.99
NLD-ILD	1.27 (0.48-3.34)	0.63	1.41 (0.53-3.71)	0.49	1.23 (0.47-3.24)	0.68
CMV D+R-	0.81 (0.31-2.15)	0.68	0.99 (0.37-2.68)	0.99	0.84 (0.35-2.21)	0.72
CMV R+	0.80 (0.34-1.91)	0.62	0.91 (0.38-2.21)	0.84	0.81 (0.34-1.94)	0.63

TABLE E5. Details of multivariable analysis results of the radiologist scoring using a Cox proportional hazards regression model for graft failure (re-transplantation or death). Each column represents a separate multivariable analysis for the radiologist variable at the column header. A p-value of <0.05 was considered significant. CLAD – chronic lung allograft dysfunction; CF – cystic fibrosis; COPD – chronic obstructive pulmonary disease; CMV – cytomegalovirus; D+R- – donor positive and recipient negative; R+ – recipient positive; GGO – ground-glass opacity texture; ILD – interstitial lung disease; NLD – native lung disease; RAD – radiologist; RET – reticular texture; RAS – restrictive allograft syndrome; UNC – unclassified; UND – undefined.

| Radiologic variable	Univariable analysis				Multivariable analysis			
	Hazard ratio	95% CI	P-value	PH validity P-value	Hazard ratio	95% CI	P-value	PH validity P-value
NL_{ML}	0.93	0.78-1.10	0.37	0.6	0.96	0.80-1.15	0.65	0.421
HL_{ML}	0.92	0.76-1.12	0.40	0.67	0.90	0.72-1.13	0.37	0.288
GGO_{ML}	1.45	1.12-1.88	**<0.001**	0.15	1.36	1.00-1.86	0.05	0.151
RET_{ML}	1.25	1.11-1.39	**<0.001**	0.54	1.20	1.05-1.37	**0.01**	0.230
PVV_{ML}	1.30	1.14-1.48	**<0.001**	0.1	1.23	1.05-1.44	**0.01**	0.054
GGO_{RAD}	1.16	1.08-1.25	**<0.001**	0.34	1.14	1.05-1.23	**<0.001**	0.434
RET_{RAD}	1.18	1.10-1.26	**<0.001**	0.37	1.17	1.08-1.27	**<0.001**	0.46
CON_{RAD}	1.26	1.12-1.41	**<0.001**	0.2	1.16	1.01-1.34	**0.04**	0.13

TABLE E6. Cox proportional hazards model assumption validity for each radiologic variable with the associated P-values. CON – consolidation; GGO – ground-glass opacity; HL – hyperlucent lung; ML – machine learning; NL – normal lung; PVV – pulmonary vessel volume; RAD – radiologist; RET – reticulation.
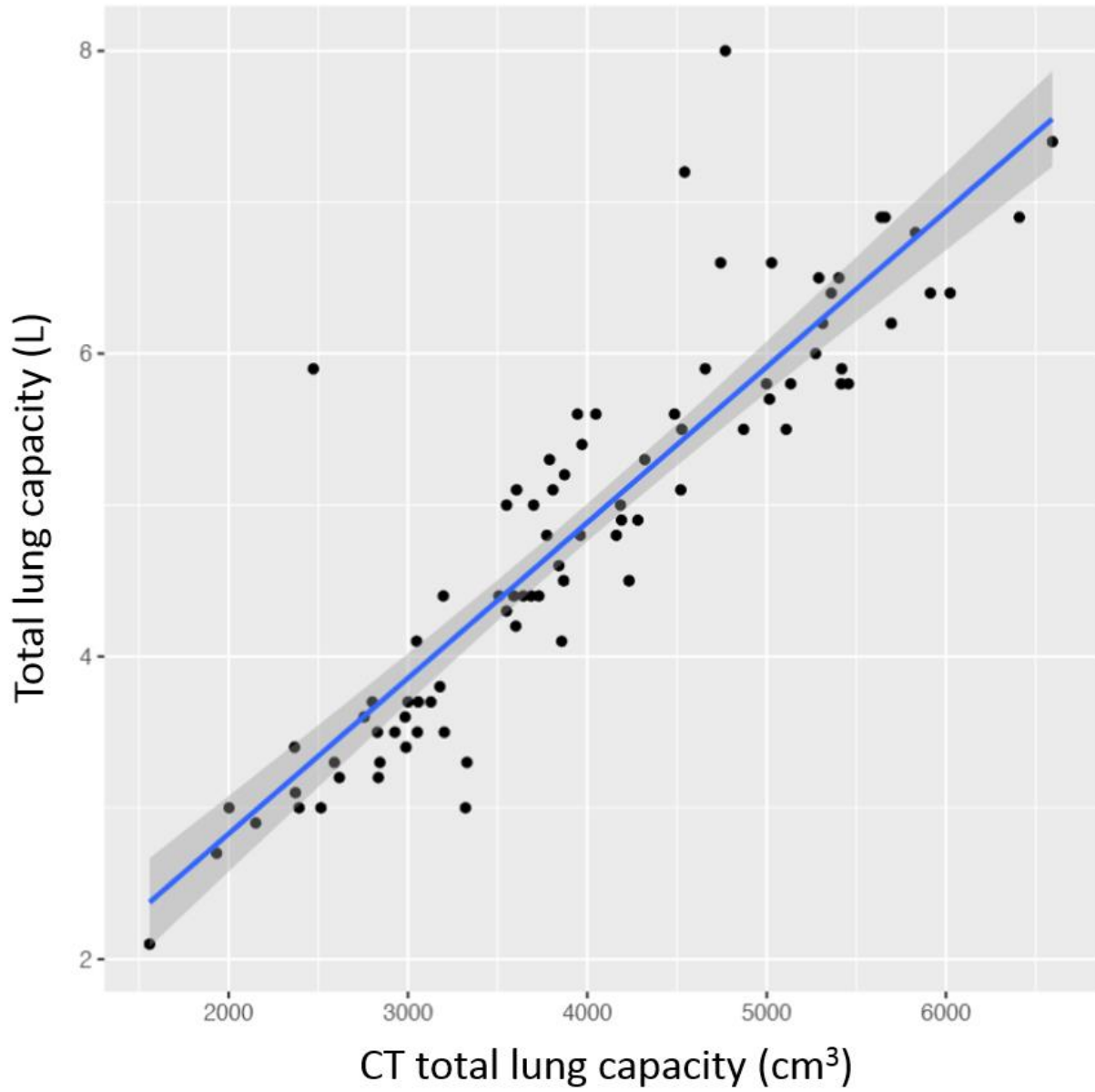
FIGURE E1. Strong correlation of total lung capacity derived by CT lung segmentation and total lung capacity derived by pulmonary function testing at the time of chronic lung allograft dysfunction diagnosis (Pearson correlation coefficient r=0.897, p<0.001).
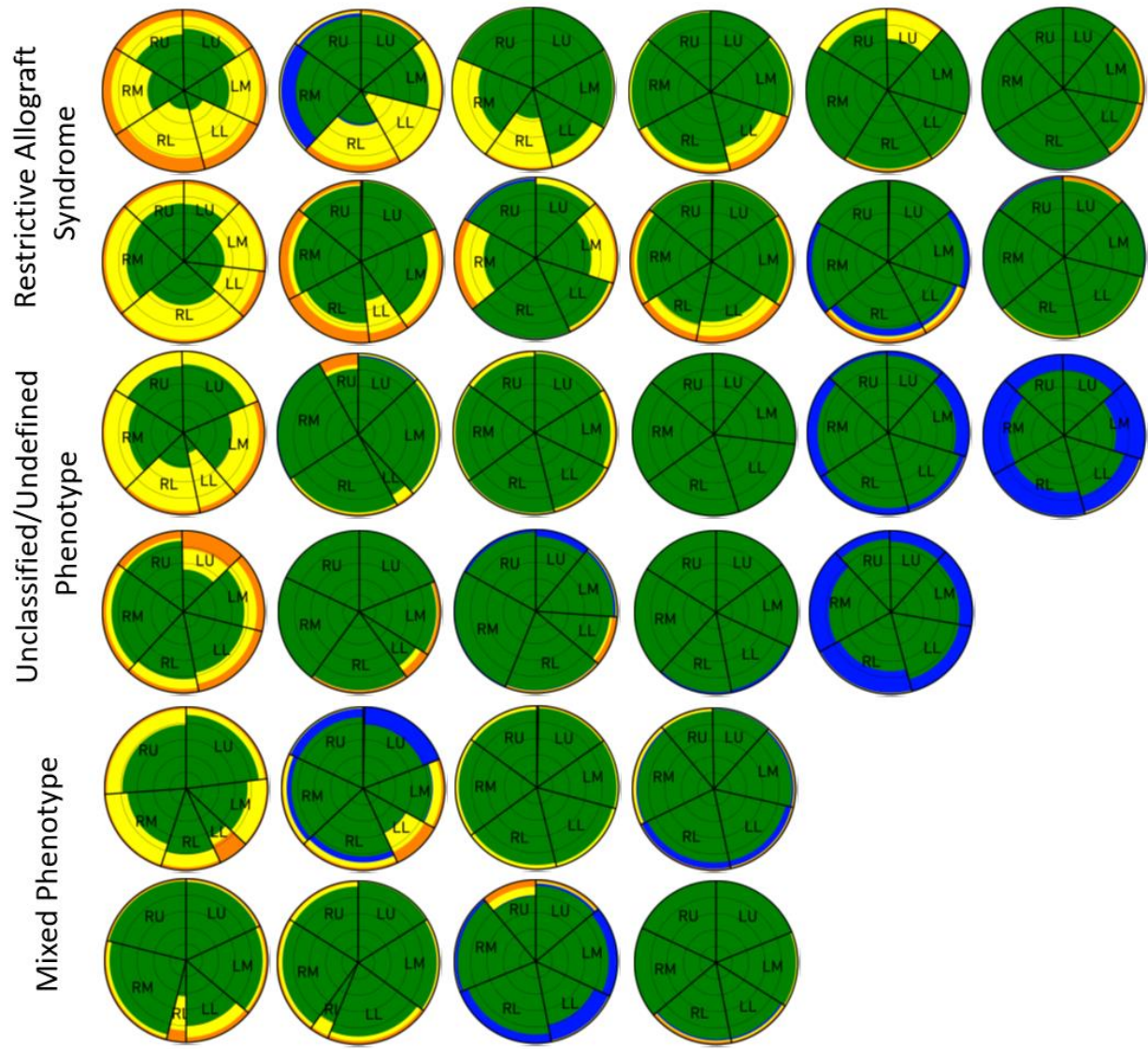
FIGURE E2. Glyph representation of the machine learning output for each cases of restrictive allograft syndrome (top two rows), unclassified or undefined phenotype (middle two rows), and mixed phenotype (bottom two rows). Glyphs are composed of five concentric rings, each representing one fifth of the lung volume, and multiple wedges each representing a lobe of the lung. Normal lung is green, hyperlucent is blue, ground-glass opacity is yellow, and reticulation is orange. Ground-glass opacity and reticulation are the predominant abnormality. LL – left lower lobe; LM – lingula; LU – left upper lobe; RL – right lower lobe; RM – right middle lobe; RU – right upper lobe.
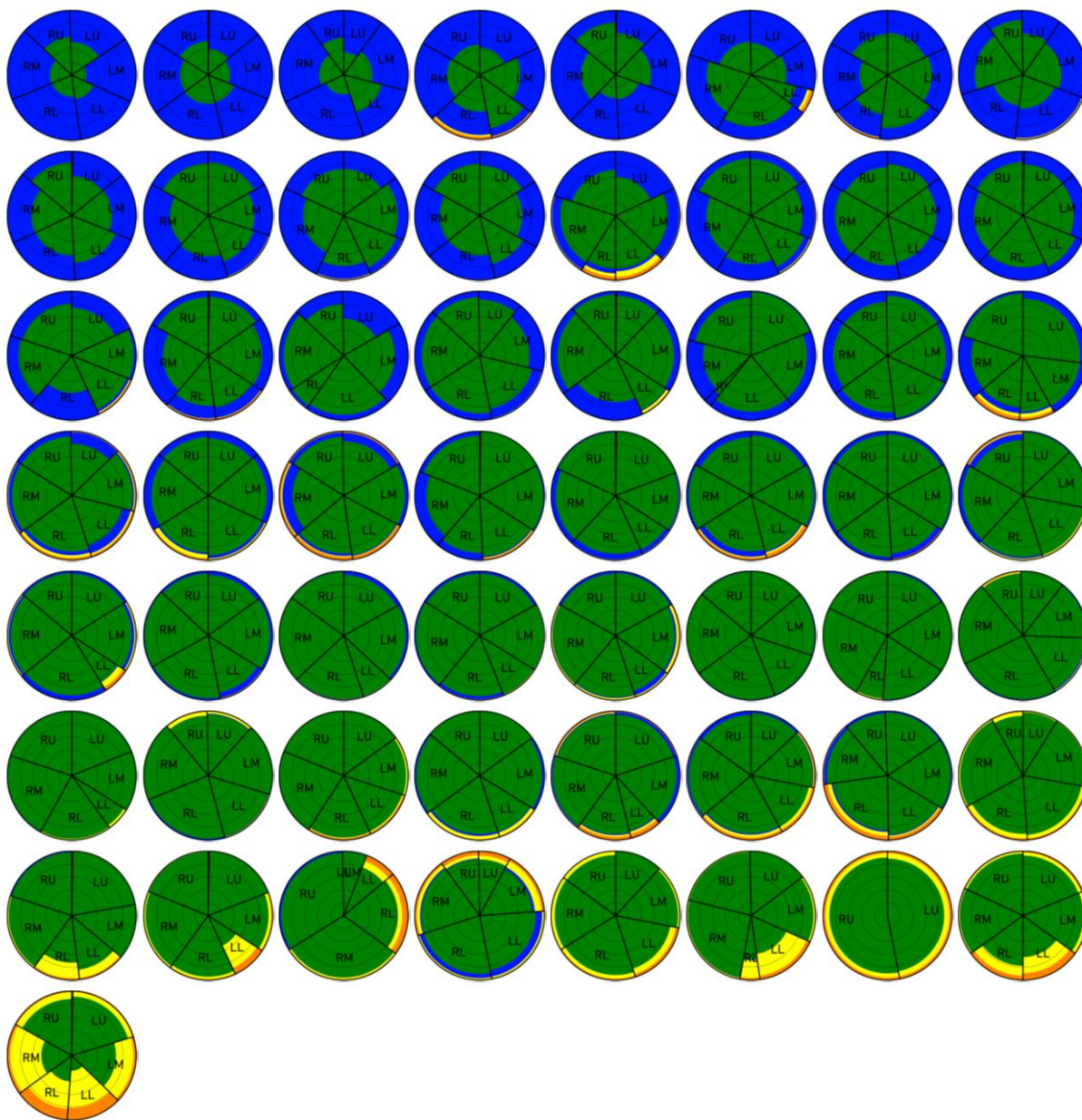
FIGURE E3. Glyph representation of the machine learning output for each case of bronchiolitis obliterans syndrome. Glyphs are composed of five concentric rings, each representing one fifth of the lung volume, and multiple wedges, each representing a lobe of the lung. Normal lung is green, hyperlucent is blue, ground-glass opacity is yellow and reticulation is orange. Hyperlucent lung is the predominant abnormality in BOS. Although pulmonary opacities (yellow and orange) are more typical of RAS in general, most of the opacities in this BOS cohort are in the lower lung zone (bottom half of the

glyph) rather than upper lung zone. Furthermore, they should persist to qualify as RAS-like and it is also possible that texture analysis is identifying more opacity than was evident to the CLAD adjudicator. BOS – bronchiolitis obliterans syndrome; LL – left lower lobe; LM – lingula; LU – left upper lobe; RAS – restrictive allograft syndrome; RL – right lower lobe; RM – right middle lobe; RU – right upper lobe.
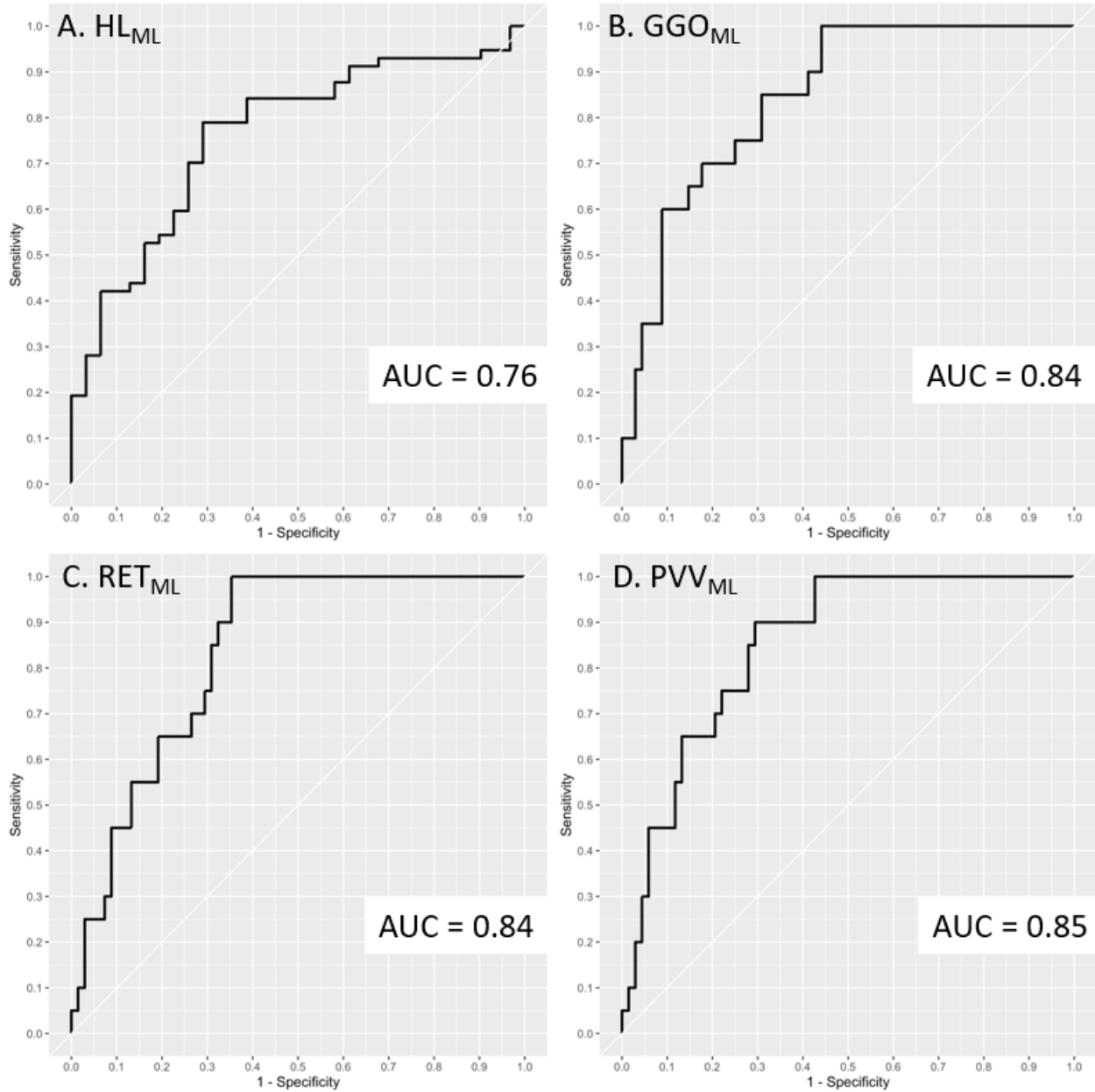
FIGURE E4. Receiver operating characteristic curves for (A) HL$_{ML}$, (B) GGO$_{ML}$, (C) RET$_{ML}$ and (D) PVV$_{ML}$. GGO – ground glass opacity texture; HL – hyperlucent texture; ML – machine learning; PVV – pulmonary vessel volume; RET – reticular texture.
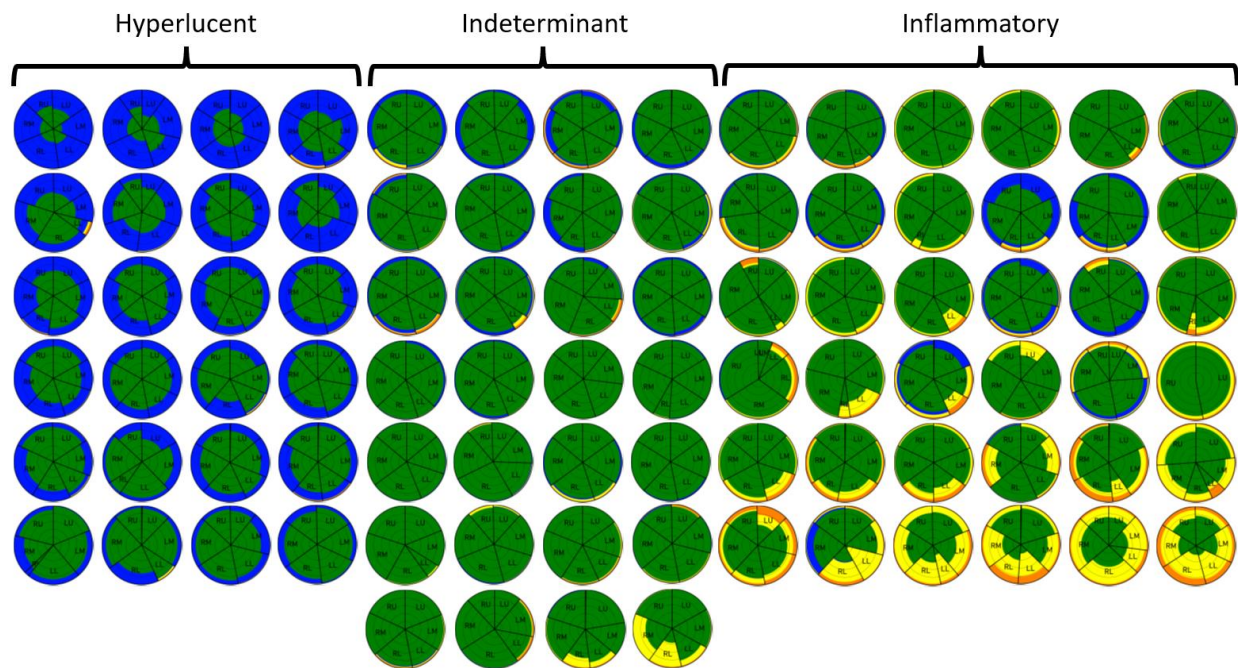
Figure E5. Glyph representation of a revised classification where the inflammatory group had GGO$_{ML}$ and RET$_{ML}$ over 1%, the hyperlucent group had over 10% HL$_{ML}$ without inflammatory features, and the indeterminant group were all remaining cases. GGO – ground glass opacity texture; HL – hyperlucent texture; ML – machine learning; RET – reticular texture.
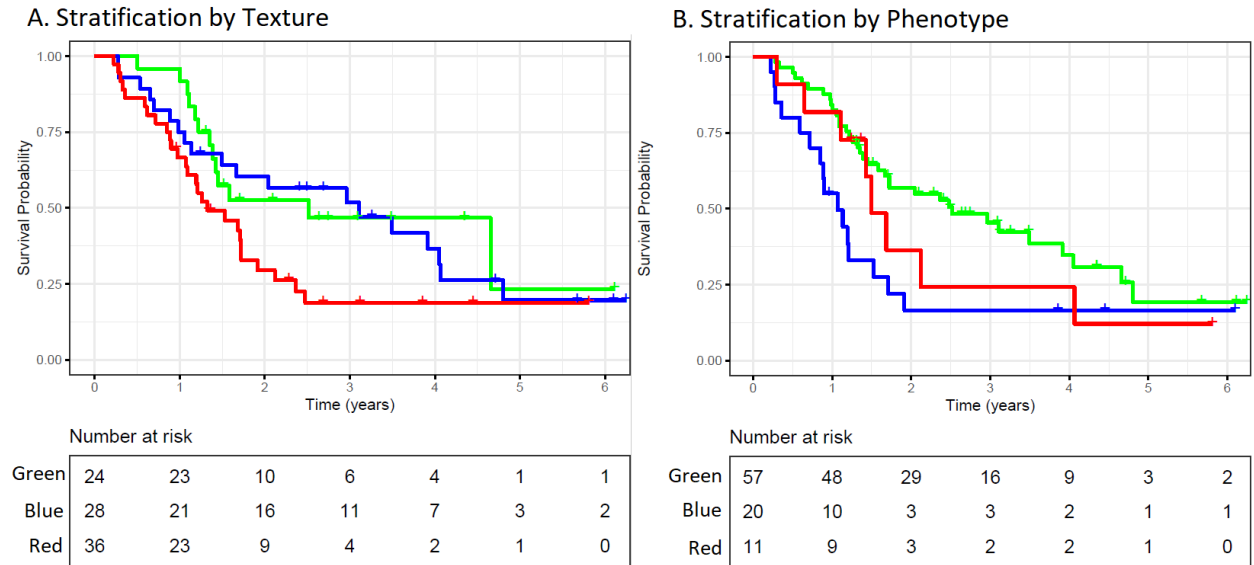
Figure E6. Kaplan-Meier curves contrasting allograft survival (death or re-transplantation) by (A) lung texture analysis in three groups: hyperlucent lung (green), indeterminate (blue) and inflammatory (red); and (B) CLAD phenotype in three groups: bronchiolitis obliterans syndrome (green), restrictive allograft syndrome/mixed (blue), and undefined/unclassified (red).
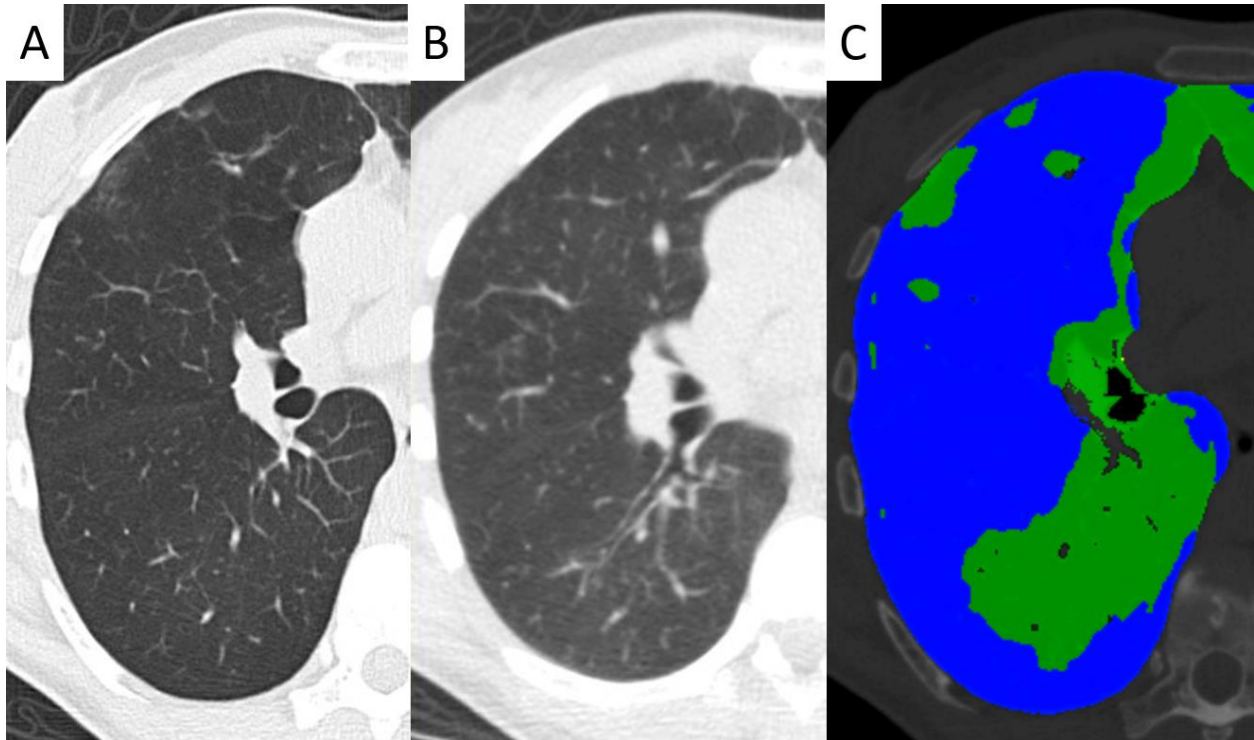
FIGURE E7. Evaluation for air trapping using paired (A) inspiratory and (B) ultra-low dose expiratory computed tomography (CT) imaging in a 30-year-old woman with bronchiolitis obliterans syndrome. Air trapping was scored as not present on these paired images, like related poor timing or effort with the expiratory acquisition. (C) Lung texture analysis of the inspiratory CT reveals extensively hyperlucent lung (blue, 56.9% of total lung volume) in contrast to normal lung texture (green).