OLYMPUS

Performance and Benefit of SeleCT™
Enhanced by Deep Learning-based Artificial Intelligence

Alexander Bankier, MD, PhD
Professor of Radiology, University of Massachusetts Medical School
Chief of Cardiothoracic Imaging Section

Charles Hatt, PhD
Adjunct Professor, University of Michigan Medical School
Director of Research, Imbio

Introduction

Emphysema severity, emphysema heterogeneity, and pulmonary fissure completeness are predictors of response to endobronchial valve reduction treatment.[1] In particular, the assessment of pulmonary fissure completeness is of paramount importance due to being a reliable indicator of collateral ventilation.[2,3] Fissure completeness and emphysema severity can be determined visually, which is time-consuming and suffers from inter- and intra-observer variability.[4,5] Traditional computer algorithms designed for the quantitative assessment of fissure completeness and emphysema severity are based on classical pattern recognition techniques and have provided good results in the past.[6,7] With the advent of deep learning-based artificial intelligence (AI) techniques, however, newer algorithms now hold promise to provide a more accurate and reliable approach to the assessment of fissure completeness and emphysema.

In general, deep learning algorithms improve their performance when provided with more data and therefore it is expected that fissure completeness and emphysema severity assessment performance will increase over time. The purpose of this study was to test the performance of the SeleCT service with Imbio, a novel AI based software package for the assessment of fissure completeness, emphysema severity, and emphysema heterogeneity in a cohort of clinical patients with emphysema deemed to be candidates for endobronchial valve placement.

Patients and patient selection

Baseline high-resolution computed tomography (CT) examinations were acquired in 479 patients as part of the EMPROVE clinical trial which generated the data for FDA approval of the Spiration® Valve System.[8] Of the 479 patients, 174 (2 withdrew before treatment) were selected for randomization as part of the study based on fissure completeness and emphysema distribution values. To select data for testing the enhanced SeleCT algorithm, 60 datasets were randomly sampled from the original 479 scans such that 30 belonged to the randomization group and 30 belonged to the non-randomization group. Each set of 30 was chosen such that the fissure completeness distributions within the testing datasets matched the fissure completeness distributions within the original study. Of the resulting 60 patients, 28 were female and 32 were male. Their mean age was 65±7 years (range, 50 to 83 years). Only the patients who were selected for randomization (30 out of 60) had pulmonary function tests (PFTs) performed at baseline. Based on those PFTs, 15 (50%) patients were categorized as GOLD 4, 14 (36%) as GOLD 3, and 1 (4%) as GOLD 2 (re-categorized as GOLD 3 at enrollment).

Continued on next page.

Performance and Benefit of SeleCT
Enhanced by Deep Learning-based Artificial Intelligence

CT acquisitions

All CT acquisitions were performed on state-of-the-art commercially available CT scanner units. Patients were scanned in the supine body position and at full suspended inspiration. Anatomical CT coverage included the entire thorax, from the thoracic inlet to the inferior costophrenic sinus. The individual CT scanner unit brands and the reconstruction kernels used were as follows: Siemens (B35f: 28 patients, B30f: 1 patient), General Electric (Standard: 13 patients), Philips (B: 13 patients), Toshiba (FC18: 4 patients, FC01: 1 patient).

Image analysis and statistical evaluation

Analysis was performed using the enhanced SeleCT service with Imbio, a software that provides objective quantification of fissure completeness and emphysema severity and heterogeneity via deep learning-based image analysis techniques. Fissure completeness was measured by computing the percentage of each lobe boundary that was separated by a complete pulmonary fissure. Each fissure was classified as "complete" if the fissure completeness was ≥90%. Emphysema percentage was computed as the percentage of voxels in each lobe that were ≤920 HU (%LAA-920). "Severe" emphysema was defined as %LAA-920 ≥40% and "heterogenous" emphysema was defined as an absolute difference in %LAA-920 ≥ 10% between the lobe of interest and the ipsilateral lobe.

In detail, the Enhanced SeleCT workflow (Figure 1) utilizes multiple deep convolutional neural networks (CNNs). The first CNN segments the pulmonary lobes and major airways from the input CT scan. A second CNN is used to create a fissure "probability map," where voxels having a high probability of belonging to a pulmonary fissure are given a value of 1.0 and voxels having a low probability of belonging to a pulmonary fissure are given a value of 0.0. Areas of the fissure surface that have corresponding high values on the fissure probability map are labeled as being complete. Otherwise, they are labeled as incomplete. The fissure completeness percentage is computed by dividing the number of complete voxels within the fissure surface by the total number of voxels.

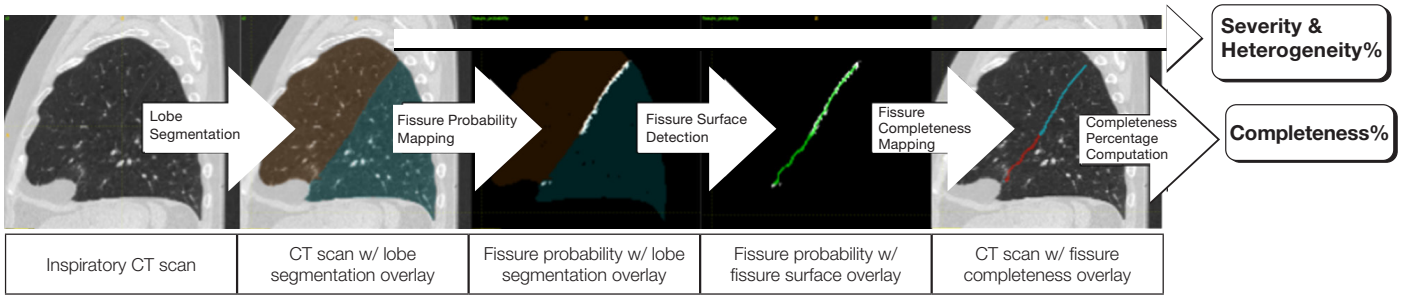

Figure 1. Fissure completeness and emphysema quantification image analysis workflow.

For software validation, the enhanced SeleCT software was compared with a prior version of the SeleCT software ("predicate" software) which computed the same measurements using different underlying image processing techniques and, in some cases, manual fissure completeness editing by a trained medical imaging analyst. A patient was deemed to be a candidate for treatment (positive) if they presented with a lobe containing a complete fissure and severe, heterogenous emphysema (per the aforementioned definitions). A patient was deemed to be a non-candidate for treatment (negative) if at least one of these conditions was not met.

Continued on next page.

Image analysis and statistical evaluation (continued)

Statistical evaluation of the findings included the calculation of:

True positives (TP):

- Fissure identified as complete by Enhanced SeleCT and the predicate software.
- Lobe identified as having severe emphysema by Enhanced SeleCT and the predicate software.
- Identification of heterogeneous emphysema by Enhanced SeleCT and the predicate software.

True negatives (TN):

- Fissure identified as incomplete by Enhanced SeleCT and the predicate software.
- Lobe identified as NOT having severe emphysema by Enhanced SeleCT and the predicate software.
- Identification of homogenous emphysema by Enhanced SeleCT and the predicate software.

False Positives (FP):

- Fissure identified as complete by Enhanced SeleCT but NOT by the predicate software.
- Lobe identified as having severe emphysema by Enhanced SeleCT but NOT the predicate software.
- Identification of heterogeneous emphysema by Enhanced SeleCT but NOT the predicate software.

False Negatives (FN):

- Fissure identified as incomplete by Enhanced SeleCT but complete by the predicate software.
- Lobe identified as NOT having severe emphysema by Enhanced SeleCT but having severe emphysema by the predicate software.
- Identification of homogenous emphysema by Enhanced SeleCT but heterogeneous emphysema by the predicate software.

From a treatment perspective (valve placement yes or no?) the implication of a FP result could be that a potentially non-qualified patient would be treated, and the implication of a FN result could be that a potentially qualified patient might not be treated.

Based on TP, TN, FP, and FN values, the test characteristics for the software were computed, both on a per-patient and on a per-fissure basis.

Continued on next page.

Performance and Benefit of SeleCT
Enhanced by Deep Learning-based Artificial Intelligence

SeleCT Test Performance

	per Fissure (n=240)	per Patient (n=60)
True positive	42/240	35/60
False positive	5/240	3/60
True negative	188/240	21/60
False negative	5/240	1/60
Sensitivity	89.36% [76.89%; 96.45%]	97.22% [85.47%; 99.93%]
Specificity	97.41% [94.06%; 99.15%]	87.50% [67.64%; 97.34%]
Accuracy	95.83% [92.47%; 97.98%]	93.33% [83.80%; 98.15%]
Positive predictive value	89.36% [77.86%; 95.25%]	92.11% [80.17%; 97.12%]
Negative predictive value	97.41% [94.26%; 98.85%]	95.45% [83.80%; 98.15%]

Table 1. Test performance for the SeleCT software on matching predicate software results on EMPROVE trial CT scans, given on a per lobe and per patient basis. Test characteristics are reported as ratios of absolute numbers and are also expressed as percentages together with their 95% confidence intervals.

SeleCT Test Performance

	Fissure Completeness	Emphysema Severity	Emphysema Heterogeneity
True positive	140/240	178/240	83/240
False positive	13/240	0/240	3/240
True negative	77/240	61/240	154/240
False negative	10/240	1/240	0/240
Sensitivity	93.33% [88.08%; 96.76%]	99.44% [96.93%; 99.98%]	100.00% [95.65%; 100.00%]
Specificity	85.56% [76.57%; 92.08%]	100.00% [94.13%; 100.00%]	98.09% [94.52%; 99.60%]
Accuracy	90.42% [85.97%; 93.83%]	99.58% [97.70%; 99.98%]	98.75% [96.39%; 99.74%]
Positive predictive value	91.50% [86.67%; 94.69%]	100.00% [100.00%;100.00%]	96.51% [90.02%; 98.83%]
Negative predictive value	88.5% [80.79%; 93.38%]	98.39% [89.62%; 99.77%]	100.00% [100.00%; 100.00%]

Table 2. Test performance for the enhanced SeleCT software on matching fissure completeness, emphysema severity, and emphysema heterogeneity results from predicate software results, given on a per lobe basis. Test characteristics are reported as ratios of absolute numbers and are also expressed as percentages together with their 95% confidence intervals.

Continued on next page.

Results

Results in absolute proportions together with the test characteristics for SeleCT are summarized in Table 1. Results for each individual criterion (fissure completeness, emphysema severity, emphysema heterogeneity) are summarized in Table 2. For all test characteristics, the performance of SeleCT was throughout very good (>80%) to excellent (>90%). Despite the relatively low patient number (n=60), the narrow 95% confidence intervals of the test characteristics indicate high reliability.

On a per-patient basis, there were only 4/60 discordant assessments: 3 FP (3/60, 5%, potentially non-qualified patient would be treated) and 1 FN (1/60, 1.66%, potentially qualified patient might not be treated). The 4 discordant determinations occurred in 4 different patients, 2 of which were due to discordant fissure completeness assessments and 2 of which were due to discordant emphysema heterogeneity assessments.

On a per-lobe basis, there were only 10/240 discordant assessments, which were equally distributed between FN (n=5) and FP (n=5) and resulted from disagreements between fissure completeness in 7/10 cases and emphysema heterogeneity in 3/10 cases. In all cases where the heterogeneity did not agree between enhanced SeleCT and the predicate software, the differences in measurements were <0.6%. The 5 FPs (5/240=2%) had the following relationships to the intended lobe for treatment: RUL (n=2), RLL (n=1), LUL (n=1), LLL (n=1). 3 FPs were related to the right minor fissure, and 2 to the left major fissure. The 5 FNs (5/240=2%) had the following relationships to the intended lobe for treatment: RUL (n=3), LUL (n=1), LLL (n=1). 3 FNs were related to the right minor fissure, and to the 2 left major fissure. Note that in the left lung the upper left and lower left lobe share the same fissure and heterogeneity calculation, resulting in discordant measurements being “double counted.”

Discussion

Our study shows that enhanced SeleCT provides robust and consistent assessment of fissure completeness and emphysema severity and heterogeneity in patients who were potential candidates for valve placement. The assessment by enhanced SeleCT was characterized by excellent overall agreement with results from the predicate software (93% to 96%) and very high to excellent sensitivity (89% to 97%) and specificity (87% to 97%). The balanced magnitudes of these test characteristics underline the overall strong performance of enhanced SeleCT in addressing both complete and incomplete fissures as well as consistently measuring quantitative emphysema. In addition, and despite the relatively low number of patients included in this study, the narrow 95% confidence intervals for all test characteristics indicate that enhanced SeleCT operates at a highly reliable level. Finally, the fact that multiple clinical CT scanner brands were included in this study suggests that the results are reliable across CT scanner platforms.

The results provided by enhanced SeleCT were equally good on a per-patient and a per-lobe assessment level. From a clinical perspective, this suggests that enhanced SeleCT allows accurate assessment of fissure integrity and emphysema both for individual lobes and for the patient overall. This aspect is important as the pre-interventional assessment of potential candidates for valve implantation increasingly moves towards a holistic approach, including the pre-interventional multidisciplinary evaluation of numerous radiological and clinical markers.[3]

The enhanced SeleCT classifications characterized as discordant based on the statistical analysis of this study deserve more detailed attention. On a per-patient basis, only 4/60 were discordant and all 4 occurred in different patients. This suggests that these assessments are non-systematic and likely within the expected range of randomness, especially when considering that the disagreements due to heterogeneity were very small but happened to be close to the 10% cut-off value. Of note, 3/4 assessments were “false-positive,” thus calling potentially untreatable lobes as treatable. From a clinical perspective, this could indicate that enhanced SeleCT tends to err towards a more tolerant side by not precluding potential candidate patients from treatment. On a per-lobe basis, the distribution of the small number of discordant assessments are not surprising. All of these occurred for the right minor and the left major

Continued on next page.

Discussion (Continued)

lobe. Both fissures often are anatomically incomplete and therefore difficult to delineate by radiologists on clinical CT examinations.[4] This is amplified by the fact that the CT appearance of these fissures is commonly degraded by artifacts resulting from cardiac and respiratory motion unless the CT examinations are gated, which was not the case in this study and which is unrealistic to expect in the clinical work-up of patients for valve placement. The margin of error of enhanced SeleCT was, thus, small and, when observed, either within the expected range or explainable by well-known anatomical or CT-technique related causes. It is important to note that from a clinical perspective, using 90% as the threshold for completeness is supported by the EMPROVE trial.

However, the clinical factors for patient eligibility should be evaluated holistically and a patient with fissure completeness slightly less than 90% still has a reasonable probability of treatment success if the patient meets all other criteria for treatment. This is an advantage that quantitative assessment has over visual assessment, as it is possible to objectively determine by how much a measurement is below or above the 90% threshold used to decide if the fissure is intact or not.

Conclusion

In conclusion, enhanced SeleCT provides a robust and consistent assessment of fissure completeness in patients with heterogeneous emphysema that are deemed candidates for valve placement. When compared to predicate software, enhanced SeleCT resulted in excellent overall diagnostic accuracy, very high to excellent sensitivity and specificity, and high reliability. Discordant assessments by SeleCT are non-random and likely reflect anatomical and CT technique related general drawbacks in fissure assessment that are known from visual fissure assessment.

Overall, these results fully justify the use of SeleCT for the clinical assessment of fissure completeness and emphysema in patients who potentially qualify for endobronchial valve reduction treatment. Given the ability of SeleCT to improve as more data is acquired, it can reasonably be expected that the results described in this study will further improve with a larger-scale clinical deployment in the future.

Performance and Benefit of SeleCT
Enhanced by Deep Learning-based Artificial Intelligence

References

1. Schuhmann M, Raffy P, Yin Y, et al. Computed tomography predictors of response to endobronchial valve lung reduction treatment. Comparison with Chartis. *Am J Respir Crit Care Med*. 2015;191(7):767-774. doi:10.1164/rccm.201407-1205OC

2. Shah PL, Herth FJ. Current status of bronchoscopic lung volume reduction with endobronchial valves. *Thorax*. 2014;69(3):280-286. doi:10.1136/thoraxjnl-2013-203743

3. Lee SW, Shin SY, Park TS, et al. Clinical Utility of Quantitative CT Analysis for Fissure Completeness in Bronchoscopic Lung Volume Reduction: Comparison between CT and Chartis™. *Korean J Radiol*. 2019;20(7):1216-1225. doi:10.3348/kjr.2018.0724

4. Lynch D a, Murphy JR, Crapo JD, et al. A Combined Pulmonary-Radiology Workshop for Visual Evaluation of COPD: Study Design, Chest CT Findings and Concordance with Quantitative Evaluation. COPD. 2012;9(2):151-159. doi:10.3109/15412555.2012.654923

5. Kent MS, Ridge C, O'Dell D, Lo P, Whyte R, Gangadharan SP. The accuracy of computed tomography to predict completeness of pulmonary fissures. A prospective study. *Ann Am Thorac Soc*. 2015;12(5):696-700. doi:10.1513/AnnalsATS.201407-343OC

6. Van Rikxoort EM, Goldin JG, Galperin-Aizenberg M, et al. A method for the automatic quantification of the completeness of pulmonary fissures: evaluation in a database of subjects with severe emphysema. *Eur Radiol*. 2012;22(2):302-309. doi:10.1007/s00330-011-2278-0

7. Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE transactions on medical imaging. 2001;20(6):490–498. doi:10.3109/15412555.2012.654923

8. Criner GJ, Delage A, Voelker K, et al. Improving Lung Function in Severe Heterogenous Emphysema with the Spiration Valve System (EMPROVE). A Multicenter, Open-Label Randomized Controlled Clinical Trial. *Am J Respir Crit Care Med*. 2019;200(11):1354-62. doi:10.1164/rccm.201902-0383OC

OLYMPUS AMERICA INC.
3500 Corporate Parkway, PO Box 610, Center Valley, PA 18034

For more information, contact your Olympus sales representative or call 800-848-9024.
www.medical.olympusamerica.com

Printed in the USA OAIRES0221WP38472