UNIVERSITY OF CALIFORNIA SAN DIEGO

Development of Quantitative Phenotyping Tools to Improve Pulmonary Thromboendarterectomy Risk-Benefit Assessment in Patients with Chronic Thromboembolic Pulmonary Hypertension

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Bioengineering

by

Elizabeth Michelle Bird

Committee in charge:

Professor Francisco Contijoch, Chair
Professor Susan Hopkins
Professor Albert Hsiao
Professor Nick Kim
Professor Atul Malhotra
Professor Elliot McVeigh

2024

The Dissertation of Elizabeth Michelle Bird is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

To Ann and Mitch Bird.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BPA	Balloon Pulmonary Angioplasty
CO	Cardiac Output
CTEPH	Chronic Thromboembolic Pulmonary Hypertension
DECTPA	Dual-Energy CT Pulmonary Angiogram
HLV	Hypoperfused Lung Volume
mPAP	Mean Pulmonary Artery Pressure
PBV	Pulmonary Blood Volume
PH	Pulmonary Hypertension
PTE	Pulmonary Thromboendarterectomy
PVR	Pulmonary Vascular Resistance
VQ	Ventilation-Perfusion

ACKNOWLEDGEMENTS

There are many individuals who I would like to thank for making this work possible. First, I would like to thank my advisor, Francisco Contijoch. I would also like to thank Atul Malhotra for his support of both my research and clinical education and efforts. Finally, I would like to thank Albert Hsiao and Samira Masoudi for their support in teaching me about machine learning model development.

Next, I would like to thank the members of my dissertation committee for sharing their time, individual expertise and feedback to shape and improve the work contained in this thesis. I would also like to thank the CTEPH team at UCSD. Specifically, I would like to thank Nick Kim, Michael Madani, David Poch, Tim Fernandes, Kim Kerr, Victor Pretorius, Demosthenes Papamatheakis, Seth Kligerman, Lewis Hahn, Peter Fedullo, Angela Bautista, Melissa Stinson, and Thao Drcar. Additionally, I appreciate the comradery and input of the members of the Contijoch Research Lab. Finally, I would like to thank the program directors, administrators, and peers in the Medical Scientist Training program at UCSD.

Chapter 1, in full, is a reprint of the material contained in "Mapping the Spatial Extent of Hypoperfusion in Chronic Thromboembolic Pulmonary Hypertension Using Dual Source CT" accepted to *Radiology: Cardiothoracic Imaging* by Elizabeth Bird, Kyle Hasenstab, Nick Kim, Michael Madani, Atul Malhotra, Lewis Hahn, Seth Kligerman, Albert Hsiao, and Francisco Contijoch. Elizabeth Bird, the dissertation author, was the primary author and investigator on this paper.

Chapter 2 is recent work that is not currently in preparation for submission as a manuscript. The dissertation author was the primary investigator and author of this work, but

this work would not have been possible without the contributions and insights of Samira Masoudi, Nick Kim, Michael Madani, Atul Malhotra, Lewis Hahn, Albert Hsiao, and Francisco Contijoch.

Chapter 3 is recent work that is not currently in preparation for submission as a manuscript. The dissertation author was the primary investigator and author of this work, but this work would not have been possible without the contributions and insights of Samira Masoudi, Nick Kim, Michael Madani, Atul Malhotra, Lewis Hahn, Albert Hsiao, and Francisco Contijoch.

VITA

2015 Bachelor of Science in Bioengineering, UC San Diego

2024 Doctor of Philosophy in Bioengineering, University of California San Diego

PUBLICATIONS

Bird E, Hasenstab K, Kim N, Madani M, Malhotra M, Hahn L, Kligerman S, Hsiao A, Contijoch F. Mapping the Spatial Extent of Hypoperfusion in Chronic Thromboembolic Pulmonary Hypertension Using Dual Source CT. Radiology Cardiothoracic Imaging. 10 Aug 2023. DOI: 10.1148/ryct.220221

ABSTRACT OF THE DISSERTATION

Development of Quantitative Phenotyping Tools to Improve Pulmonary Thromboendarterectomy Risk-Benefit Assessment in Patients with Chronic Thromboembolic Pulmonary Hypertension

by

Elizabeth Michelle Bird

Doctor of Philosophy in Bioengineering

University of California San Diego, 2024

Professor Francisco Contijoch, Chair

Chronic Thromboembolic Pulmonary Hypertension (CTEPH) is a disease where patients have unresolved clots in their pulmonary vasculature. Patients can be successfully

treated by Pulmonary Thromboendarterectomy (PTE) when the pulmonary hemodynamic impairment is due to surgically accessible CTEPH disease burden (fibrotic clot and resulting downstream perfusion deficits). Unfortunately, up to 10 - 50% of patients who undergo the surgery will have residual pulmonary hypertension, which carries higher complication and mortality rates. Residual pulmonary hypertension is thought to occur in patients with either microvascular remodeling, mixed pulmonary hypertension types, or chronic clot too distal in the vasculature for complete surgical removal. As such, key preoperative questions to identify patients likely to benefit from PTE are (1) the extent to which their vascular obstructions and associated perfusion deficits are the drivers of hemodynamic impairment and (2) whether the obstructions are surgically accessible for removal (aka located in lobar to sub- segmental vessels).

Both of the above questions are currently assessed by expert visual interpretation of imaging studies alongside hemodynamic measures prior to PTE surgery. Unfortunately, there are currently no quantitative tools to aid these evaluations. As a result, the impact of specific lesions, the total disease burden, and the amount of surgically-accessible disease on PTE risk-benefit determination is unclear. Thus, there is a need to develop tools to systematically and objectively assess CTEPH patients being evaluated for PTE surgery.

This work aims to develop quantitative imaging measures to assess CTEPH characteristics important for surgical planning. To accomplish this, first we focus on the ability of Dual Energy CT Pulmonary Angiograms to evaluate agreement between CTEPH disease burden and hemodynamic severity. Next, we evaluate the ability of Machine Learning Analysis of Ventilation Perfusion Images predict surgical disease level, an indication of clot location with important accessibility implications.

INTRODUCTION

Chronic Thromboembolic Pulmonary Hypertension

Chronic Thromboembolic Pulmonary Hypertension (CTEPH) is a specific type of pulmonary hypertension, falling into group 4 of the 5 different pulmonary hypertension types (**Figure 1)**[1]. CTEPH is a rare form of pulmonary hypertension, affecting about 30 – 50 people per million in the United States[2]. CTEPH is due to vascular obstructions that do not resolve in the pulmonary vasculature, leading to pulmonary hypertension (mean pulmonary artery pressure/mPAP>20 mmHg, Pulmonary Vascular Resistance/PVR>240 dyn · s · cm^{-5}, and Pulmonary Capillary Wedge Pressure/PCWP<15 mmHg)[3], right heart failure, and high mortality when not treated[4,5]. Obstructions typically occur in both lungs, and the pattern of obstruction in one lung does not have to be reflective of the pattern of obstruction in the other lung. These vascular obstructions have a high amount of variability in their size, number, severity of obstruction and location, which can influence treatment selection for a given patient[6].

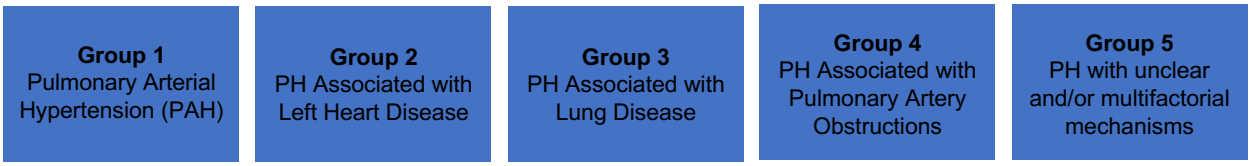


Figure 1 Different Pulmonary Hypertension Groups

CTEPH Treatment via Pulmonary Thromboendarterectomy Surgery

When a patient with CTEPH's pulmonary hypertension is predominantly caused by remodeled, chronic vascular obstructions in the pulmonary vasculature, removal of these obstructions can be an effective treatment[6]. Pulmonary Thromboendarterectomy (PTE) is a surgery that treats CTEPH by removing the chronic vascular obstructions from a patient's pulmonary arterial tree[7,8] **(Figure 2)**. PTE significantly improves mortality and reduces

pulmonary hypertension, making it the treatment of choice for patients who are good PTE candidates[9–11].

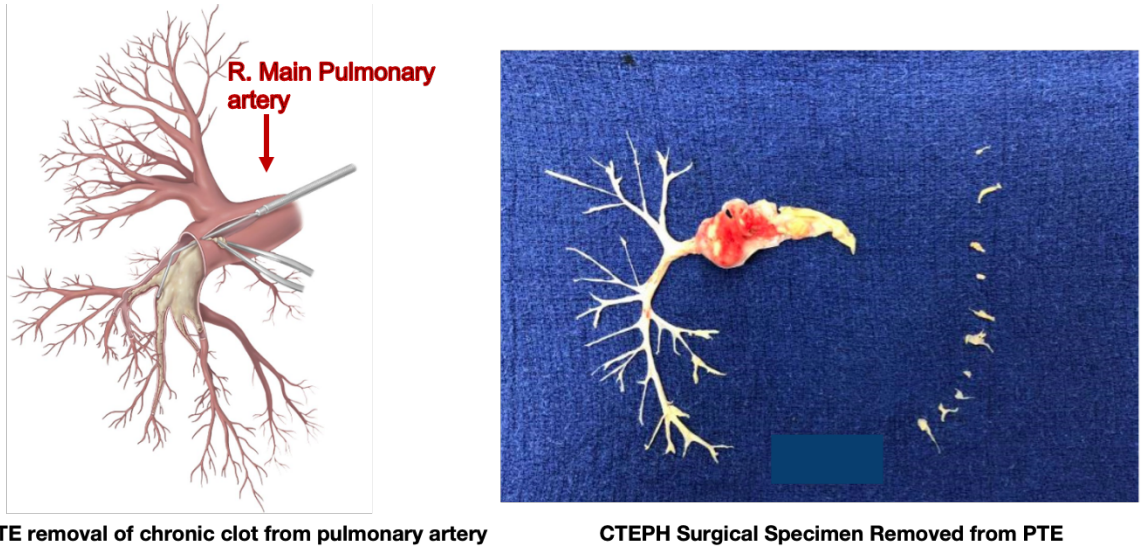


Figure 2 PTE Surgery Removes Chronic Clot from the pulmonary arteries (Image modified from Delcroix et al [12])

Identification of good PTE candidates is dependent on multiple factors. First, it is important that patients will be able to undergo cardiovascular surgery without significant risk of morbidity and mortality. Currently, the Society of Thoracic Surgeons has developed a quantitative morbidity and mortality model and affiliated risk calculator that addresses this concern. This calculator uses patient demographic information, comorbidities, and organ function to calculate individualized risk for common cardiothoracic procedures[13,14]. Second, it is important to determine if a patients symptoms and hemodynamic severity are primarily due to CTEPH fibrotic clot[6,15]. This includes determining if a patient is likely to have a mixed pulmonary hypertension type, for example a CTEPH patient who also shows features of Pulmonary Arterial Hypertension and has a history of methamphetamine use. Patients can also have CTEPH microvascular remodeling and disease as a result of pulmonary blood flow changes that accompany macrovascular fibrotic clots[4]. PTE removes macrovascular fibrotic clot and does

not address microvascular or nonfibrotic clot contributions to pulmonary hypertension. Third and finally, it is important for the fibrotic clot to be surgically accessible, as successful PTE surgery requires the surgeon to access and remove the chronic clot from the pulmonary arterial tree. Location of clot origin in the pulmonary arterial tree is described using the UCSD Surgical Disease Level **(Figure 3)**. Currently, disease as distal as the segmental and subsegmental vasculature (UCSD surgical disease level 3 and 4) can be surgically removed during PTE. Removal of this distal disease is more surgically challenging, and can only be successfully accomplished at a few surgical centers[6,16,17].

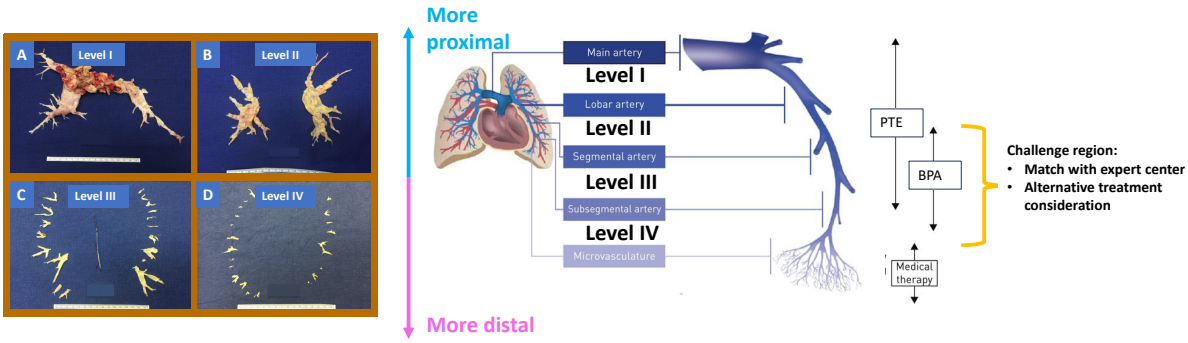


Figure 3 UCSD Surgical Disease Level defines the location of fibrotic clot origin and is important for surgical accessibility assessment. Modified from Madani et al and Delcroix et al [6,12].

Patients who are not good candidates for PTE can potentially benefit from alternative treatments, such as balloon pulmonary angioplasty (BPA) or medical therapy with pulmonary hypertension medications. Patients with macrovascular fibrotic clots that are too distal to be surgically accessible can be good candidates for BPA, whereas patients who have significant microvascular remodeling or mixed pulmonary hypertension types may be good candidates for medical therapy alone or in conjunction with PTE and/or BPA[12,18]. As seen in **Figure 3**, the selection of patients for treatment at the segmental and subsegmental level begins to be more complicated, as patients in this region may benefit best from PTE if their case is addressed at an expert center, or they may benefit from an alternative treatment such as PTE in combination with

BPA, BPA alone or medical therapy alone. It is important to best match patients to a given therapy and appropriate center to carry out that therapy in order to avoid suboptimal treatment outcomes and the requirement for additional procedures to correct suboptimal outcomes.

Selection of Good PTE Candidates

PTE candidacy criteria are evaluated in CTEPH patients during diagnosis and clinical workup by a team of multidisciplinary experts[19]. Determination of patient general risk for cardiac surgery is accomplished using the Society of Thoracic Surgeons Risk Score[13,14]. Assessment of the second criteria, determining a patient's pulmonary hypertension is likely primarily due to fibrotic clot, is accomplished by integrating imaging and invasive pulmonary hemodynamic assessment. Specifically, clinicians assess that a) the disease burden (lung parenchyma perfusion deficit and pulmonary arterial tree fibrotic clot) seen on imaging (Ventilation-Perfusion Studies for perfusion deficit; Dual Energy-CT Pulmonary Angiograms and Catheter Derived Digital Subtraction Angiograms for both perfusion deficit and fibrotic clot visualization) agree with one another, reducing suspicion for alternative causes of pulmonary hypertension. Clinicians also assess that (b) the visualized disease burden agrees with the severity of pulmonary hypertension, reducing the likelihood that there are significant microvascular contributions to disease. For the third criteria, clinicians again integrate clinical and imaging information. Clinicians use imaging information to assess where clot is likely to be in the vascular tree based on the location of perfusion deficit and fibrotic clot disease burden. But because fibrotic clot visualization is not perfect, clinicians also use known distal disease risk factors to increase their suspicion that a patient may have distal disease[6,12,15,20].

Quantitative Support Needed for Pre PTE Evaluation

Both surgical accessibility and agreement between disease burden and hemodynamic severity are assessed by expert interpretation of imaging studies alongside hemodynamic measures and patient clinical data prior to PTE surgery. Currently, imaging evaluations and integration with clinical data are not supported with quantitative analysis, and therefore can be subjective. As a result, the impact of specific lesions, the total disease burden, and the amount of surgically-accessible disease on PTE risk and benefit has not been quantitatively evaluated. This lack of quantitative evaluation prevents systematic assessment of PTE benefit for individual patients and limits risk-benefit analysis. Further, the subjective nature of the surgical criteria leads to center-to-center (and even expert-to-expert) variability in patient selection for PTE surgery, which precludes comparison of different treatment strategies within and across centers. There is a need to develop tools that can systematically and objectively assess CTEPH patients being evaluated for PTE surgery, which the work contained in this thesis focuses on addressing.

Dissertation Overview

This dissertation addresses the need for quantitative tools that can assess CTEPH patients prior to PTE surgery in the following chapters. The first chapter focuses on the assessment of disease burden and hemodynamic severity. In this chapter we discuss our developed of an automated metric that captures hypoperfusion from Dual-Energy CT images. Our metric, called Hypoperfused Lung Volume or HLV, captures the total spatial amount of hypoperfusion. This metric distinguishes CTEPH patients from controls, agrees with expert visual assessment, and agrees with hemodynamic severity of CTEPH in patients who underwent PTE.

The second chapter focuses on the preoperative prediction of surgical accessibility. In this chapter we investigate the use of machine learning to analyze ventilation perfusion images

and from this analysis predict disease level in the left lung. We investigate two different approaches to predict disease level from these models (classifying all cases vs classifying only cases that we can make confident predictions in). We identified two models that are able to classify patient left lungs as proximal, segmental, or (sub)segmental.

CHAPTER 1: DETECTION OF PERFUSION DEFICIT IN CTEPH

1.1. Abstract

The purpose of this study was to assess if a novel, automated method to spatially delineate and quantify the extent of hypoperfusion on multi-energy CT angiograms can aid evaluation of chronic thromboembolic pulmonary hypertension (CTEPH) disease severity. DECT angiograms obtained between January 2018 and December 2020 in 51 patients with CTEPH (mean age, 47 ± [SD] 17 years, 27 females) were retrospectively compared with 110 controls with no imaging findings suggestive of pulmonary vascular abnormalities (mean age, 51±16 years, 81 females). Parenchymal iodine values were automatically isolated using deep learning lobar lung segmentations. Low iodine concentration (defined as $\leq$ 0 mg/mL) was used to delineate areas of hypoperfusion and calculate hypoperfused lung volume (HLV). Receiver operating characteristic (ROC) curves, correlations with preoperative and postoperative changes in invasive hemodynamics, and comparison with visual assessment of lobar hypoperfusion by two expert readers were evaluated. Global HLV correctly separated patients with CTEPH from controls (area under the ROC curve=0.84, 10% HLV cutoff: 90% sensitivity, 72% accuracy, and 64% specificity) and correlated moderately with hemodynamic severity at time of imaging (pulmonary vascular resistance [PVR], ρ=0.67, $p<0.001$) and change after surgical treatment (ΔPVR ρ=-0.61, $p<0.001$). In patients surgically classified as having segmental disease, global HLV correlated with pre-operative PVR ($\rho = 0.81$) and post-operative change in PVR (ρ=-0.70). Lobar HLV correlated moderately with expert reader lobar assessment (ρ_{HLV}=0.71 for reader 1; ρ_{HLV}=0.67 for reader 2). In conclusion, automated quantification of hypoperfused areas in patients with CTEPH can be performed from clinical multi-energy CT examinations and may aid clinical evaluation, particularly in patients with segmental level disease.

1.2. Introduction

Obstruction of the pulmonary arteries by chronic thromboemboli leads to chronic thromboembolic pulmonary hypertension (CTEPH)[3,21,22]. Obstructions vary in number and location, leading to perfusion deficits in the pulmonary parenchyma with varying spatial extent and spatial distributions. Agreement between location(s) of arterial obstructions, the extent and location of hypoperfusion, and the severity of pulmonary hypertension are important factors in diagnostic and treatment evaluations [6,23].

Dual-energy CT (DECT) pulmonary angiography (DECTPA) offers a non-invasive approach to evaluate both arterial obstructions and hypoperfusion [24–30]. Average global iodine concentration (termed Pulmonary Blood Volume or PBV) measured from DECTPA images is reduced in patients with acute pulmonary embolisms (PE) and CTEPH [31]. Additionally, in patients with CTEPH, reduced PBV has also been correlated with hemodynamic severity [32–34]. In addition to global assessment, PBV and DECTPA iodine maps have proven useful in aiding visual detection of individual perfusion deficits by expert readers [25,30,35]. However, to date, DECTPA images have not been utilized to automatically delineate perfusion deficits in a pixel-fashion and quantify the spatial extent of impairment.

The objective of this study was to further DECTPA evaluation of pulmonary perfusion by developing and validating an automated, quantitative approach to define, on a per-pixel basis, regions of hypoperfusion in patients with CTEPH. Our hypothesis is that the spatial extent of hypoperfusion measured with our approach will agree with presurgical disease severity, changes in hemodynamics after surgical intervention, and with expert radiologist visual assessment. In particular, we expect pixelwise evaluation to be particularly helpful in patients with chronic thromboembolic disease that originates in the segmental pulmonary arteries.

1.3. Methods

1.3.1. Patient Selection

Under institutional review board-approved (#191797) waiver of informed consent and in compliance with the Health Insurance Portability and Accountability Act, 395 consecutive dual-energy acquisitions between 1/1/18 and 12/31/20 on a single CT scanner were retrospectively evaluated (by author FC, 12 years experience in cardiovascular imaging) in an unmatched case-control fashion. Scans were acquired for evaluation of CTEPH (n=120) or suspicion of acute PE (n=275) (**Figure 4**).

Error! Reference source not found.

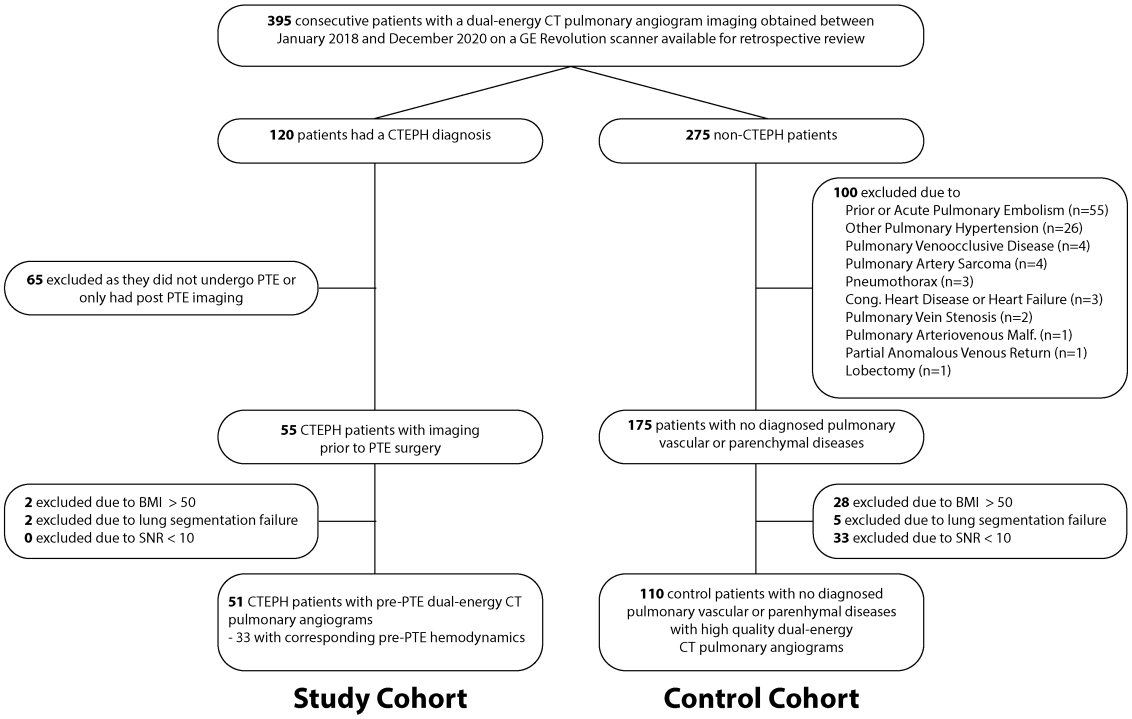


Figure 4: Flow diagram of study design

The CTEPH cohort **(Figure 4)** was comprised of patients who underwent pulmonary thromboendarterectomy (PTE) surgery for definitive treatment of CTEPH. Patient diagnosis and treatment selection was determined by consensus of a multidisciplinary team of cardiothoracic

surgeons, vascular disease expert pulmonologists, and fellowship trained cardiothoracic radiologists. Clinicians assessed CTEPH diagnosis using the criteria from Galiè et al, which include mean pulmonary artery pressure (mPAP) >25mmHg, pulmonary artery wedge pressure ≤15 mmHg, pulmonary vascular resistance ≥240 dyn·s·cm^{-5}, CTEPH-specific imaging signs, and ventilation-perfusion mismatch[3]. Patients diagnosed with CTEPH were excluded if they did not undergo PTE (including those treated medically or by balloon pulmonary angioplasty), lacked pre-operative dual-energy imaging (n=65), had body mass index (BMI) >50 kg/m^2 (which reduces image quality[36]; n=2), or had automatic lobar segmentation measurements (see *Hypoperfused Lung Volume Assessment)* that failed visual inspection (n=2; blinded visual review by author EB, 5 years experience in cardiovascular imaging). The final CTEPH cohort included 51 patients. Median time between DECT imaging and PTE surgery was 5 days (43/51 within 30 days, range: 2-372 days). A subset of the patients with CTEPH (n=33) had pre- and post-PTE invasive hemodynamic values available for review, which were used for comparison with CT-based evaluation (discussed in *Invasive Hemodynamics*).

Table 1 Summary of control and CTEPH cohort characteristics

	Controls n=110	**CTEPH** n=51	p-value
Age, years	51 ± 16	47 ± 17	0.14
Sex, female (%)	81 (73.6)	27 (52.9)*	0.01
BMI, kg/m2	39.2 ± 6.6	31.1 ± 7.5*	<0.001
Race			0.01
African American	5 (5)	12 (24)*	
Asian	7 (6)	1 (2)	
Pacific Islander	1 (1)	2 (4)	
White	63 (57)	25 (49)	
Other or Mixed$^+$	33 (30)	11 (22)	
Unknown	1 (1)	0 (0)	

The control cohort was comprised of patients found to have no imaging findings suggestive of pulmonary vascular abnormalities after multi-energy CT imaging was performed to rule out acute PE. Patients were excluded if the radiologic report or patient medical record mentioned findings consistent with any PE (acute, subacute, or chronic; n=55), pulmonary hypertension (n=26), pulmonary venoocclusive disease (n=4), pulmonary artery sarcoma (n=4), pneumothorax (n=3), congenital heart disease (n=3), pulmonary vein stenosis (n=2), pulmonary arteriovenous malformation (n=1), partial anomalous venous return (n=1), or prior lobectomy (n=1) (see **Figure 4**). As with the CTEPH cohort, studies were excluded for patient BMI>50 kg/m^2 (n=28), poor dual-energy image quality (defined as pulmonary artery iodine signal-to-noise ratio of <10[35]; n=33), or if automatic lobar segmentation failed visual inspection (n=5). The final control cohort included 110 patients.

1.3.2. Image Acquisition

Multi-energy CT pulmonary angiograms for both cohorts were acquired using the same dual-energy protocol on a single single-source Revolution CT scanner (GE Medical Systems, Chicago, IL). The acquisition rapidly switched kVP (80 to 140 kVp), and data were acquired helically with an 80mm detector width, median pitch of 1.38 (range: 0.98-1.38), and 0.5 s revolution time. Median tube current was 485 mAs (range: 240-630 mAs). Patients were administered iohexol contrast (Omipaque) with a concentration of 350mg/mL based on BMI (BMI<20: 60mL at 4mL/s, BMI 20-29.9: 75 mL at 5 mL/s, BMI 30-39.9: 80mL at 5.5mL/s, BMI> 40: 90mL at 6mL/s; median volume 75 mL, range: 50-150 mL). Imaging was timed for pulmonary arterial phase imaging using “smart prep” bolus tracking, with images acquired 8 seconds after main pulmonary artery enhancement at the T4 level reached 125 HU[37]. Iodine-water material decompositions yielding iodine concentration in mg/mL were generated using the

Gemstone Spectral Imaging software (GE Healthcare, Chicago IL). All images were reconstructed on a 512x512 image grid with xy-resolution of 0.7±0.1 mm (range: 0.5-0.9 mm) and section thickness of 1.25 mm. The typical field of view was 345x345 mm (range: 250-470 mm). Virtual mono-energetic images were reconstructed from the dual-energy acquisition to simulate a 70 keV acquisition.

1.3.3 Hypoperfused Lung Volume (HLV) Assessment

Pixel-wise classification of lung parenchyma as hypoperfused was obtained from the multi-energy pulmonary angiograms to calculate metrics of global and lobar spatial hypoperfusion in a completely automated fashion. Automated processing generated iodine concentration maps for the lung parenchyma. First, lung lobes were automatically segmented from virtual monoenergetic images **(Figure 5A, 5B)** using a recently published machine learning algorithm[38]. The lobar segmentations were used to isolate the iodine signal (concentration in mg/mL) from iodine-water maps **(Figure 5C)**, yielding lung parenchyma iodine concentration images (**Figure 5D**). The segmentation excluded large pulmonary vessels, bones, and airways, but smaller pulmonary vessels were removed separately. Specifically, pixels that were two standard deviations higher than the mean parenchymal concentration were removed. In addition, imperfect iodine-water separation (due to limited spectral separation of the incident x-ray beam, photon starvation, and artifacts from highly attenuating structures such as metal [39]) can lead to negative iodine concentrations independent of parenchyma perfusion, so pixels with iodine concentration ≤ -3 mg/mL were also removed.

After automated segmentation and isolation of the lung parenchyma iodine images, pixels with an iodine concentration ($I_{mg/mL}$) ≤ 0 mg/mL **(Figure 5E, blue pixels)** were classified as

hypoperfused. All other pixels were assigned a value of 0. This classification is shown in Equation 1 below.

$$f(I_{mg/mL}) = \begin{cases} 1, & I_{\frac{mg}{mL}} \leq 0\ mg/mL \\ 0, & I_{\frac{mg}{mL}} > 0\ mg/mL \end{cases} \quad \textbf{(1)}$$

Hypoperfused lung volume (HLV) was then calculated as the total number of hypoperfused parenchymal pixels divided by the total number of pixels (n) in the region of interest:

$$HLV = \frac{\sum_{i=1}^{n} f(I_{i,mg/mL})}{n} \quad \textbf{(2)}$$

In addition to global HLV, lobar HLV values (**Figure 5F**) were calculated by dividing the total number of hypoperfused pixels in each lobe by the volume of that lobe (and not the total lung volume). Therefore, global HLV represents a weighted average of the lobar HLV values.

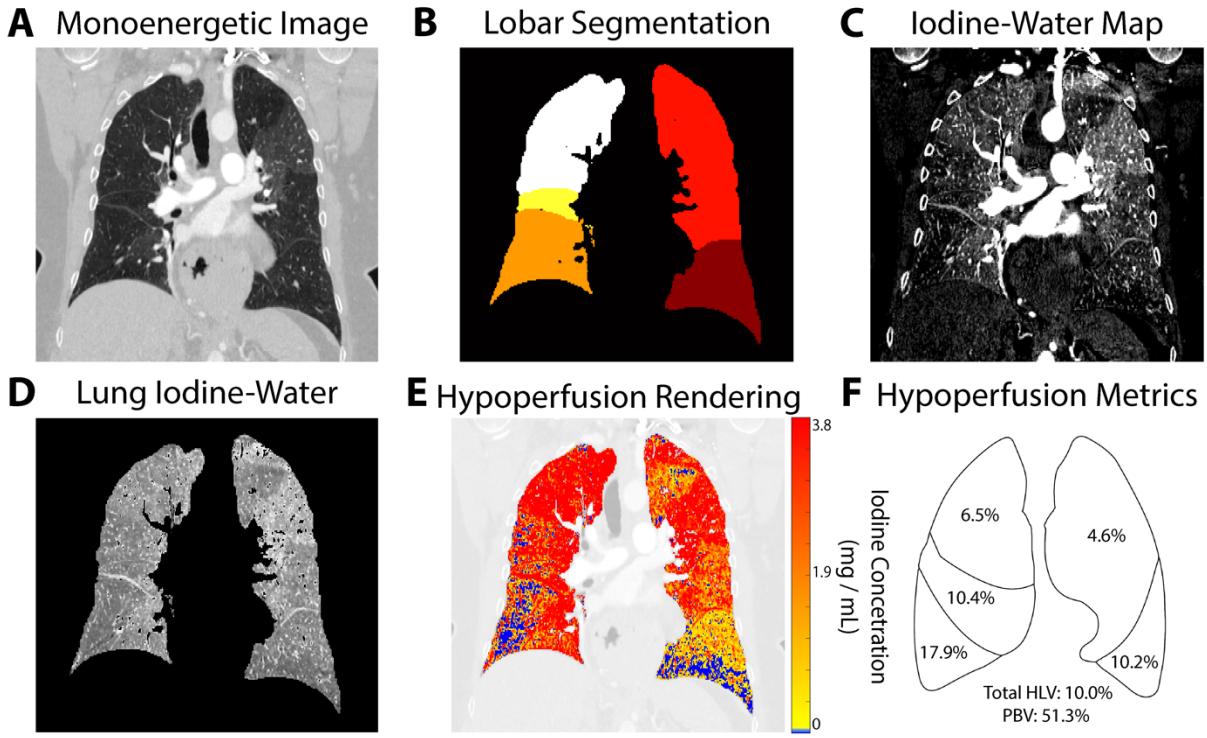


Figure 5 Automated lobe segmentation and summation of pixels classified as hypoperfused allows automatic quantification of HLV

For comparison, global PBV was measured according to the method described by Meinel et al using the same iodine concentration images and segmentations used for HLV analysis[33]. Briefly, 15% of the pulmonary artery iodine concentration was used to normalize parenchymal lung iodine concentrations. Mean pulmonary artery iodine concentration ($PA_{mg/mL}$) was measured for each patient using a 4 cm^2 circular region of interest placed in the pulmonary artery trunk (drawn by author EB) using Horos open-source Digital Imaging and Communications in Medicine (DICOM) viewer. This normalization by pulmonary artery iodine concentration yields a map of $I_{normalized}$ values as shown by Equation 3.

$$I_{normalized} = \frac{I_{mg/mL}}{0.15 \times PA_{mg/mL}} \quad \textbf{(3)}$$

Global PBV was then calculated as the average of $I_{normalized}$ over the whole, normalized parenchymal lung (**Equation 4**) while lobar PBV was measured by averaging across an individual lobe.

$$PBV = \frac{\sum_{i=1}^{n} I_{i,normalized}}{n} \quad \textbf{(4)}$$

Equations 1-4 illustrate the similarities and differences between PBV and HLV. Specifically, HLV measures the fraction of a region of interest considered hypoperfused, while PBV is the spatial average of $I_{normalized}$. Thus, HLV reflects the extent of hypoperfusion at a given severity (or greater) while PBV averages together $I_{normalized}$ values from normal, hyperperfused, and hypoperfused areas.

1.3.4 Invasive Hemodynamics and Correlation with Perfusion

For most of the patients with CTEPH (n=44/51), systolic pressure, diastolic pressure, mPAP, and thermodilution-derived cardiac output (CO) before and after PTE surgery were obtained as part of routine clinical care. Pulmonary vascular resistance (PVR) was calculated preoperatively using right heart catheterization-measured pulmonary artery wedge pressure (PAWP) according to the following equation: PVR=(mPAP – PAWP)/CO. Postoperative PVR was calculated using the central venous pressure to estimate PAWP, as PAWP is not measured by Swan-Ganz catheter in the intensive care unit at our institution[1]. When multiple hemodynamics were obtained, pre-operative values obtained just before the PTE surgery and post-operative hemodynamics just prior to Swan-Ganz catheter removal were selected. Patient hemodynamics were excluded if they had a change in pulmonary hypertension medications after PTE surgery to avoid the confounding effects of medications on the relationship between HLV and pulmonary hemodynamics. Eleven out of the 44 patients with hemodynamics had pulmonary hypertension medication changes after PTE. This resulted in a cohort of n=33 patients with both pre and post PTE hemodynamic measurements without a change in pulmonary hypertension medications. The median time between hemodynamic recordings and imaging was 2 days, with 26/33 patients having presurgical imaging within 30 days of hemodynamic recordings (range 0-326 days). Median time between surgery and post-operative hemodynamics was 1 day, (range 0-8 days, **Table 2)**.

Table 2 CTEPH cohort characteristics

Parameter	Value	
Time between CT imaging and surgery, days	5 [3 – 8]	
Time between Pre-Op Hemodynamic recordings and CT	2[1 - 24.5]	
Time between Post-Op Hemodynamic Recordings and Surgery	1[1 – 3]	
Pre- and Post-Op Hemodynamic recordings, n (%)	33 (65)	
Pulmonary Hypertension Medication Changes, n (%)	11 (22)	
NYHA Functional Class, n (%)		
1	1 (2)	
2	5 (10)	
3	32 (64)	
4	4 (8)	
Undefined	9 (21)	
Disease Level, n (%)	Right	Left
0	0(0)	3 (5.9)
1	17 (33.3)	6 (11.8)
2	23 (45.1)	25 (49.0)
3	11 (21.6)	17 (33.3)
4	0 (0)	0 (0)
Hemodynamics	Pre-Surgery	Post-Surgery
mPAP, mmHg	45 [36 - 54]	22 [18 - 26]
PVR, dyn s cm^{-5}	552 [346 - 758]	225 $\pm$ 108
CO, L/min	4.86 $\pm$ 1.53	5.75 $\pm$ 1.25

1.3.5. Visual assessment of hypoperfusion and comparison with HLV

Two fellowship-trained cardiothoracic radiologists (author SK: 13 years of post-cardiovascular radiology fellowship experience, author LH: 3 years post cardiovascular

radiology fellowship experience) blinded to HLV and PB scores performed independent, semi-quantitative visual evaluation of hypoperfusion for all patients with CTEPH. Iodine-water images were reviewed and scored based on the percentage of each lobe considered to be hypoperfused. Readers were asked to categorize each lobe as having either no (0% of lobe affected), minimal (1% to <25% of lobe affected), mild (25 to<50% of lobe affected), moderate (50% to <75% of lobe affected), severe (75% to<100% of lobe affected), or complete (100% of lobe affected) hypoperfusion. Both readers had clinical experience evaluating CTEPH multi-energy CT scans as attending radiologists at our institution but were not given specific instructions nor training cases for this task.

1.3.6. Surgically Defined Segmental Disease Subgroup Analysis

HLV's ability to assess spatial hypoperfusion deficits in patients with chronic obstructions originating in the segmental pulmonary vasculature (surgically defined disease level 3) (5) was evaluated using classifications obtained during surgery. At the time of PTE surgery, each lung of each CTEPH case was classified as having disease located at the main pulmonary or lobar artery level (University of California San Diego (UCSD) surgical disease level classification 1 or 2), at the segmental level (UCSD surgical disease level classification 3), or at the subsegmental level (UCSD surgical disease level 4) by surgeons blinded to this analysis (5). The surgically defined segmental disease subgroup contained patients who had one or both lungs classified as having disease level 3. No patients in our cohort had disease level 4. Global HLV's performance in patients with surgically defined segmental disease and correlation with invasive hemodynamics, as well as lobar HLV's correlation with expert visual assessment were measured. Of the 51 patients with CTEPH analyzed, 32 had surgically defined segmental disease. Of the 33

patients with CTEPH with complete hemodynamics, 12 had surgically defined segmental disease.

1.3.7. Cases series to highlight the utility of HLV

Three cases are shown to demonstrate how HLV can capture known patterns of CTEPH impairment (n=2) and how HLV can help identify patients with significant obstructions despite high PBV (n=1). For each case, a representative coronal slice of the HLV map is presented alongside lobar scores, reader visual assessment scores, and the patient's subsequent surgical specimen.

1.3.8. Statistics

Demographic measures, image acquisition parameters, and invasive hemodynamics were tested for normality using the Shapiro-Wilks test. Normally distributed variables are reported as mean ± standard deviation, and non-normally distributed variables are reported as median [IQR from Q1 to Q3]. The Student t-test and one-way ANOVA were used to assess normally distributed variables, while Wilcoxon rank sum and Kruskal-Wallis tests were used for non-normally distributed variables with p=0.05. Receiver operating characteristic curves (ROC) curves were used to identify both an HLV and PBV cutoff that best separated CTEPH scans from control scans. Given that the CTEPH cohort was comprised of patients with CTEPH necessitating surgical treatment, we evaluated global HLV and PBV cutoff performance at 90% sensitivity to minimize false negatives. ROC curves were compared using the Mann-Whitney statistic [40,41]. Differences in global HLV and PBV cutoff specificity were compared with the McNemar test, while differences in accuracy were compared using z-statistic[42]. The agreement between global HLV and global PBV was determined via the Pearson correlation coefficient. Correlation of preoperative hemodynamics (mPAP, PVR, and CO) and post-operative change in

hemodynamics (ΔmPAP, ΔPVR, and ΔCO) for both global HLV and PBV were also evaluated and compared using the Pearson correlation coefficient. Correlation between reader scores and lobar HLV for each reader was evaluated using the Spearman correlation coefficient. Forward multiple regression was performed to evaluate the complementary nature of HLV and PBV in predicting preoperative hemodynamics (mPAP, PVR, and CO). Correlations were classified as little to no relationship ($0 \leq \rho < 0.25$); fair ($0.25 \leq \rho < 0.5$); moderate ($0.5 \leq \rho < 0.75$); or excellent ($\rho \geq 0.75$)[43]. Correlation coefficients in dependent samples were compared using the t-score[44,45]. Agreement between reader visual assessments was evaluated using intraclass correlation coefficient (ICC). ICCs were classified as little to no relationship ($0 \leq ICC < 0.5$); moderate ($0.5 \leq ICC < 0.75$); good ($0.75 \leq ICC < 0.9$); or excellent ($\rho \geq 0.9$)[46]. Analysis was performed in MATLAB 2021a (MathWorks, Natick, MA). Multiple regression was performed in SPSS (IBM SPSS Statistics Version 28.0.1.1)[43,47]. Statistics were performed by authors EB and FC.

1.4. Results

1.4.1 Patient Cohorts

The CTEPH cohort comprised 51 patients, and the control cohort comprised 110 patients. Controls and patients with CTEPH (**Table 1)** were similar in age (51 ± 16 years vs 47 ± 17 years, respectively; p =0.14). Patients with CTEPH had lower BMI than controls (31.1 ± 7.5 kg/m^2 vs 39.2 ± 6.6 kg/m^2, p<0.001), and more patients in the CTEPH cohort self-identified as African American (n=12, 24% vs n=6, 5%, p=0.03). The control group had more females (n=78, 71% vs n=27, 53%, p=0.03). There was a significant difference in the age of females vs males in the control cohort (48 ± 16 years vs 58 ± 13 years; p=0.01) and no evidence of a difference in

the CTEPH cohort (44±16 vs 50±17; p= 0.17). For both cohorts, there was no evidence of significant differences between female and male patients in terms of BMI or race.

Patients with CTEPH cohort had a mean preoperative mPAP of 45 [36-54] mmHg, PVR of 552 [346–758] dyn·s·cm^{-5}, and CO of 4.86±1.53 L/min **(Table 2)**. PTE decreased mean mPAP to 22 [18-26] mmHg and mean PVR to 225±108 dyn·s·cm^{-5} and increased mean CO to 5.75±1.25 L/min. The majority of patients with CTEPH were New York Heart Association Heart Failure Functional Class 3 (n=32, 64%) prior to PTE.

Right lungs had CTEPH surgical disease level 1 (n=17, 33.3%), disease level 2 (n=23, 45.1%), and disease level 3 (n=11, 21.6%). Left lungs had CTEPH surgical disease level 0 (n=3, 5.9%), level 1 (n=6, 11.8%), disease level 2 (n=25, 49.0%), and disease level 3 (n=17, 33.3%). No patients had disease isolated to the subsegmental vasculature (level 4) in either the left or right lung.

1.4.2. Quantification of Hypoperfusion Lung Volume and Comparison to PBV

Global HLV (area under the ROC curve [AUC]=0.84, 95% CI: 0.81-0.87) and PBV (AUC=0.79, 95% CI: 0.75–0.82) separated patients with CTEPH from controls (**Figure 6A**). Detection of CTEPH cases from controls with 90% (46/51) sensitivity corresponded to global HLV>10% and global PBV<80.5%. Global HLV>10% had 72% (116/161) accuracy and 64% (70/110) specificity. Global PBV<80.5% had similar accuracy (63%, 102/161; p=0.10) but significantly lower specificity (51%, 56/110; p<0.001) (**Table 3**). Global HLV and PBV had a moderate negative correlation (ρ=-0.72, p<0.001) (**Figure 6B**). Of the five patients with CTEPH with global PBV within the normal range (>80.5%), three had global HLV indicating disease (HLV>10%). Surgery led to a sizeable decrease in PVR (237-298 dyn·s·cm^{-5}) in these three patients.

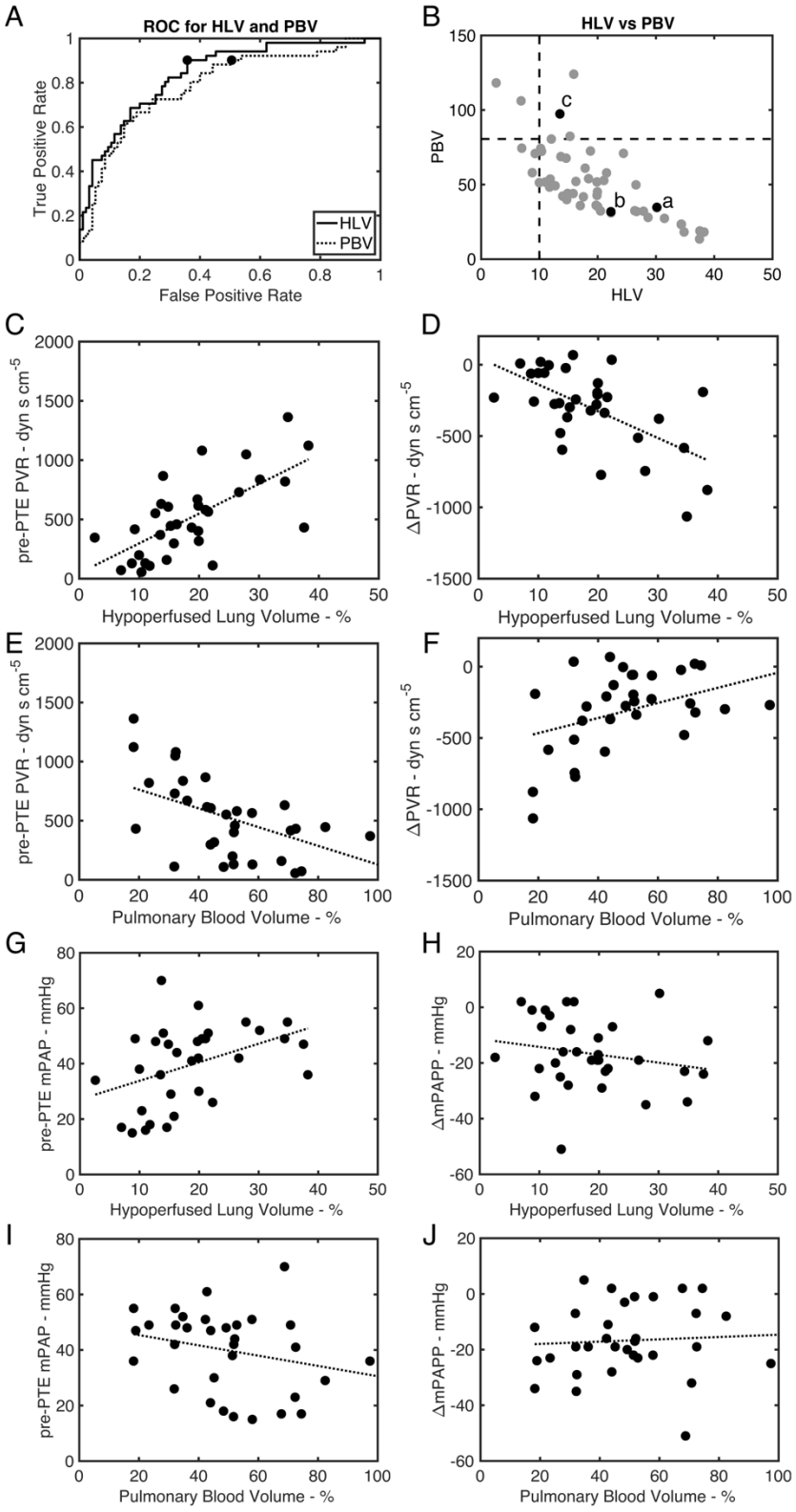


Figure 6 Dual-energy metrics are broadly in agreement and can complement hemodynamic assessment

Table 3 Separation of control patients and patients with CTEPH using global HLV and PBV.

Metric	AUC	Threshold	Accuracy	Sensitivity	Specificity
HLV	0.837	>10.0	116/161 (72%)	46/51 (90%)	70/110 (64%)
PBV	0.785	<80.5	102/161 (64%)	46/51 (90%)	56/110 (51%)
p-value	0.63	N/A	0.10	N/A	<0.001

1.4.3. Correlation of Hypoperfusion Metrics with Hemodynamics

Pre-operative PVR correlated moderately with global HLV (ρ=0.67, p<0.001) and mean PBV (ρ=-0.54, p=0.001) (**Table 4**). Similarly, ΔPVR after PTE correlated moderately with global HLV (ρ=-0.61, p<0.001) and fairly with PBV (ρ=0.43, p=0.01) (**Figure 6C-F**). Although a better degree of correlation was observed with global HLV compared with PBV, the t-score associated with this difference did not reach statistical significance for either preoperative PVR (p=0.15) or ΔPVR (p=0.08) metrics.

Table 4 Correlation of global HLV and PBV with pre-operative hemodynamics and change in hemodynamics after PTE

	HLV		PBV	
	ρ	p-value	ρ	p-value
All patients with hemodynamics(n=33)				
Pre-Operative				
PVR	0.67	<0.001	-0.54	0.001
mPAP	0.42	0.02	-0.29	0.10
CO	-0.52	0.002	0.37	0.03
Change after Surgery				
ΔPVR	-0.61	<0.001	0.43	0.01
ΔmPAP	-0.20	0.27	0.07	0.68
ΔCO	0.32	0.07	-0.03	0.85
Only those with Segmental Disease (n=12)				
Pre-Operative				
PVR	0.81	0.002	-0.48	0.12
mPAP	0.43	0.16	-0.17	0.60
CO	-0.80	0.002	0.54	0.07
Change after Surgery				
ΔPVR	-0.70	0.01	0.31	0.33
ΔmPAP	-0.29	0.37	0.00	0.99
ΔCO	0.45	0.14	0.1	0.76

Preoperative mPAP correlated fairly well with global HLV (ρ=0.42, p=0.02) and showed no correlation with PBV (ρ=-0.29, p=0.10). Neither global HLV (ρ=-0.20, p=0.27) nor PBV (ρ=0.07, p=0.68) was correlated with ΔmPAP after surgery (**Figure 6G-J**). Correlation of pre-operative CO with global HLV (ρ=-0.52, p=0.002) and PBV (ρ=0.37, p=0.03) were not different (p=0.16). Likewise, ΔCO after surgery was not correlated with global HLV (ρ=0.45, p=0.14) or PBV (ρ=0.1, p=0.76).

The addition of PBV to a linear regression using HLV (via forward multiple linear regression) did not improve model performance for PVR (p=0.85), mPAP (p=0.76), CO (p=0.75), or ΔPVR (p=0.68).Note, there were significant correlations between pre-operative hemodynamics; specifically, PVR and mPAP (ρ=0.72 , p<0.001) and PVR and CO (ρ=-0.76 , p<0.001). However, mPAP and CO were not significantly correlated (ρ=-0.32 , p=0.07).

1.4.4. Expert Visual Assessment of Lobar Hypoperfusion and Comparison to Lobar HLV

For visual assessment, 255 lobes (51 patients with CTEPH each with 5 lung lobes scored) were independently assessed by each reader. Reader 1 identified 247/255 lobes as hypoperfused (16 minimal, 27 mild, 82 moderate, and 93 severe, 29 complete) in the CTEPH cohort **(Figure 7A),** while reader 2 identified 225/255 lobes as hypoperfused (31 minimal, 46 mild, 58 moderate, and 65 severe, 25 complete) **(Figure 7B)**. The HLV of lobes graded as completely, severely, moderately, and mildly hypoperfused were significantly different (p <0.001) for both expert readers.

Visual assessments by reader 1 and reader 2 correlated moderately with lobar HLV (ρ= 0.71 and ρ= 0.67; p<0.001 for both) and moderately with lobar PBV (ρ= -0.57 and ρ= -0.68; p <0.001 for both). For reader 1, visual assessment was more strongly correlated with lobar HLV than with PBV (*t*-statistic=5.2, p<0.001). For reader 2, there was no evidence of a difference between correlations for lobar HLV and PBV (*t*-statistic=- 0.29, p=0.61). Agreement between the two readers was moderate, with an ICC of 0.61 (p<0.001); 60% (133/255) of lobes had matching classifications **(Figure 7C)** and 85% (218/255 lobes) had reader scores within 1 grade. Of the 37 disagreements of more than 1 grade, 36 occurred in lobes that reader 2 identified as having no to mild hypoperfusion.

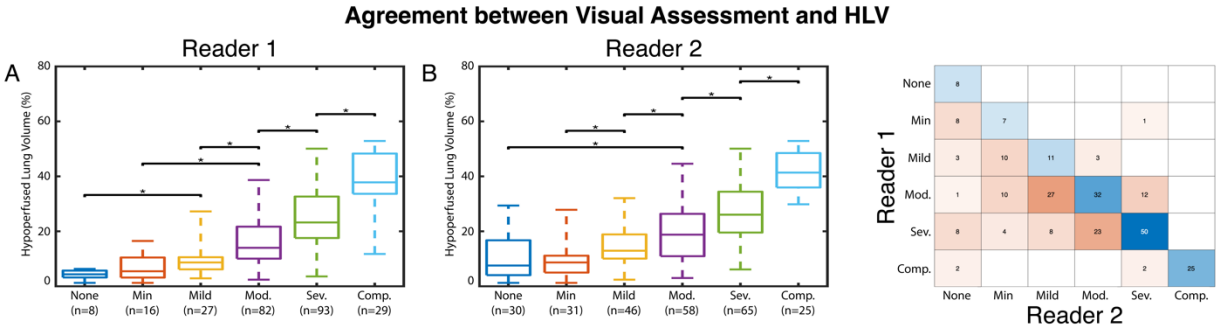


Figure 7 Lobar HLV percent hypoperfusion agrees with lobar hypoperfusion severity scores from expert radiologists

1.4.5. Surgically Defined Segmental Disease Subgroup Analysis

Separation of patients with CTEPH surgically defined as level 3 (originating in segmental vessels) (n=19) from controls had similar performance as the overall CTEPH cohort. A global HLV threshold of >10.0% led to a sensitivity of 95% (18/19) and accuracy of 67% (87/129), which differ from the sensitivity (p=0.52) or accuracy (p=0.67) found for the overall CTEPH cohort. Global PBV threshold of <80.5% led to a sensitivity of 95% (18/19) and accuracy of 57% (73/129), with both values similar to those found for patients with CTEPH of any disease level

(sensitivity, p=0.52; accuracy p=0.56). Global HLV and PBV correlated moderately (ρ=-0.71, p<0.001) in patients with surgically defined segmental CTEPH, which agreed (p=0.96) with the correlation observed in the overall CTEPH cohort.

In patients with segmental disease and invasive hemodynamics (n=12), global HLV strongly correlated with pre-operative PVR (ρ=0.81, p=0.002) and moderately correlated with ΔPVR after PTE (ρ=-0.70, p=0.01) (**Table 4**). Global PBV did not significantly correlate with any hemodynamics in patients with segmental disease.

Visual assessment of surgically defined segmental CTEPH moderately correlated with lobar HLV for reader 1 (ρ=0.68, p<0.001) and reader 2 (ρ=0.63, p<0.001). These correlation coefficients were similar to those calculated for visual assessment for all disease levels (reader 1: p=0.54; reader 2 p=0.36). Similarly, the correlation of lobar PBV with visual assessment in segmental CTEPH was similar to the overall CTEPH cohort (reader 1: ρ=-0.49, p<0.001, $p_{segmental\_vs\_all}$=0.56; reader 2: ρ=-0.60, p<0.001, $p_{segmental\_vs\_all}$=0.21).

1.4.6. Example Cases

Example cases (labeled as a-c in **Figure 6b**) were chosen to highlight the use of HLV in specific clinical scenarios. Patient A in **Figure 8** has bilateral, predominantly lower lobe hypoperfusion, while Patient B presents with a unilateral pattern of disease. Both patients have similar global PBV values (A: 34.7%, B: 31.8%) but differ in their global HLV (A: 30.2%, B: 22.3%) and lobar HLV values. Readers 1 and 2 agreed on the visual evaluation of both patients, and the rank order for the reader scores matches the rank order for the lobar HLV scores in all lobes except the left upper lobe of Patient A.

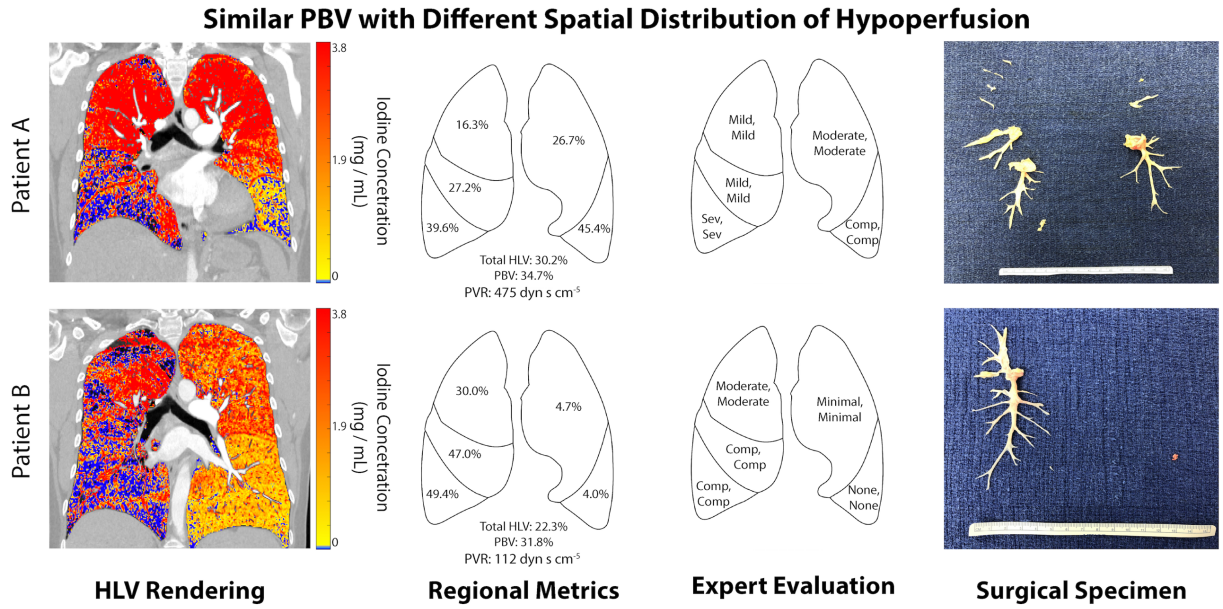


Figure 8 HLV can characterize spatial patterns of chronic thromboembolic regional perfusion deficits.

Figure 9 demonstrates findings in a patient who had a global PBV of 97.4% (normal >80.5%) despite having hemodynamic impairment and removed surgical specimen. This patient had a global HLV of 13.5% (>10.0% cutoff). The readers agreed on the severity of hypoperfusion of the lobes relative to one another, but reader 1 graded the hypoperfusion as more severe than reader 2.

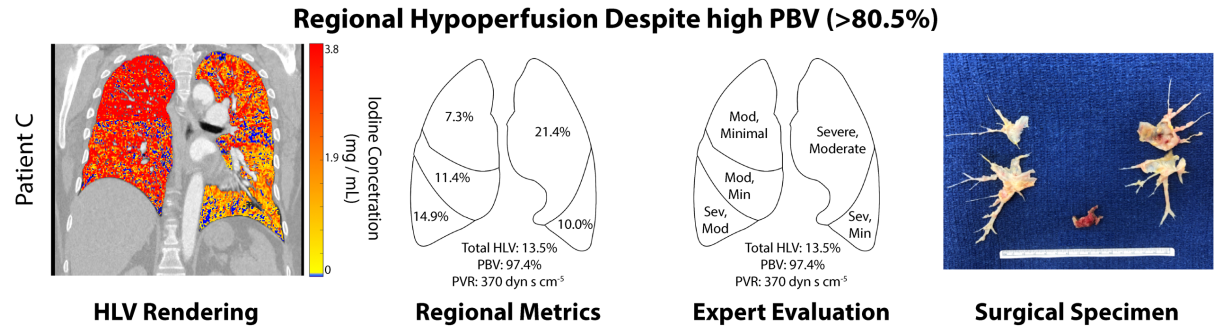


Figure 9 HLV can visualize and quantify small regional perfusion deficits in patients with high PBV.

1.5. Discussion

We have developed and validated an automated, quantitative approach to define, on a per-pixel basis, regions of hypoperfusion in patients with CTEPH using DECT and evaluated its potential clinical use. The developed metric, global HLV, correctly classified patients with CTEPH and controls (AUC: 0.84), and had higher specificity relative to global PBV (64% VS 51%). Global HLV correlated with PVR at the time of imaging (preoperatively, ρ=0.67) and predicted change in PVR after surgery in the entire cohort (ρ=-0.61) as well as in the subset of patients who had surgically defined segmental disease (ρ=-0.70). Lobar HLV agreed with visual assessment by two expert radiologists (ρ=0.71 and ρ=0.67).

Throughout the manuscript, HLV findings are presented alongside a previously-published, dual-energy based measurement of global mean lung parenchymal perfusion (PBV) [25,27,32–34]. PBV and HLV are expected to provide complementary information as HLV measures the spatial extent of impairment while PBV is an average of parenchymal iodine concentration. As we have shown, impairments limited to a small portion of the lung may not affect the global mean iodine concentration but may result in pixels identified as being hypoperfused. We highlight how HLV mapping agrees with known CTEPH phenotypes. This could aid clinical evaluation, as the surgical technique and expertise required varies based on phenotype[48]. Further, our approach to measure HLV results in pixel-wise, lobar, and global values of hypoperfusion that can be easily integrated with other patient data. Future work is planned to prospectively evaluate the use of these HLV metrics and determine an optimal way to combine PBV, HLV, and other patient data to further improve clinical assessment.

HLV correlated with preoperative and post-operative change in hemodynamic severity, in the form of PVR, in patients with surgically defined disease originating in the main or lobar

artery as well as patients with disease originating in the segmental arteries. While HLV may potentially help identify patients who are likely to have excellent or limited response pre-operatively (by identifying preoperative features of surgical success), few patients in our cohort had what would be considered limited surgical response [7,49]. Thus, the ability to detect these rare events is left for future work. Cases of primarily segmental or subsegmental level disease are technically more challenging surgically, and patients may not receive adequate or thorough clearance except in the highest volume endarterectomy centers. As a retrospective study, we did not target enrollment of patients with these specific subtypes. Further, we did not evaluate non-operated cases such as patients treated with balloon pulmonary angioplasty, medications, or a combination of treatment modalities. Therefore, the use of HLV in these cohorts should be investigated in future work.

Visual evaluations of lobar hypoperfusion by two expert readers moderately agreed with one another and with lobar HLV. 85% (217/255) of lobes were scored as having the same severity or differed by only 1 category. Readers did not undergo specific training for this visual scoring task (outside their normal clinical training), which may explain variation in grading. However, this variation reflects typical clinical practice in which individual readers may vary in sensitivity and specificity regarding hypoperfusion, particularly given relatively recent clinical adoption of dual energy imaging. The moderate agreement of HLV with both readers suggests that lobar HLV could serve to support and adjudicate reader evaluations, as having an external quantitative metric has shown to improve agreement in visual assessment for other pulmonary visual assessment tasks[50]. However, further studies are needed to evaluate these potential clinical uses of HLV.

The correlation between global PBV and PVR in our study (ρ=-0.54) was similar to that found by Takagi et al[34] (ρ=0.47) and higher than that reported by Meinel et al[33] (ρ=-0.20). Both Takagi et al (ρ=0.48) and Meinel et al (ρ=-0.57) reported correlations between global PBV and mPAP, which were not observed in this study. These differences may be due to the differences in the CTEPH samples studied. In our study, all patients with CTEPH underwent PTE surgery; however, Takagi et al included patients who underwent non-surgical treatments (medical therapy and balloon pulmonary angioplasty), and Meinel et al studied patients with CTEPH regardless of intervention. Further, we found that global HLV did not correlate with ΔCO after surgery. This may be due to evaluation of post-operative hemodynamics while patients recovered in the intensive care unit (typically within 7 days of surgery). Long-term post-operative hemodynamic evaluation should be evaluated in future work.

Our study had several limitations. First, given the observational and retrospective nature of the study, patients in the control cohort had clinical symptoms that merited CT investigation even though they did not have pulmonary findings. It is possible that undetected pulmonary or vascular disease in control patients could have impacted iodine values and limited HLV specificity. For example, we did not exclude patients with small airway disease, which can lead to hypoperfusion. Additionally, the study did not evaluate other types of pulmonary hypertension. The study's retrospective design and the resulting effects on cohort differences is also a limitation. Despite excluding patients with BMI>50 kg/m^2, control patients had significantly higher BMIs than patients with CTEPH; this may be attributed to dual energy CT being used for improved imaging of suspected PE in patients with high BMI at our institution. However, the higher BMI in the control cohort further highlights the robustness of the approach, as high BMI could decrease parenchymal iodine (and increase HLV) in control patients, making

it more difficult to separate CTEPH and control cases. We also observed racial and sex differences between the cohorts, which can be attributed to differences between patients referred for PTE surgery (a national referral population) and patients imaged for suspicion of PE (our local population).

Additionally, iodine concentration maps were subject to material decomposition errors, leading to negative iodine concentration values. In thoracic dual energy CT images, it is common to have highly negative pixels due to streak artifacts from regions with high contrast concentrations, such as in the superior vena cava. Additionally, motion artifacts from cardiac and diaphragmatic motion can lead to photon starvation artifacts [39,51]. We aimed to minimize these artifacts by using a lower intensity threshold (-3mg/mL iodine) to remove pixels with values likely due to artifact. Dual energy material decomposition can also have a bias in the iodine concentration. In phantom imaging studies, negative biases of -1.5 mg/mL have been reported. This bias supports our hypoperfusion threshold of 0mg/mL as capturing pixels with poor perfusion.

Furthermore, while our study evaluated a considerable number of patients with CTEPH who underwent surgical treatment, inclusion of more patients would improve subgroup analysis and generalizability of this study. The study design did not control for the level of CTEPH disease removed at the time of PTE surgery nor the timing of CT imaging, hemodynamic assessment, and surgery. As a result, our cohort did not include patients with disease originating in the subsegmental arteries (surgical disease level 4) and may be limited by changes between imaging and hemodynamic measurement. Thus, future investigation in a broader patient cohort reflecting all disease levels, is warranted. Additionally, the single site design reduced variability

in CTEPH diagnostic and treatment criteria, and our results should be confirmed in a validation cohort as well as at other institutions.

Another limitation is that we compared HLV and PBV with only visual evaluation by expert readers. Future work should compare HLV and PBV with detailed analysis of surgically removed specimens and evaluate changes in HLV and PBV after PTE surgery for additional clinical validation. Unfortunately, the surgical specimens we evaluated did not document whether specific vessels were open, partially obstructed, or completely obstructed. This precluded quantitative comparisons between imaging metrics and surgical findings. In future work, we aim to assess the agreement between areas detected using HLV and presence of surgically confirmed specimens. Finally, post-operative imaging was not obtained in our patients. Longitudinal DECTPA imaging would allow for confirmation that areas of hypoperfusion improve with surgery.

1.6. Conclusion

In conclusion, dual-energy perfusion imaging of the lungs can be combined with lobar segmentation to automatically visualize perfusion deficits and provide pixel-wise quantification of the spatial extent of hypoperfusion. The spatial extent of hypoperfusion, global HLV, was able to separate patients with CTEPH from controls and correlated with invasive PVR and change in PVR with surgery, even in patients with surgically defined segmental disease. Lobar HLV values agreed with expert visual assessment. These findings suggest that HLV evaluation could provide quantitative, imaging-derived, perfusion metrics to evaluate disease progression or response either independently or when combined with other patient clinical and imaging data, warranting further investigation.

1.7. Acknowledgements

Chapter 1, in full, is a reprint of the material contained in “Mapping the Spatial Extent of Hypoperfusion in Chronic Thromboembolic Pulmonary Hypertension Using Dual Source CT” accepted to (published online 10 August 2023) *Radiology: Cardiothoracic Imaging* by Elizabeth Bird, Kyle Hasenstab, Nick Kim, Michael Madani, Atul Malhotra, Lewis Hahn, Seth Kligerman, Albert Hsiao, and Francisco Contijoch. Elizabeth Bird, the dissertation author, was the primary author and investigator on this paper.

CHAPTER 2: QUANTIFICATION OF FIBROTIC CLOT IN CTEPH

2.1. Abstract

The purpose of this study was to determine if quantitative evaluation of fibrotic clot burden in the both the large pulmonary arteries and segmental arteries can aid pre PTE evaluation of chronic thromboembolic pulmonary hypertension (CTEPH) disease severity. DECT angiograms in 32 patients with CTEPH (mean age, 47 ± 18 years, 15 females) had fibrotic clot location and obstruction severity annotated by thoracic radiologists as part of normal CTEPH clinical evaluation. Fibrotic clot annotations were retrospectively converted into a fibrotic clot score, which was a percent obstruction of the lung vasculature determined based on the location and severity of the annotated obstructions for a given patient. Fibrotic clot scores correlated moderately with pre-operative PVR ($\rho=0.60$, $p<0.001$) and pre-operative CO ($\rho=-0.56$, $p<0.001$), but only fairly with pre-operative mPAP ($\rho=0.43$, $p<0.001$). Fibrotic clot score achieved only fair correlation with post-operative change in PVR ($\rho=-0.49$ $p=0.004$) and change in CO ($\rho=0.49$, $p<0.001$), and had no correlation with post-operative change in mPAP. Both fibrotic clot score and HLV had moderate strength correlations with pre-operative PVR and fair correlations with post-operative change in PVR, but only fibrotic clot score had a significant correlation with both pre-operative CO and post-operative change in CO. Fibrotic clot score and Hypoperfused Lung Volume (HLV) significantly correlated with one another, with a moderate strength of correlation ($\rho=0.52$, $p=0.002$). Fibrotic clot score that takes into account vascular obstruction down to the segmental level and agrees with CTEPH pre-operative and post operative pulmonary hemodynamic severity, as well as perfusion deficit measured as HLV.

2.2. Introduction

As mentioned in chapter 1, patients with CTEPH that are thought to be good PTE candidates have agreement between their disease burden and the hemodynamic severity of their pulmonary hypertension[6,23]. Disease burden includes both the fibrotic clots that a patient has in their pulmonary vasculature, as well as the related pulmonary parenchymal perfusion deficits. PTE primarily directly intervenes on chronic thrombotic obstructive sources of pulmonary hypertension (fibrotic clots), so elevations in pulmonary pressures due to microvascular disease or other causes of pulmonary hypertension are less likely to be improved. These non-chronic thrombotic sources of pulmonary hypertension are suspected when patients have pulmonary hypertension that is out of proportion to their fibrotic clot burden or when patients have perfusion deficits that do not correspond to fibrotic clot obstruction. Because pulmonary thromboendarterectomy surgery treats the elevated pulmonary pressures of CTEPH through removal of a patient's fibrotic clot, it is important that the fibrotic clot and pulmonary parenchymal perfusion deficit, as well as hemodynamic severity, all agree with one another.

Contrast enhanced, monoenergetic CT pulmonary angiograms (CTPAs) offer high resolution approach to evaluate fibrotic clot location and obstruction severity within the pulmonary arterial tree. Additionally, CTPAs can be obtained as part of a dual-energy CT pulmonary angiogram acquisition (see chapter 1), which provides more complete disease burden information with the addition of pulmonary parenchymal perfusion information in addition to intravascular fibrotic clot information. Vascular obstruction due acute pulmonary embolism has been quantified previously on CTPA using the Qanadli score, which captures the proportion of the lung vasculature that is obstructed based on obstruction location and severity, with consideration of central and peripheral vessels (down to the segmental and subsegmental level)[52]. In patients with

acute PE, this score has been found to agree with DECT perfusion deficit findings [53,54] In CTEPH patients however, modified Qanadli scores have shown no or only fair correlation with either pre-operative or post-operative pulmonary hemodynamics[55,56]. This is likely because, to date, no study calculates fibrotic clot score that captures both central and peripheral clot location in a cohort of CTEPH patients that has undergone PTE surgery. Without such a study, agreement between fibrotic clot score and perfusion deficit in patients who have undergone PTE can not be assessed.

The objective of this study was to further pre-operative disease burden evaluation in CTEPH patients. First, we aim to determine if quantitative CT pulmonary angiogram fibrotic clot scoring provides utility in prediction of CTEPH pre-operative and post-operative pulmonary hemodynamic severity. Next, we assess if CT pulmonary angiogram fibrotic clot scores agree with, and provide complimentary information to, perfusion deficit scores obtained from dual energy CT pulmonary angiograms. Our hypothesis is that the fibrotic clot score will agree with presurgical disease severity, changes in hemodynamics after surgical intervention, and perfusion deficit measured as Hypoperfused Lung Volume (HLV).

2.3. Methods

2.3.1. Patient Selection

Patient selection was performed under institutional review board-approved (#191797) waiver of informed consent and in compliance with the Health Insurance Portability and Accountability Act. Our cohort was comprised of patients diagnosed with CTEPH who had undergone per-operative CT pulmonary angiogram under specific CTEPH clinical protocol, which includes annotation of the main pulmonary arteries down to the subsegmental arteries fellowship trained thoracic radiologist. Patient diagnosis and treatment selection was determined

by consensus of a multidisciplinary team of cardiothoracic surgeons, vascular disease expert pulmonologists, and fellowship trained cardiothoracic radiologists. Clinicians assessed CTEPH diagnosis using the criteria from Galiè et al, which include mean pulmonary artery pressure (mPAP) >25mmHg, pulmonary artery wedge pressure ≤15 mmHg, pulmonary vascular resistance ≥240 dyn·s·cm^{-5}, CTEPH-specific imaging signs, and ventilation-perfusion mismatch[3]. While more current criteria defines pulmonary hypertension as mPAP ≥ 20mmHg, the Gailé criteria were used as they reflected the pulmonary artery pressure criteria at the time that the cases were evaluated. Patients were included if they had also obtained a dual-energy CT pulmonary angiogram at the time of annotated contrast CT pulmonary angiogram, had preoperative pulmonary hemodynamics measured by right heart catheterization, and had undergone PTE surgery at our institution for definitive treatment of CTEPH. Patients were excluded if they were taking pre PTE pulmonary hypertension medications, or had CT pulmonary angiograms that had missing vessel by vessel annotations. This led to a cohorts of 32 patients (similar but not identical to that in chapter 1) with CT pulmonary angiogram annotations, dual energy CT iodine maps, and pre- and post-operative pulmonary hemodynamics. Patient characteristics are summarized in **Table 5**.

Table 5 Demographics of patients undergoing fibrotic clot scoring

	CTEPH (n=32)
Age, years	47 ± 18
Sex, female, n (%)	15 (46.9)
BMI, kg/m2	32.8 ± 7.5
Race, n (%)	
African American	8 (25)
Asian	0 (0)
Pacific Islander	2 (6)
White	25 (47)
Other, Mixed, or Unknown	7 (22)

2.3.2. Image Acquisition

CT pulmonary angiograms, including mono-energetic contrast enhanced CT Pulmonary angiogram as well as multi energy CT pulmonary angiogram, were acquired using the same dual-energy protocol on a single single-source Revolution CT scanner (GE Medical Systems, Chicago, IL). The acquisition rapidly switched kVP (80 to 140 kVp), and data were acquired helically with an 80mm detector width, median pitch of 1.38 (range: 0.98-1.38), and 0.5 s revolution time. Median tube current was 485 mAs (range: 240-630 mAs). Patients were administered iohexol contrast (Omipaque) with a concentration of 350mg/mL based on BMI (BMI<20: 60mL at 4mL/s, BMI 20-29.9: 75 mL at 5 mL/s, BMI 30-39.9: 80mL at 5.5mL/s, BMI> 40: 90mL at 6mL/s; median volume 75 mL, range: 50-150 mL). Imaging was timed for pulmonary arterial phase imaging using “smart prep” bolus tracking, with images acquired 8 seconds after main pulmonary artery enhancement at the T4 level reached 125 HU[37]. Iodine-water material decompositions yielding iodine concentration in mg/mL were generated using the Gemstone Spectral Imaging software (GE Healthcare, Chicago IL). All images were

reconstructed on a 512x512 image grid with xy-resolution of 0.7±0.1 mm (range: 0.5-0.9 mm) and section thickness of 1.25 mm. The typical field of view was 345x345 mm (range: 250-470 mm). Virtual mono-energetic images were reconstructed from the dual-energy acquisition to simulate a 70 keV acquisition.

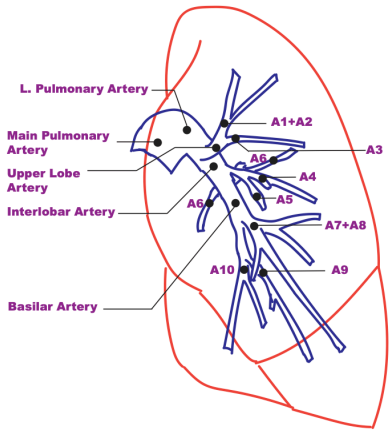
2.3.3. CT Pulmonary Angiogram Clinical Assessment and Vessel Annotation

CTPAs were assessed by fellowship trained thoracic radiologists performing standard clinical reads in accordance with the CTEPH CTPA protocol at our institution whereby radiologists annotated all vessels in the vasculature to state the location and severity of fibrotic clot obstruction. Vessels annotated included the main pulmonary artery, the left and right main pulmonary arteries, lobar arteries, segmental arteries and subsegmental arteries. Vessel severity was annotated to include whether an obstruction was partial or complete. Complete obstruction was indicated on annotations as abrupt vascular narrowing, no vessel opacity, or complete occlusive thrombus indicating complete obstruction. Partial obstruction was indicated on annotation by the presence of vascular webs or bands, eccentric thrombus with lumen narrowing, or mild- severe vascular narrowing without direct clot visualization. No obstruction was indicated on annotation by normal contrast opacity, a finding of no obstruction in the vessel, or a lack of annotated clot annotation in a given vessel.

2.3.4 Fibrotic Clot Score Calculation

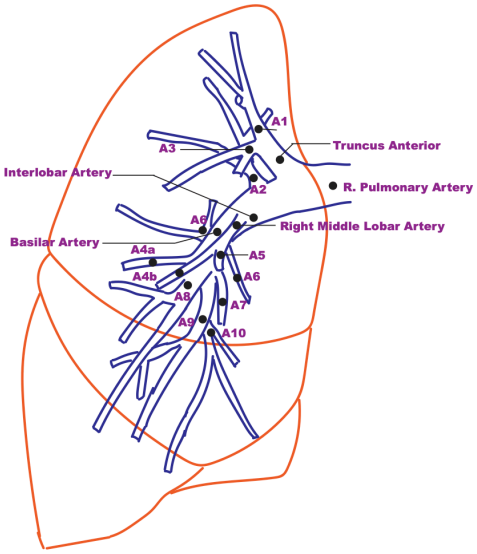
Vessels were scored using the Qanadli method, which captures the percentage of the segmental arteries that are obstructed in the arterial tree[52]. **Figures 10** and **11** highlight the segmental arteries for the left and right lung, respectively. Scores were determined first by assessing if the vascular territory of each segment was completely obstructed, partially obstructed or unobstructed. Obstruction status for each vessel was determined from radiologist

annotations. A segment was determined to be completely obstructed if any artery directly upstream of it (eg Main Pulmonary Artery for any of the segments; R. Basilar Artery obstruction for the Right Lower Lobe Medial Segmental Artery; etc.) or the segment itself had a complete occlusion. A segment was determined to be partially obstructed if it or any artery directly upstream of the segment itself had a partial occlusion. Segments were also determined to be partially obstructed if there was a subsegmental occlusion in the region of that segment, but this was not reflected in our cohort as there were no patients with subsegmental disease (**Table 6**). Segments deemed completely obstructed were given a score of 2, partially obstructed were given a score of 1, and unobstructed a score of 0. After each segment had been given a score, the total obstruction for each individual could be summed and divided by 40 (the maximum score if all segments in the bilateral lungs were affected) and multiplied by 100%. This gave a final fibrotic clot score that was the percentage of the segmental arterial tree that was obstructed relative to the whole segmental arterial tree.



Lobe	Segment Name	Weight
Upper	Apicoposterior -A1+A2	2
	Anterior - A3	1
Upper-Lingula	Superior - A4	1
	Inferior - A5	1
Lower	Superior - A6	1
	Anteromedial - A7+A8	2
	Lateral - A9	1
	Posterior - A10	1

Figure 10 Arterial Tree of the Left Lung Including Segmental Arteries



Lobe	Segment Name	Weight
Upper	Apical - A1	1
	Posterior - A2	1
	Anterior - A3	1
Middle	Lateral - A4	1
	Medial - A5	1
Lower	Superior - A6	1
	Medial - A7	1
	Anterior - A8	1
	Lateral - A9	1
	Posterior - A10	1

Figure 11 Arterial Tree of the Right Lung Including Segmental Arteries

2.3.5. Hypopefused Lung Volume Assessment

Hypoperfused lung volume, a pixel-wise classification and quantification of hypoperfused lung parenchyma to obtain a global hypoperfusion metric, was obtained from the multi-energy pulmonary angiograms using the methods described in detail in chapter 1. Briefly, lung parenchyma iodine concentration maps were automatically generated using automated segmentations from the lung parenchyma of the corresponding virtual monoenergetic images[38]. Smaller pulmonary vessels were removed next based on thresholded iodine values (pixels that were two standard deviations higher than the mean parenchymal concentration), as automatic segmentation previously excluded large pulmonary vessels, bones, and airways. Lung parenchyma pixels with an iodine concentration ≤ 0 mg/mL were classified as hypoperfused. All other pixels were assigned a value of 0. Hypoperfused lung volume (HLV) was then calculated

as the total number of hypoperfused parenchymal pixels divided by the total number of pixels in the region of interest.

2.3.6 Invasive Hemodynamics and Correlation with Perfusion

Mean pulmonary artery pressure and thermodilution-derived cardiac output (CO) before and after PTE surgery were obtained as part of routine clinical care. Pulmonary vascular resistance (PVR) was calculated preoperatively using right heart catheterization-measured pulmonary artery wedge pressure (PAWP) according to the following equation: PVR=(mPAP – PAWP)/CO. Postoperative PVR was calculated using the central venous pressure to estimate PAWP, as PAWP is not measured by Swan-Ganz catheter in the intensive care unit at our institution[1]. When multiple hemodynamics were obtained, pre-operative values obtained just before the PTE surgery and post-operative hemodynamics just prior to Swan-Ganz catheter removal were selected. Patients were only included in this study cohort if they had no change in pulmonary hypertension medications after PTE surgery to avoid the confounding effects of medications on the relationship between fibrotic clot score and pulmonary hemodynamics. Mean hemodynamics and CTEPH characteristics of the patient cohort are summarized in **Table 6**.

Table 6 Cohort hemodynamic and CTEPH characteristics

Parameter	Value	
NYHA Functional Class, n (%)		
1	0 (0)	
2	1 (3)	
3	23 (72)	
4	2 (6)	
Undefined	6 (19)	
Disease Level, n (%)	Right	Left
0	0 (0)	2 (6)
1	10 (31)	2 (12)
2	15 (47)	17 (57)
3	7 (22)	11 (37)
4	0 (0)	0 (0)
Hemodynamics	Pre-Surgery	Post-Surgery
mPAP, mmHg	$\underline{41 +}$ 13	21.5 [17.5-25.5]
PVR, dyn s cm^{-5}	$\underline{496 +}$ 287	211 [138-285]
CO, L/min	2.28 [1.73-2.83]	5.87 $\underline{+}$ 1.12

2.3.7. Statistics

Demographic measures and invasive hemodynamics were tested for normality using the Shapiro-Wilks test. Normally distributed variables are reported as mean ± standard deviation, and non-normally distributed variables are reported as median [IQR from Q1 to Q3]. Correlation of preoperative hemodynamics (mPAP, PVR, and CO) and post-operative change in hemodynamics (ΔmPAP, ΔPVR, and ΔCO) with fibrotic clot score and HLV were evaluated and compared using the spearman correlation coefficient. The agreement between fibrotic clot score and HLV was determined via the spearman correlation coefficient. Correlations were classified as little to no relationship ($0 \leq \rho < 0.25$); fair ($0.25 \leq \rho < 0.5$); moderate ($0.5 \leq \rho < 0.75$); or excellent ($\rho \geq 0.75$)[43]. Analysis was performed in MATLAB 2021a (MathWorks, Natick, MA). Multiple regression was performed in SPSS (IBM SPSS Statistics Version 28.0.1.1)[43,47].

2.4. Results

2.4.1 Correlation Between Fibrotic Clot Score and Invasive Hemodynamics

A summary of all the correlations between both fibrotic clot score and HLV with pulmonary hemodynamics are shown in **Table 7**.

Table 7 Summary of correlations between pulmonary hemodynamics and fibrotic clot score

	Fibrotic Clot Score		HLV	
	ρ	**p-value**	**ρ**	**p-value**
All patients with hemodynamics				
Pre-Operative				
PVR	0.60	<0.001	0.60	<0.001
mPAP	0.43	0.01	0.43	0.01
CO	-0.56	<0.001	-0.39	0.03
Changer after Surgery				
ΔPVR	-0.49	0.004	-0.46	0.008
ΔmPAP	0.00	1.00	-0.16	0.37
ΔCO	0.49	<0.001	0.21	0.25

Pulmonary Vascular Resistance

Fibrotic clot scores correlated moderately with pre-operative PVR (ρ=0.60, p<0.001, **Figure 12A**) and fairly with post-operative change in PVR (ρ=-0.49, p-0.004, **Figure 12B**). HLV also correlated moderately with pre-operative PVR (ρ=0.60, p<0.001) and fairly with post-operative change in PVR (ρ=-0.46, p-0.008).

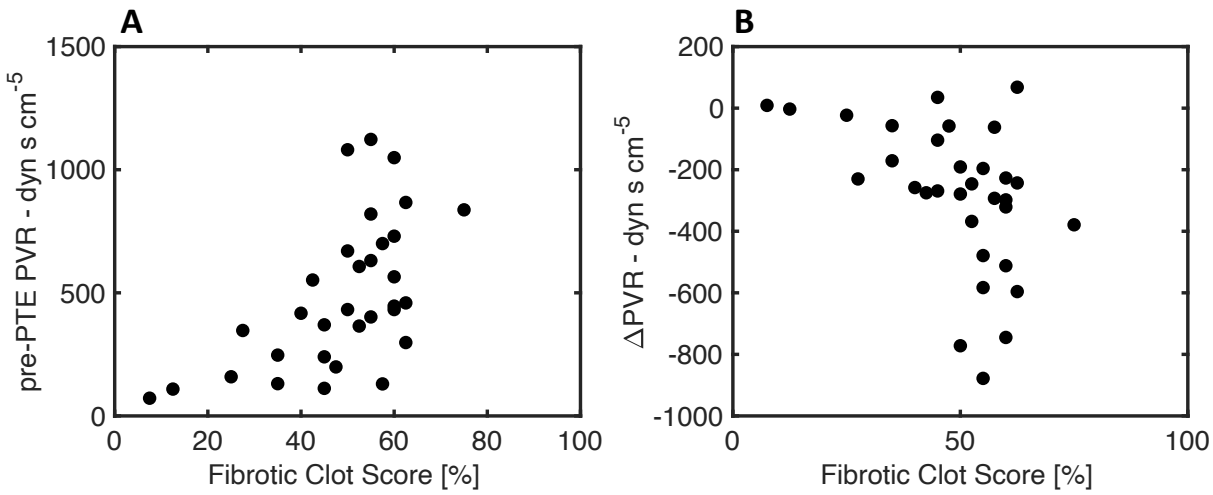


Figure 12 Fibrotic Clot Score Correlation with Pulmonary Vascular Resistance

Cardiac Output

Fibrotic clot scores correlated moderately with pre-operative CO (ρ=-0.56, p<0.001, **Figure 13A**) and fairly with post-operative change in CO (ρ=0.49, p<0.001, **Figure 13B**). HLV also correlated only fairly with pre-operative CO (ρ=-0.39, p=0.03) and did not correlate with post-operative change in CO (ρ=0.21, p=0.25).

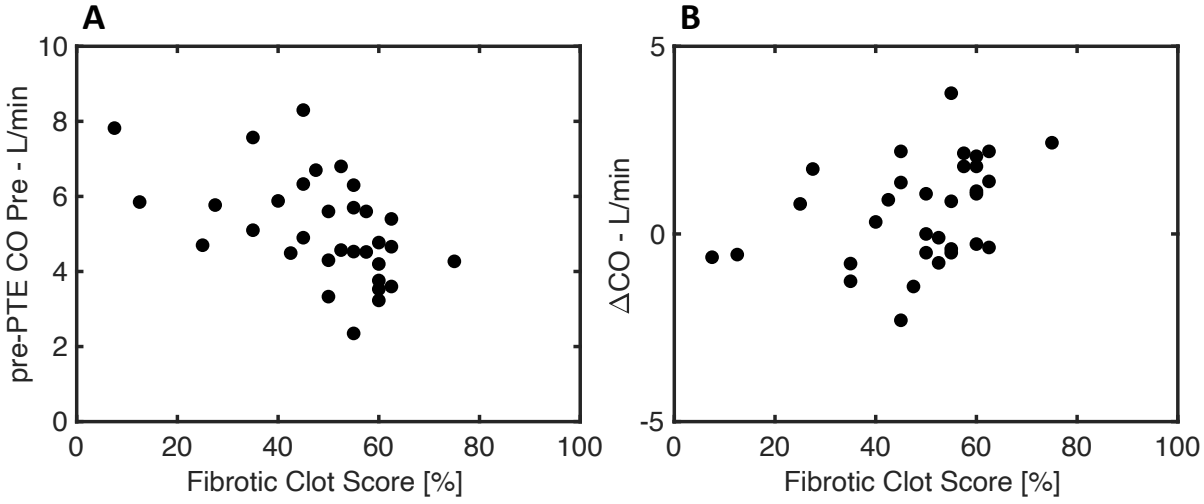


Figure 13 Fibrotic Clot Score Correlation with Cardiac Output

Mean Pulmonary Artery Pressure

Fibrotic clot scores correlated fairly with pre-operative mPAP (ρ=0.43, p=0.01, **Figure 14A**). HLV also corelated fairly with pre-operative mPAP(ρ=0.43 p=0.03) . Neither fibrotic clot score (ρ=0.00, p=0.9996, **Figure 14B**) nor HLV (ρ=-0.16, p=0.37) correlated with post-operative change in mPAP.

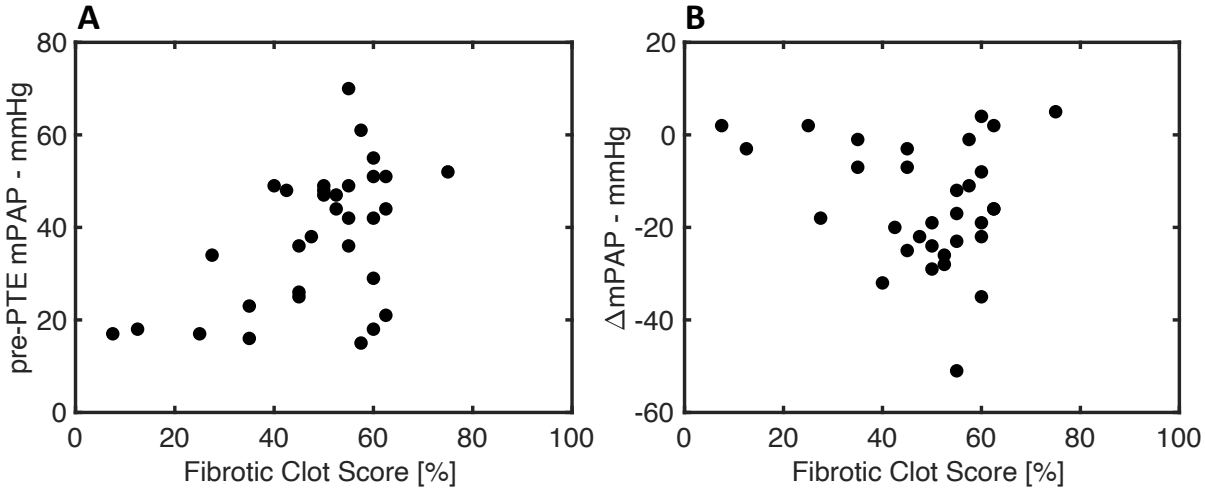


Figure 14 Fibrotic Clot Score Correlation with Mean Pulmonary Artery Pressure

2.4.2 Correlation between Fibrotic Clot Score and Hypoperfused Lung Volume

Fibrotic clot score correlated moderately with HLV, with spearman correlation coefficient of ρ=0.52 (p=0.02, **Figure 15**).

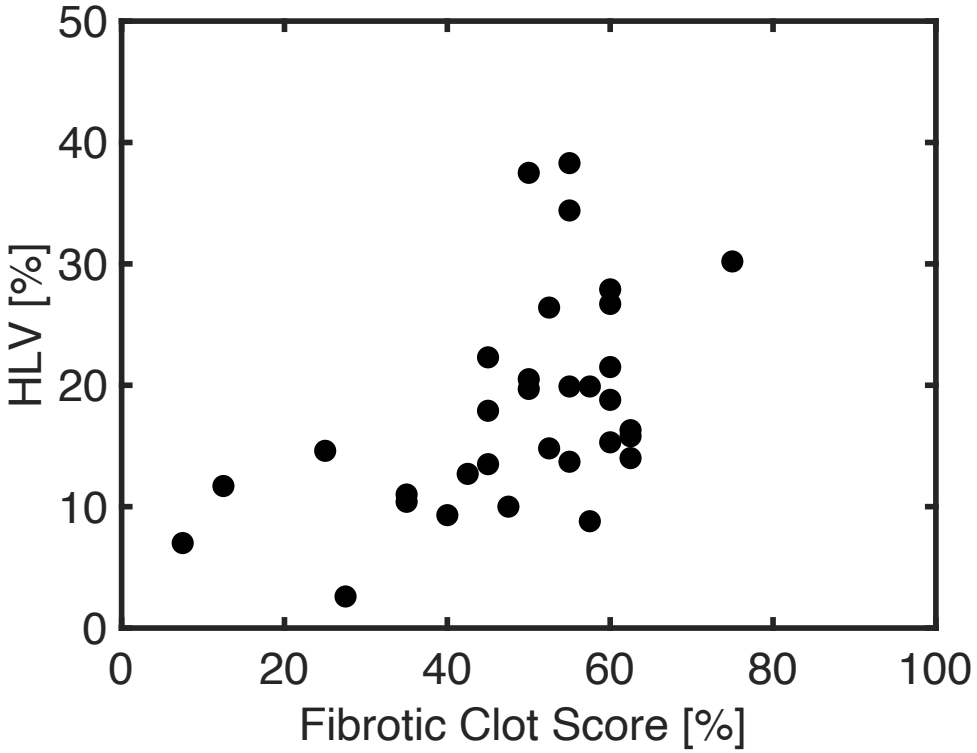


Figure 15 Fibrotic Clot Score Correlation with Hypoperfused Lung Volume

2.5. Discussion

We have quantified fibrotic clot amount and location and investigated its agreement with both perfusion deficit and invasive pulmonary hemodynamics in patients with CTEPH who have undergone PTE. As expected in a CTEPH population of good PTE candidates (most patients with post-operative PVR < 450 dyn·s·cm$^{-5)}$, fibrotic clot score, a measure of disease burden, agrees with CTEPH pulmonary hypertension hemodynamic severity (moderate correlation with pre-operative PVR and fair correlation with pre-operative CO and mPAP). Additionally, fibrotic clot score agrees with the quantified perfusion deficit, HLV, with a moderate strength spearman correlation coefficient. Fibrotic clot score also correlates with post-operative improvement in pulmonary hemodynamics, with a fair correlation with both post-operative change in PVR and post-operative change in CO.

Our fibrotic clot scores agree with invasive pulmonary hemodynamics better than fibrotic clot scores reported previously. Heinrich et al calculated a Qanadli based fibrotic clot score in 60 CTEPH patients who underwent PTE and had preoperative CT pulmonary angiograms[55]. They found no correlation between their fibrotic clot score and pre-operative pulmonary hemodynamics, including PVR (ρ=0.03, p-0.85) and mPAP (ρ=0.21, p=0.11). In contrast, they did see a fair correlation between their fibrotic clot score and post-operative PVR (ρ=-0.36, p=0.007), similar to our findings. The difference in observed fibrotic clot score correlation with preoperative hemodynamics is likely due to methodology differences and timing of the studies. The fibrotic clot score calculated by Heinrich et al still calculated the proportion of the segmental vasculature affected, but they did not score the segmental or subsegmental vessels directly. Instead, they calculated clot score as only the "central" obstructions (proximal to the segmental arteries) that would affect a given segmental territory. This is likely because Heinrich et al performed their 2005 study on patients scanned between 1996 and 2002, representing a different CTEPH cohort, with segmental clots that were thought to be surgically inaccessible. Their metric of peripheral scar and perfusion score, which are two peripheral metrics including a mix of perfusion and fibrotic clot measures, also supports the idea that our differences in hemodynamic correlations are likely due differences in direct segmental evaluation. Both of these metrics correlated with preoperative mPAP ($\rho_{\text{peripheral scar}}$=0.32, $p_{\text{peripheral scar}}$=0.017; $\rho_{\text{perfusion}}$=0.66, $p_{\text{perfusion}}$<0.001) and PVR ($\rho_{\text{peripheral scar}}$=0.44, $p_{\text{peripheral scar}}$<0.001; $\rho_{\text{perfusion}}$=0.76, $p_{\text{perfusion}}$<0.001). Post-operative PVR also fairly correlated with the peripheral scar score and perfusion score ($\rho_{\text{peripheral scar}}$=0.26, $p_{\text{peripheral scar}}$=0.047; $\rho_{\text{perfusion}}$=0.37, $p_{\text{perfusion}}$=0.007).

Our cohort has a stronger fibrotic clot score correlation with pulmonary hemodynamics than another more recent CTEPH with fibrotic clot scores analyzed by Abozeed et al[56]. Our cohorts

has similar agreement between fibrotic clot score and pre-operative mPAP, with both cohorts showing fair correlations (ρ=0.28, p=0.018 for Abozeed et al). Our cohort had a stronger, moderate correlation with pre-operative PVR, compared to only a fair correlation in the Abozeed et al study (ρ=0.27, p=0.026).Post-operative hemodynamics are not comparable between the two groups as the patients in Abozeed et al did not undergo PTE surgery. These differences in hemodynamic correlations are likely due to two main methodological differences. First, as mentioned above, the patients in the Abozeed et al cohort did not under go treatment by PTE. This may reflect that the patients were not ideal PTE candidates. Additionally, the fibrotic clot scoring method used by Abozeed et al did not distinguish partial vs complete obstructions, nor did it represent the portion of the segmental vessels obstructed (instead representing total cumulative clot in the lung, giving rise to higher clot scores).

Our cohort had moderate agreement between fibrotic clot score and HLV, our perfusion deficit metric discussed in chapter 1. As most of our patients had good PTE outcome (post-operative PVR < 450 $dyn \cdot s \cdot cm^{-5}$), our cohort represents a cohort of CTEPH patients that are good PTE candidates. This quantitatively supports the PTE selection criteria that patients who are good PTE candidates should have agreement with their fibrotic clot and their perfusion deficit[6]. Additionally, our fibrotic clot score and perfusion deficit metric (HLV) both agree with pre-operative PVR and post-operative change in PVR. This further quantitatively supports the PTE selection criteria that good PTE candidates will have agreement between their CTEPH disease burden (fibrotic clot and perfusion deficit) and their pulmonary hypertension severity (PVR). Given that fibrotic clot score agrees with but is not fully explained by HLV (moderate not strong correlation), future work should focus on investigating how fibrotic clot score and HLV can be combined to improve pre PTE assessment. Specifically, more detailed analysis of

the regional instead of global agreement between fibrotic clot score and HLV should be investigated. This could facilitate the creation of a combined fibrotic clot and perfusion deficit disease burden score that may further aid prediction of pulmonary hemodynamic severity. Despite our studies strengths, there were multiple limitations in our study. First, the patients with CTEPH represented in our cohort are a small number of patients relative to the annual cohort that undergoes PTE at UCSD. While CT pulmonary angiograms occur for most patients currently as part of routine CTEPH clinical evaluation, dual-energy CT studies (enabling HLV calculation and comparison) are not part of the current standard clinical practice. Additionally, while our cohort does contain patients with segmental disease, there are no patients with disease level 4 represented in our cohort. Further, our fibrotic clot score was comprised of annotations provided by clinical readers, who were not aware of our research question given the retrospective nature of the study. While this may lead to more fibrotic clot score variability, it can also be seen as an advantage, as our fibrotic clot scores represent annotations obtained from a real clinical work flow. Finally, bronchiole or coronary artery collateral flow to the pulmonary parenchyma was not considered in our study. Bronchiole artery collaterals can directly affect pulmonary blood flow by supplying blood to parenchyma downstream of a clot. In addition to having diagnostic utility, they have also are associated with lower post-operative PVR and lower mortality rate after PEA.[57–60]. Collateral flow circumventing fibrotic clot may be an important consideration for future work.

2.6. Conclusion

In conclusion, fibrotic clot score from pre-operative CT pulmonary angiograms can quantitatively capture CTEPH disease burden severity. In patients who have undergone PTE,

fibrotic clot score moderately agrees with pre-operative PVR and CO, and fairly agrees with post-operative change in CO and PVR. Fibrotic clot score also moderately agrees with our perfusion deficit metric of disease burden, HLV. These findings suggest that fibrotic clot score evaluation could provide quantitative, imaging derived metrics to evaluate CTEPH disease severity or PTE response. Further investigation into the utility of fibrotic clot score independently and in combination with HLV for assessment of disease severity and PTE planning is warranted.

2.7. Acknowledgements

Chapter 2 contains work that is not currently in preparation for submission as a manuscript. The dissertation author was the primary investigator and author of this work. This work was supported with contributions from undergraduate researcher Sicong Shen, who helped calculate fibrotic clot scores. This work would not have been possible without the contributions and insights of Nick Kim, Michael Madani, Atul Malhotra, Lewis Hahn, Seth Kligerman, Albert Hsiao, and Francisco Contijoch.

CHAPTER 3: PREDICTION OF DISEASE LOCATION IN CTEPH

3.1. Abstract

The purpose of this study was to assess if UCSD surgical disease level could be predicted preoperatively from Ventilation-Perfusion (VQ) Imaging in patients with CTEPH. A training set of 625 pre-operative pulmonary ventilation-perfusion images in patients with CTEPH who underwent PTE (median age 55 with interquartile range 31-79 years, 289 females) was used to train three different transfer-learning initialized convolutional neural networks (MobileNetV2, VGG-16, and Xception) with linear regression output layer to predict left lung UCSD surgical disease level. Network disease level prediction performance was evaluated using the validation cohort (154 cases, median age 54.5 years with interquartile range 52-78 years, 61 females). Classification accuracy of all VQ cases (proximal vs disease level 3 vs disease level 4) was not better than naïve prediction (selection of most common class, 55%) accuracy for any model, but confidence based prediction on only a subset of cases performed better than naïve accuracy for all models ($p<0.001$). Confident cases classification had 71% accuracy with prediction in 14% of cases for VGG-16 and 69% accuracy with prediction in 27% of cases for Xception, and had moderate agreement with known surgical disease level ($ICC_{VGG-16}=0.70$ $p_{VGG-16}<0.001$; $ICC_{XCEPTION}=0.64$ $p_{XCEPTION}<0.001$). Likely thresholds increased the percentage of cases that were classified (57% for VGG-16 and 67% for Xception), while still performing better than naïve accuracy (67% for VGG-16 and 67% for Xception), and maintaining moderate agreement with known disease level ($ICC_{VGG-16}=0.55$ $p_{VGG-16}<0.001$; $ICC_{XCEPTION}=0.55$ $p_{XCEPTION}<0.001$). In conclusion, left lung UCSD surgical disease level can be predicted from VQ images. VQ based machine learning predictions of disease level could aid clinician presurgical assessment of

CTEPH, and serve as the foundation for future multi-modality models to pre-operatively predict CTEPH disease level.

3.2. Introduction

Pulmonary Thromboendarterectomy (PTE) is the preferred treatment for patients with CTEPH, because of its significant improvement in patient mortality and resolution or drastic mitigation of pulmonary hypertension[8,61]. One of the factors dictating successful PTE outcome is the location and depth in the vascular tree of a patient's chronic vascular obstructions, which influences a surgeons ability to reach and completely remove them[6,15]. Different patients have a variety of chronic obstruction locations and origins, and not all patients will have surgically accessible disease. Therefore, preoperative assessment of patient chronic clot location is critical.

A standardized metric of clot location is the UCSD surgical disease level, which specifically described the location of chronic clot origin within the vascular tree of a given lung. Surgical disease level ranges from 1 to 4, with the origin of clot being the main pulmonary arteries (disease level 1), the lobar pulmonary arteries (disease level 2), the segmental pulmonary arteries (disease level 3), or the subsegmental arteries (disease level 4)[6]. Surgical disease level is definitively determined for the right and left lung (independent value for each lung) at the time of PTE surgery, based on the chronic obstruction location the surgeon encounters. Unfortunately, disease level determination during surgery limits its utility preoperative decision making and planning. Current preoperative estimation of clot location and disease level is accomplished by a multidisciplinary team (pulmonologists, radiologists, surgeons, and specialized practitioners) through integration of multiple data sources (imaging and clinical) for a given patient[8,19]. Unfortunately, though, there is no quantitative tool to assist with clinician prediction of disease level.

Prior studies have shown relationships between different preoperative findings, and patient disease location[18,48,62,63]. For relationships identified in the clinical data, several general modeling strategies exist to integrate this data into a single disease level prediction, although none of these strategies have been published at this time. However, imaging data provides more of a challenge, despite the high resolution blood flow and perfusion information (visual features) it provides. Specifically, it is challenging to convert visual features into numeric ones that can predict disease location on their own or be readily integrated with clinical data for multimodality predictions. One available strategy is the conversion of detailed, expert-determined annotations into numeric values representing important disease location features and predicting disease level from these values[64]. This strategy is not ideal, it increases the workload of the clinical team and depends on the availability of subspecialized readers. We would like to streamline the process of clinical image feature detection and disease level prediction, without increasing clinician workload.

Machine learning offers an approach to identify and predict on imaging features indicative of fibrotic clot location without requiring expert image annotation. Machine learning strategies have previously been used on problems analogous to disease location prediction, with the successful identification of clot presence or absence in patients with suspected acute Pulmonary Embolism[65–68]. One imaging modality used in the prediction of acute PE, planar Ventilation-Perfusion (VQ) studies, is obtained in almost all patients with suspected CTEPH as part diagnostic workup and surgical planning[12,20]. Thus, there is a large number of pre PTE VQ studies in patients with CTEPH. The high availability of studies and utility in analogous machine learning predictions suggests that machine learning based prediction of CTEPH disease location from VQ studies could be feasible.

The objective of the work contained in this chapter was to further pre PTE assessment of disease location through development of systematic and quantitative approach to predict disease level from pre-operative imaging in CTEPH patients. Our hypothesis was that a machine learning model can predict surgical disease level from pre PTE VQ scans, and that these predictions would agree with the surgically confirmed disease level.

3.3. Methods

3.3.1. Cohort Selection

Under IRB approved (#191797) waiver of informed consent, 797 consecutive CTEPH cases treated by Pulmonary Thromboendarterectomy (PTE) performed at our institution between April 2013 and June 2018 were retrospectively evaluated (by author F.C.). Cases were excluded if they were missing disease level information (n=6) or if their Ventilation perfusion Scans had a different arrangement than those obtained at our institution (n=12). This resulted in 779 cases for model training and evaluation. Patient CTEPH diagnosis and PTE eligibility was determined by consensus of a multidisciplinary team which included cardiothoracic surgeons, vascular disease expert pulmonologists, and fellowship trained cardiothoracic radiologists. CTEPH diagnosis was based on the criteria from Galiè et al[3]. These criteria require pulmonary hypertension evaluation using right heart catheterization, with mean pulmonary artery pressure (mPAP) >25mmHg, pulmonary artery wedge pressure ≤15 mmHg, pulmonary vascular resistance ≥240 dyn s cm-5. Additionally, diagnostic criteria require CTEPH specific imaging signs, including ventilation-perfusion mismatch seen on Ventilation-Perfusion SPECT Imaging.

Our cohort (**Figure 16**) was comprised of patients who underwent PTE surgery for definitive treatment of CTEPH and found to have bilateral disease during PTE, and had pre PTE VQ scans obtained within 90 days of PTE (n=686 VQ scans). These patients were randomly split

into a training cohort (n=532 cases), used to determine the model parameters that best predicted disease level, and a validation cohort (n=154 cases) used to monitor training progress and evaluate model performance. While not reflective of the overall CTEPH cohort, cases that had VQ scans obtained greater than 90 days prior to PTE (n=74) and cases with only unilateral disease level (n=19) were included in the training cohort to increase number of training cases. This led to an 80/20 split between training and validation cohorts (n=625 training cases and n=154 validation cases).

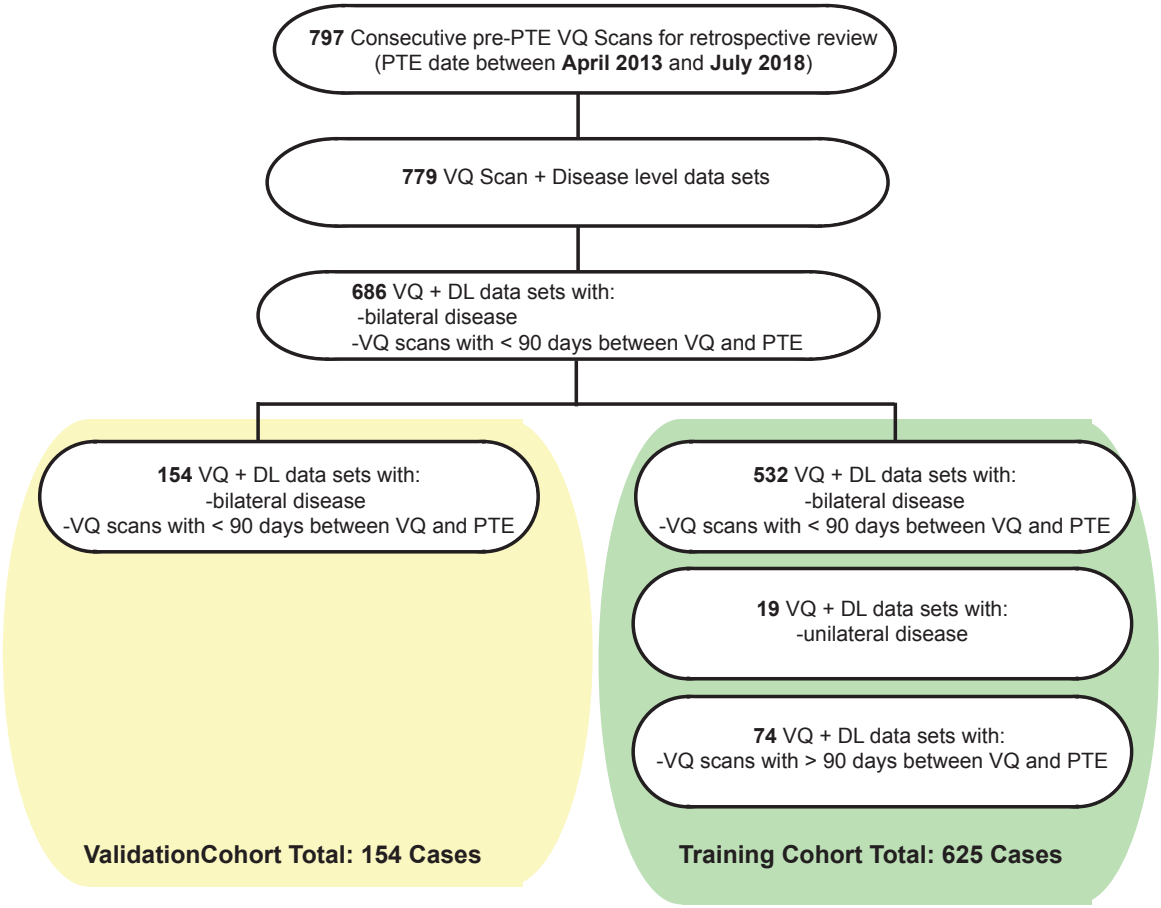


Figure 16. VQ cases in validation and training cohort

3.3.2. Image Acquisition and preprocessing

Ventilation Perfusion SPECT images were acquired during workup for suspected CTEPH to rule out other forms of PH [3,12]. Planar VQ images were acquired using Xe-133 (ventilation) and Tc-99 macroaggregated albumin (perfusion) on a gamma camera. Ventilation Images were acquired with patients upright after inhalation of Xe-133 gas from a non-rebreathing mouthpiece for the wash-in, equilibrium and 2 washout phases of gas inhalation. After ventilation image acquisition, planar perfusion images of 8 standard views (Anterior, Posterior, Left Posterior Oblique, Right Posterior Oblique, Left Lateral, Right Lateral, Left Antior Oblique and Right Anterior Oblique) were acquired after intravenous administration of Technetium-99 microaggregated albumin, also with patients in the upright position. All images within an individual Ventilation-Perfusion Scan study were compiled into a single 1584 x 1024 pixel image containing 4 ventilation images, 8 perfusion views, and header information. These images were deidentified using the Virtual Research Desktop provided through the Altman Clinical Translational Research Center at UC San Diego into 1584 x 768 images containing only the 4 ventilation images and 8 perfusion images without personal health information. Images were preprocessed before neural network training to create a 244 pixel x 244 pixel x 3 channel input image with the anterior perfusion view in channel 1, posterior perfusion view in channel two and wash-in ventilation image in channel 3 (**Figure 17**, Step 1. VQ image preprocessing). Posterior perfusion images and ventilation images were flipped across the vertical access, leading to all images having the left lung on the same side of the image. Images were normalized by dividing by 255 (the maximum pixel value), leading to all pixel values falling in the 0 – 1 range.

3.3.3. Surgical Disease Level Determination

Patients diagnosed with CTEPH and selected as good PTE candidates underwent PTE as is described in detail in both Madani et al and Gernhofer and Pretorius [6,8]. During PTE surgery, the surgeon assigns the surgical disease level for each lung based on the most proximal location or most proximal origin of fibrotic clot. The location of most proximal fibrotic clot or fibrotic clot origin corresponding to each disease level are as follows: disease level 1 has disease originating in the left or right main pulmonary, disease level 2 has disease originating in a lobar or the interlobar pulmonary artery, disease level 3 had disease originating in a segmental pulmonary artery and disease level 4 has disease originating in a subsegmental pulmonary artery [6]. Disease level is determined independently for each lung, and lungs can have different disease levels from one another. Likewise, it is also possible for one lung to have disease when the other lung does not. The majority of patients who undergo bilateral PTE have surgical disease level assessed for both lungs, but patients who have undergone minimally invasive PTE (included in the training but not validation cohort) have the surgical disease level for only one lung assessed. The surgically defined disease level is used as the ground truth fibrotic clot location. For compatibility with CNN training, known disease levels were arranged in the following ordinal categories: Disease level 1, Disease Level 2, Disease Level 3, Disease Level 4 and No Disease. These ordinal categories were mapped to a value between 0 and 1 prior to being used for CNN training (Disease level 1=0, Disease Level 2=0.25, Disease Level 3=0.5, Disease Level 4=0.75 and No Disease=1).

3.3.4. Model Training

Three different CNN architectures, VGG-16, Xception and MobileNetV2, were trained to predict the disease level of the left lung using transfer learning. Networks were all initialized

with ImageNet weights and had their classification/fully connected layers removed and replaced by a single neuron regression output layer with linear activation function (**Figure 17**, 2. CNN analysis of VG Image). No layers were frozen, giving weights in all layers of the network the opportunity to update with training. Input images of 224x224x3 containing the Anterior Perfusion Image, Posterior Perfusion Image and Ventilation perfusion image were used to train the network. Image augmentation using a random combination of rotation (≤ 20 degrees), shifting (≤ 20 pixels), and cropping (≤ 20 pixels) was used to supplement the 625 training cases. The training data sets used were the same for all 3 models, with the only differences being differences in random data augmentation. Training was accomplished using a weighted mean absolute error loss function, ADAM optimizer, and a learning rate schedule initialized at 8 x 10^{-4} that reduced by 0.8 times the prior learning rate after 10 epochs with no improvement in loss. Networks training was stopped after networks trained for 200 epochs or after 50 epochs without validation loss improvement, whichever happened first. Models were trained on a GPU using Keras from Tensorflow (version 2.2.0).

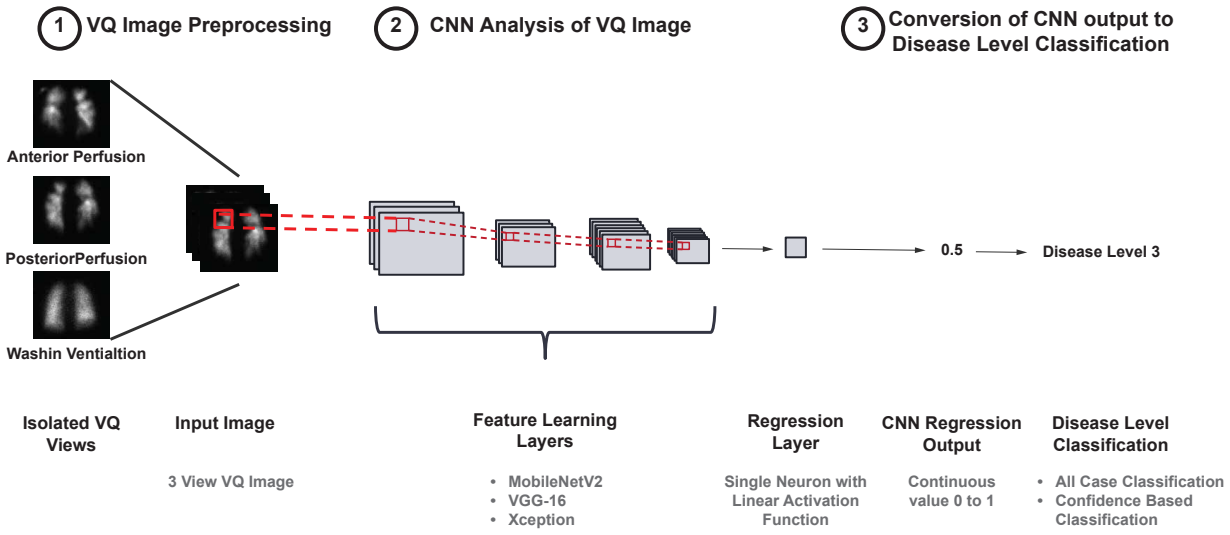


Figure 17 Analysis pipeline to predict disease level from VQ images

3.3.5. Model Evaluation

CNN performance was evaluated using the validation data set (same individual cases for all 3 networks) that was also used to monitor network training. First, correlation between raw CNN regression output (continuous value between 0.0 and 1.0; prior to conversion to disease level classification of 1,2,3, or 4) and known surgical disease level was assessed. Next, CNN regression outputs differences between group known disease level groups were assessed. Group comparisons included cases with surgically defined proximal (disease level 1 or 2) vs segmental or subsegmental (disease levels 3 or 4; referred to as (sub)segmental disease for the remainder of the chapter) disease, as well as cases with adjacent disease levels (ex: disease level 2 vs 3). Initial ability of the CNN output to separate a) proximal vs (sub)segmental cases and b) disease level 3 vs disease level 4 cases was evaluated using the Receiver Operating Characteristic Curve (ROC). ROC curves were used to select different thresholds to classify CNN regression outputs into a disease level prediction. CNN outputs were classified into proximal, disease level 3 or disease level 4. Disease levels 1 and 2 were not separated out, as the main increase in surgical case difficulty occurs with disease level 3 and disease level 4 [16].Two different classification approaches were investigated: an all case classification approach and a confidence based classification approach. The agreement between these classification approaches and known disease level was evaluated and accuracy of the classifications were compared to the accuracy of a naive prediction. The naïve prediction was the classification of all cases as the most common disease level.

3..3.6. All Case Classification

The first evaluated classification method classified all validation cases. Thresholds for classification of CNN outputs into disease level were selected from the highest accuracy

operating point of the ROC curves. First, cases were classified into proximal or (sub)segmental using the highest accuracy ROC operating point for the separation of proximal vs (sub)segmental cases. If CNN regression output was able to successfully separate disease level 3 from disease level 4 in isolated (sub)segmental cases (based on ROC analysis), then the highest accuracy operating point was used to define the CNN output threshold for disease level 3 vs 4. In the case that multiple operating points had the same accuracy, the point with the higher positive predictive value was used (**Figure 18**).

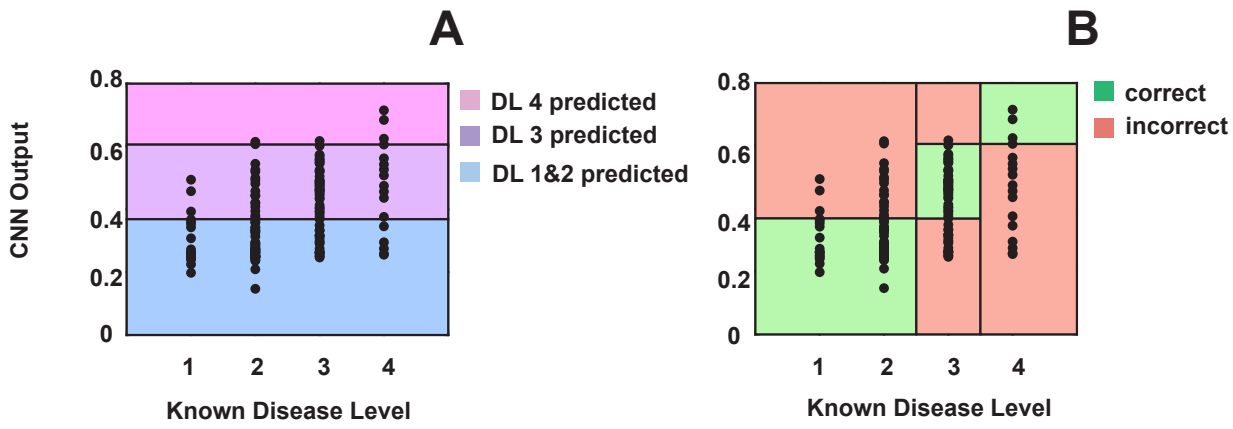


Figure 18 All case classification example

3.3.7. Confidence Based Classification

Under the second evaluated classification method, confidence-based classification, cases were classified only if their CNN regression output values were within a range of values that had a high predictive value for a given disease level (Proximal, (sub)segmental/disease level 3, and disease level 4). Predicted values that did not meet the threshold for classifying patients as likely proximal or likely distal were not classified. Two different confidence levels were used with this approach, “confident” and “likely”. The range of values was defined based on the predictive values of different points along the ROC curve. Confident thresholds were defined as the highest

accuracy ROC operating point with predictive value close to 90% (range 87.5-92.5%). If there was no predictive value between 87.5-92.5%, then the range was expanded to 85%-95%, and finally 85%-100%. Likely thresholds were defined as the highest accuracy ROC operating point with predictive value close to 80% (range 77.5-82.5%). If there was no predictive value between 77.5-82.5%, then the range was expanded to 75%-85%, and finally 70%-85%. The thresholds defining the CNN output values as "confident" or "likely" were chosen based on the positive and negative predictive values of different operating points along the ROC curve (**Figure 19**).

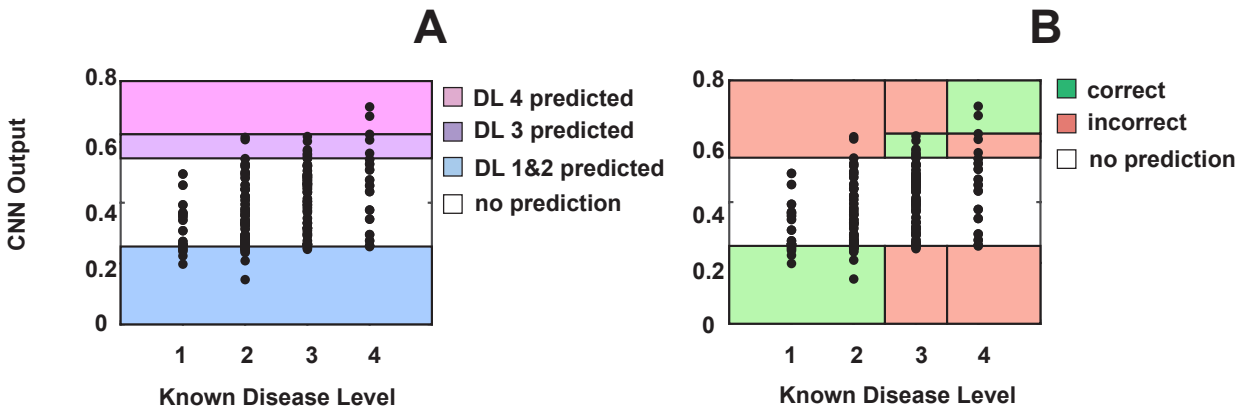


Figure 19 Confidence based classification example

3..3.8. Statistics

Demographic measures and disease level distribution in the left lung were compared between the training and validation cohort. Parameters were tested for normality using the Shapiro-Wilks test. Normally distributed variables are reported as mean ± standard deviation and non-normally distributed variables are reported as median [interquartile range from q1 to q3]. Categorical variables are reported as percentage of the group and number within the group. Student's t-test and one-way ANOVA were used to assess normally distributed variables, while Wilcoxon rank sum and Kruskal-Wallis tests were used for non-normally distributed variables with p=0.05. Agreement between CNN output predicted disease level value and known surgical

disease level was determined via Spearman's correlation coefficient. Correlations were classified as little to no relationship ($0 \leq \rho < 0.25$); fair ($0.25 \leq \rho < 0.5$); moderate ($0.5 \leq \rho < 0.75$); or excellent ($\rho \geq 0.75$)[43]. Comparison of CNN output predicted disease level value in different known disease level groups was accomplished with one-way ANOVA followed by student's t-test for individual group differences (including of known proximal vs know (sub)segmental studies) with p=0.125 (Bonferonni correction for 4 comparisons). ROC curves identified the cutoffs used to classify predictions as proximal vs distal, including a single threshold to classify all patients and 4 thresholds to identify confident proximal, likely proximal, confident (sub)segmental and likely (sub)segmental classifications. The area under each receiver operating curve and its 95% confidence interval were calculated. Accuracy, percent classified, sensitivity and specificity for each classification strategy/mode was compared within a given model type using the Z statistic. Analysis was performed in MATLAB 2021b (MathWorks, Natick, MA).

3.4. Results

3.4.1. CTEPH cohort demographics

The training cohort cases were 55 [31-79] years old, 46% female (n=289). Pre PTE hemodynamics in the training cohort were 44 [29-59] mmHg for mPAP, 569 [98.5-1050] $dyn \cdot s \cdot cm^{-5}$ for PVR, and 4.6 [3.8-5.4] L/min. The training cohort (n=625) and validation cohort (n=154) were similar in age, percent female, and preoperative hemodynamics (**Table 8**). The validation cohort cases were 54.5 [52-78] years old, 40% female (n=61). Pre PTE hemodynamics in the training cohort were 43$\pm$12 mmHg for mPAP, 550 [137-963] $dyn \cdot s \cdot cm^{-5}$ for PVR, and 4.3 [2.5-6.1] L/min. The training cohort and validation cohort were similar in age, percent female, and preoperative hemodynamics (**Table 8**).

Aside from the exclusion of No Disease cases from the validation cohort, the disease level distributions were similar between the validation and training cohorts. In the training cohort, the disease level distribution was 11% (n=74) disease level 1, 39% (n=233) disease level 2, 34% (n=215) disease level 3, 13% (n=84) disease level 4, and 3% (n=19) No Disease. In the validation cohort, the disease level distribution was 15% (n=18) disease level 1, 40% (n=61) disease level 2, 34% (n=58) disease level 3, and 10% (n=17) disease level 4. Because the largest percentage of validation cases fell in disease level 1 and 2, the naïve prediction accuracy was 55% for proximal vs (sub)segmental classification and also proximal vs disease level 3 vs disease level 4 classification.

Table 8. Training and validation cohort characteristics

	Training	Val	p-value
Age	55 [31-79]	54.5 [52-78]	0.63
Sex	46% (n=289)	40% (n=61)	0.14
mPAP	44 [29-59]	43±12	0.53
PVR	569 [98.5-1040]	550 [137-963]	0.59
CO	4.6 [3.8 - 5.4]	4.3 [2.5-6.1]	0.09
Disease Level 1	11% (n=74)	15% (n=18)	0.81
Disease Level 2	39% (n=233)	40% (n=61)	
Disease Level 3	34% (n=215)	34% (n=58)	
Disease Level 4	13% (n=84)	10% (n=17)	
No Disease	3% (n=19)	-	

3.4.2. Agreement of CNN Linear Regression Output with Known Disease Level

CNN outputs of all three models had a fair correlation with the true disease level values in the validation cohort (MobileNetV2 ρ=0.37, $p < 0.001$; VGG-16 ρ=0.36, $p < 0.001$; Xception ρ=0.45, $p < 0.001$). Boxplots of CNN output for each known disease level are shown in **Figure 20**. Using MobileNetV2 (**Figure 20A**), the median and interquartile range for CNN output was 0.36 [0.24 -0.47], 0.36 [0.23-0.50], 0.43 [0.28-0.59], and 0.50 [0.38-0.62] for known surgical disease levels 1-4, respectively. There was a significant difference in the CNN output for

proximal vs (sub)segmental disease (p <0.001) as well as in known surgical disease levels 2 vs 3 (p=0.003). There was not a significant difference in CNN disease level outputs in known surgical disease levels 3 vs 4 (p=0.054). Like MobileNetV2, VGG-16 (**Figure 20B**) disease level outputs were also significantly different between proximal vs (sub)segmental disease (p< 0.001) and surgically known disease level 2 vs 3 (p=0.004), with median outputs of 0.31 [0.22-0.40], 0.31 [0.16-0.46], 0.39 [0.21-0.57] , and 0.53 [0.21-0.85] for known surgical disease levels 1-4, respectively. The difference in output between known surgical disease levels 3 and 4 (p=0.0138) was not different after Bonferoni correction (requiring p< 0.0125 for significance). For Xception (**Figure 20C**), the median and interquartile range for CNN output was 0.27 [0.15 -0.38], 0.28 [0.16-0.42], 0.44 [0.26-0.62], and 0.47 [0.23-0.72] for known surgical disease levels 1-4, respectively. Like MobileNetV2 and VGG-16, there was a significant difference in the CNN predicted disease level outputs in proximal vs (sub)segmental disease (p <0.001) as well as in known surgical disease levels 2 vs 3 (p <0.001). There was not a significant difference in CNN disease level outputs in known surgical disease levels 3 vs 4 (p=0.35). There was not a significant difference in CNN disease level outputs in known surgical disease level 1 vs 2 for any of the CNN Models (MobileNetV2 p=0.43; VGG-16 p= 0.61; Xception p=0.35).

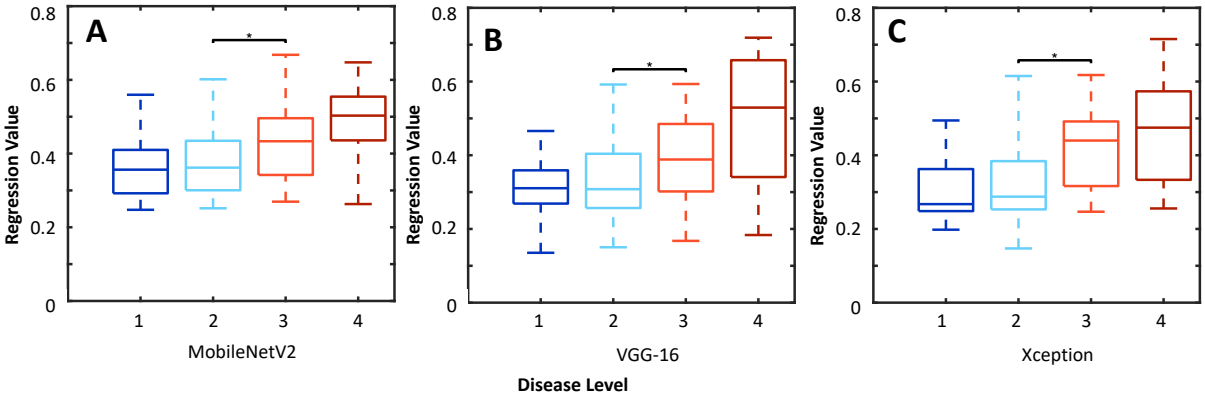


Figure 20 CNN regression output of three different models for each known disease level

3.4.3. Receiver Operating Characteristic Curves

The receiver operating characteristic curves for separation of proximal and (sub)segmental disease based on CNN output is shown in **Figure 21**. All three models separate proximal and (sub)segmental disease better than chance (AUC >0.5). AUC and 95% Confidence intervals were 0.70 (0.66-0.74), 0.69 (0.65-0.73), and 0.76 (0.72–0.79) for MobileNetV2, VGG-16, and Xception, respectively.

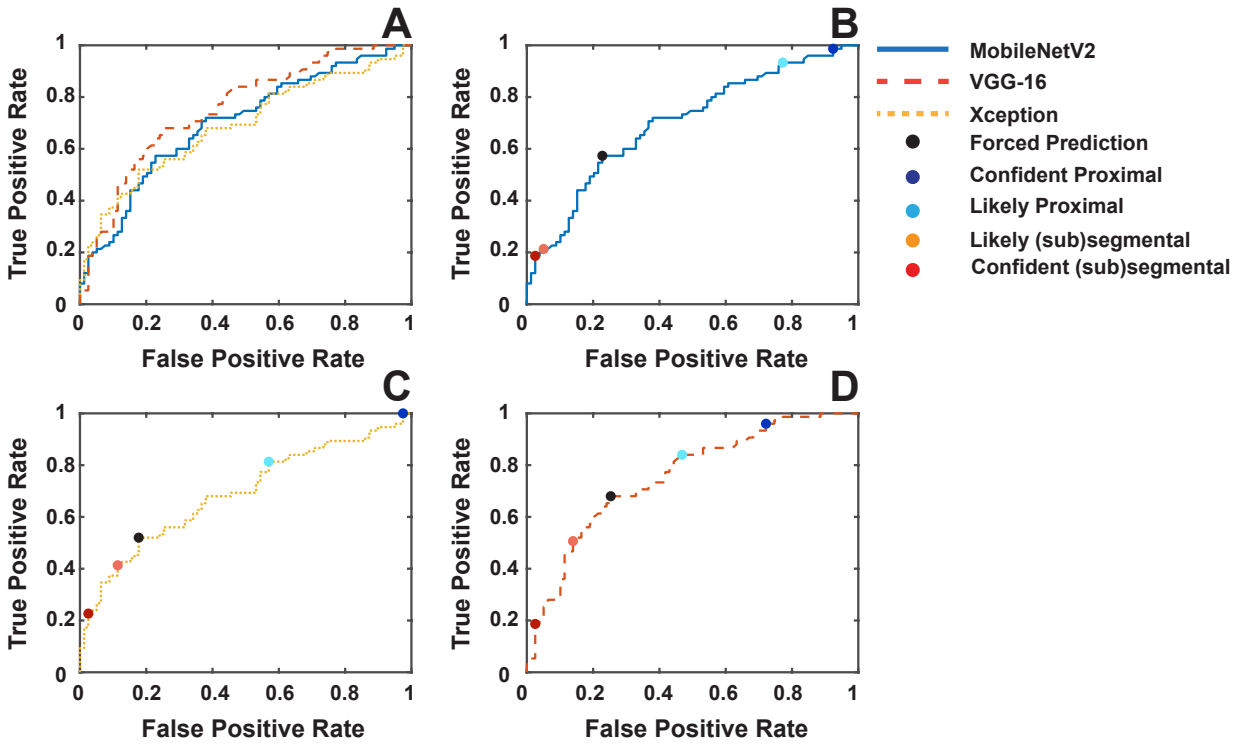


Figure 21 ROC analysis showing separation of proximal vs (sub)segmental cases in all three CNN models

The receiver operating characteristic curves for the separation of disease level 3 vs 4 in known (sub)segmental disease based on CNN output is shown in **Figure 22.** VGG-16 (yellow line) and Xception (red line) CNN output separated known (sub)segmental cases into disease level 3 vs 4, with AUC and 95% CI of 0.62 (0.68 – 0.56), and 0.60 (0.65 – 0.56) for VGG-16 and Xception models, respectively. MobileNetV2 (blue line) was not able to separate known

(sub)segmental patients into disease levels 3 and 4 (AUC and 95% Confidence Interval of 0.56 , 0.47 – 0.65).

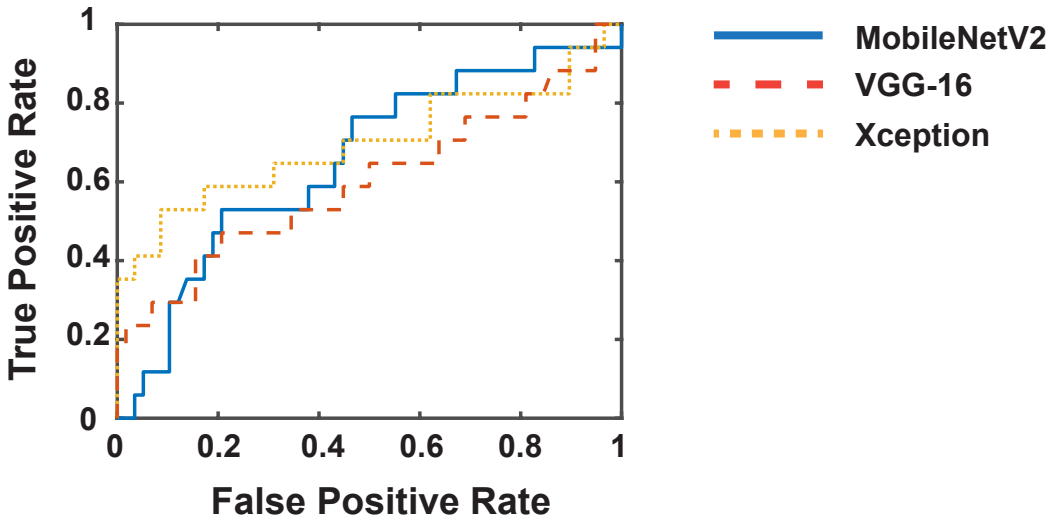


Figure 22 ROC curve of CNN output shows VGG-16 and Xception but not MobileNetV2 separate disease level 3 vs 4

3.4.4. All Case Classification

MobileNetV2: The MobileNetV2 all case classification CNN output thresholds (**Table 9**) were>0.369 for (sub)segmental classification (**Figure 21B**). Because MobileNetV2 could not separate disease levels 3 and 4 in (sub)segmental cases, there is no identified threshold for classifying disease level 3 from 4. Classification of all validation samples as proximal or (sub)segmental had an accuracy of 68%, sensitivity of 68%, and specificity of 47% (**Table 10**). Accuracy for proximal vs (sub)segmental accuracy was better than naïve prediction accuracy ($p<0.02$).

VGG-16: The VGG-16 all case classification CNN output thresholds (**Table 9**) were>0.414 for proximal vs (sub)segmental classification (accuracy 68%, sensitivity 52%, specificity of 82%; **Figure 21C, Table 10**) and>0.616 for disease level 4 classification (disease level 3 vs 4 accuracy 86%, sensitivity 35%, specificity of 98%). The operating point accuracy for all case classification into proximal or (sub)segmental performed better than naïve classification (p=0.02). Classification of CNN output for all cases into predicted disease level (proximal, 3 or 4) had low agreement with ground truth surgical disease level (ICC= 0.44 p<0.001). All cases classification accuracy was 64% for VGG-16 **(Table 11)**, which was not statistically different than the naïve accuracy of 55% (p=0.12).

Xception: The Xception all case classification CNN output thresholds (**Table 9**) were>0.369 for proximal vs (sub)segmental classification (proximal vs (sub)segmental accuracy 71%, sensitivity 68%, specificity of 75%; **Figure 21D, Table 10**) and>0.606 for disease level 4 classification (accuracy 81%, sensitivity 24%, specificity of 100%) for disease level 4 classification. The operating point accuracy for all case classification into proximal or (sub)segmental performed better than naïve classification (p=0.004). Classification of CNN output for all cases into predicted disease level (proximal, 3 or 4) had low agreement with ground truth surgical disease level (ICC=0.37 p <0.001). All cases classification accuracy was 64% for Xception (**Table 11**), which was not statistically different than the naïve accuracy of 55% (p=0.47).

Table 9 Thresholds for disease level classification from CNN output

		(Sub)segmental Threshold	NPV	Proximal Threshold	PPV	Disease Level 4 Threshold	PPV
MobileNetV2	**All Cases**	0.433	-	0.433	-	-	-
	Confident Cases	0.552	0.88	0.270	0.86	-	-
	Likely Cases	0.536	0.80	0.295	0.78	-	-
VGG-16	**All Cases**	0.414	-	0.414	-	0.616	
	Confident Cases	0.503	0.89	0.168	1.00	0.593	0.86
	Likely Cases	0.452	0.78	0.296	0.71	0.587	0.78
Xception	**All Cases**	0.369	-	0.369	-	0.606	
	Confident Cases	0.546	0.88	0.256	0.88	0.625	1.00
	Likely Cases	0.450	0.78	0.291	0.78	0.606	0.80

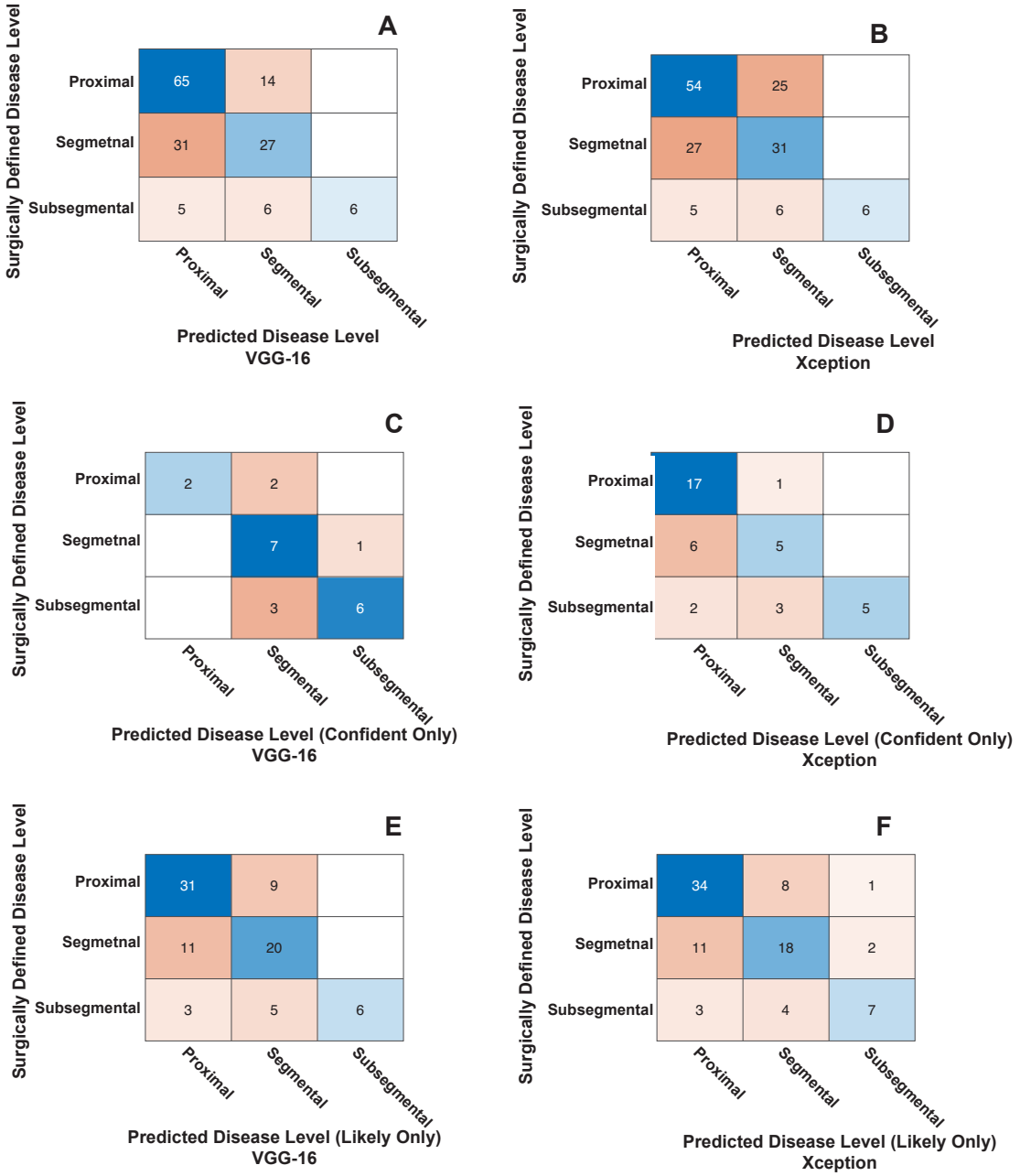


Figure 23 Classification of proximal disease level, disease level 3, and disease level 4 for all case and confidence-based classification strategies

3.4.5. Confidence Based Classification

MobileNetV2: MobileNetV2's confident thresholds were<0.267 for proximal classification and>0.552 for (sub)segmental classification (**Table 9**). Because MobileNetV2 could not separate disease levels 3 and 4 in (sub)segmental cases, there is no identified threshold for classifying disease level 3 from 4. Confident thresholds classified 15% of cases as proximal or (sub)segmental with 87% accuracy, 88% sensitivity and 86% specificity. Confident threshold classification of proximal vs (sub)segmental cases had higher sensitivity (p=0.006), but similar accuracy (p=0.058) and specificity (p=0.88) compared to MobileNetV2 all case classification (**Table 10**).

MobileNetV2's likely thresholds were <0.295 for proximal classification and>0.536 for (sub)segmental classification (**Table 9**). Because MobileNetV2 could not separate disease levels 3 and 4 in (sub)segmental cases, there is no identified threshold for classifying disease level 3 from 4. Likely thresholds classified 28% of cases as proximal or (sub)segmental with 79% accuracy, 80% sensitivity and 78% specificity. Likely threshold classification of proximal vs (sub)segmental cases had higher sensitivity (p=0.007), but similar accuracy (p=0.14) and specificity (p=0.36) compared to MobileNetV2 all case classification (**Table 10**).

VGG-16: VGG-16's confident thresholds were<0.168 for proximal classification,>0.503 for (sub)segmental classification, and>0.593 for disease level 4 classification (**Table 9**). Proximal vs (sub)segmental classification thresholds (**Table 10**) had higher accuracy (p=0.03), sensitivity (p=0.001), and specificity (p=0.04) than the all case proximal vs (sub)segmental threshold. Confident thresholds classified 14% of cases as proximal, disease level 3 or disease

level 4 with 71% accuracy, which was significantly better than the naïve prediction accuracy of 55% ($p<0.001$;**Table 11**). Confident only classification moderately agreed with known surgical disease level (ICC=0.70 $p<0.001$).

VGG-16's likely thresholds were<0.296 for proximal classification and>0.452 for (sub)segmental classification, and>0.587 for disease level 4 classification (**Table 9**). Proximal vs (sub)segmental classification thresholds (**Table 10**) had higher sensitivity ($p<0.001$) and specificity ($p=0.04$), but not accuracy ($p=0.30$) compared to the all case proximal vs (sub)segmental threshold. Likely thresholds classified 57% of cases as proximal, disease level 3 or disease level 4 with 67% accuracy, which was significantly better than the naïve prediction accuracy of 55% ($p<0.001$; **Table 11**). Likely case classification moderately agreed with known surgical disease level (ICC=0.55 $p<0.001$).

Xception: Xception's confident thresholds were<0.256 for proximal classification,>0.546 for (sub)segmental classification, and>0.625 for disease level 4 classification (**Table 9**). Proximal vs (sub)segmental classification thresholds (**Table 10**) had higher accuracy ($p=0.03$) and sensitivity ($p=0.01$), but not specificity ($p=0.07$) compared to the all case proximal vs (sub)segmental threshold. Confident thresholds classified 27% of cases as proximal, disease level 3 or disease level 4 with 69% accuracy, which was significantly better than the naïve prediction accuracy of 55% ($p<0.001$; **Table 11**). Confident only classification moderately agreed with known surgical disease level (ICC=0.64 $p<0.001$).

Xception's likely thresholds were<0.291 for proximal classification and>0.450 for (sub)segmental classification, and>0.606 for disease level 4 classification (**Table 9**). Proximal vs

(sub)segmental classification thresholds (**Table 10**) had similar accuracy (p=0.26), sensitivity (p=0.10), and specificity (p=0.57) as the all case proximal vs (sub)segmental threshold. Likely thresholds classified 67% of cases as proximal, disease level 3 or disease level 4 with 67% accuracy, which was significantly better than the naïve prediction accuracy of 55% (p<0.001; **Table 11**). Likely case classification moderately agreed with known surgical disease level (ICC=0.55 p<0.001).

Table 10 Proximal vs (sub)segmental operating point comparison for all case vs confidence-based classification methods

	Proximal vs (sub)segmental classification	Percent Classified	Accuracy (Predicted)	Accuracy (All)	Sen	Spec
MobileNet V2	**All Cases**	100	0.68	0.68	0.57	0.77
	Confident Cases	15	0.87 (0.06)	0.13	0.88 (0.01)	0.86 (0.36)
	Likely Cases	28	0.79 (0.14)	0.22	0.80 (0.01)	0.78 (0.88)
VGG-16	**All Cases**	100	0.68	0.68	0.52	0.82
	Confident Cases	14	0.90 (0.03)	0.13	0.89 (0.001)	1.00 (0.04)
	Likely Cases	57	0.74 (0.30)	0.42	0.78 (< 0.001)	0.71 (0.04)
Xception	**All Cases**	100	0.71	0.71	0.68	0.75
	Confident Cases	27	0.88 (0.03)	0.24	0.88 (0.01)	0.88 (0.07)
	Likely Cases	67	0.78 (0.26)	0.52	0.78 (0.10)	0.78 (0.57)

Table 11 Classification of proximal, disease level 3 and disease level 4 using all case or confidence based classification

	Proximal vs DL 3 vs DL 4 classification	% Classified	Accuracy	ICC
VGG-16	**All Cases**	100	0.64	0.44
	Confident Cases	14	0.71	0.70
	Likely Cases	57	0.67	0.55
Xception	**All Cases**	100	0.59	0.37
	Confident Cases	27	0.69	0.64
	Likely Cases	67	0.67	0.55

3.5. Discussion

This chapter outlines the development, training and evaluation of a machine learning approach to predict left lung CTEPH disease level from VQ images. All three models were able to distinguish proximal from (sub)segmental disease (disease level 1 and 2 vs disease level 3 and 4) based on AUC. All three models had all case proximal vs (sub)segmental classification that predicted with better accuracy than the naïve prediction. Confidence based predictions into proximal disease, segmental disease, or subsegmental disease moderately agreed with surgical disease level and had prediction accuracy better than naïve prediction for VGG-16 and Xception. We will focus on Xception for the remainder of the discussion, as it classified the highest number of samples using the confidence based thersholds

All case classification into proximal or (sub)segmental disease shows similar accuracy compared to clinician prediction from a prior study of clinician disease level prediction before PTE (**Table 12**)[69]. Clinician prediction of disease level as proximal vs (sub)segmental in a series of 50 consecutive cases that included all available clinical data had an accuracy of 68%, with 72% of (sub)segmental cases correctly classified and 59% of proximal cases correctly classified in the left lung[69]. The class balance in the left lung of these cases was 34% proximal, 64%

(sub)segmental, and 2% with no disease (excluded in the proximal vs (sub)segmental accuracy). Xception had similar performance when classifying all cases (55% proximal and 45% (sub)segmental), with overall accuracy of 71%, 68% (sub)segmental accuracy, and 75% proximal accuracy for the left lung. Both clinician and Xception had increased difficulty agreeing with surgical disease level when predicting on all cases. There was slight agreement (based on reported kohen's kappa greater than 0 but less than 0.2) between left lung surgical disease level and clinician predicted disease level was when predicting 5 disease locations in the 50 cases (no disease or disease level 1 -4). Similarly, Xception had slight agreement when classifying all cases as proximal, disease level 3 or disease level 4. In contrast, confidence-based classifications of Xception and VGG-16 outputs moderately agreed with surgical disease level. Given that clinician agreement only improved to fair when classifying cases as proximal or (sub)segmental, future work should assess if confidence-based predictions from machine learning models used in conjunction with clinician expertise can aid clinician prediction.

Table 12 Previously published disease level prediction from Pirompanich et al[69].

Left Lung Known Cases (n=50)	Number predicted correctly
Disease Level 1	2/4 (50%)
Disease Level 2	3/13 (23%)
Disease Level 3	9/25 (36%)
Disease Level 4	5/7 (71%)
No Disease	1/1 (100%)

Left Lung Known Cases (n=49)	Number predicted correctly
Proximal	10/17 (59%)
Sub(segmental)	23/32 (72%)

While current disease level classification performance is promising, further evaluation of our models is needed. First, the evaluation presented in this chapter focuses exclusively on the validation cases used to monitor training progress of the networks. While the network did not train specifically on the validation cases, they did influence the choice of model architecture and

when to stop network training. Future work will utilize the 2019 and onward cases to have a hold out test cohort for evaluation. Identification of a test cohort that underwent PTE at a different institutions is also necessary. Second, the clinical utility of these models depends on their ability to aid clinician evaluation. While comparison to published clinician accuracy aids machine learning model evaluation, direct comparison between our model and clinician prediction on the same cases is needed and will be included in future work. Finally, our models currently only predict on the left lung. Development and evaluation of models that predict on disease level in the right lung is also required.

Our models used only VQ images when predicting disease level. Despite their low spatial resolution, proximal vs (sub)segmental disease could be predicted with better than naïve accuracy when performing all case or confidence based predictions using Xception and confidence based predictions for both MobilenetV2 and VGG-16. Additionally, our VQ only, all case predictions with Xception had similar performance with previously reported clinician performance using all data modalities. This performance using only VQ scans is promising for further model development and clinical implementation. First, many higher resolution images are available, such as CT Pulmonary Angiograms which allow direct visualization of vessel obstruction. Integrating the machine learning based prediction of disease level from these images has the potential to improve predictions over VQ images alone. Addition of patient clinical history and cardiopulmonary data also have the potential to improve machine learning prediction of disease level. Specifically, age, female sex, BMI, DVT history, identified coagulopathy, prior intravenous device, or use of pulmonary hypertension medication[16]. The use of exclusively VQ images also provides exciting clinical potential given VQ's location in the CTEPH and diagnostic work up pathway. Ventilation perfusion scans are one of the first imaging modalities

acquired in the diagnosis of CTEPH, since they are required to confirm the presence of vascular obstruction (perfusion deficit without corresponding ventilation deficit/ventilation-perfusion mismatch) and rule out other causes of pulmonary hypertension. Combining this early diagnostic time point with confidence-based prediction of disease could provide a method to triage high confidence distal patients to expert centers earlier.

3.6. Conclusion

In conclusion, machine learning models can be used to predict left lung CTEPH disease location as disease level from VQ scans. Proximal vs (sub)segmental disease can be identified in all VQ cases when using all three models, and confident cases when using Xception, VGG-16, and MobileNetV2. Confidence based predictions using Xception and VGG-16 classify over half of cases and can further split (sub)segmental into specific segmental vs subsegmental groups. These findings suggest VQ based machine learning predictions of disease level could aid clinician presurgical assessment of CTEPH patients, and serve as the foundation for future models that can combine multiple data modalities to predict CTEPH disease location.

3.7. Acknowledgements

Chapter 3 is recent work that is not currently in preparation for submission as a manuscript. The dissertation author was the primary investigator and author of this work, but this work would not have been possible without the contributions and insights of Samira Masoudi, Nick Kim, Michael Madani, Atul Malhotra, Lewis Hahn, Albert Hsiao, and Francisco Contijoch.

REFERENCES

1. Humbert, M., Kovacs, G., Hoeper, M., Badagliacca, R., Berger, R.. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur. Respir. J.* **61**, 2200879 (2023).
2. Hoeper, M., Kramer, T., Pan Z., Eichstaedt, C., Speisshoefer, J. Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. *Eur. Respir. J.* **50**, 1700740 (2017).
3. Galiè, N., Humbert, M., Vachiery, J., Gibbs, S., Lang, I., Torbicki, A., Simonneau, G., Peacock, A., Noordegraaf, A., Beghetti, M., Ghofrani, A., Sanchez, M., Hansmann, G., Klepetko, W., Lancellotti, P., Matucci, M., McDonagh, T., Pierard, L., Trindade, P., Zompatori,M., Hoepe*, M.* 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS)Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). *Eur. Respir. J.* **46**, 903–975 (2015).
4. Simonneau, G., Torbicki, A., Dorfmüller, P. & Kim, N. The pathophysiology of chronic thromboembolic pulmonary hypertension. *Eur. Respir. Rev.* **26**, 160112 (2017).
5. Lang, I. M., Dorfmüller, P. & Noordegraaf, A. V. The Pathobiology of Chronic Thromboembolic Pulmonary Hypertension. *Ann. Am. Thorac. Soc.* **13**, S215–S221 (2016).

6. Madani, M., Mayer, E., Fadel, E. & Jenkins, D. P. Pulmonary Endarterectomy. Patient Selection, Technical Challenges, and Outcomes. *Ann. Am. Thorac. Soc.* **13**, S240–S247 (2016).
7. Madani, M. Surgical Treatment of Chronic Thromboembolic Pulmonary Hypertension: Pulmonary Thromboendarterectomy. *Methodist DeBakey Cardiovasc. J.* **12**, 213 (2016).
8. Gernhofer, Y. & Pretorius, V. Operative Technique in Pulmonary Thromboendarterectomy. *Oper. Tech. Thorac. Cardiovasc. Surg.* **24**, 219–236 (2019).
9. Madani, M. AUger, W., Pretorius, V., Sakakibara, N., Kerr, K., Kim, N., Fedullo, P., Jamieson, S. Pulmonary Endarterectomy: Recent Changes in a Single Institution's Experience of More Than 2,700 Patients. *Ann. Thorac. Surg.* **94**, 97–103 (2012).
10. Delcroix, M. Lang, I, Pepke-Zaba, J., Jansa, P., D'Armini, A., Snijder, R., Bresser, P., Torbicki, A., Mellemkjaer, S., Lewczuk., J, Simkova, I., Barberà, J., de Perrot, M., Hoeper, M., Gaine, S., Speich, R., Gomez-Sanchez, M., Kovacs, G., Jaïs, X., Ambroz, D., Treacy, C., Morsolini, M., Jenkins, D., Lindner, J., Dartevelle, P., Mayer, E., Simonneau, *G.* Long-Term Outcome of Patients With Chronic Thromboembolic Pulmonary Hypertension: Results From an International Prospective Registry. *Circulation* **133**, 859–871 (2016).
11. Quadery, S., Swift, A., Billings, C., Thompson, A., Elliot, C., Hurdman, J., Athanasios Charalampopoulos, A., Sabroe, I., Armstrong, I., Hamilton, N, Sephton, P., Garrad, S., Pepke-Zaba, J., Jenkins, D., Screaton, N., Rothman, N., Lawrie, A., Cleveland, T., Thomas, S., Rajaram, S., Hill, C., Davies, C., Johns, C., Wild, S., Condliffe, R., Kiely, D. The impact of patient choice on survival in chronic thromboembolic pulmonary hypertension. *Eur. Respir. J.* **52**, 1800589 (2018).

12. Delcroix, M., Torbicki, A., Gopalan, D., Sitbon, O., Klok, F., Lang, I., Jenkins, D., Kim, N., Humbert, M., Jais, X., Noordegraaf., A, Pepke-Zaba, J., Brénot, P., Dorfmuller, P., Fadel, E., Ghofrani, H., Hoeper, M., Jansa, P., Madani, M., Matsubara, H., Ogo, T., Grünig, E., D'Armini, A., Galie, N., Meyer, B., Corkery, P., Meszaros, G., Mayer, M., Simonneau, G. ERS statement on chronic thromboembolic pulmonary hypertension. *Eur. Respir. J.* **57**, 2002828 (2021).
13. O'Brien, S., Feng, L., He, X., Xian, Y., Jacobs, J., Badhwar, V., Kurlansky, P., Furnary, A., Cleveland Jr, J., Lobdell, K., Vassileva, C., Wyler von Ballmoos, M., Thourani, V., Rankin, J., Edgerton, J., D'Agostino, R., Desai, N., Edwards, F., Shahian, D. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2—Statistical Methods and Results. *Ann. Thorac. Surg.* **105**, 1419–1428 (2018).
14. Shahian, D., Jacobs, J., Badhwar, V., Kurlansky, P., Furnary, A., Cleveland Jr, J., Lobdell, K., Vassileva, C., Wyler von Ballmoos, M., Thourani, V., Rankin, J., Edgerton, J., D'Agostino, R., Desai, N., Feng, L., He, X., O'Brien, S. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1—Background, Design Considerations, and Model Development. *Ann. Thorac. Surg.* **105**, 1411–1418 (2018).
15. Kim, N., Delcroix, M., Jenkins, D., Channick, R., Dartevelle, P., Jansa, P., Lang, I., Madani, M., Ogino, H., Pengo, V., Mayer, E. Chronic Thromboembolic Pulmonary Hypertension. *J. Am. Coll. Cardiol.* **62**, D92–D99 (2013).
16. Fernandes, T. M. Kim, N., Kerr, K., Auger, W., Fedullo, P., Poch, D., Yang, J., Papamatheakis, D., Alotaibi, M., Bautista, M., Pretorius, V., Madani, M. Distal vessel pulmonary thromboendarterectomy: Results from a single institution. *J. Heart Lung Transplant.* S1053249823015681 (2023) doi:10.1016/j.healun.2023.02.1500.

17. Jenkins, D., Biederman, A., D'Armini, A., Dartevelle, P., Gan, H., Klepetko, W., Lindner, J., Mayer, E., Madani, M. Operability assessment in CTEPH: Lessons from the CHEST-1 study. *J. Thorac. Cardiovasc. Surg.* **152**, 669-674.e3 (2016).
18. Poch, D., Mahmud, E., Patel, M., Papamatheakis, D., Fernandes, T., Kerr, K., Yang, J., Pretorius, V., Madani, M., Kim, N. Patient selection for balloon pulmonary angioplasty: Six-year results from a high volume PTE surgical center. *Pulm. Circ.* **12**, (2022).
19. Klok, F., Delcroix, M. Bogaard, H. Chronic thromboembolic pulmonary hypertension from the perspective of patients with pulmonary embolism. *J. Thromb. Haemost.* **16**, 1040–1051 (2018).
20. Hahn, L., Papamatheakis, D., Fernandes, T., Poch , D., Yang, J., Shen, J., Hoh, C., Hsiao, A., Kerr, K., Pretorius, V., Madani. M., Kim N., Kligerman, S. Multidisciplinary Approach to Chronic Thromboembolic Pulmonary Hypertension: Role of Radiologists. *RadioGraphics* **43**, e220078 (2023).
21. Fedullo, P., Auger, W., Kerr, K. Rubin, L. Chronic Thromborembolic Pulmonary Hypertension. *N. Engl. J. Med.* **345**, 1465–72 (2001).
22. Kim, N., Delcroix, M., Jais , X., Madani, M., Matsubara, H., Mayer, E., Ogo, T., Tapson, V., Ghofrani, H., Jenkins, D. Chronic thromboembolic pulmonary hypertension. *Eur. Respir. J.* **53**, 1801915 (2019).
23. Kim, N., Delcroix, M., Jenkins, D., Channick, R., Dartevelle, P., Jansa, P., Lang, I., Madani, M., Ogino, H., Pengo, V., Mayer, E. Chronic Thromboembolic Pulmonary Hypertension. *J. Am. Coll. Cardiol.* **62**, D92–D99 (2013).

24. Fuld, M., Halaweish, A., Haynes, S., Divekar, A., Guo, J., Hoffman, E. Pulmonary Perfused Blood Volume with Dual-Energy CT as Surrogate for Pulmonary Perfusion Assessed with Dynamic Multidetector CT. *Radiology* **267**, 747–756 (2013).
25. Dournes, G., Verdier, D., Montaudon, M., Bullier, E., Rivière, A., Dromer, C., Picard, F., Billes, M., Corneloup, O., Laurent, F., Lederlin, M. Dual-energy CT perfusion and angiography in chronic thromboembolic pulmonary hypertension: diagnostic accuracy and concordance with radionuclide scintigraphy. *Eur. Radiol.* **24**, 42–51 (2014).
26. Kim, S., Hur, J., Kim, Y., Lee, H., Hong, Y., Choi, B. Dual-energy CT for differentiating acute and chronic pulmonary thromboembolism: an initial experience. *Int. J. Cardiovasc. Imaging* **30**, 113–120 (2014).
27. Masy, M., Giordano, J., Petyt, G., Hossein-Foucher, C., Duhamel, A., Kyheng, M., De Groote, P., Fertin, M., Lamblin, N., Bervar, J., Remy, J., Remy-Jardin, M. Dual-energy CT (DECT) lung perfusion in pulmonary hypertension: concordance rate with V/Q scintigraphy in diagnosing chronic thromboembolic pulmonary hypertension (CTEPH). *Eur. Radiol.* **28**, 5100–5110 (2018).
28. Thieme, S., Becker, C., Hacker, M., Nikolaou, K., Reiser, M., Johnson, T. Dual energy CT for the assessment of lung perfusion—Correlation to scintigraphy. *Eur. J. Radiol.* **68**, 369–374 (2008).
29. Koike, H., Sueyoshi, E., Sakamoto, I., Uetani, M., Nakata, T., Maemura, K. Correlation between lung perfusion blood volume and SPECT images in patients with chronic thromboembolic pulmonary hypertension by balloon pulmonary angioplasty. *Clin. Imaging* **49**, 80–86 (2018).

30. Kunihiro, Y., Okada, M., Matsunaga, N., Sano, Y., Kudomi, S., Suga, K., Kido, S. Dual-energy perfusion CT of non-diseased lung segments using dual-source CT: correlation with perfusion SPECT. *Jpn. J. Radiol.* **31**, 99–104 (2013).
31. Nallasamy, N., Bullen, J., Karim, W., Heresi, G. Renapurkar, R. Evaluation of Vascular Parameters in Patients With Pulmonary Thromboembolic Disease Using Dual-energy Computed Tomography. *J. Thorac. Imaging* **34**, 367–372 (2019).
32. Hoey, E., Mirsadraee, S., Pepke-Zaba, J., Jenkins, D., Gopalan, D., Screaton, N. Dual-Energy CT Angiography for Assessment of Regional Pulmonary Perfusion in Patients With Chronic Thromboembolic Pulmonary Hypertension: Initial Experience. *Am. J. Roentgenol.* **196**, 524–532 (2011).
33. Meinel, F., Graef, A., THierfelder, K., Armbruster, M., Schild, C., Neurohr, C., Reiser, M., Johnson, T. Automated Quantification of Pulmonary Perfused Blood Volume by Dual-Energy CTPA in Chronic Thromboembolic Pulmonary Hypertension. *RöFo - Fortschritte Auf Dem Geb. Röntgenstrahlen Bildgeb. Verfahr.* **186**, 151–156 (2013).
34. Takagi, H., Ota, H., Sugimura, K., Otani, K., Tominaga, J., Aoki, T., Tatebe, S., Miura, M., Yamamoto, S., Sato, H., Yaoita, N., Suzuki, H., Shimokawa, H., Takase, K. Dual-energy CT to estimate clinical severity of chronic thromboembolic pulmonary hypertension: Comparison with invasive right heart catheterization. *Eur. J. Radiol.* **85**, 1574–1580 (2016).
35. Koike, H., Sueyoshi, E., Sakamoto, I., Uetani, M., Nakata, T., Maemura, K. Comparative clinical and predictive value of lung perfusion blood volume CT, lung perfusion SPECT and catheter pulmonary angiography images in patients with chronic thromboembolic pulmonary hypertension before and after balloon pulmonary angioplasty. *Eur. Radiol.* **28**, 5091–5099 (2018).

36. Atwi, N., Smith, D., Flores, C., Dharaiya, E., Danrad, R., Kambadakone, A., Toshav, A. Dual-energy CT in the obese: a preliminary retrospective review to evaluate quality and feasibility of the single-source dual-detector implementation. *Abdom. Radiol.* **44**, 783–789 (2019).
37. Hahn, L., Hall, K., Alebdi, T., Kligerman, S., Hsiao, A. Automated Deep Learning Analysis for Quality Improvement of CT Pulmonary Angiography. *Radiol. Artif. Intell.* **4**, e210162 (2022).
38. Hasenstab, K., Yuan, N., Retson, T., Conrad, D., Kligerman, S., Lynch, D., Hsiao, A. Automated CT Staging of Chronic Obstructive Pulmonary Disease Severity for Predicting Disease Progression and Mortality with a Deep Learning Convolutional Neural Network. *Radiol. Cardiothorac. Imaging* **3**, e200477 (2021).
39. Haruhiko, M., Tanaka, I., Fukui, R., Shen, Y., Ishikawa, T., Tate, E., Ueno, E. Dual-Energy Spectral CT: Various Clinical Vascular Applications. *RadioGraphics* **36**, 1215–1232 (2016).
40. Hanley, J. & McNeil, B. The meaning and use of the area under a reciever oiperatin charachteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
41. DeLong, E., DeLong, D., Clarke-Pearson, D. Comparing the Areas under Two or More Correlated Reciever Operating Charachterisitic Curves: A Nonparametric Approach. *Biometrics* **44**, 837.
42. Li, J. & Fine, J. On sample sensitivity and specificity in prospective diagnostic accuracy studies. *Satistics Med.* **23**, 2537–2550 (2004).
43. Portney, L. & Watkins, M. *Foundations of Clinical Research: Applications to Practice*. (Pearson/Prentice Hall, 2009).

44. Williams, E. The Comparison of Regression Variables. *J. R. Stat. Soc. Ser. B Methodol.* **21**, 396–399 (1959).

45. Steiger, J. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* **87**, 245–251 (1980).

46. Koo, T. & Li, M. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **15**, 155–163 (2016).

47. Navarro, D. *The F-test as model comparison. In Learning sstatistics with R: A tutorial for psychology students and other beginners*. (CC BY-SA, 2018).

48. D'Armini, A. M. *et al.* Pulmonary endarterectomy for distal chronic thromboembolic pulmonary hypertension. *J. Thorac. Cardiovasc. Surg.* **148**, 1005-1012.e2 (2014).

49. Mahmud, E., Madani, M., Kim, N., Poch, D., Ang, L., Behnamfar, O., Patel, M., Auger, W. Chronic Thromboembolic Pulmonary Hypertension. *J. Am. Coll. Cardiol.* **71**, 2468–2486 (2018).

50. Retson, T., Hasenstab, K., Kligerman, S., Jacobs, K., Yen, A., Brouha, S., Hahn, L., Hsiao, A. Reader Perceptions and Impact of AI on CT Assessment of Air Trapping. *Radiol. Artif. Intell.* **4**, e210160 (2022).

51. Parakh, A., Lennartz, S., An, C., Rajiah, P., Yeh, B., Simeone, F., Sahani, D., Kambadakone, A. Dual-Energy CT Images: Pearls and Pitfalls. *RadioGraphics* **41**, 98–119 (2021).

52. Qanadli, S., Hajjam, M., Vieillard-Baron, A., Joseph, T., Mesurolle, B., Olivia, V., Barré, O., Bruckert, F., Dubourg, O., Lacombe, P. New CT Index to Quantify Arterial Obstruction in Pulmonary Embolism: Comparison with Angiographic Index and Echocardiography. *Am. J. Roentgenol.* **176**, 1415–1420 (2001).

53. Lee, H., Wanderley, M., Cardinal da Silva Rubin, V., Clara Tude Rodrigues, A., Rocha Diniz, A., Rodrigues Parga, J., Britto Passos Amato, M. Lobar pulmonary perfusion quantification with dual-energy CT angiography: Interlobar variability and relationship with regional clot burden in pulmonary embolism. *Eur. J. Radiol. Open* **9**, 100428 (2022).
54. Kunihiro, Y., Okada, M., Matsunaga, N. Evaluation of a proper cutoff value on quantitative dual-energy perfusion CT for the assessment of acute pulmonary thromboembolism. *Acta Radiol.* **58**, 1061–1067 (2017).
55. Heinrich, M., Uder, M., Tscholl, D., Grgic, A., Kramann, B., Schäfers, H. CT Scan Findings in Chronic Thromboembolic Pulmonary Hypertension. *Chest* **127**, 1606–1613 (2005).
56. Abozeed, M., Conic, S., Bullen, J., Rizk, A., Bin Saeedan, M., Karim, W., Heresi, G., Renapurkar, R. Dual energy CT based scoring in chronic thromboembolic pulmonary hypertension and correlation with clinical and hemodynamic parameters: a retrospective cross-sectional study. *Cardiovasc. Diagn. Ther.* **12**, 305–313 (2022).
57. Shimizu, H., Tanabe, N., Terada, J., Masuda, M., Sakao, S., Kasahara, Y., Takiguchi, Y., Tatsumi, K., Kuriyama, T. Dilatation of Bronchial Arteries Correlates With Extent of Central Disease in Patients With Chronic Thromboembolic Pulmonary Hypertension. *Circ. J.* **72**, 1136–1141 (2008).
58. Delcroix, M., Vonk Noordegraaf, A., Fadel, E., Lang, I., Simonneau, G., Naeije, R. Vascular and right ventricular remodelling in chronic thromboembolic pulmonary hypertension. *Eur. Respir. J.* **41**, 224–232 (2013).
59. Kauczor, H., Schwickert, H., Mayer, E., Schweden, F., Schild, H., Thelen, M. Spiral CT of Bronchial Ateries in Chronic Thromboembolism. *J. Comput. Assist. Tomogr.* **18**, 855–861 (1994).

60. Lee, N., Blanchard, D., Knowlton, K., McDivit, A., Pretorius, V., Madani, M., Fedullo, P., Kerr, K., Kim, N., Poch, D., Auger, W., Daniels, L. Prevalence of Coronary Artery–Pulmonary Artery Collaterals in Patients with Chronic Thromboembolic Pulmonary Hypertension. *Pulm. Circ.* **5**, 313–321 (2015).
61. Jenkins, D. Pulmonary endarterectomy: the potentially curative treatment for patients with chronic thromboembolic pulmonary hypertension. *Eur. Respir. Rev.* **24**, 263–271 (2015).
62. Suntharalingam, J., Machado, R., Sharples, L., Toshner, M., Sheares, K., Hughes, R., Jenkins, D., Trembath, R., Morrell, N., Pepke-Zaba, J. Demographic features, BMPR2 status and outcomes in distal chronic thromboembolic pulmonary hypertension. *Thorax* **62**, 617–622 (2007).
63. Ruigrok, D., Meijboom, L., Westerhof, B., Huis In 't Veld, A., van der Bruggen, C., Marcus, J., Nossent, E., Vonk Noordegraaf, A., Symersky, P., Bogaard, H. Right Ventricular Load and Function in Chronic Thromboembolic Pulmonary Hypertension: Differences between Proximal and Distal Chronic Thromboembolic Pulmonary Hypertension. *Am. J. Respir. Crit. Care Med.* **199**, 1163–1166 (2019).
64. Frauenfelder, T., Mcinnis, M., Eberhard, M., De Perrot, M., Ulrich, S., Inci, I., Opitz, I. Predictive value of CT pulmonary angiography to assess surgical accessibility for pulmonary endarterectomy in chronic thromboembolic pulmonary hypertension (CTEPH) patients. *Eur. Respir. J.* **58**, (2021).
65. Remy-Jardin, M., Faivre, J., Kaergel, R., Hutt, A., Felloni, P., Khung, So., Lejeune, A., Giordano, J., Remy, J. Machine Learning and Deep Neural Network Applications in the Thorax: Pulmonary Embolism, Chronic Thromboembolic Pulmonary Hypertension, Aorta, and Chronic Obstructive Pulmonary Disease. *J. Thorac. Imaging* **35**, S40–S48 (2020).

66. Scott, A. Neural Network Ventilation-Perfusion. (1993).

67. Yang, X., Lin, Y., Su, J., Wang, X., Li, X., Lin, J., Cheng, K. A Two-Stage Convolutional Neural Network for Pulmonary Embolism Detection From CTPA Images. *IEEE Access* **7**, 84849–84857 (2019).

68. Holst, H., Aström, K., Järund, A., Palmer, J., Heyden, A., Kahl, F., Tägil, K., Evander, E., Sparr, G., Edenbrandt, L. Automated interpretation of ventilation-perfusion lung scintigrams for the diagnosis of pulmonary embolism using artificial neural networks. *Eur. J. Nucl. Med. Mol. Imaging* **27**, 400–406 (2000).

69. Pirompanich, P., Fernandes, T., Kerr, K., Madani, M., Papamatheakis, D., Poch, D., Pretorius, V., Kim, N. Accuracy of pre-operative specimen level prediction in chronic thromboembolic pulmonary hypertension (CTEPH) patients undergoing pulmonary thromboendarterectomy (PTE). *Eur. Respir. J.* **56**, 1538 (2020).