

Original Research

Computed Tomography Features of Lung Structure Have Utility for Differentiating Malignant and Benign Pulmonary Nodules

Johanna M. Uthoff, PhD^{1,2,3} Sarah L. Mott, MS³ Jared Larson, BS¹ Christine M. Neslund-Dudas, PhD^{4,5}
Ann G. Schwartz, PhD, MPH⁶ Jessica C. Sieren, PhD^{1,2,3} and the COPDGene® Investigators

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is a known comorbidity for lung cancer independent of smoking history. Quantitative computed tomography (qCT) imaging features related to COPD have shown promise in the assessment of lung cancer risk. We hypothesize that qCT features from the lung, lobe, and airway tree related to the location of the pulmonary nodule can be used to provide informative malignancy risk assessment.

Methods: A total of 183 qCT features were extracted from 278 individuals with a solitary pulmonary nodule of known diagnosis (71 malignant, 207 benign). These included histogram and airway characteristics of the lungs, lobe, and segmental paths. Performances of the least absolute shrinkage and selection operator (LASSO) regression analysis and an ensemble of neural networks (ENN) were compared for feature set selection and classification on a testing cohort of 49 additional individuals (15 malignant, 34 benign).

Results: The LASSO and ENN methods produced different feature sets for classification with LASSO selecting fewer qCT features (7) than the ENN (17). The LASSO model with the highest performing training area under the curve (AUC) (0.80) incorporated automatically extracted features and reader-measured nodule diameter with a testing AUC of 0.62. The ENN model with the highest performing AUC (0.77) also incorporated qCT and reader diameter but maintained higher testing performance AUC (0.79).

Conclusions: Automatically extracted qCT imaging features of the lung can be informative of the differentiation between individuals with malignant pulmonary nodules and those with benign pulmonary nodules, without requiring nodule segmentation and analysis.

Abbreviations: chronic obstructive pulmonary disease, **COPD**; quantitative computed tomography, **qCT**; least absolute shrinkage and selection operator, **LASSO**; ensemble of neural networks, **ENN**; area under the curve, **AUC**; computed tomography, **CT**; COPD Genetic Epidemiology, **COPDGene**®; Inflammation, Health, and Lung Epidemiology study, **INHALE**; National Lung Screening Trial, **NLST**; Global initiative for chronic Obstructive Lung Disease, **GOLD**; pulmonary function tests, **PFTs**; forced vital capacity, **FVC**; forced expiratory volume in 1 second, **FEV₁**; segmental airway paths, **sAP**; coefficient of variation, **CV**; Hounsfield unit, **HU**; area under the receiver operating characteristic curve, **AUC-ROC**; information optimization, **IO**; Response Evaluation Criteria in Solid Tumors, **RECIST**

Funding Support: This work was supported in part by a National Cancer Institute Core grant (P30CA086862) to the University of Iowa Holden Comprehensive Cancer Center, an American Lung Association Dissertation Award (LH-574107), and an American Lung Association Cancer Discovery Award (grant LCD-220717-N). The COPDGene study (NCT00608764) is also supported by the COPD Foundation through contributions made to an Industry Advisory Committee comprised of AstraZeneca, Boehringer-Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion. The INHALE study was supported by Award Number R01CA141769 and P30CA022453 from the National Cancer Institute, Health and Human Services Award HHSN26120130011I and the Herrick Foundation.

Date of Acceptance: January 12, 2022 | **Published Online Date:** January 12, 2022

Citation: Uthoff JM, Mott SL, Larson J, et al; the COPDGene Investigators. Computed tomography features of lung structure have utility for differentiating malignant and benign pulmonary nodules. *Chronic Obstr Pulm Dis.* 2022;9(2):154-164. doi: <https://doi.org/10.15326/jcopdf.2021.0271>

1. Department of Radiology, University of Iowa, Iowa City, Iowa, United States
2. Department of Biomedical Engineering, University of Iowa, Iowa City, Iowa, United States
3. Holden Comprehensive Cancer Center, University of Iowa, Iowa City, Iowa, United States
4. Department of Public Health Sciences, Henry Ford Health System, Detroit, Michigan, United States
5. Henry Ford Cancer Institute, Henry Ford Health System, Detroit, Michigan, United States
6. Karmanos Cancer Institute, Wayne State University, Detroit, Michigan, United States

Address correspondence to:

Jessica C. Sieren, PhD
 University of Iowa
 Department of Radiology
 200 Hawkins Drive, CC704GH
 Iowa City, IA 52242
 Phone: (319) 356-1407
 Email: Jessica-sieren@uiowa.edu

Keywords:

quantitative computed tomography; chronic obstructive pulmonary disease; cancer risk assessment; machine learning; artificial intelligence

Introduction

Chronic obstructive pulmonary disease (COPD) is characterized by obstructive lung function. Evaluated with medical imaging, COPD is heterogenous with varying presentations of structural changes in the lung parenchyma and airways. COPD is a risk factor for lung cancer development, independent of smoking history.^{1,2} Prior studies that included pulmonary function testing and qualitative imaging assessment using computed tomography (CT)²⁻⁴ have shown links between COPD features and risk of lung cancer. There has been limited published research into the overlap of the COPD-related quantitative-CT (qCT) measures and pulmonary nodules and risk of lung cancer. Most studies have compared individuals with a lung cancer diagnosis to individuals without pulmonary nodules.^{2,5-10} Extent of emphysema in the lungs has been shown to be a positive predictor of lung cancer^{5-7,9-11}; qCT airway measures have also been assessed for potential predictive benefit, yet no significant discriminatory ability has been demonstrated.^{7,11} However, many of these studies have not focused on controls with nodules. Here, we investigate a population with identified nodules that

present a significant post-CT imaging challenge in the selection of the appropriate follow-up procedure.

We investigate the utility of objective and automatically obtained qCT metrics in predicting individuals with lung cancer on a cohort of scans all of which include solitary pulmonary nodules ≥ 4 mm. Here, the nodule was not extracted or segmented from the scans prior to qCT feature extraction. We hypothesize that qCT features from the lung, lobe, and airway tree related to the location of the pulmonary nodule can be used to provide informative malignancy risk assessment, without requiring segmentation of the pulmonary nodule. Ultimately, for a risk assessment pipeline to be most clinically helpful, there should be the requirement for as little human effort as possible. The performance of the least absolute shrinkage and selection operator (LASSO) regression analysis and an ensemble of neural networks (ENN) were compared for feature set selection and classification on qCT features and demographical/clinical characteristics.

Materials and Methods

Study Population

The study cohort was comprised of individuals retrospectively collected with institutional review board approval from 3 prospective research studies: the COPD Genetic Epidemiology Study (COPDGene[®]),¹² the Inflammation, Health, and Lung Epidemiology Study (INHALE),² and the National Lung Screening Trial (NLST).¹³ Inspiratory CT data was collected from multiple institutions following a standardized protocol. Demographic and basic clinical features were obtained from parent studies (Table 1).

The established study cohort included 327 individuals with solitary pulmonary nodules (86 with primary lung cancer diagnosis) who underwent CT prior to diagnosis. The criterion for inclusion in this study was based on the availability of diagnostic information: malignant nodules were confirmed on histopathology, and benign nodules were diagnosed by histopathology and/or stability/resolution on more than 24 months follow-up. Diagnosis of COPD was not an inclusion criterion of the study, with 42% of the cohort having no COPD (Global initiative for chronic Obstructive Lung Disease¹⁴ [GOLD] 0) and 58% having COPD (GOLD 1–4). The individuals were sectioned into a training cohort (n=278) and a testing cohort (n=49), using class-persevering random selection.

Table 1. Individual Demographics From the Training and Testing Cohort

	Malignant	Benign	<i>p</i>
Training			
Individuals	71	207	-
Age, years (mean ± SD)	64.3 ± 10.1	62.2 ± 8.3	0.07
Sex (female: male)	41:30	101:106	0.19
Smoking History (yes: no)	63:8	204:3	<0.01
Any Cancer History (yes: no)	12:59	9:198	0.07
Family Cancer History (yes: no)	37:34	85:122	0.11
Family Lung Cancer History (yes: no)	18:53	68:139	0.24
Cessation Time, years (mean ± SD)	2.1 ± 5.4	3.6 ± 6.2	0.03
Diameter, mm (mean ± SD)	14.7 ± 8.3	9.11 ± 5.0	<0.01
GOLD Stage			
0	34	78	0.20
1	8	44	
2	18	42	
3	7	32	
4	4	10	
Testing			
Individuals	15	34	-
Age, years (mean ± SD)	60.4 ± 7.9	61.7 ± 9.9	0.48
Sex (female: male)	10:5	21:13	0.74
Smoking History (yes: no)	14:1	33:1	0.52
Any Cancer History (yes: no)	3:12	4:30	0.66
Family Cancer History (yes: no)	6:9	16:18	0.40
Family Lung Cancer History (yes: no)	9:6	12:22	0.75
Cessation Time, years (mean ± SD)	2.1 ± 5.4	3.6 ± 6.2	0.24
Diameter, mm (mean ± SD)	15.1 ± 8.3	10.4 ± 6.4	0.11
GOLD Stage			
0	10	14	0.16
1	2	6	
2	2	9	
3	0	5	
4	1	0	

Family Lung Cancer History refers to any first degree relative

SD=standard deviation; GOLD=Global initiative for chronic Obstructive Lung Disease

Feature Groups

Two feature groups were collected and analysed for predictive capabilities: (1) Clinical and (2) Imaging qCT. Clinical features required input from a human based on recollection of the patient, clinical testing, or image reader assessment. Imaging qCT features

were automatically extracted from the CT datasets using Apollo software suite (VIDA Diagnostics).

Clinical Features:

A subset of this study's cohort was previously used to investigate the utility and consistency of post-imaging mathematical prediction models for the differentiation between malignant and benign lung nodules.¹⁵ The clinical predictive values collected were included as clinical features including: 7 measures of individual-provided historical information (age, sex, smoking history, cessation time, history of cancer, family history cancer, family history lung cancer), radiologist-reported maximum in-plane diameter, and 4 pulmonary function tests (PFTs) including forced vital capacity (FVC) percent predicted, forced expiratory volume at 1 second (FEV₁) percent predicted, FEV₁/FVC, and GOLD stage.

Imaging Features:

In total, 183 qCT measures were available for model development. qCT characteristics of the parenchyma and airways (segmental branches) were extracted (Apollo, VIDA Diagnostics). Parenchyma measurements were calculated across the whole lung (Lung) or isolated from the lobe containing the pulmonary nodule (Lobe). Measures from the segmental bronchi were incorporated across the whole lung (Lung), from 5 selected airway pathways passing through RB1, RB4, RB10, LB1 and LB10 (segmental airway paths [sAP]) and limited only to the bronchi in the lung lobe containing the pulmonary nodule (Lobe). The diversity in measurements throughout the lung was captured by incorporating the coefficient of variation (CV) in the measurements across all lobes, calculated using the "raster" package in the R statistical programming language.¹⁶

Parenchymal characteristics were captured through Hounsfield unit (HU) intensity, from the lung tissue including HU histogram mean, standard deviation, and skew. Areas of low tissue density were captured as the percentage of lung volume less than -910 HU (*percentage below -910 HU*), and high tissue density regions were represented as the percentage of lung volume above 0 HU (*percentage above 0 HU*). Total, tissue, and air volumes were also included.

Airway characteristics included wall thickness, diameters, perimeters, areas, and wall area fraction. Another commonly utilized airway

metric of wall thickness at a standardized internal airway perimeter of 10mm (*PiIO*) was included.

Application of Statistical and Machine Learning Techniques

Chi-squared or Fisher's exact test (if >20% of cells had expected frequencies <5) were used to compare categorical variables, and Wilcoxon rank sum tests were used to compare continuous variables between patients based on diagnosis each for the training and testing sets. The models were developed using 3 feature pools: qCT (imaging features only), Clinical (clinical features only), and qCT+Clinical (qCT features alongside the clinical features). Model performance was estimated with area under the receiver operating characteristic curve (AUC-ROC) and compared with the method of DeLong. AUC-ROC is also known as the c-statistic and represents the probability that a higher risk will be assigned to a randomly chosen true high-risk case (here, cancer) than a randomly chosen low risk case (here, non-cancer).¹⁷⁻¹⁹

Multivariable Model Development:

Using the training dataset, LASSO models were applied to identify prognostic predictors of nodule malignancy status. LASSO model performance on the training set was estimated with 100 repeats of 10-fold cross-validation. The model derived in the building phase was applied to the testing dataset. AUC-ROC was used as the measure of model performance.

Ensemble Neural Network Model Application:

The machine learning pipeline described in Uthoff et al was implemented with the following modifications.¹⁵ Lung parenchyma and airway image segmentation and feature extraction was implemented using the Apollo software. No segmentation or feature extraction of the pulmonary nodule(s) was required. Highly correlated features were reduced to a single representative feature through k-medoid clustering and mutual information optimization (IO) was used to rank informative predictors for feature selection. The selected feature set is used to train an ensemble of neural networks (ENN). We apply the techniques developed for feature set reduction, selection, and classification to feature groups qCT and qCT+Clinical. The final trained models (development) were applied to the independent testing cohort.

Results

Statistical and Machine Learning Technique Results

The set selection methods (multivariable and IO) were applied to the feature pools qCT and qCT+Clinical. Table 2 indicates the features selected in each of the models. In total, 30 features were selected as predictors in 1 or more of the models. There was minimal overlap between the features selected by multivariable and IO methods (2 features overlap – *lobe percentage above 0 HU and diameter*). Table 3 shows the performance of the feature set selection methods (LASSO, IO) and classification methods (LASSO, ENN).

Multivariable Analysis Selects Diameter and qCT Features for Highest Training Performance:

The multivariable analysis yielded a model that incorporated qCT features from the airway tree, whole lung, and lobe (Table 2). Selecting only automatically extracted imaging features for model development included 7 measures, with a training AUC-ROC of 0.80 and a testing AUC-ROC of 0.56 – indicating overfitting of the training dataset (Table 3). Allowing the model to select clinically ascribed features from radiologist or participant input produced a model that included diameter and 4 qCT features, with a training AUC-ROC of 0.80 and testing AUC-ROC of 0.62; the improvement in testing AUC-ROC could point to size bias within our cohort. A model developed using only clinical/demographical features included only the diameter, with a training AUC-ROC of 0.70 and testing AUC-ROC of 0.64.

Ensemble Neural Network Schema Improves Testing Performance of Multivariable Selected Imaging Features:

The features selected through multivariable model development were used to train ENN models. The resulting training performance was similar to the multivariable model training performance, however, the resulting testing performance from the ENN model was higher (AUC-ROC of 0.60) for the qCT-only features model compared to the multivariable testing (AUC-ROC of 0.56). This indicates the ENN classification performance is more stable on new cases.

Table 2. Selected Features for Each of the Developed Ensemble Neural Network Models with Odds Ratio

Location	Category	Feature	OR	L-I	IO-I	L-CI	IO-CI	L-C
Nodule	Clinical	Diameter	-			X	X	X
Lung	Airway	std_dev_MaxWallThickness	1.06	X		X		
Lung	Airway	std_dev_AvgWallThickness	0.67	X		X		
Lung	Airway	pi10	1.51		X		X	
sAP	Airway	avg_WallAreaFraction	0.73		X		X	
sAP	Airway	avg_MinorInnerDiam	1.10				X	
sAP	Airway	std_dev_InnerPerimeter	0.79		X			
sAP	Airway	avg_MinWallThickness	0.79		X		X	
sAP	Airway	std_dev_MaxWallThickness	0.66	X		X		
sAP	CV	avg_OuterArea	1.20		X		X	
sAP	CV	avg_MinorOuterDiam	1.09				X	
sAP	CV	std_dev_AvgWallThickness	1.16		X			
sAP	CV	avg_MaxWallThickness	1.29		X			
Lobe	Histogram	total_volume_cm3	0.73		X		X	
Lobe	Histogram	mean_HU	1.67		X		X	
Lobe	Histogram	std_dev_HU	2.52	X		X		
Lobe	Histogram	skew	0.49				X	
Lobe	Histogram	percentage_above_0_HU	1.47	X	X		X	
Lobe	Histogram	percentage_below_-910_HU	0.75		X			
Lobe	Airway	avg_WallAreaFraction	0.61				X	
Lobe	Airway	avg_MinorOuterDiam	1.04				X	
Lobe	Airway	std_dev_MajorOuterDiam	0.55				X	
Lobe	Airway	std_dev_MinWallThickness	0.74		X			
Lobe	Airway	std_dev_AvgWallThickness	0.67	X				
Lobe	CV	total_volume_cm3	1.25		X		X	
Lobe	CV	skew	1.36		X			
Lobe	CV	std_dev_MajorOuterDiam	0.83	X				
Lobe	CV	avg_InnerPerimeter	0.94				X	
Lobe	CV	avg_InnerEquivalentCircleDiam	0.94		X			
Lobe	CV	std_dev_AvgWallThickness	0.98		X			
Lobe	CV	std_dev_MinWallThickness	0.82		X			
Count				7	17	5	16	1

OR=odds ratio; L-I=LASSO (least absolute shrinkage and selection operator regression analysis) selection of quantitative computed tomography (qCT) derived features; IO-I=information optimization selection of qCT derived features; L-CI=LASSO selection of qCT+clinical features; L-C=LASSO selection of clinical features only; std dev=standard deviation; sAP=segmental airway paths; avg=average; diam=diameter; CV=coefficient of variation; HU=Hounsfield unit

Combination Medoids-Information Optimization and Ensemble Neural Network Less Likely to Over-train than Multivariable Approach:

While the multivariable selection method for features obtained higher training AUC-ROCs for both feature pools (qCT+Clinical and qCT-only), the testing AUCs for these models implied potential overfitting, particularly with the

qCT-only features (Table 3). This overfitting persisted in the ENN trained sample with the multivariable selected features indicating the overfitting is occurring during feature set selection. In contrast, the Medoids-IO selection followed by ENN model development obtained testing AUCs that were more consistent with the training AUCs. Delong's analysis of the testing AUCs demonstrated that the qCT+Clinical Medoids-IO selected ENN model

Table 3. Performance Results From The Developed Models Using Quantitative Computed Tomography and/or Clinical Characteristics

Feature Groups	Selection	Model	Training AUC-ROC	Testing AUC-ROC
qCT	LASSO	LASSO	0.80	0.56
	LASSO	ENN	0.81	0.60
	IO	ENN	0.74	0.74
qCT + Clinical	LASSO	LASSO	0.80	0.62
	LASSO	ENN	0.81	0.62
	IO	ENN	0.77	0.79
Clinical	LASSO	LASSO	0.70	0.64

AUC-ROC=area-under-curve of receiver-operator characteristic; qCT=quantitative computed tomography; LASSO=least absolute shrinkage and selection operator regression analysis; ENN=ensemble of artificial neural networks; IO=information optimization

(testing AUC-ROC of 0.79) was significantly better on new cases than the qCT+Clinical Multivariable selected ENN model ($p \leq 0.01$) and the qCT-only Multivariable selected ENN model ($p = 0.01$). While the qCT-only IO selected ENN model also had a high testing AUC-ROC of 0.74, it was not statistically better than the qCT+Clinical Multivariable selected ENN model ($p = 0.14$) or the qCT-only Multivariable selected ENN model ($p = 0.08$).

Quantitative Imaging Feature Importance

Nodule Diameter is the Only Clinical Feature Selected by Both Multivariable and Information Optimization Approaches:

The nodule diameter, measured as the Response Evaluation Criteria In Solid Tumor (RECIST) diameter, was the only clinical characteristic selected by the models. This could be due to the size bias that is seen in the cohort of pulmonary nodules –with cancerous nodules tending on average to be larger than their benign counterparts. The LASSO model using diameter achieved a training AUC-ROC of 0.70 and testing AUC-ROC of 0.64. Diameter was selected in both LASSO and IO selection methods. The addition of clinical characteristics did not significantly improve the performance of either LASSO or IO set selection in ENN development ($p > 0.05$). The addition of clinical characteristics in IO set selection did significantly improve the testing AUC-ROC over LASSO set selection without clinical features ($p = 0.02$).

K-medoids Clustering Interrogates Feature Correlations:

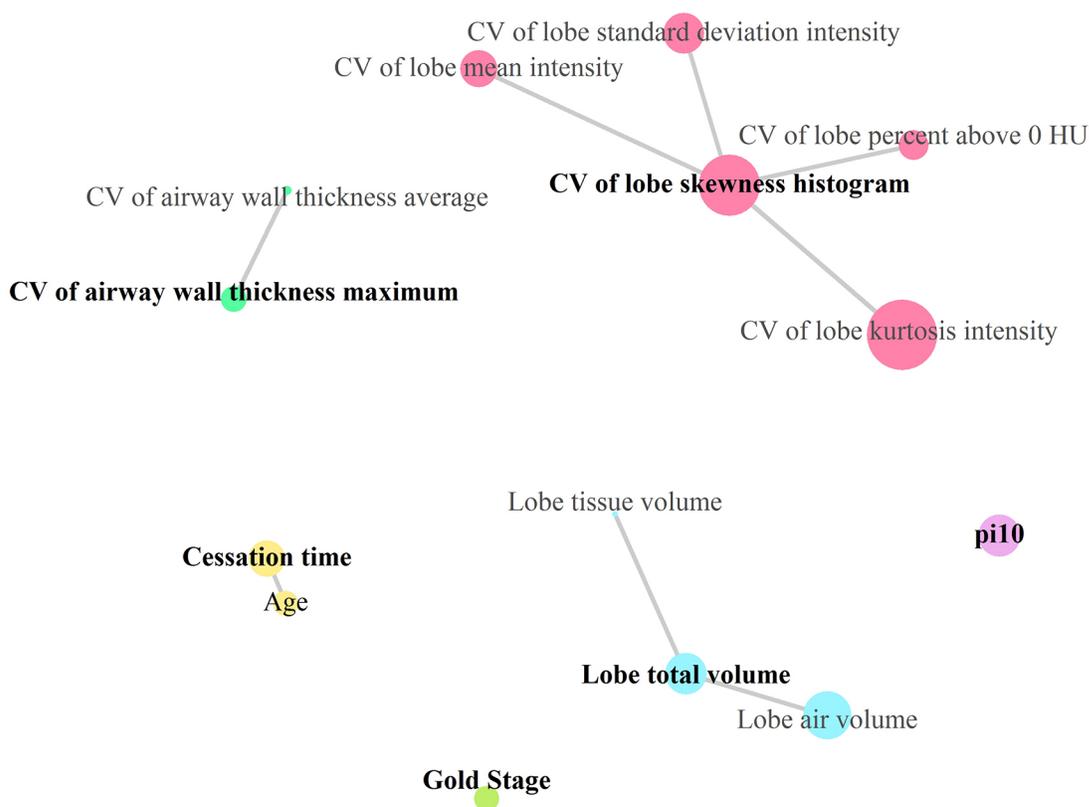
For the full training dataset, k-medoid clustering with the optimal k produced 37 clusters. Example clustering of qCT features and clinical features is shown in Figure 1. Across 10-fold cross validation, clustering of the qCT features using k-medoids method showed that 21 were stable (medoids in all 10 folds) and an additional 6 (27 total) were semi-stable (medoids in at least 8 of the 10 folds). Of note, the *lobe standard deviation HU* qCT feature, which was selected in both the multivariable models, was not selected as a medoid in any of the 10-folds; instead, it was either a member of the cluster with representative *lobe mean HU* (8 folds) or *lobe skew* (2 folds).

Multivariable and Information Optimization Select Vastly Different qCT Features:

Comparing the features selected by the 2 methods, only *nodule diameter* and the *lobe percentage above 0 HU* were common (Table 2). This is likely largely due to the feature-set reduction by k-medoids performed prior to IO-set selection. Of the features selected by the multivariable model, only 2 (*diameter* and *lobe percentage above 0 HU*) were selected as a medoid during 10-fold cross-validation k-medoids clustering. As such, none of the remaining multivariable selected features were available for IO-set selection. The IO method selected more CV features than the LASSO selection method which only selected the *CV of the Standard Deviation of the Major Outer Diameter* using the qCT feature pool. The selected CV features were predominately from the airways.

Features Related to the Location of the Pulmonary Nodule:

Table 2 illustrates that qCT features derived from the lobe containing the pulmonary nodule are incorporated to differing degrees across the model configurations. The qCT Medoids-IO selected ENN model included the most lobar features (n=10), while the qCT+Clinical Multivariate selected ENN model included the least lobe derived features (n=1). The qCT+Clinical Medoids-IO selected ENN model, which achieved the highest testing performance (AUC-ROC=0.79), included 9 lobar qCT features; 4 features from the parenchyma histogram, 3 features from airways, and 2 CV features.

Figure 1. Example of Clustering Arrangement for Select Medoid Features^a

^aselect medoid features are in bold

Clusters are color coded. Lines indicate strength of correlation between features within a cluster. The size of the feature point indicates the information theory metric (larger circle means the feature shares more information with diagnosis).

QCT=quantitative computed tomography; CV=coefficient of variation; GOLD=Global initiative for chronic Obstructive Lung Disease

Discussion

This study has demonstrated the potential richness in extra-nodular, automatically extracted imaging-derived features for the distinction between individuals with malignant and benign pulmonary nodules. It has highlighted the utility of more advanced methods of qCT feature selection for less overtraining. Here, we extracted and included in the analysis many features in an exploratory manner instead of selecting a small subset of specific features as has predominately been reported previously. As such, we have found that the ENN system with IO was advantageous for the exploratory manner of feature selection.

Prior works with qCT have primarily focused on the associations with lung cancer, irrespective of pulmonary nodule presence. These works have reported mixed results

in the benefit of whole lung qCT features for lung cancer risk assessment. Studies by Carr et al, Chubachi et al, and Gagnet et al indicated that increased low attenuation area percentages, indicative of CT-characterized emphysema, were higher in individuals who developed lung cancer.^{5,6} Studies by Gierada et al and Wille et al did not find statistically significant differences but indicated that emphysema was more frequently seen and at a higher grade in individuals with lung cancer.^{7,8} Bae et al investigated whole lung and lobar qCT emphysema ratios, finding the odds of lung cancer increased in lobes with more severe emphysema.⁹ However, work from Wilson, Maldonado, and Johannessen showed no statistical evidence in quantitative lung parenchyma and/or airway measures and risk of lung cancer.^{3,4,11} Schwartz et al found in multivariable modelling only the expiratory qCT measure, percentage below -856 HU, and

PFT characteristics were independent predictors of lung cancer risk.² In contrast to the prior publications, this study consisted only of individuals with a pulmonary nodule (diagnosed malignant or benign tumor), a subset that may be highly relevant in lung cancer screening programs where solid pulmonary nodules require rigorous follow-up. One qCT feature associated with emphysema or low tissue density regions, *percentage below -910 HU*, was selected as a feature but only in 1 model. However, associated information may have been captured indirectly by the selected features, *mean HU*, *standard deviation of HU*, and the *skew*.

This study did not segment the pulmonary nodule or remove it from the analysis. Previous machine learning works have looked at qCT features extracted solely from the nodule²⁰⁻²⁴ and from the lung tissue immediately surrounding the nodule^{15,24,25} with high performance in training (AUC-ROC=0.84-1.0). The features extracted for these tools incorporated intensity, shape, size, border, and texture; a recent multi-institutional study by the quantitative imaging network has demonstrated that quantification of the same image features can vary between groups.²⁶ In this study, we extracted qCT measures using the Apollo proprietary software package (Food and Drug Administration approved, Conformité Européenne certified for clinical use in the United States and Europe); this allows for consistency in the method of quantification of features beyond what can be expected for in-house developed features.

The best performing model from this study, qCT+Clinical Medoids-IO selected ENN model, included many features from the lobe of the lung that contained the pulmonary nodule. This is supportive of the study hypothesis that qCT features from the lung, lobe, and airway tree related to the location of the pulmonary nodule can be used to provide informative malignancy risk assessment, without requiring segmentation of the pulmonary nodule. The high predominance of airway features selected in all the models was somewhat unexpected. This warrants further investigation as it is not clear if these features are differentiated due to the presence of cancer or potentially indicative of structural airway features of populations more at risk for lung cancer development.

It was surprising that diameter was the only clinical feature selected, particularly as the features included have been utilized in previously published lung cancer risk predictions.²⁷⁻³² There is a known trend towards

larger nodules having increased risk of malignancy, so much so that size is a main factor in the work-up recommendations for screening and incidental nodules.^{27,33} However, with broader implementation of CT-based lung cancer screening, including collection of other contributing demographical and clinical factors could provide more impact for other risk factors. Also, the increased use of lung cancer screening and the associated mechanisms for structured reporting of patient data could standardize these factors further, allowing for potentially more useful information than is currently gathered.

From a clinical workflow standpoint, a fully automated tool which does not require human interaction for nodule identification and segmentation would be advantageous. With the increased use of CT for lung cancer screening and for longitudinal tracking of incidental nodules, the workload of imaging data requiring radiology reads is increased; therefore, for a tool to be useful it must be easy to implement and require as little added effort as possible. Here, we demonstrated the potential improvement in testing AUC-ROC by incorporating automated qCT features (AUC-ROC=0.74, 0.79) over solely clinical characteristics (AUC-ROC=0.64).

The limitations of this study included the retrospective collection, with a focus on solid solitary pulmonary nodules. The cohort used in this study has a size bias between malignant and benign classes, and the nodule was not excluded (segmented) from the analysis. However, the difference between the average malignant nodule diameter (mean 14.7mm) and benign nodule diameter (mean 9.11mm) is considered very small when placed in the context of whole lung structure assessment. We did not run separate analysis based on race which could in the future provide more accurate predictions as was shown previously in an African American population.³⁴ We only included inspiratory scans in the analysis, and there are known expiratory scan measures (percentage below -856HU) which could further assist in the differentiation between malignant and benign cases.

In conclusion, this work presents the potential value of COPD-related qCT features from the lung, combined with information optimization feature selection and ensemble neural net classification for distinguishing malignant from benign lesions. While performance of this developed technique was not perfect in the task of correctly identifying all lung cancer cases, it could have potential utility in reducing the clinical follow-up in individuals with benign CT-identified pulmonary nodules.

Acknowledgments

Author Contributions: JU, SLM, and JCS wrote the main manuscript text, with editing provided by JL, CND, AGS. JU, SLM, CND, AGS, and JCS contributed to data collection. JU, SLM, JL, and JCS performed data analysis and figure creation. All authors reviewed the manuscript.

Data Availability: The CT datasets utilized in this study are available from the originating clinical trials: COPD Genetic Epidemiology Study (COPDGene),¹² Inflammation, Health, and Lung Epidemiology

Study (INHALE),² and the National Lung Screening Trial (NLST).¹³ The image feature data generated during the current study are available from the corresponding author on reasonable request.

Declaration of Interest

JCS has a family member that is a shareholder and receives compensation from VIDA Diagnostics. All other authors have nothing to declare.

References

1. Wasswa-Kintu S, Gan WQ, Man SF, Pare PD, Sin DD. Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: a systematic review and meta-analysis. *Thorax*. 2005;60(7):570-575. doi: <https://doi.org/10.1136/thx.2004.037135>

2. Schwartz AG, Lusk CM, Wenzlaff AS, et al. Risk of lung cancer associated with COPD phenotype based on quantitative image analysis. *Cancer Epidemiol Biomarkers Prev*. 2016;25(9):1341-1347. doi: <https://doi.org/10.1158/1055-9965.EPI-16-0176>

3. Wilson DO, Leader JK, Fuhrman CR, Reilly JJ, Scirba FC, Weissfeld JL. Quantitative computed tomography analysis, airflow obstruction, and lung cancer in the Pittsburgh lung screening study. *J Thorac Oncol*. 2011;6(7):1200-1205. doi: <https://doi.org/10.1097/JTO.0b013e318219aa93>

4. Maldonado F, Bartholmai BJ, Swensen SJ, Midthun DE, Decker PA, Jett JR. Are airflow obstruction and radiographic evidence of emphysema risk factors for lung cancer? A nested case-control study using quantitative emphysema analysis. *Chest*. 2010;138(6):1295-1302. doi: <https://doi.org/10.1378/chest.09-2567>

5. Chubachi S, Takahashi S, Tsutsumi A, et al. Radiologic features of precancerous areas of the lungs in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*. 2017;12:1613-1624. doi: <https://doi.org/10.2147/COPD.S132709>

6. Gagnat AA, Gjerdevik M, Gallefoss F, Coxson HO, Gulsvik A, Bakke P. Incidence of non-pulmonary cancer and lung cancer by amount of emphysema and airway wall thickness: a community-based cohort. *Eur Respir J*. 2017;49(5):1601162. doi: <https://doi.org/10.1183/13993003.01162-2016>

7. Gierada DS, Guniganti P, Newman BJ, et al. Quantitative CT assessment of emphysema and airways in relation to lung cancer risk. *Radiology*. 2011;261(3):950-959. doi: <https://doi.org/10.1148/radiol.11110542>

8. Wille MM, Thomsen LH, Petersen J, et al. Visual assessment of early emphysema and interstitial abnormalities on CT is useful in lung cancer risk analysis. *Eur Radiol*. 2016;26(2):487-494. doi: <https://doi.org/10.1007/s00330-015-3826-9>

9. Bae K, Jeon KN, Lee SJ, et al. Severity of pulmonary emphysema and lung cancer: analysis using quantitative lobar emphysema scoring. *Medicine*. 2016;95(48):e5494. doi: <https://doi.org/10.1097/MD.0000000000005494>

10. Carr LL, Jacobson S, Lynch DA, et al. Features of COPD as predictors of lung cancer. *Chest*. 2018;153(6):1326-1335. doi: <https://doi.org/10.1016/j.chest.2018.01.049>

11. Johannessen A, Skorge TD, Bottai M, et al. Mortality by level of emphysema and airway wall thickness. *Am J Respir Crit Care Med*. 2013;187(6):602-608. doi: <https://doi.org/10.1164/rccm.201209-1722OC>

12. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010;7(1):32-43. doi: <https://doi.org/10.3109/15412550903499522>

13. Aberle DR, Adams AM, Berg CD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395-409. doi: <https://doi.org/10.1056/NEJMoa1102873>

14. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of COPD, 2021 report. GOLD website. Published 2021. Accessed December 2021. <https://goldcopd.org>.

15. Uthoff J, Stephens MJ, Newell JDJ, et al. Machine learning approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT. *J Med Phys*. 2019;46(7):3207-3216. doi: <https://doi.org/10.1002/mp.13592>

16. The R Foundation (R). A language and environment for statistical computing. R website. Published 2020. Accessed December 2021. <https://www.R-project.org/>

17. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. *CJEM*. 2006;8(1):19-20. doi: <https://doi.org/10.1017/S1481803500013336>

18. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: Cohen W, Moore A, eds. *Proceedings of the 23rd International Conference on Machine Learning*. Association for Computing Machinery; 2016:233-240. Accessed December 2021. <https://dl.acm.org/doi/10.1145/1143844.1143874#d3069651e1>

19. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77. doi: <https://doi.org/10.1186/1471-2105-12-77>

20. Ferreira JR, Oliveira MC, de Azevedo-Marques PM. Characterization of pulmonary nodules based on features of margin sharpness and texture. *J Digit Imaging*. 2018;31(4):451-463. doi: <https://doi.org/10.1007/s10278-017-0029-8>

21. Lee MC, Boroczky L, Sungur-Stasik K, et al. Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction. *Artif Intell Med*. 2010;50(1):43-53. doi: <https://doi.org/10.1016/j.artmed.2010.04.011>

-
22. Way TW, Sahiner B, Chan HP, et al. Computer-aided diagnosis of pulmonary nodules on CT scans: improvement of classification performance with nodule surface features. *Med Phys*. 2009;36(7):3086-3098.
doi: <https://doi.org/10.1118/1.3140589>
-
23. Zhu Y, Tan Y, Hua Y, Wang M, Zhang G, Zhang J. Feature selection and performance evaluation of support vector machine (SVM)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography. *J Digit Imaging*. 2010;23(1):51-65.
doi: <https://doi.org/10.1007/s10278-009-9185-9>
-
24. Dilger SKN, Uthoff J, Judisch A, et al. Improved pulmonary nodule classification utilizing quantitative lung parenchyma features. *J Med Imaging*. 2015;2(4):041004.
doi: <https://doi.org/10.1117/1.JMI.2.4.041004>
-
25. Huang P, Park S, Yan R, et al. Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study. *Radiology*. 2017;286(1):286-295.
doi: <https://doi.org/10.1148/radiol.2017162725>
-
26. Kalpathy-Cramer J, Mamomov A, Zhao B, et al. Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features. *Tomography*. 2016;2(4):430-437. doi: <https://doi.org/10.18383/j.tom.2016.00235>
-
27. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology*. 2017;284(1):228-243.
doi: <https://doi.org/10.1148/radiol.2017161659>
-
28. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules. Application to small radiologically indeterminate nodules. *Arch Intern Med*. 1997;157(8):849-855.
doi: <https://doi.org/10.1001/archinte.157.8.849>
-
29. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med*. 2013;369(10):910-919.
doi: <https://doi.org/10.1056/NEJMoa1214726>
-
30. Li Y, Wang J. A mathematical model for predicting malignancy of solitary pulmonary nodules. *World J Surg*. 2012;36(4):830-835.
doi: <https://doi.org/10.1007/s00268-012-1449-8>
-
31. Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest*. 2007;131(2):383-388.
doi: <https://doi.org/10.1378/chest.06-1261>
-
32. Kim H, Park CM, Goo JM, Wildberger JE, Kauczor H-U. Quantitative computed tomography imaging biomarkers in the diagnosis and management of lung cancer. *Invest Radiol*. 2015;50(9):571-583.
doi: <https://doi.org/10.1097/RLI.000000000000152>
-
33. Baldwin DR, Callister ME. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. *Thorax*. 2015;70(8):794-798.
doi: <https://doi.org/10.1136/thoraxjnl-2015-207221>
-
34. Etzel CJ, Kachroo S, Liu M, et al. Development and validation of a lung cancer risk prediction model for African Americans. *Cancer Prev Res (Phila)*. 2008;1(4):255-265.
doi: <https://doi.org/10.1158/1940-6207.CAPR-08-0082>