CT Image Parameter Estimation Using PCA-Based Deep Learning in Chronic Obstructive Pulmonary Disease

Amir Moslemi MSc,[1] Wan C. Tan MD,[2] Jean Bourbeau MD[3,4], James C. Hogg MD, PhD,[2] Harvey O. Coxson PhD[2] and Miranda Kirby PhD[1,2]

[1]Department of Physics, Ryerson University, Toronto, ON, Canada; [2]Centre for Heart Lung Innovation, University of British Columbia, Vancouver, BC, Canada; [3]Montreal Chest Institute of the Royal Victoria Hospital, McGill University Health Centre; [4]Respiratory Epidemiology and Clinical Research Unit, Research Institute of McGill University Health Centre, Canada.

Introduction

- Chronic obstructive pulmonary disease (COPD) is a debilitating and progressive lung disease caused primarily by cigarette smoking (1)
- It is now estimated that one in four will be diagnosed and receive medical attention for COPD in their lifetime (2)
- Smokers have evidence of abnormalities on computed tomography (CT), before airflow limitation can be detected and COPD is diagnosed (3)
- Although CT is not part of COPD standard-of-care, predicting at risk individuals with structural abnormalities may enable earlier detection.
- Principle component analysis (PCA)-based deep learning (4) enables dimension reduction of the feature matrix and training data efficiently.

Objective & Hypothesis

Objective

To use PCA-based deep learning to estimate CT image features in at risk smokers using only conventional clinical measurements.

Hypothesis

We hypothesize that CT image features that reflect emphysema and airway remodeling can be estimated by demographic and lung function measurements using PCA-based deep learning.

Methods

- Participants from the Canadian Cohort of Obstructive Lung Disease (CanCOLD) (5) study were evaluated. A total of n=250 at risk participants were used for testing. The remaining CanCOLD cohort were used for training.
- A total of eight clinical (age, sex, BMI, smoking status, pack years, prior hospitalization and history of bronchiectasis, heart disease, hypertension or diabetes) and nine pulmonary function test measurements (FEV_1, FVC, FEV_1/FVC, $FEF_{25\text{-}75}$, TLC, RV, FRC, RV/TLC and DL_{CO}) were investigated.
- CT measurements included: airway wall thickness (Pi10) and emphysema (low attenuation areas below -950HU - LAA_{950}) generating using VIDA Diagnostics Inc.
- PCA method for feature selection followed by Z-score normalization was used.
- A deep neural network was trained using 3-fully connected layers with "Swish" activation for hidden layers. K-fold cross validation and 20% drop out were used to tune number of epochs and neural network topology, respectively.
- Mean absolute percentage error (MAPE) is the percentage of the deviation of the predicted from true values.

Results

Table 1. Subject Demographics and Pulmonary Function Test Measurements

Parameter (±SD unless specified)	Never-smokers (n=266)	At Risk (n=365)	GOLD I (n=349)	GOLD II+ (n=266)
Age, yrs	66 (9.97)	66 (9.3)	68 (9.7)$^{*\$}$	66 (10.46)^{o}
Female Sex, n (%)	138 (51.88)	153 (41.92)*	118 (33.81)$^{*\$}$	115 (43.23)^{o}
Pack-years, yrs	0 (0)	19.26(20.47)*	18.21(22.98)*	28.32(26.36)$^{*\$o}$
BMI, kg/m^2	27.18 (5.06)	28.14 (5.15)	27.07 (4.46)$^{\$}$	27.84 (5.47)
Current-smoker, n (%)	0(0)	69 (18.9)	49 (14.04)*	66 (24.81)*
FEV_{1}, $\%_{pred}$	106 (15)	103 (14)	96 (12)	69 (8)$^{*\$o}$
FEV_1/FVC, %	77.6 (4.65)	76.78 (4.58)	64.5 (4.65)$^{*\$}$	56.93 (9.47)$^{*\$o}$

SD=standard deviation; BMI=body mass index; FEV_1=forced expiratory volume in one second; $\%_{pred}$=percent predicted; FVC=forced vital capacity. Significance of difference ($p<0.05$): * Significantly different from Never-smoker; $ Significantly different from At Risk; o Significantly different from GOLD I.

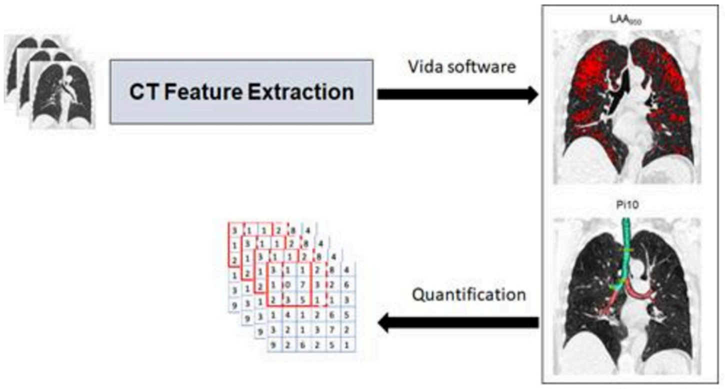

Figure 1. Feature Extraction and Quantification. The images were first analyzed using VIDA Diagnostic Inc. software to generate quantitative CT features.

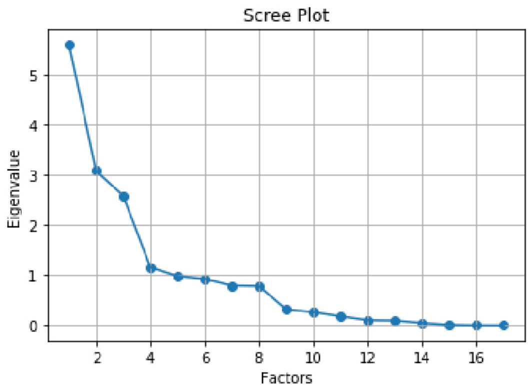

Figure 2. The plot shows the eigen-value of each principle component in the feature map. The Kaiser rule selects every principle component with eigen-values greater than one.

Table 2. The Prediction of LAA_{950} and Pi10 was Assessed by Mean Absolute Percentage Error (MAPE) and R-square.

Target	Actual value	Predicted value	MAPE (%)	R-square
LAA_{950}, %	69.8±9.3	69.5±9.0	1.1	0.84
Pi10, mm	3.97±0.15	4.00±0.1	3.7	0.51

Table 3. The Four Top Features in Terms of Eigenvalue, Variance and Cumulative Variance.

Selected feature	Eigenvalue	Variance (%)	Cumulative Variance (%)
FEV_1	5.6	20	20
FRC	3	18	38
TLC	2.6	13	51
RV	1.2	10	61

Summary of Results & Discussion

- The four top features that were selected were pulmonary function measurements: FEV_1, FRC, TLC and RV. These four features contain 61% of variance.
- The MAPE for LAA_{950} was 1.1% and for Pi10 was 3.7%. The r-square for the goodness of fit of model was 0.51 and 0.84 for Pi-10 and LAA_{950}, respectively.
  - Therefore, lung function measurements explained 84% of the variance in LAA_{950}. On the other hand, the estimation of Pi10 (airway wall thickness) explained only 51% and is therefore not predicted well with conventional clinical measurements.
- The estimated / actual LAA_{950} and Pi10 measurements were 69.5±9.0% / 69.8±9.3% and 4.00±0.15mm / 3.97±0.15mm, respectively.
- The dimensionality reduction by using PCA increases the performance of the neural network. PCA assumes that eigenvectors of the feature map are linear combinations of the feature matrix and so it may miss the nonlinear features.

Conclusions

- CT parameters can be estimated using only clinical and spirometry measurements.
- Estimation of structural abnormalities may allow at risk individuals to be stratified to receive CT imaging to confirm abnormalities, and enable more careful follow-up / management to reduce adverse health-related outcomes.

References

1. Sorano JB, Abajobir AA, Abate KH, et al. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Respir Med. Elsevier; 2017;5(9):691–706.
2. Gershon AS, Warner L, Cascagnette P, et. al. Lifetime risk of developing chronic obstructive pulmonary disease: a longitudinal population study. Lancet. 2011;378(9795):991–996.
3. Regan EA, Lynch DA, Curran-Everett D, et. al. Clinical and Radiologic Disease in Smokers With Normal Spirometry. JAMA Intern Med. 2015 Sep 1; 175(9): 1539–1549.
4. Haykin S. Neural Network and Learning Machine. 2009 by Pearson Education, Inc., Upper Saddle River, New Jersey 07458.
5. Bourbeau J, Tan WC, Benedetti, A, et. al. Canadian Cohort Obstructive Lung Disease (CanCOLD): Fulfilling the Need for Longitudinal Observational Studies in COPD. COPD. 2014 Apr;11(2):125-32.

Acknowledgements

We would like to thank all researchers, staff and participants that are part of the CanCOLD study, as well as our sponsors. This research was funded by NSERC. Dr. Kirby gratefully acknowledges support from the Parker B. Francis Fellowship Program and the Canada Research Chair Program (Tier II). Dr. Kirby is a consultant for VIDA Diagnostics Inc..