Emphysema Quantification on Cardiac CT Scans Using Hidden Markov Measure Field Model: The MESA Lung Study

Jie Yang[1], Elsa D. Angelini[1], Pallavi P. Balte[2], Eric A. Hoffman[3,4],
Colin O. Wu[5], Bharath A. Venkatesh[6], R. Graham Barr[2,7],
and Andrew F. Laine[1(✉)]

[1] Department of Biomedical Engineering, Columbia University, New York, NY, USA
laine@columbia.edu
[2] Department of Medicine, Columbia University Medical Center,
New York, NY, USA
[3] Department of Radiology, University of Iowa, Iowa City, IA, USA
[4] Department of Biomedical Engineering, University of Iowa, Iowa City, IA, USA
[5] Office of Biostatistics Research, National Heart, Lung and Blood Institute,
Bethesda, MD, USA
[6] Department of Radiology, Johns Hopkins University, Baltimore, MD, USA
[7] Department of Epidemiology, Columbia University Medical Center,
New York, NY, USA

Abstract. Cardiac computed tomography (CT) scans include approximately 2/3 of the lung and can be obtained with low radiation exposure. Large cohorts of population-based research studies reported high correlations of emphysema quantification between full-lung (FL) and cardiac CT scans, using thresholding-based measurements. This work extends a hidden Markov measure field (HMMF) model-based segmentation method for automated emphysema quantification on cardiac CT scans. We show that the HMMF-based method, when compared with several types of thresholding, provides more reproducible emphysema segmentation on repeated cardiac scans, and more consistent measurements between longitudinal cardiac and FL scans from a diverse pool of scanner types and thousands of subjects with ten thousands of scans.

1 Introduction

Pulmonary emphysema is defined by a loss of lung tissue in the absence of fibrosis, and overlaps considerably with chronic obstructive pulmonary disease (COPD). Full-lung (FL) quantitative computed tomography (CT) imaging is commonly used to measure a continuous score of the extent of emphysema-like lung tissue, which has been shown to be reproducible [1], and correlates well with respiratory symptoms [2]. Cardiac CT scans, which are commonly used for the assessment of coronary artery calcium scores to predict cardiac events [3], include about 70 % of the lung volume, and can be obtained with low radiation exposure. Despite missing apical and basal *individual* measurements, emphysema

quantification on cardiac CT were shown to have high reproducibility and high correlation with FL measures [4], and correlate well with risk factors of lung disease and mortality [5] at the *population*-based level. With the availability of large-scale well characterized cardiac CT databases such as the Multi-Ethnic Study of Atherosclerosis (MESA) [6,7], emphysema quantification on cardiac scans has now been actively used in various population-based studies [8].

However, currently used methods for emphysema quantification on cardiac scans rely on measuring the percentage of lung volume (referred to as $\%emph$) with intensity values below a fixed threshold. Although thresholding-based methods are commonly used in research, they can be very sensitive to factors leading to variation in image quality and voxel intensity distributions, including variations in scanner type, reconstruction kernel, radiation dose and slice thickness.

To study $\%emph$ on heterogeneous datasets of FL scans, density correction [9], noise filtering [10,11] and reconstruction-kernel adaptation [12] have been proposed. These approaches consider only a part of the sources of variation, and their applicability to cardiac scans has not been demonstrated. Superiority of a segmentation method based on Hidden Markov Measure Field (HMMF) to thresholding-based measures and these correction methods was demonstrated in [13,14] on FL scans. In this work, we propose to adapt the parameterization of the HMMF model to cardiac scans from 6,814 subjects in the longitudinal MESA Lung Study. Results compare HMMF and thresholding-based $\%emph$ measures for three metrics: (1) intra-cardiac scan reproducibility, (2) longitudinal correlation on "normal" subjects (never-smokers without respiratory symptoms or disease [15]), and (3) emphysema progression on "normal" and "disease" subjects.

2 Method

2.1 Data

The MESA Study consists of 6,814 subjects screened with cardiac CT scans at baseline (Exam 1, 2000–2002), and with follow-up scans in Exam 2 to 4 (2002–2008). Most subjects had two repeated cardiac scans per visit (same scanner). Among these subjects, 3,965 were enrolled in the MESA Lung Study and underwent FL scans in Exam 5 (2010–2012). Cardiac scans were collected using either an EBT scanner from GE, or six types of MDCT scanners from GE or Siemens (details in [4]). The average slice thickness is 2.82 mm, and in-plane resolution is in the range [0.44, 0.78] mm. Lung segmentation was performed with the APOLLO® software (VIDA Diagnostics, Iowa). FL scans were cropped (removing apical and basal lung) to match the cardiac scans field of view. Longitudinal correlation of lung volumes in incremental cardiac exams is in the range [0.84, 0.95]. Cardiac scans were acquired at full inspiration with cardiac and respiration gating, while FL scans were acquired at full inspiration without cardiac gating.

For this study, we selected a random subset of 10,000 pairs of repeated cardiac scans with one in each pair considered as the "best" scan in terms of inflation or scan quality [8]. Out of these 10,000 pairs, 379 pairs were discarded due to corruption in one scan during image reconstruction or storage, detected via
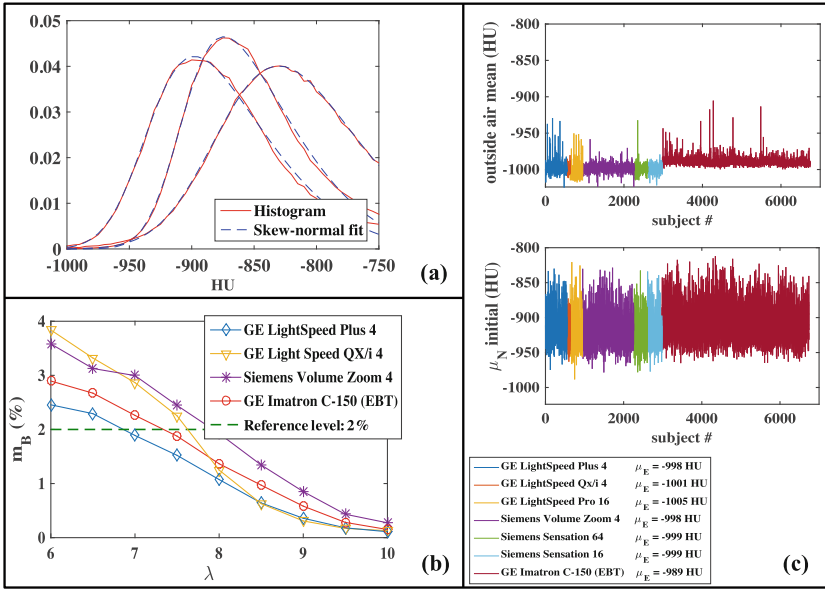
Fig. 1. (a) Illustration of fitting lung-field intensity with skew-normal distribution on three cardiac scans. (b) Population average of $\%emph_{\mathrm{HMMF}}$ ($m_B(\lambda)$) measured from normal subjects on four baseline cardiac scanners (S_B) versus λ values. (c) [top] Outside air mean value (HU) per subject and per scanner used to tune μ_E; [bottom] Initial μ_N value (HU) per subject and per scanner.

Table 1. Year and number of MESA cardiac/FL CT scans evaluated.

MESA Exam #	1	2	3	4	5
Year start-end	2000-02	2002-04	2004-05	2005-08	2010-12
# of subjects available in MESA	6,814	2,955	2,929	1,406	3,965
# of normal subjects evaluated	741	261	307	141	827
Total # of scans evaluated	6,088 $(\times 2)$	1,164 $(\times 2)$	1,645 $(\times 2)$	724 $(\times 2)$	2,984

abnormally high values of mean and standard deviation of outside air voxel intensities (cf. Fig. 1(c) for ranges of normal values). The selected subset involves 6,552 subjects, among which 2,984 subjects had a FL scan in Exam 5, and 827 are "normals", as detailed in Table 1. Overall, we processed a grand total of 9,621 pairs of repeated cardiac scans, 3,508 pairs of "best" longitudinal cardiac scans, and 5,134 pairs of "best" cardiac-FL scans.

2.2   HMMF-Based Emphysema Segmentation

The HMMF-based method [13] enforces spatial coherence of the segmentation, and relies on parametric models of intensity distributions within emphysematous and normal lung tissue. It uses a Gaussian distribution $N_E(\theta_E)$ for the emphysema class and a skew-normal distribution $N_N(\theta_N)$ for normal lung tissue. We found the skew-normal distribution model to be applicable to cardiac

scans. Figure 1(a) gives examples of histogram fitting results for three cardiac scans from normal subjects.

For a given image $I : \Omega \rightarrow \mathbb{R}$, the HMMF estimates on Ω the continuous-valued measure field $q \in [0, 1]$ by maximizing the posterior distribution P for q and the associated parameter vector $\theta = [\theta_E, \theta_N]$ expressed as:

$$P(q, \theta | I) = \frac{1}{R} P(I | q, \theta) P_q(q) P_\theta(\theta) \tag{1}$$

where R is a normalization constant. The Markov random field (MRF) variable q is a vector $q = [q_E, q_N]$, representing the intermediate labeling of both classes. Emphysema voxels are selected as $\{v \in \Omega | q_E(v) > q_N(v)\}$, from which $\%emph_{\mathrm{HMMF}}$ is computed. The distribution $P_q(q)$ enforces spatial regularity via Markovian regularization on neighborhood cliques C and involves a weight parameter λ in the potential of the Gibbs distribution. The likelihood $P(I | q, \theta)$ requires initialization of parameter values for both classes, which are tuned in this work to handle the heterogeneity of the dataset, as described below.

Parameter Tuning for Cardiac Scans. The parameters of intensity distributions are $\theta_E = [\mu_E, \sigma_E]$, $\theta_N = [\mu_N, \sigma_N, \alpha_N]$ where μ denotes the mean, σ the standard deviation and α the skewness of respective classes.

Likelihood for Normal Lung Tissue: The standard deviation σ_N and the skewness α_N are assumed to be sensitive to scanner-specific image variations. They are tuned separately for each scanner type by averaging on the subpopulation of normal subjects, after fitting their intensity histograms. The initial value of mean μ_N is sensitive to inflation level and morphology and therefore made subject-specific via fitting individual intensity histograms with the pre-fixed σ_N and α_N. Measured initial μ_N values are plotted in Fig. 1(c).

Likelihood for Emphysema Class: The initial value of mean μ_E is set to the average scanner-specific outside air mean intensity value, learned on a subpopulation of both normal and disease subjects from each scanner type, and illustrated in Fig. 1(c). The standard deviation σ_E is set to be equal to the scanner-specific σ_N since the value of σ is mainly affected by image quality.

Cliques: To handle the slice thickness change from FL (mean 0.65 mm) to cardiac CT (mean 2.82 mm), the spatial clique is set to 8-connected neighborhoods in 2-D planes instead of 26-connected 3-D cliques used in [13].

Regularization Weight λ: The regularization weight λ is made scanner-specific to adapt to image quality and noise level. We note m_X the population average of $\%emph_{\mathrm{HMMF}}$ measures on normal subjects using scanners in category X. There are three scanner categories: scanners used only at baseline (S_B), scanners used at baseline and some follow-up times (S_{BF}) and scanners used only at follow-up (S_F). For each scanner in S_B and S_{BF}, we chose, via Bootstrapping, the λ_B values that returns $m_B(\lambda_B) = 2\%$ (i.e. a small arbitrary value). The selection process is illustrated in Fig. 1(b). For scanners in S_{BF}, the same λ_B

values are used at follow-up times, leading to population $\%emph_{\mathrm{HMMF}}$ averages $m_{BF}\ (\lambda_B)$. Finally, the λ_F are chosen such that $m_{BF}(\lambda_B) = m_F(\lambda_F)$.

Parameter Tuning for FL Scans. Parameters for the segmentation of FL scans with HMMF were tuned similarly to [13], except for λ and the initial values of μ_N and μ_E. In [13], scans reconstructed with a smooth kernel were used as a reference to set λ for noisier reconstructions. In this work, having only one reconstruction per scan, we propose to use the progression rate of $\%emph$ measured on longitudinal cardiac scans from the subpopulation of normal subjects. We set $m_{FL}(\lambda_{FL}) = m_{pr}$ with m_{pr} the predicted normal population average of $\%emph$ at the time of acquisition of the FL scans, based on linear interpolation of anterior progression rates. This leads to λ_{FL} in the range [3, 3.5] for the different scanners, which is quite different from the range of λ values tuned on cardiac scans (cf. Fig. 1(b)).

2.3   Quantification via Thresholding ($\%emph_{-950}$)

Standard thresholding-based measures $\%emph_{-950}$ were obtained for comparison, using a threshold value of reference T_{ref}. Among standard values used by radiologists, $T_{ref} = -950$HU was found to generate higher intra-class correlation and lower maximal differences on a subpopulation of repeated cardiac scans.

For reproducibility testing on repeated cardiac scans (same scanner), an additional measure $\%emph_{-950G}$ was generated after Gaussian filtering, which was shown to reduce image noise-level effect in previous studies [13]. The scale parameter of the Gaussian filter is tuned in the same manner as λ for the HMMF (i.e. matching average values of $\%emph_{-950G}$ on normal subpopulations with the reference values). This leads to scale parameter values in the range [0.075, 0.175].

For longitudinal correlations, an additional measure $\%emph_{-950C}$ was computed correcting T_{ref} with respect to the scanner-dependent bias observed on mean outside air intensity values (μ_E), as: $T_{ref} = -950 + (\mu_E - (-1000))$ HU.

3   Experimental Results

3.1   Intra-Cardiac Scan Reproducibility

Intraclass Correlation (ICC) on Repeated Cardiac Scans. Scatter plots and ICC (average over Exams 1–4) of $\%emph$ in 9,621 pairs of repeat cardiac scans are shown in Fig. 2(a). All three measures show high reproducibility (ICC > 0.98). $\%emph_{-950G}$ provides minor improvement compared with $\%emph_{-950}$, which may be explained by the low noise level in MESA cardiac scans.

Spatial Overlap of Emphysema Masks on Repeated Cardiac Scans. Lung masks of repeated cardiac scans were registered with FSL [16], using a similarity transform (7 degrees of freedom). Spatial overlap of emphysema was measured
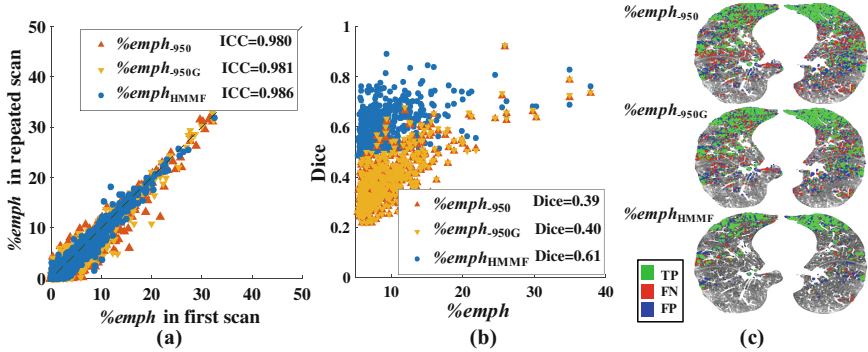
Fig. 2. Reproducibility of %*emph* measures on repeated cardiac scans: (a) Intraclass correlation (ICC) (N = 9,621); (b) Dice of emphysema mask overlap for disease subjects (N = 471); (c) Example of emphysema spatial overlap on a baseline axial slice from a pair of repeated cardiac scans (TP = true positive, FN = false negative, FP = false positive.)

with the Dice coefficient, on subjects with $\%emph_{-950} > 5\%$ (N = 471). Scatter plots and average values of Dice are reported in Fig. 2(b). Except for very few cases, HMMF returned higher overlap measures than thresholding, with an average Dice = 0.61, which is comparable to the value achieved on FL scans (0.62) [14]. Figure 2(c) gives an example of spatial overlaps of emphysema segmented on a pair of repeated cardiac scans, where there is less disagreement with HMMF.

3.2 Longitudinal Correlation and Progression of %*emph*

Pairwise Correlation on Longitudinal Cardiac Scans. For longitudinal cardiac scans, we correlated all baseline scans and follow-up scans acquired within a time interval of 48 months, in the population of normal subjects, who are expected to have little emphysema progression over time (only due to aging). Figure 3(a) shows that $\%emph_{HMMF}$ measures return the highest pair-wise correlations on longitudinal cardiac scans, followed by $\%emph_{-950C}$ measures.

Emphysema Progression. Differential %*emph* scores Δ were computed at follow-up times t to evaluate emphysema progression, as: $\Delta(t) = \%emph(t) - \%emph(\text{baseline})$. Mean values and standard errors of the mean of Δ for 87 normal subjects and 238 disease subjects who have three longitudinal cardiac scans and one FL scan are shown in Fig. 3(b).

The $\%emph_{HMMF}$ measures progressed steadily along cardiac and FL (measuring on cardiac field of view) scans, and at different rates for normal and disease populations. The $\%emph_{-950C}$ measures progressed steadily across cardiac scans but decreased from cardiac to FL scans, which indicates that a single threshold is not able to provide consistency between cardiac and FL scans. Furthermore, thresholding-based measurements on cardiac scans show similar progression rates in normal and disease populations, which is not what is expected.
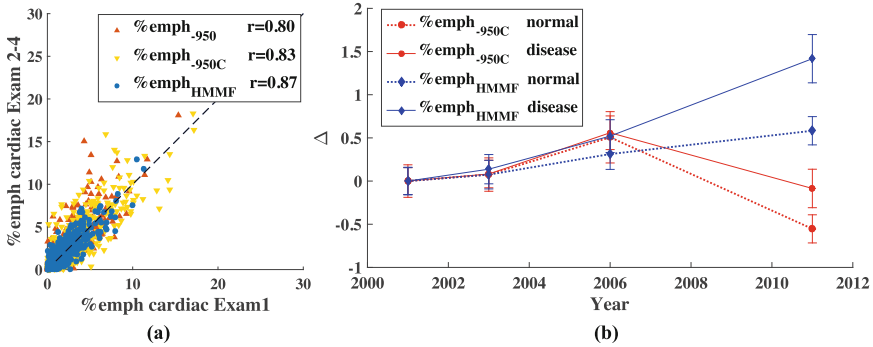
Fig. 3. (a) %*emph* measures on longitudinal cardiac scans of normal subjects (N = 478; r = pairwise Pearson correlation); (b) Mean and standard error of the mean of emphysema progression Δ (normal: N = 87, disease: N = 238).

Finally, we tested mixed linear regression models on all longitudinal scans to assess the progression of %*emph*$_{HMMF}$ and %*emph*$_{-950C}$ over time after adjusting for demographic and scanner related factors. The initial model (model 1) includes age at baseline, gender, race, height, weight, BMI, baseline smoking pack years, current cigarettes smoking per day, scanner type, and voxel size. In the subsequent model (model 2), to assess the effect modification for some demographic factors (including age at baseline, gender, race, baseline smoking pack years and current cigarettes smoking per day), their interaction terms with time (starting from the baseline) were added. In model 2 we observed that progression of %*emph*$_{HMMF}$ was higher with higher baseline age ($p = 0.0001$), baseline smoking pack years ($p < 0.0001$) and current cigarettes smoking per day ($p = 0.03$). These findings were not significant for %*emph*$_{-950C}$ except for baseline smoking pack years ($p = 0.0016$). Additionally, both models demonstrated that the effects of scanner types in cardiac scans are attenuated for %*emph*$_{HMMF}$ when compared with %*emph*$_{-950C}$.

4   Discussions and Conclusions

This study introduced a dedicated parameter tuning framework to enable the use of an automated HMMF segmentation method to quantify emphysema in a robust and reproducible manner on a large dataset of cardiac CT scans from multiple scanners. While thresholding compared well with HMMF segmentation for intraclass correlation on repeated cardiac scans, only HMMF was able to provide high spatial overlaps of emphysema segmentations on repeated cardiac scans, consistent longitudinal measures between cardiac and FL scans, attenuated scanner effects on population-wide analysis of emphysema progression rates, and clear discrimination of emphysema progression rates between normal and disease subjects. Exploiting HMMF segmentation to quantify emphysema on

low-dose cardiac CT scans has great potentials given their very large incidence in clinical routine.

References

1. Mets, O.M., De Jong, P.A., Van Ginneken, B., et al.: Quantitative computed tomography in COPD: possibilities and limitations. Lung **190**(2), 133–145 (2012)
2. Kirby, M., Pike, D., Sin, D.D., et al.: COPD: Do imaging measurements of emphysema and airway disease explain symptoms and exercise capacity? Radiology, p. 150037 (2015)
3. Detrano, R., Guerci, A.D., Carr, J.J., et al.: Coronary calcium as a predictor of coronary events in four racial or ethnic groups. NEJM **358**(13), 1336–1345 (2008)
4. Hoffman, E.A., Jiang, R., Baumhauer, H., et al.: Reproducibility and validity oflung density measures from cardiac CT scansthe multi-ethnic study ofatherosclerosis (MESA) lung study 1. Acad. Radiol. **16**(6), 689–699 (2009)
5. Oelsner, E.C., Hoffman, E.A., Folsom, A.R., et al.: Association between emphysema-like lung on cardiac CT and mortality in persons without airflow obstruction: A cohort study. Ann. Intern. Med. **161**(12), 863–873 (2014)
6. Bild, D.E., Bluemke, D.A., Burke, G.L., et al.: Multi-ethnic study of atherosclerosis: objectives and design. Am. J. Epidemiol. **156**, 871–881 (2002)
7. The MESA website. https://mesa-nhlbi.org/
8. Barr, R.G., Bluemke, D.A., Ahmed, F.S., et al.: Percent emphysema, airflow obstruction, and impaired left ventricular filling. NEJM **362**(3), 217–227 (2010)
9. Kim, S.S., Seo, J.B., Kim, N., et al.: Improved correlation between CT emphysema quantification and pulmonary function test by density correction of volumetric CT data based on air and aortic density. Eur. Radiol. **83**(1), 57–63 (2014)
10. Schilham, A.M.R., van Ginneken, B., Gietema, H., et al.: Local noise weighted filtering for emphysema scoring of low-dose CT images. IEEE TMI **25**(4), 451–463 (2006)
11. Gallardoestrella, L., Lynch, D.A., Prokop, M., et al.: Normalizing computed tomography data reconstructed with different filter kernels: effect on emphysema quantification. Eur. Radiol. **26**, 478–486 (2016)
12. Bartel, S.T., Bierhals, A.J., Pilgram, T.K., et al.: Equating quantitative emphysema measurements on different CT image reconstructions. J. Med. Phys. **38**(8), 4894–4902 (2011)
13. Hame, Y., Angelini, E.D., Hoffman, E., et al.: Adaptive quantification and longitudinal analysis of pulmonary emphysema with a hidden markov measure field model. IEEE TMI **33**(7), 1527–1540 (2014)
14. Hame, Y., Angelini, E.D., Barr, R.G., et al.: Equating emphysema scores and segmentations across CT reconstructions: a comparison study. In: ISBI 2015, pp. 629–632. IEEE
15. Hoffman, E.A., Ahmed, F.S., Baumhauer, H., et al.: Variation in the percent of emphysema-like lung in a healthy, nonsmoking multiethnic sample. the MESA lung study. Ann. Am. Thorac. Soc. **11**(6), 898–907 (2014)
16. Smith, S.M., Jenkinson, M., Woolrich, M.W., et al.: Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage **23**, S208–S219 (2004)