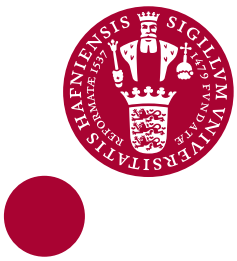

Quantifying emphysema extent from weakly labeled CT scans of the lungs using label proportions learning

Ørting, Silas Nyboe; Petersen, Jens; Wille, Mathilde; Thomsen, Laura; de Bruijne, Marleen

Published in:
The Sixth International Workshop on Pulmonary Image Analysis

Publication date:
2016

Document version
Publisher's PDF, also known as Version of record

Quantifying Emphysema Extent from Weakly Labeled CT Scans of the Lungs using Label Proportions Learning

Silas Nyboe Ørting[1], Jens Petersen[1], Mathilde M W Wille[2], Laura H. Thomsen[3], and Marleen de Bruijne[1,4]

[1] Department of Computer Science, University of Copenhagen, Denmark, silas@di.ku.dk
[2] Department of Diagnostic Imaging, Section of Radiology, Nordsjællands Hospital, Denmark
[3] Department of Respiratory Medicine, Gentofte Hospital, Denmark
[4] Department of Radiology and Medical Informatics, Erasmus MC Rotterdam, The Netherlands

Abstract. Quantification of emphysema extent is important in diagnosing and monitoring patients with chronic obstructive pulmonary disease (COPD). Several studies have shown that emphysema quantification by supervised texture classification is more robust and accurate than traditional densitometry. Current techniques require highly time consuming manual annotations of patches or use only weak labels indicating overall disease status (e.g, COPD or healthy). We show how visual scoring of regional emphysema extent can be exploited in a learning with label proportions (LLP) framework to both predict presence of emphysema in smaller patches and estimate regional extent. We evaluate performance on 195 visually scored CT scans and achieve an intraclass correlation of 0.72 (0.65–0.78) between predicted region extent and expert raters. To our knowledge this is the first time that LLP methods have been applied to medical imaging data.

1 Introduction

Emphysema is a central structural abnormality in patients suffering from chronic obstructive pulmonary disease (COPD), a leading cause of death worldwide. Emphysema is characterized by destruction of lung tissue and entrapment of air in affected regions. Quantifying emphysema extent is useful for monitoring progression [11] and in the search for genetic associations with COPD [1].

Emphysema is visible in chest CT scans and standard methods for CT-based assessment of emphysema are densitometry and visual scoring by experts. Densitometry provides an objective measure of emphysema, but is vulnerable to noise and cannot be used to distinguish emphysema sub-types. Visual scoring provide information about emphysema sub-type along with estimates of emphysema extent, but suffers from inter-observer variability and is time consuming. A recent machine learning approach used expert annotations of CT patches for

predicting emphysema sub-type and severity [2]. Region based visual scoring is less time-consuming than annotating patches and more clinically relevant [11], making it more realistic to obtain large data sets.

In this work we classify patches of CT scans by learning emphysema patterns from visual scoring of regional emphysema extent. In this type of visual scoring, the lungs are divided into six regions, the upper, middle and lower regions of the right and left lungs, and each region is assigned a percentage interval indicating the extent of emphysema in the region.

We view this learning problem as an instance of learning with label proportions (LLP). LLP is a relatively new learning setting first introduced by [6] as an extension of multiple instance learning (MIL) to proportion labels. In both MIL and LLP we are concerned with bags of instances, e.g. a collection of patches from a CT scan, and we wish to predict unknown instance labels from known bag labels. The difference between MIL and LLP is that MIL learns from binary bag labels, e.g. COPD versus no-COPD as in [3, 10] and LLP learns from proportion labels that indicate the proportion of instances in a bag with a certain label. Bag proportion labels provide more information about instance labels than binary bag labels and LLP methods attempt to use the extra information to improve performance.

Several LLP methods have been proposed, Kück and de Freitas [6] develop a graphical model where both instance labels and true bag proportions are treated as unknowns; Yu et al. [12] adapt support vector machines to LLP, and present a method for iteratively optimizing instance and bag loss; Patrini et al. [7] present Laplacian Mean Map and show that aggregate statistics can be sufficient for optimizing a large class of loss functions.

In this work we adapt cluster model selection (CMS) [9] to the problem of learning from visual scoring of emphysema. CMS searches for a clustering of patches that match known region labels. A part of the search is reshaping the feature space to improve clustering, and this feature space optimization together with the fact that no assumptions are made for the bag loss makes CMS attractive. We reformulate the CMS problem so it is straightforward to use a non-standard bag loss and contribute an interval bag loss for visually scored intervals of emphysema extent. We replace the feature weight optimization method with CMA-ES, a state-of-the-art method for black-box optimization and evaluate the method on visually scored CT scans. To our knowledge this is the first time that regional visual scoring of emphysema has been used to train a classifier, and the first time that LLP has been applied to medical image data.

2 Methods

Based on previous work by [10] and [3] for predicting COPD from CT scans, we take a texture-analysis approach to characterizing emphysema patterns. Each patch is represented by a collection of histograms of filter responses. The filters are multi-scale Gaussians and combinations of derivatives of Gaussians. A summary of the used filters is given in Table 1 and a thorough description of the

filters can be found in [10]. The filters are applied at scales $\sigma \in \{1.2, 2.4, 4.8\}$ mm, a subset of those used in [10] chosen as a compromise between feature space dimension and expressiveness.

Table 1. Multi-scale filters for analyzing lung texture. I is an image and G_σ is a Gaussian with scale σ. The asterisk $*$ indicates convolution. The Hessian is the matrix of second order partial derivatives of I, where the partial derivatives are computed by convolution with a corresponding partial derivative of a Gaussian

Feature name	Definition	Feature name	Definition
Gaussian blur	$G_\sigma * I$	Laplacian of Gaussian	$\sum_{i=1}^{3} \lambda_i$
Gradient magnitude	$\|\|\nabla G_\sigma * I\|\|$	Gaussian curvature	$\prod_{i=1}^{3} \lambda_i$
Eigenvalues of the Hessian	$\|\lambda_1\| \geq \|\lambda_2\| \geq \|\lambda_3\|$	Frobenius norm	$\sqrt{\sum_{i=1}^{3} \lambda_i^2}$

2.1 Cluster Model Selection

Cluster model selection (CMS) introduced by [9] is a machine learning method for learning from label proportions (LLP). Let $\mathcal{X}^d$ be a d-dimensional feature space, in our case it is the d filter responses, and $x \in \mathcal{X}^d$ an instance or patch. A bag $G_i \in \mathcal{X}^{l \times d}$ is a set of l patches from a lung region and $Y_G^i \in \mathcal{Y}$ is a bag label indicating the extent of emphysema in the region. Here $\mathcal{Y} = \{[I_{low}, I_{high}] | I_{low} < I_{high},\ I_{low}, I_{high} \in [0,1]\}$ is the set of closed intervals on the closed unit interval $[0,1]$. In LLP we have a set of m bags $G = \{G_1, G_2, \ldots G_m\}$ with associated bag labels $Y_G = \{Y_G^1, Y_G^2, \ldots Y_G^m\}$ and we want to predict a binary label for each patch indicating if emphysema is present.

Cluster model selection is a data-driven approach based on clustering. A cluster model in this context is a partitioning of $X = \{G_1 \cup G_2 \cup \ldots G_m\}$ into k clusters $S = \{S_1, S_2, \ldots S_k\}$ with a cluster labeling $Y_S \in \{0,1\}^k$ indicating if a cluster is an emphysema cluster. An instance $x \in S_i$ inherits the label of S_i and a bag label can be estimated as the mean instance label over all instances in the bag. The cluster model problem can be defined as

$$\arg\min_{w, \tilde{Y}_S} \frac{1}{m} \sum_{i=1}^{m} L(Y_G^i, \tilde{Y}_G^i) \ , \tag{1}$$

where $w \in [0,1]^d$ is a weighting of features and $\tilde{Y}_G$ the estimated bag labels derived from the cluster labeling $\tilde{Y}_S$. L is a bag loss function that measures the loss incurred by predicting $\tilde{Y}_G^i$ when the real bag label is Y_G^i.

Optimizing (1) is done by splitting it in smaller steps. For a given feature weight vector w we find a clustering S^w by minimizing the within-cluster distance

to the cluster center

$$S^w = \arg\min_S \sum_i^k \sum_{x \in S_i} d_P(x, \mu_i | w) \ , \tag{2}$$

where μ_i is the mean of instances in cluster S_i and d_P is a weighted patch distance defined by

$$d_P(x, y|w) = \sum_{i=1}^{d} w_i d_H(x_i, y_i) \ , \tag{3}$$

where d_H is a histogram distance function. Following [10] we use the earth movers' distance to measure histogram distance. Minimizing (2) is NP-hard and we use the k-means algorithm to find an approximate solution.

Cluster Labeling. The original CMS formulation considers real valued label proportions and uses a loss function with potentially[5] multiple "sub-optimal" global minima. The problem is that several terms are combined as a product, so if any term is zero the other terms can be arbitrarily large. While the loss function cannot distinguish between the cases where all terms are zero and one term is zero, it is unreasonable to consider the two cases equally good solutions. Here we contribute an interval bag loss more suitable for our purpose, and while it also has potential for multiple global minima, due to the interval bag labels, but all the global minima are "equally optimal" from the definition of the loss function.

For a clustering S we search for the cluster labeling that minimizes the bag loss L. Let $I = [I_{low}, I_{high}]$ be the known interval label and $p \in [0, 1]$ the predicted label. We define the bag loss

$$L(I^i, p_i) = \begin{cases} I^i_{low} - p_i & \text{if } p_i < I^i_{low} \\ p_i - I^i_{high} & \text{if } p_i > I^i_{high} \\ 0 & \text{otherwise} \end{cases} \ . \tag{4}$$

$L(I^i, p_i)$ is zero when p_i is inside the interval and equal to the shortest absolute distance from the interval otherwise.

The instances from each bag G_i are distributed over the clustering S and we define a matrix M that maps cluster labels to bag labels, such that M_{ij} is the proportion of instances from G_i that belongs to cluster S_j. This allows us to formulate the labeling problem as

$$\arg\min_{Y_S} \sum_i^m L(I^i, (MY_S)_i), \ s.t. \forall j \in [1:k]. \ 0 \leq Y^j_S \leq 1 \ . \tag{5}$$

Solving (5) is NP-hard for binary cluster labels, so we use a greedy heuristic. We start by assigning all clusters label zero, then we search for the best labeling

[5] It is potentially, because it depends on the clustering - some clusterings have a unique global minima

when only one cluster is labeled one. From a cluster labeling with i clusters labeled one, we search for the best labeling with $i+1$ clusters labeled one. The labeling is stopped when there is no longer an improvement in (5).

Feature Weight Optimization Clustering and cluster labeling is wrapped in a black-box optimization over w. The original formulation of CMS uses a simple genetic algorithm which we have replaced with state-of-the in black-box optimization, CMA-ES. Originally proposed in [5], CMA-ES is a genetic algorithm that works by generating a set of candidate weight vectors W from a multivariate Gaussian distribution with mean m and co-variance C. For each $w' \in W$ we evaluate the fitness of w' by optimizing (1) with $w = w'$. The candidate weights are then ranked and used to update m and C before a new set of candidates are generated. The process is iterated until convergence or a maximum number of iterations is reached.

3 Experiments and Results

The method is evaluated on low-dose CT scans from the Danish Lung Cancer Screening Trial [8]. Visual scoring of emphysema is performed by two raters using the method described in [11]. Each rater assigns one of seven labels to the upper, middle and lower regions of each lung. The labels {0%, 1–5%, 6–25%, 26–50%, 51–75%, 76–100%} indicate the percentage of the region affected by emphysema. Three data sets have been defined A_{train}, A_{validate}, A_{test} with respective sizes of 193, 195, 195 scans. Each data set was initially 200 scans, matching the data sets defined in [10], but some scans were excluded because they were not visually scored. A set of 50 patches with a size of approximately $21 \times 21 \times 21\text{mm}^3$ were sampled from each region of the lungs and aggregated into bags. Emphysema is commonly characterized by the appearance of tissue destruction in lobules, which are about 10–25mm in diameter [4], and the patch size has been chosen to approximately match the size of lobules. Each bag was labeled by combining the extent of both raters, such that the combined interval is the smallest interval containing the interval of both raters. We assume that the extent labels can be interpreted as the proportion of patches containing emphysema.

Model training is a two-step procedure, in the first step we train several models on A_{train} and use predictions on A_{validate} to choose parameters. In the second step we train on A_{train} combined with A_{validate} using the selected parameters and use predictions on A_{test} to estimate the performance of the model.

Choosing Parameters. A separate classifier was trained on each of the six regions and the number of clusters was set to $k = [5, 10, 15, 20, 25, 30]$ for each classifier, giving a total of 36 models. The performance of each model was estimated on A_{validate} by calculating mean absolute error (MAE) from the reference intervals and intraclass correlation (ICC). To calculate ICC we converted CMS predictions to interval midpoints and used the average interval midpoint of the

raters. MAE stabilized around 0.01 for all regions for $k \geq 20$. ICC was highest in the upper regions and values for the right and left upper regions are given in Table 2. ICC was poor in the lower (≤ 0.24) and middle (≤ 0.31) regions for all values of k. Prevalence of emphysema and rater agreement is generally highest in the upper regions (Average prevalence in upper, middle, lower: 26%, 20%, 12%. Average ICC in upper, middle, lower: 0.81, 0.65, 0.51). We focus on the upper regions in the following analysis because learning from the lower prevalence and rater agreement in the middle and lower regions appear to be a much harder problem, which we leave for future work.

Table 2. Intraclass correlation for parameter selection. The best values are shown along with the number of clusters in the model. ICC is calculated with a two-way model and measures consistency. Avg refers to the average of R1 and R2

Region	Number of clusters	Raters	ICC (CI)
Right upper	20	R1/R2	0.83 (0.79–0.87)
		Avg/CMS	0.62 (0.53–0.70)
Left upper	15	R1/R2	0.78 (0.72–0.83)
		Avg/CMS	0.53 (0.43–0.63)

Region Prediction. We use the selected parameters to train two new models on the combined data $A_{\text{combined}} = A_{\text{train}} \cup A_{\text{validate}}$. Performance of the four models, two trained on A_{train} and two trained on A_{combined}, is evaluated on A_{test} by calculating ICC, using the same procedure for converting predictions as for parameter selection. Performance scores are summarized in Table 3, and we see that ICC in the upper right region improves when training on the larger data set, while ICC decrease in the upper left region. We also note that performance in upper left on A_{test} is much lower than on A_{validate} indicating overfitting in the parameter selection.

Reduced data set for training A potential issue when applying CMS to this data is that the proportion of non-emphysema bags is large ($> 70\%$) and only very few bags have a label proportion larger than 25%. This gives a highly skewed data set where less than 10% of instances contain emphysema. It is possible that the skewed data makes it difficult to identify emphysema clusters because all clusters will contain mostly non-emphysema instances.

To investigate this hypothesis we re-run the above experiment, but use only bags with emphysema for training. This gives a less skewed data set, but the proportion of emphysema instances is still less than 25% . First we train on a reduced version of A_{train} and use performance on the full version of A_{validate} for parameter selection. Then we train a new model using the selected parameter

Table 3. Agreement between raters and model predictions on A_{test}. 95% confidence intervals are shown for ICC. ICC measures consistency and is calculated with a two-way model

Region	Raters	ICC on A_{test}	
		A_{train}	A_{combined}
Right upper	R1/R2	0.82 (0.76–0.86)	0.82 (0.76–0.86)
	R1/CMS	0.67 (0.59–0.74)	0.71 (0.63–0.77)
	R2/CMS	0.54 (0.44–0.64)	0.58 (0.48–0.66)
	Avg/CMS	0.64 (0.55–0.71)	0.67 (0.59–0.74)
Left upper	R1/R2	0.81 (0.75–0.85)	0.81 (0.75–0.85)
	R1/CMS	0.38 (0.25–0.49)	0.38 (0.25–0.49)
	R2/CMS	0.37 (0.24–0.49)	0.31 (0.18–0.43)
	Avg/CMS	0.40 (0.27–0.51)	0.36 (0.23–0.48)

on the reduced version of A_{combined} and measure performance on the full version of A_{test}.

Performance on A_{validate} is summarized in table 4 where we again see best performance in the upper right region. Performance on A_{test} is summarized in table 5 and again we see indication of overfitting in the parameter selection. Training on the reduced A_{combined} result in large improvements over training on the reduced A_{train}, beating performance when training on the full data.

Table 4. Intraclass correlation for parameter selection using reduced training data. Best values are shown with the number of clusters in the model. ICC is calculated with a two-way model and measures consistency. Avg refers to the average of R1 and R2

Region	Number of clusters	Raters	ICC (CI)
Right upper	30	R1/R2	0.83 (0.79–0.87)
		Avg/CMS	0.73 (0.65–0.79)
Left upper	20	R1/R2	0.78 (0.72–0.83)
		Avg/CMS	0.56 (0.46–0.65)

It is interesting to note that ICC between CMS and Avg is larger than ICC between CMS and any of the raters when training on the reduced data. Training on the full data shows highest ICC between CMS and R1 in three out of four cases. Estimates from R1 is generally a bit lower than from R2, so underestimating emphysema should give a better ICC with R1 than with R2 and Avg. This indicates that training on the reduced data overcomes a problem of underestimation present when training on the full data.

Table 5. Agreement between raters and model predictions on A_{test} using reduced training data. 95% confidence intervals are shown for ICC. ICC measures consistency and is calculated with a two-way model

Region	Raters	ICC on A_{test}	
		A_{train}	A_{combined}
Right upper	R1/R2	0.82 (0.76–0.86)	0.82 (0.76–0.86)
	R1/CMS	0.61 (0.52–0.69)	0.68 (0.60–0.75)
	R2/CMS	0.64 (0.55–0.71)	0.69 (0.61–0.75)
	Avg/CMS	0.66 (0.57–0.73)	0.72 (0.65–0.78)
Left upper	R1/R2	0.81 (0.75–0.85)	0.81 (0.75–0.85)
	R1/CMS	0.45 (0.33–0.56)	0.59 (0.49–0.67)
	R2/CMS	0.45 (0.33–0.55)	0.60 (0.50–0.68)
	Avg/CMS	0.47 (0.36–0.58)	0.63 (0.53–0.70)

Patch Prediction. We inspected patch predictions visually. Figure 1 shows slices and patch predictions for two subjects. Top row shows a case where raters and prediction agree and bottom row shows a case where prediction is larger than raters. In the case with agreement we see that patches classified as not emphysema contain little to no emphysema, while patches classified as emphysema contain large areas with clear tissue destruction. It appears that emphysema patches are in an area with a large degree of paraseptal emphysema, while not-emphysema patches are in an area with a small degree of centrilobular emphysema. In the case of larger predicted extent it appears that there is a small decrease in density in the upper part of the region compared to the lower part. The patches predicted as emphysema are in the upper part and appear to contain some tissue destruction.

4 Discussion and Conclusion

The agreement in the upper right region shows that CMS can estimate emphysema extent, which is clinically more relevant than predicting COPD presence considered in [10] and [3].

The performance improvement when training only on emphysema bags indicates that subsampling training data to achieve a more balanced data set is beneficial for CMS. The tendency to overfit, suggested by the performance decrease from A_{validate} to A_{test}, indicate that removing all non-emphysema bags is detrimental to performance. Future work could investigate how to determine the optimal mix of bags. It is possible that performance in the middle and lower regions could be improved in the same manner, but the very low prevalence in the lower regions could result in overfitting because the amount of training data is too small to be representative of the full data set. Another approach is to train on data from several regions, either by combining a couple of regions or using all six regions.

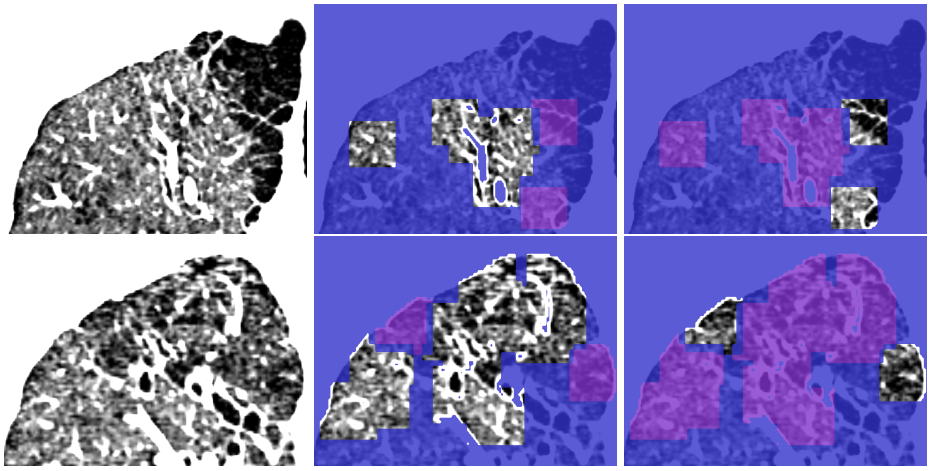

Fig. 1. Patch prediction in upper right region. Top: Rated as 26-50% extent, predicted as 26% extent. Bottom: Rated as 0% extent, predicted as 10% extent. Left: Intensity rescaled coronal slice. Center: Blue regions are not labeled. Purple patches are labeled as emphysema and non-colored patches as not emphysema. Right: Blue regions are not labeled. Purple patches are labeled as not emphysema and non-colored patches as emphysema

The increased performance when training on A_{combined} versus training on A_{train} indicates that improving performance could be a matter of increasing the amount of training data. However increasing the amount of training runs counter to one of the primary objectives of weak label learning, that of reducing the burden of labeling training data, and future work should consider the trade off between labeling burden and performance.

The inspected patch predictions show that patches with severe emphysema are likely to be labeled emphysema, while regions with mild emphysema tend to be labeled not emphysema. It is unlikely that we can account for the heterogeneity of emphysema with binary patch labels alone, and an alternative is to assign continuous labels indicating emphysema extent in the patch. This would allow us to rank patches and could be interesting as a tool for studying progression of emphysema. An interesting possibility suggested by the patch predictions for the region assessed as having 0% emphysema, is that the method is more sensitive to some mild cases of emphysema than the raters. If this is true, the approach could become a valuable tool for early detection of emphysema.

In this work we have focused on predicting emphysema extent without considering emphysema sub-type. Sub-type information is clinically interesting and a model that simultaneously predicts extent and sub-type is a future goal. Emphysema sub-types appear differently in CT scans, centrilobular emphysema is diffuse with small holes spread out over the affected area and paraseptal emphysema is more clearly defined with large bounded regions of complete tissue destruction. Simply extending cluster labels to {no-emphysema, centrilobular,

paraseptal} could improve performance, because it is likely that some patches with centrilobular emphysema are more similar to patches without emphysema than to patches with paraseptal emphysema.

There is, to our knowledge, no previous work that attempts to learn from the kind of visual assessment we consider here. The patch-based classifier from [2] uses a different labeling scheme with six classes (three severities of centrilobular, one panlobular, one pleural-based and one non-emphysema), and the evaluation metrics are also different making it difficult to compare. It appears that the biggest problem for [2] is distinguishing mild and severe cases of centrilobular emphysema. This suggests that replacing binary labels indicating presence with categorical labels indicating severity might not be enough to model emphysema severity and a continuous severity score could be the way forward.

In conclusion, we show that visual scoring of emphysema extent in regions can be used for training an LLP method to predict both region extent and presence of emphysema in patches. The results also show that predictions correlate poorly with raters when training on data where emphysema prevalence is very low and rater agreement is low to moderate.

References

1. Castaldi, P.J., Cho, M.H., San José Estépar, R., McDonald, M.L.N., Laird, N., Beaty, T.H., Washko, G., Crapo, J.D., Silverman, E.K.: Genome-wide association identifies regulatory Loci associated with distinct local histogram emphysema patterns. AM J RESP CRIT CARE 190(4), 399–409 (2014)
2. Castaldi, P.J., San José Estépar, R., Mendoza, C.S., Hersh, C.P., Laird, N., Crapo, J.D., Lynch, D.A., Silverman, E.K., Washko, G.R.: Distinct quantitative computed tomography emphysema patterns are associated with physiology and function in smokers. AM J RESP CRIT CARE 188(9), 1083–1090 (Aug 2013)
3. Cheplygina, V., Sørensen, L., Tax, D., Pedersen, J., Loog, M., de Bruijne, M.: Classification of COPD with Multiple Instance Learning. INT C PATT RECOG. pp. 1508–1513 (Aug 2014)
4. Hansell, D.M., Bankier, A.A., MacMahon, H., McLoud, T.C., Müller, N.L., Remy, J.: Fleischner society: Glossary of terms for thoracic imaging. Radiology 246(3), 697–722 (2008)
5. Hansen, N., Ostermeier, A.: Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In: IEEE C EVOL COMPUTAT on. pp. 312–317. IEEE (1996)
6. Kuck, H., de Freitas, N.: Learning about individuals from group statistics. UNCERTAIN ARTIF INTELL. pp. 332–339. AUAI Press, Arlington, Virginia (2005)
7. Patrini, G., Nock, R., Caetano, T., Rivera, P.: (almost) no label no cry. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) ADV NEUR IN 27, pp. 190–198. Curran Associates, Inc. (2014)
8. Pedersen, J.H., Ashraf, H., Dirksen, A., Bach, K., Hansen, H., Toennesen, P., Thorsen, H., Brodersen, J., Skov, B.G., Døssing, M., Mortensen, J., Richter, K., Clementsen, P., Seersholm, N.: The Danish randomized lung cancer CT screening trial–overall design and results of the prevalence round. J THORAC ONCOL 4(5) (2009)

9. Stolpe, M., Morik, K.: Learning from label proportions by optimizing cluster model selection. LECT NOTES ARTIF INT, vol. 6913, pp. 349–364. Springer Berlin Heidelberg (2011)
10. Sørensen, L., Nielsen, M., Lo, P., Ashraf, H., Pedersen, J., de Bruijne, M.: Texture-based analysis of COPD: A data-driven approach. IEEE T MED IMAGING 31(1), 70–78 (Jan 2012)
11. Wille, M.M., Thomsen, L.H., Dirksen, A., Petersen, J., Pedersen, J.H., Shaker, S.B.: Emphysema progression is visually detectable in low-dose CT in continuous but not in former smokers. Eur Radiol 24(11), 2692–2699 (Nov 2014)
12. Yu, F.X., Liu, D., Kumar, S., Jebara, T., Chang, S.: $\propto$SVM for learning with label proportions. CoRR abs/1306.0886 (2013), http://arxiv.org/abs/1306.0886