Making Sense of Large Data Sets without Annotations: Analyzing Age–related Correlations from Lung CT Scans

Yashin Dicente Cid[a b], Artem Mamonov[c], Andrew Beers[c], Armin Thomas[c], Vassili Kovalev[d], Jayashree Kalpathy–Cramer[c], and Henning Müller[a b c]

[a]University of Applied Sciences Western Switzerland (HES–SO), Sierre, Switzerland;
[b]University of Geneva, Switzerland;
[c]Martinos Center for Biomedical Imaging, Charlestown, MA, USA;
[d]National Academy of Sciences, Minsk, Belarus

ABSTRACT

The analysis of large data sets can help to gain knowledge about specific organs or on specific diseases, just as big data analysis does in many non–medical areas. This article aims to gain information from 3D volumes, so the visual content of lung CT scans of a large number of patients. In the case of the described data set, only little annotation is available on the patients that were all part of an ongoing screening program and besides age and gender no information on the patient and the findings was available for this work. This is a scenario that can happen regularly as image data sets are produced and become available in increasingly large quantities but manual annotations are often not available and also clinical data such as text reports are often harder to share. We extracted a set of visual features from 12,414 CT scans of 9,348 patients that had CT scans of the lung taken in the context of a national lung screening program in Belarus. Lung fields were segmented by two segmentation algorithms and only cases where both algorithms were able to find left and right lung and had a Dice coefficient above 0.95 were analyzed. This assures that only segmentations of good quality were used to extract features of the lung. Patients ranged in age from 0 to 106 years. Data analysis shows that age can be predicted with a fairly high accuracy for persons under 15 years. Relatively good results were also obtained between 30 and 65 years where a steady trend is seen. For young adults and older people the results are not as good as variability is very high in these groups. Several visualizations of the data show the evolution patters of the lung texture, size and density with age. The experiments allow learning the evolution of the lung and the gained results show that even with limited meta–data we can extract interesting information from large–scale visual data. These age–related changes (for example of the lung volume, the density histogram of the tissue) can also be taken into account for the interpretation of new cases. The database used includes patients that had suspicions on a chest X–ray, so it is not a group of healthy people, and only tendencies and not a model of a healthy lung at a specific age can be derived.

Keywords: Big data, Lung segmentation, Lung tissue analysis

1. INTRODUCTION

Medical imaging plays an important role in diagnosis and treatment planning. The number of images produced in hospitals has increased strongly and actually covers a very large part of world storage.[1] Image sharing for research has been strongly encouraged[2] and NIH–funded (National Institutes of Health) projects are usually required to release research data after the end of projects. Many of the images related to cancer are for example made available in the TCIA* (The Cancer Imaging Archive) or the TCGA (The Cancer Genome Atlas).[3] One of the problems with many data sets is that there is often limited annotation available, both in terms of regions of interest marked in the images but also in terms of available clinical meta–data, as acquiring these is expensive. Often, the data sets can only be used in a very specific way and can not be transferred to other areas of applications. Another problem with data sets is that imaging modalities change over time and also the used protocols in clinical routine are regularly adapted. This means that older data sets might be of interest scientifically but besides main concepts the clinical usefulness can be limited.

Further author information: Send correspondence to Y.D.C. Email: yashin.dicente@hevs.ch

*http://www.cancerimagingarchive.net/ as of November 28 2016

Annotated data sets and scientific challenges have changed research both for the non–medical[4] and also for the medical field,[5] as such challenges really allow to identify the best techniques on the same data and same tasks, making quality of approaches really comparable. A problem with manual annotations of image data is that these annotations are expensive to obtain and thus usually do not scale well for large data sets. They are often only available for a small subset of the data. Crowdsourcing can help with the problem of the cost of manual annotations and it has increasingly been used for medical imaging,[6–8] as well. This can lower the costs but requires very well defined tasks and also strict quality control.

Most existing data sets and scientific challenges concentrate on a single organ, for example the liver[5] or the brain.[9] Such challenges have had a major impact in medical imaging and an increasingly large number of challenges has been proposed. One of the few data sets available with region–based annotations for several organs was made available in the context of the VISCERAL project.[10] The fact that 20 organs were annotated in four modalities (MR and CT with and without contrast agent) limited the total number of cases per modality and organ to around 60 cases. A silver corpus could be created since a cloud–based evaluation was used and the executables of the participating research groups were available to be run on more data sets. This silver corpus is based on a label fusion of the algorithms.[11] This scales also to extremely large data sets and limits the cost of manual annotations. Comparisons showed that the quality of this annotation was high.

Lung CT images have been the target for analysis and decision support for a long period of time both for nodule detection[12] and also for the diagnosis of interstitial lung diseases.[13] Many techniques have developed texture descriptors to characterize lung tissue but only rarely clinical data such as age are also taken into account.[14, 15] To the best of our knowledge no system analyzes the deviation from a healthy norm. Databases of lung CT images for both nodule detection[16] and interstitial lung diseases[17] are available. For all these approaches it could be interesting to analyze a reference of a healthy lung of the same age to analyze deviation from this norm.

In,[18] a model for a healthy organ and person was proposed at a specific age based on a study of a cohort of patients. Such a model of healthy patients at a specific age can look at the deviation from the norm and if longitudinal data are available then also the evolution of this deviation can be analyzed to develop patient–specific trajectories. Other studies built age–related models for brain volumetry[19] and Alzheimer diagnosis.[20] The prediction of the age was also attempted using hand MRI images and applying deep learning in.[21]

The objective of this article is to explore a data set of CT images with a large distribution of age ranges, something rare for lung CTs, as usually such studies include mainly older persons, for example for lung screening of risk patients as in the NLST (National Lung Screening Trial) database[†]. Visual features linked to texture, lung volume, shape and grey level distribution were analyzed and correlated to age and gender, the only two meta–data that were available for this study. Prediction of age only from the CT scan is a secondary objective and is described in the text. Overarching purpose of this article is to explore the visual content of large data sets and what we can learn from such data sets even when limited meta–data are available. In this case the influence of aging on a set of visual features of the image data is explored.

This article is organized as follows: section 2 describes the data set used for the study and the main techniques used to analyze the data. Section 3 shows the results of the experiments and tries to visualize the obtained result to make the main outcomes easy to understand. The article finishes with a discussion and then a conclusions section.

2. METHODS

This section describes the details concerning the data set and the techniques employed to carry out the experiments.

[†] `https://www.cancer.gov/types/lung/research/nlst` as of November 28 2016

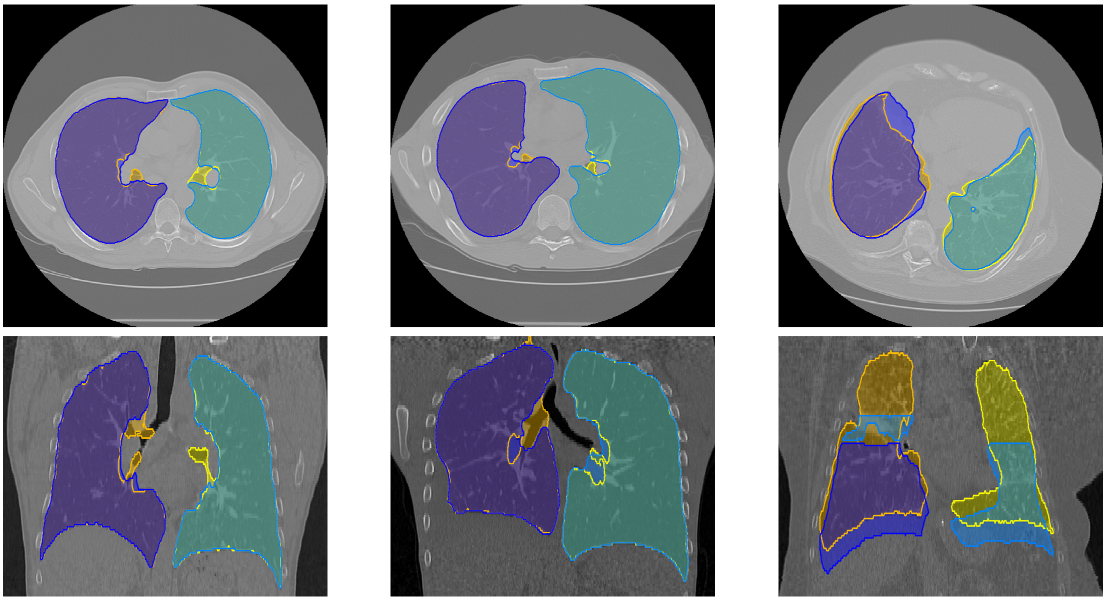


Figure 1: Three examples of the lung segmentation obtained by the two segmentation methods (axial and coronal views). The CIP segmentation is colored in dark blue (right lung) and light blue (left lung) and the DBC segmentation in orange (right lung) and yellow (left lung). When comparing both segmentations, examples (a) and (b) had a Dice coefficient of 0.9920 and 0.9943 respectively. On the other hand, (c) had a Dice coefficient of 0.6788 and was thus discarded from the final database.

2.1 Database

The database contains 14,285 volumetric CT series of 9,406 patients stored in 1,468,012 DICOM files. Images are part of a national lung screening program in Belarus, where lung cancer rates have been high since the Chernobyl disaster. They include a subset of consecutive cases referred to the Scientific and Practical Center for Pulmonology and Tuberculosis, Minsk, Belarus over a period of 8 years. The use of the data for research is approved by the institutional ethics commission. Imaging was performed using a LightSpeed Pro 16 scanner from General Electric with typical slice thickness of 2.5 mm but with some variations. Screening is usually first done with chest X–rays and persons with abnormal chest X–rays are getting a CT exam. Thus, the patients are generally not healthy patients but do have some abnormal patterns in the chest X–ray. The age of the patients at acquisition time ranges from 0 to 106. A subset of the data with tuberculosis patients are available from and can be queried at[‡]. The number of slices per series varied between 1 and 688.

Preparation of the database For some patients several image series (original, primary, derived, etc.) were available. We decided on the following rules for inclusion of series into our study to avoid some redundant information and comparable complete lung volumes: The series has to contain images of the type original and has to contain at least 30 slices. The threshold on the number of slices takes into account that there are young patients with smaller volumes but also that we want to have only full lung volumes if possible. These limits were based on the analysis of a subset of the data. After this preparation the database consists of 9,348 patients and 12,414 series.

2.2 Data analysis and age prediction

2.2.1 Lung segmentation

Two lung segmentation algorithms were used on the 12,414 series. The algorithm developped by Dicente et al. in[22] achieved the best lung segmentation performance in the VISCERAL Challenge[23] and was thus chosen.

[‡] `http://tuberculosis.by/`, `http://imlab.grid.by/` as of November 28 2016

This algorithm applies a binary clustering on the voxels of the CT. The two classes correspond to the two peaks of the intensity histogram of the full CT volume. This clustering produces a binary image based on the density of the material in each voxel. The algorithm then segments the largest non–dense 3D–connected region corresponding to the lungs. The final segmentation is refined by morphological operations. We refer to this technique as *Density–based Binary Clustering* (DBC) segmentation in the following subsections. The second algorithm is part of the Chest Imaging Platform (CIP)[24] of 3D Slicer[§] and is commonly used in lung analysis. Both segmentations provide *right* and *left* lung labels and are available as open source, so results can easily be reproduced.

2.2.2 Subset of volumes analyzed

For the final selection of lung volumes we chose series where both segmentations provided left and right lung to compare features on both lungs separately. In several cases such a separation was not done by the CIP. We also only took lung volumes where the Dice coefficient between the two segmentations was at least 0.95 in each lung (to avoid extreme cases and poor lung segmentations), and the full lung volume was covered in the CT volume. This step removes cases with mistakes, that are incomplete and extreme cases, for example cases with a single lung. Figure 1 shows three examples with the masks obtained from both methods. Example (c) corresponds to a discarded case where CIP did not read the CT volume correctly. To ensure the full lung volume is present, we checked whether any of the segmentations contained more than 0.5 % of the pixels in the superior or inferior slice of the CT volume. This was also set heuristically and only a limited number of volumes were excluded in this way. Two series were finally excluded because they presented an unusual lung segmentation volume of more than 27 liters, and two other series were removed for not having the gender specified. It is also possible that the remaining volumes contain a few mistakes in the data, so age or gender.

After this collection cleaning, the analyzed data set contains 6,395 patients and 8,034 series, with 3,491 males (54.59 %) and 2,904 females (45.41 %). The age range was still from 0 to 106 years. Figure 2 shows the distribution of the number of patients with respect to the age, so very few patients above 85 years and a smaller number of patients under 20 years with a peak population between 45 and 60 years. Considering this distribution, all patients above 85 years were considered to belong to a single class (85+) in the following sections, as otherwise too few cases are present for any proper visualization or prediction.

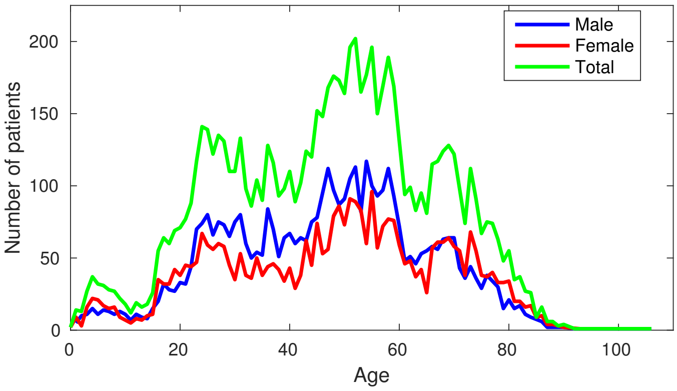


Figure 2: Distribution of the number of patients with respect to age.

2.2.3 Visual characteristics extracted from the images

Several basic visual characteristics were computed on the final data set for each segmentation algorithm. The visual features include texture, volume, shape and grey level distribution. The volume of each lung was calculated in liters and the ratio between right and left lung volume was also used as a characteristic. For describing the grey

[§] `https://www.slicer.org/` as of November 28 2016 level distribution, the mean, standard deviation, skewness, and kurtosis of the Houndsfield Unit (HU) histogram inside each lung were computed. A limited grey level histogram of the HUs in each lung was calculated using the following bins: [-1000,-950) characterizing emphysema, [-950, -650) and [-650, -300) characterizing generally healthy tissue, [-300, 0) to detect the consolidation peak, [0, 100) to identify vessels, and [100, 600) for soft tissue.

Texture features based on Gray Level Co–occurrence Matrices (GLCMs)[25] were used, such as contrast, correlation, energy and homogeneity with four directions averaged and a displacement of a single pixel. They were computed for each 2D–slice over the area segmented as lung and averaged across each full lung volume.

2.2.4 B–spline approximation

We chose the DBC segmentation to analyze the shape of the features extracted along the age. In the case of the volume of the lungs we divided the data set by gender and right/left lung. We considered the value of the other features aggregated on both lungs and did not differentiate by gender.

For each feature we computed the B–spline curve that best fits the distribution using the weighted least square approximation as follows: For a feature F, let us consider $y_{i,p}$ the value of F for a patient p in a given age x_i. Patients with ages above 85 were collapsed into one age category, (85+), and thus $x_i \in [0, 85]$. Let n_i be the number of values $y_{i,p}$ for a given age x_i (that is equivalent to the number of patients with age x_i) and let $\mathcal{Y}_i$ be the distribution of $y_{i,p}$. We define $\bar{y}_i$ as the mean of the values $y_{i,p}$ inside the first and third quartile ($q_{1,i}$ and $q_{3,i}$) of $\mathcal{Y}_i$. Then, for a given degree d, we find the B–spline S_d that minimizes the following error:

$$E_d = \sum_i w_i(\bar{y}_i - S_d(x_i))^2 \tag{1}$$

where the weight w_i is defined as $w_i = |n_i \frac{\bar{y}_i}{(q_{1,i} - q_{3,i})}|$. w_i encodes the sparsity of the data $y_{i,p}$ for a given x_i. If the difference between $q_{1,i}$ and $q_{3,i}$ with respect to $\bar{y}_i$ is small, and the number of patients with age x_i is high, this means that the data $y_{i,p}$ are strongly centered towards $\bar{y}_i$, i.e. $\bar{y}_i$ is a good estimator of $y_{i,p}$. In this case, w_i is high, forcing the B–spline S_d to be close to $\bar{y}_i$.

We increased the degree d until $|\frac{E_{d-1} - E_d}{E_{d-1}}| < 1\%$. The later was considered the degree of the best B–spline regression curve S. Finally, we added to our analysis the absolute correlation coefficient of S with respect to the age, and the mean relative error of S with respect to $\bar{y}_i$. This is:

$$ME_r = mean\left(\sum_i \left|\frac{\bar{y}_i - S(x_i)}{\bar{y}_i}\right|\right) \tag{2}$$

2.2.5 Machine learning for age prediction

A Random Forest Regressor was used on the final data set. The 120 features were regressed against age. Feature reduction was performed via inspection of correlation matrices to reduce the data set to 35 features without a change in performance.

The cleaned data set of 8,034 scans was divided by gender and then further divided into a training and a test data set. Training and test sets have equal number of patients at every age level. The separation into training and test set is random. Some patients had multiple scans, so training and testing data sets keep each patients' scans entirely in one group or the other. Again, ages above 85 were collapsed into one age category, (85+). Training and test sets had approximately 1,800 and 2,200 scans each by the end of this process for the male and female group respectively. The random forest algorithm was run 200 times with variable numbers of sub–trees (10 times for each number of sub–trees, ranging from 1 to 20).

3. EXPERIMENTAL RESULTS

In terms of lung segmentation, the CIP algorithm in its standard form had frequent problems separating left and right lung in 3,839 cases. An initial analysis showed that CIP did not identify each lung when the left and right lung were connected by the parenchyma.

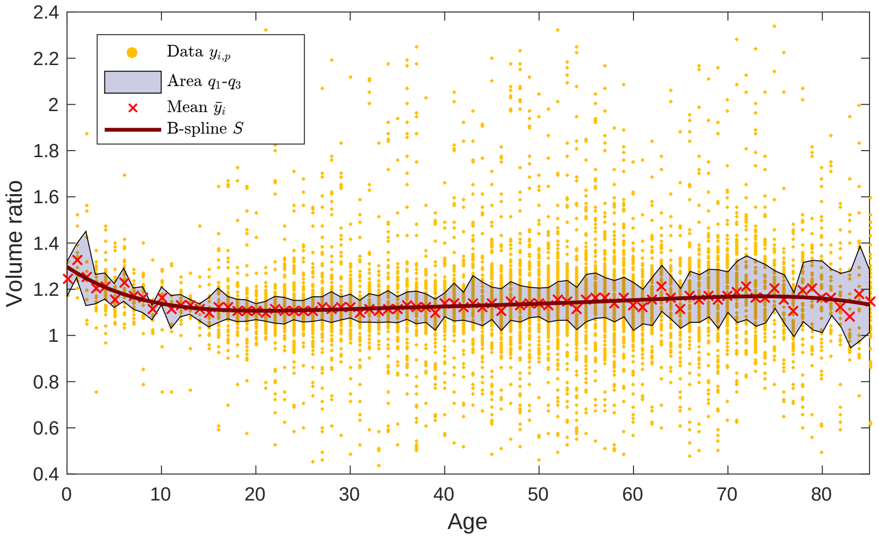


Figure 3: B–spline generation for the feature measuring the ratio between right and left lung.

Figure 3 shows the elements involved in the generation of the B–spline regression curve explained in Section 2.2.4 for the feature measuring the volume ratio between right and left lung. These elements are: the data regarding the feature, the area and the mean between the first and the third quartiles, and the B–spline regression curve generated. In the following figures, only the the mean values $\bar{y}_i$ and the regression curve S are shown.

Figure 4 shows the evolution of the volume of right and left lung for males and females. Each lung presents a similar curve increasing until an age of 20 years, then remaining stable until approximately 60 years of age and decreasing afterwards. As expected, male lungs are larger than female and right lungs are bigger than left lungs.

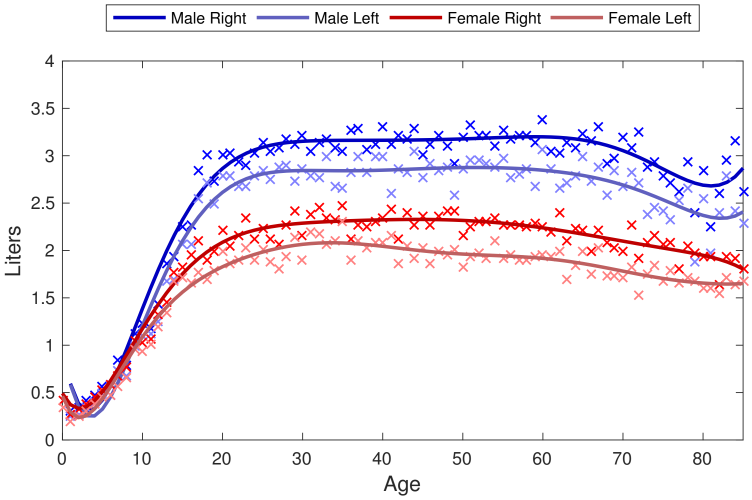


Figure 4: Evolution of the average lung volume with respect to the age.

In the case of the GLCM features presented in Figure 5, the homogeneity behaves stable with respect to the age, showing a slightly negative slope. On the other hand, the contrast shows a more unstable behavior. Between 25 and 80 years of age the contrast presents generally higher values than outside this range. The correlation feature

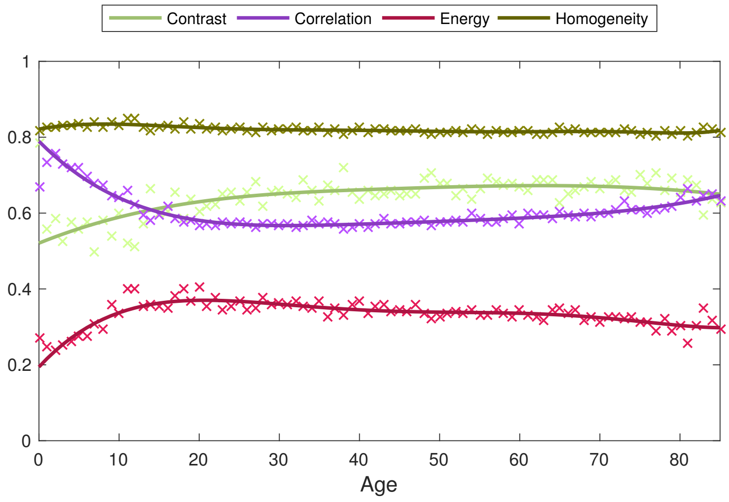


Figure 5: Evolution of the different GLCM features with respect to the age.

has a clear negative slope from 2 to 14 years old. After 60 this slope becomes positive and from 14 to 60, the tendency is almost flat. Finally, the energy presents trends for almost every age. Between 0 and 20 it shows a positive slope with respect to age, while the slope is negative after 20 and until 85.

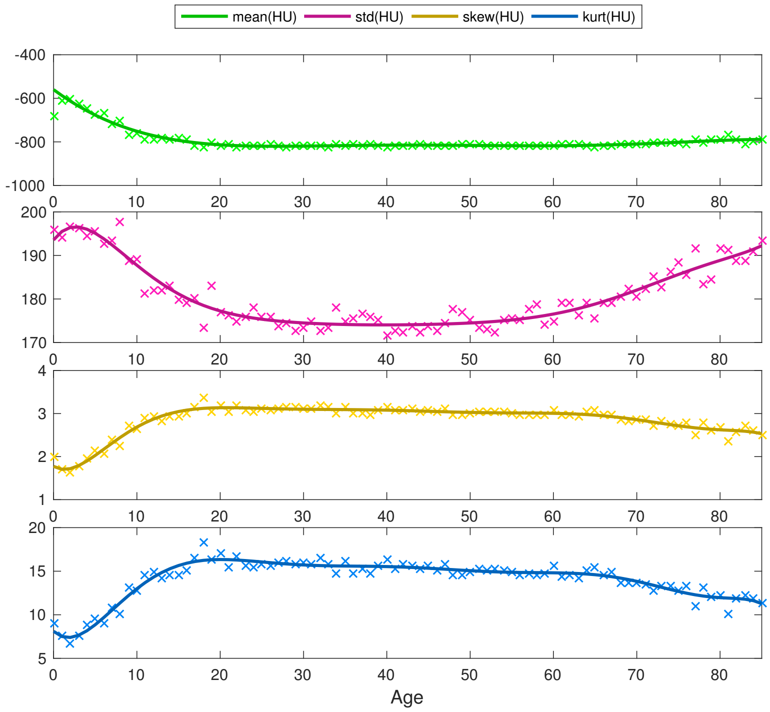


Figure 6: Evolution of the first four statistical moments of the HU distribution inside the lungs with respect to age.

The evolution of the first four statistics computed on the HU distribution are shown in Figure 6. The mean

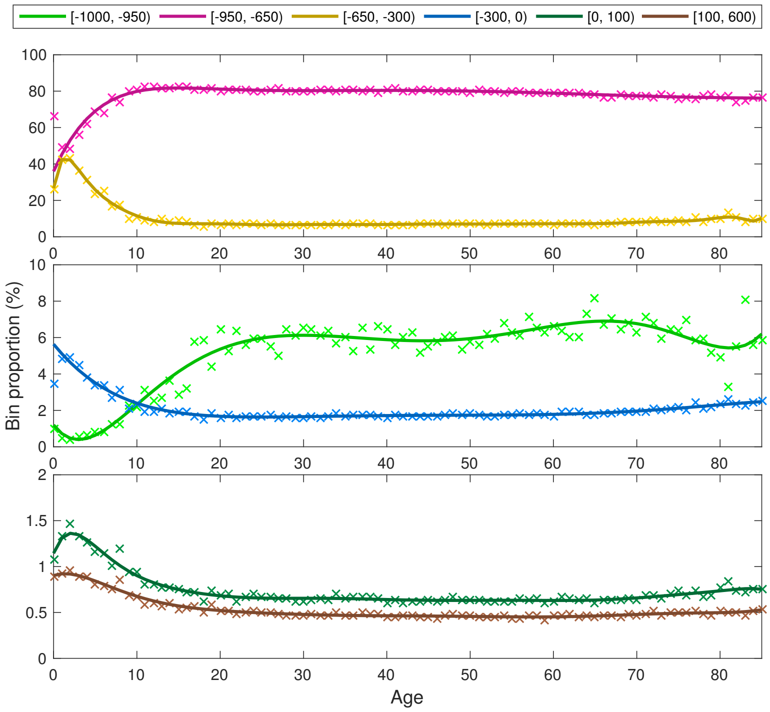


Figure 7: Evolution of the HU histogram bins with respect to age. The different bins are grouped into three subplots due to their different scales.

decreases between 0 and 20 years, and remains relatively stable afterwards. The standard deviation presents a quite symmetric behavior in the intervals [0,40] and [40,85], with a softer slope in the second interval. This seems like a good predictor for early and late age groups as there is a high gradient.

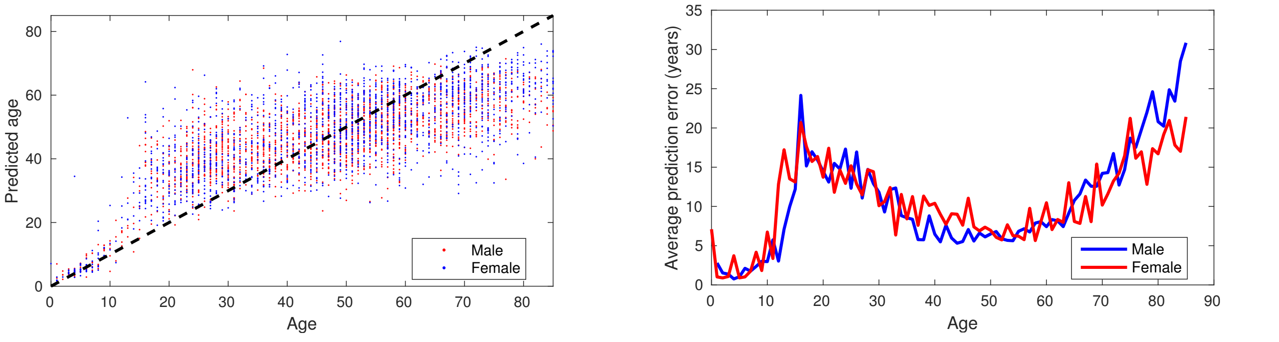
There are clear trends between 1 and 15 years of age when using the histogram of the HUs (see Figure 7). The evolution of the consolidation peak is inversely proportional to the evolution of the healthy tissue bin ([-950, -650) and [-650, -300)), the latter presenting a negative slope with respect to age. In the same range there is an increment of the [-1000,-950) bin that continues until 20 years of age. The other bins (healthy tissue, vessels, and soft tissue) present a negative slope in this range (0 to 20). Between 20 and 85 it is possible to see a slow positive slope in the bins characterizing emphysema, [-1000, -950), and both bins identifying healthy tissue, [-1000, -950) and [-650, -300). While in the same range the consolidation peak bin ([-950, -650)) presents a negative slope. The behavior of the vessels and soft tissue bins remains very stable after 20 years of age.

Table 1 shows the degree, mean relative error and absolute correlation coefficient with respect to the age of the B–spline approximation for each of the features extracted (see Section 2.2.4). In the case of the volume features, the best degree of the B–spline S is between 9 and 10, with a mean relative error between 5.11 and 8.38 % and a correlation coefficient from 0.5008 to 0.6146 with respect to the age. When analyzing the behavior of the ratio between lungs, a lower degree was needed to fit the data (6) and the mean relative error was of only 1.28 %. However, as the ratio is almost constant along the ages, its absolute correlation coefficient is only 0.0189, this being the smallest value of all the features extracted. On the opposite site, the GLCM features homogeneity and contrast present the highest correlation coefficients, 0.8899 and 0.7883 respectively. Moreover, the homogeneity is the feature presenting the smallest mean relative error, 0.54 %, and the contrast is the feature with the smallest degree, 4. Clearly each feature evolves differently with respect to the age, requiring B–splines of degrees between 4 and 18 to fit their evolution. Figure 8 shows the basic Random Forest prediction, so on one axis the real age and on the other the predicted age. Age prediction was very accurate in the first 12 years

	Feature	B–Spline degree (d)	Relative error (ME_r)	Age correlation
Volume	Male Right	9	0.0689	0.6146
	Male Left	9	0.0838	0.5822
	Female Right	9	0.0511	0.5502
	Female Left	10	0.0613	0.5008
	Right/Left ratio	6	0.0128	0.0189
Statistics	mean(HU)	7	0.0103	0.5163
	std(HU)	9	0.0090	0.0996
	skew(HU)	10	0.0228	0.2137
	kurt(HU)	10	0.0356	0.0839
Histogram	Bin [-1000,-950)	9	0.1080	0.7231
	Bin [-950,-650)	8	0.0167	0.2586
	Bin [-650,-300)	18	0.0534	0.4467
	Bin [-300,0)	7	0.0482	0.3554
	Bin [0,100)	13	0.0317	0.5594
	Bin [100,600)	10	0.0294	0.6524
GLCMs	Contrast	4	0.0367	0.7883
	Correlation	5	0.0143	0.3167
	Energy	6	0.0396	0.0488
	Homogeneity	8	0.0054	0.8899

Table 1: Scores of the B–spline approximation method for each of the features extracted on the DBC segmentation. Except from the volume features, the features where averaged over both lungs, and no division by gender was done.

of life, with an average error of 2.2 years. Ages from 13 to 35 and over 65 were the most difficult to predict with prediction errors of 13.5 years and 18.4 years respectively. Accuracy was better from ages 36 to 64, with an average prediction error of 7.4 years, which is relatively close looking at the wide range. Accuracy is not significantly different between males and females.



(a) Actual age vs. predicted age.

(b) Average prediction error for each age using all patients of the test set.

Figure 8: Random forest prediction results.

4. DISCUSSION

The results show that even with extremely limited meta–data, such as age and gender, it is possible to analyze and visualize visual properties of lung CT images. When analyzing a large number of cases as in the work described, the robustness of the segmentation algorithm is extremely important to work really on correct tissue and not create a bias due to wrong segmentation algorithms. Even a comparison between the segmentation algorithms is possible when only analyzing and judging those cases with large differences. In this study it was only observed that CIP had trouble separating left and right lung. Manually analyzing a few other cases with differences we found that they were due to dense tissue at the boundaries and other systematic differences between the algorithms. To be on the safe side, for this study all critical cases were removed.

Several visual features such as lung volume present a strong correlation with respect to age. Lung volume for example increases strongly at a young age and then slower until approximately 20 years of age when things stabilize. At a later age, so between 50 and 60, the volumes starts to decrease first slow and then a bit faster. Women have a smaller volume than men. The left lung is much smaller than the right. All these things can easily be explained with the lung first growing and then at a later age decreasing. Men being taller also explains the average size difference. The left lung surrounding the heart has only two lobes and is thus smaller than the right. The ratio between the two decreases relatively strongly at a young age (before 12 years), which can be explained with the proportion of the heart with respect to the size of the body at an early age. It decreases again at a higher age.

Even when the GLCM homogeneity does not seem to have any major change in respect to age, the correlation coefficient shows that it varies in concordance to the age. The latter combined with the small mean relative error makes it a good candidate estimator of the age. Correlation strongly decreases until 20 and then slowly increases again after around 60. Energy increases until 15 and then slowly after around 25, so an inverse correlation. Contrast first extends faster, then slower until around 60 and then slowly decreases. Many of these processes can likely be linked to the organs forming and the vessels inside the lung that change with age.

The four histogram statistics all follow a relatively similar pattern, even though in opposite directions, having a high gradient from 0 to 15 or 20, then a relative stability until around 60 and then a degradation. This is also linked to the bin histogram that shows that the less dense healthy tissue develops before 12 years of age and is denser before this age. The emphysema range or air in the lung is low in the beginning and then increases quickly until 20 years. This likely also covers large vessels in the lung, that only contain visible air, when actually big enough. Denser areas similar to fibrosis is existent in children and then it decreases when grow up but later it appear as denser areas again from around 50.

The errors in the age prediction that peak between 15 and 30 can likely be explained with the evolution of the lung that has a different speed in different people. This might also depend on people doing sports, where the lung can likely develop strongly for more years, even when already grown up. Then between 30 and 60 the tendencies are relatively stable. There is high variability in the population itself but all global tendencies seem stable. At a higher age the differences get much stronger as degradation can be very fast but healthy persons might keep a lung at the same state as a much younger person. These effects are likely accentuated, as this data set does not contain healthy persons. For healthy persons the differences might not be as marked. Disease onset can be at many different ages, so patients with diseases will have change in pattern at very different ages, really making a prediction difficult.

5. CONCLUSIONS

Analysis of very large data sets has started in the medical field and much can be learned from such large amounts of images, even when the amount of available meta–data is limited, as in our case. Additional meta–data such as diagnosis codes would have definitely been a very good addition to this data set, as it allows to detect also some systematic patterns linked to disease groups and separate them. Basic tools such as segmentation tools can be evaluated by comparing overlap and then manually checking differences but also looking at failed volumes, as in our case. This allows to test the basic tools and avoid mistakes due to problems in the organ segmentation. The clear correlation of several visual features with age shows interesting evolutions and can help in interpreting such lung images. Much can be learned on a population even if the distance from the norm is known and obviously even more when clinical data are available. It is clear that the data set described here does not include many healthy persons and thus only tendencies can be analyzed and not a healthy model at a specific age, which would be extremely interesting for the lung or other organs. Comparison between several cohorts should also help to learn different trends in particular populations. Comparing for example the Belarus cohort with the NLST cohort of mainly smokers and people at risk of lung cancer will be an interesting comparison. It is rare to have such data sets that include a large number of young persons. Particularly the analysis of this age group shows the evolution of the organs and how this can be measured with several visual features. Big data analysis has several challenges in developing standardized pipelines but if well done, much can be learned from the data.

ACKNOWLEDGMENTS

This work was partly supported by the Swiss National Science Foundation in the PH4D project (grant agreement 320030-146804), and by the NIH/NCI grants U24CA180927, U24CA180918, and U01CA154601.

REFERENCES

1. "Riding the wave: How europe can gain from the rising tide of dcientific data." Submission to the European Comission, available online at `http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf`, October 2010.
2. M. W. Vannier and R. M. Summers, "Sharing images," *Radiology* **228**, pp. 23–25, 2003.
3. D. A. Gutman, J. Cobb, D. Somanna, Y. Park, F. Wang, T. Kurc, J. H. Saltz, D. J. Brat, L. A. D. Cooper, and J. Kong, "Cancer digital slide archive: an informatics resource to support integrated in silico analysis of tcga pathology data," *Journal of the American Medical Informatics Association* **20**(6), pp. 1091–1098, 2013.
4. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds., pp. 1097–1105, Curran Associates, Inc., 2012.
5. T. Heimann, B. Van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, *et al.*, "Comparison and evaluation of methods for liver segmentation from CT datasets," *Medical Imaging, IEEE Transactions on* **28**(8), pp. 1251–1265, 2009.
6. D. Mitry, T. Peto, S. Hayat, J. E. Morgan, K.-T. Khaw, and P. J. Foster, "Crowdsourcing as a novel technique for retinal fundus photography classification: Analysis of images in the epic norfolk cohort on behalf of the UK biobank eye and vision consortium," *PLOS ONE* **8**(8), 2013.
7. L. Maier-Hein, S. Mersmann, D. Kondermann, C. Stock, H. G. Kenngott, A. Sanchez, M. Wagner, A. Preukschas, A.-L. Wekerle, S. Helfert, *et al.*, "Crowdsourcing for reference correspondence generation in endoscopic images," in *International Conference on Medical Image Computing and Computer–Assisted Intervention*, pp. 349–356, Springer, 2014.
8. A. Foncubierta-Rodríguez and H. Müller, "Ground Truth Generation in Medical Imaging: A Crowdsourcing Based Iterative Approach," in *Workshop on Crowdsourcing for Multimedia, ACM Multimedia*, oct 2012.
9. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *Medical Imaging, IEEE Transactions on* **34**(10), pp. 1993–2024, 2015.
10. O. Jimenez-del-Toro, H. Müller, M. Krenn, K. Gruenberg, A. A. Taha, M. Winterstein, I. Eggel, A. Foncubierta-Rodríguez, O. Goksel, A. Jakab, G. Kontokotsios, G. Langs, B. Menze, T. Salas Fernandez, R. Schaer, A. Walleyo, M.-A. Weber, Y. Dicente Cid, T. Gass, M. Heinrich, F. Jia, F. Kahl, R. Kechichian, D. Mai, A. B. Spanier, G. Vincent, C. Wang, D. Wyeth, and A. Hanbury, "Cloud–based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL Anatomy Benchmarks," *IEEE Transactions on Medical Imaging* **35**(11), pp. 2459–2475, 2016.

11. M. Krenn, M. Dorfer, O. Jimenez-del-Toro, H. Müller, B. Menze, M.-A. Weber, A. Hanbury, and G. Langs, "Creating a Large–Scale Silver Corpus from Multiple Algorithmic Segmentations," in *MICCAI Workshop on Medical Computer Vision: Algorithms for Big Data, MCV 2015*, **8848**, pp. 163–170, Springer, 2014.
12. M. N. Gurcan, B. Sahiner, N. Petrick, H.-P. Chan, E. A. Kazerooni, P. N. Cascade, and L. Hadjiiski, "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of computer–aided diagnosis system," *Medical Physics* **29**(1), pp. 2552–2558, 2002.
13. A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori, "Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment," *Radiology* **228**, pp. 265–270, July 2003.
14. A. Depeursinge, D. Van De Ville, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, "Near–affine–invariant texture learning for lung tissue analysis using isotropic wavelet frames," *IEEE Transactions on Information Technology in BioMedicine* **16**, pp. 665–675, July 2012.
15. Y. Xu, E. J. R. van Beek, Y. Hwanjo, J. Guo, G. McLennan, and E. A. Hoffman, "Computer–aided classification of interstitial lung diseases via MDCT: 3D adaptive multiple feature method (3D AMFM)," *Academic Radiology* **13**, pp. 969–978, August 2006.
16. M. F. McNitt-Gray, S. G. Armato, C. R. Meyer, A. P. Reeves, G. McLennan, R. C. Pais, J. Freymann, M. S. Brown, R. M. Engelmann, P. H. Bland, G. E. Laderach, C. Piker, J. Guo, Z. Towfic, D. Qing, D. F. Yankelevitz, D. R. Aberle, E. J. R. van Beek, H. MacMahon, E. A. Kazerooni, B. Y. Croft, and L. P. Clarke, "The lung image database consortium (LIDC) data collection process for nodule detection and annotation," *Academic Radiology* **14**, pp. 1464–1474, December 2007.
17. A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller, "Building a reference multimedia database for interstitial lung diseases," *Computerized Medical Imaging and Graphics* **36**, pp. 227–238, April 2012.
18. D. W. Belsky, A. Caspi, R. Houts, H. J. Cohen, D. L. Corcoran, A. Danese, H. Harrington, S. Israel, M. E. Levine, J. D. Schaefer, K. Sugden, B. Williams, A. I. Yashin, R. Poulton, and T. E. Moffitt, "Quantification of biological aging in young adults," *Proceedings of the National Academy of Sciences* **112**(30), pp. E4104–E4110, 2015.
19. Y. Huo, K. Aboud, H. Kang, L. E. Cutting, and B. Landman, "Mapping lifetime brain volumetry with covariate-adjusted restricted cubic spline regression from cross-sectional multi-site MRI," in *Medical Image Computing and Computer–Assisted Interventions, MICCAI*, 2016 (accepted).
20. J. Zhang, M. Liu, L. An, Y. Gao, and D. Shen, "Landmark-based alzheimers disease diagnosis using longitudinal structural MR images," in *MICCAI Workshop on Machine Learning in Medical Imaging, MLMI 2016*, 2016 (accepted).
21. D. Štern, C. Payer, V. Lepetit, and M. Urschler, "Automated age estimation from hand MRI volumes using deep learning," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, eds., pp. 194–202, Springer International Publishing, (Cham), 2016.
22. Y. Dicente Cid, O. Jimenez-del-Toro, A. Depeursinge, and H. Müller, "Efficient and fully automatic segmentation of the lungs in CT volumes," in *Proceedings of the VISCERAL Challenge at ISBI*, O. Orcun Goksel, Jimenez-del-Toro, A. Foncubierta-Rodriguez, and H. Müller, eds., *CEUR Workshop Proceedings*, Apr 2015.
23. O. Goksel, A. Foncubierta-Rodríguez, O. Jimenez-del-Toro, H. Müller, G. Langs, M.-A. Weber, B. Menze, I. Eggel, K. Gruenberg, M. Winterstein, M. Holzer, M. Krenn, G. Kontokotsios, S. Metallidis, R. Schaer, A. A. Taha, A. Jakab, T. Salas Fernandez, and A. Hanbury, "Overview of the VISCERAL challenge at ISBI 2015," in *Proceedings of the VISCERAL Challenge at ISBI*, O. Goksel *et al.*, eds., *CEUR Workshop Proceedings*, pp. 6–11, Apr 2015.
24. R. San Jose Estepar, J. C. Ross, R. Harmouche, J. Onieva, A. A. Diaz, G. R. Washko, and R. S. J. Estepar, *Chest Imaging Platform: an open-source library and workstation for quantitative chest imaging*, pp. A4975–A4975. Am Thoracic Soc, 2015.
25. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics* **3**, pp. 610–621, November 1973.