

# Regression as a Tool to Measure Segmentation Quality and Preliminary Indicator of Diseased Lungs

Norliza Mohd. Noor<sup>1</sup>(✉), Omar Mohd. Rijal<sup>2</sup>, Joel Chia Ming Than<sup>1</sup>,  
Rosminah M. Kassim<sup>3</sup>, and Ashari Yunus<sup>4</sup>

<sup>1</sup> UTM Razak School of Engineering and Advanced Technology,  
Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia  
norliza@utm.my

<sup>2</sup> Institute of Mathematical Sciences, University of Malaya, Kuala Lumpur, Malaysia

<sup>3</sup> Department of Diagnostic Imaging, Kuala Lumpur Hospital,  
Kuala Lumpur, Malaysia

<sup>4</sup> Institute of Respiratory Medicine, Kuala Lumpur, Malaysia

**Abstract.** Segmentation of the lung from HRCT Thorax images was studied. An automatic method of determining segmentation area is proposed. High quality of segmentation is considered achieved when the segmented area from the proposed algorithm is almost identical to the area obtained from the manual tracings by lung expert (ground truth). High correlation between the two types of segmented areas showed that regression may be used as a tool to measure segmentation quality. Supplementary information may also be obtained from the regression plot. Prediction interval may be used as a possible indicator of diseased whilst outliers may show or indicate low segmentation quality or a possible severity of the disease.

**Keywords:** Lung segmentation · Regression · Segmentation quality · High resolution computed tomography

## 1 Introduction

Lung segmentation is defined as the separation of the lung regions from background regions to obtain lung region defined by a boundary. Most segmentation techniques can be attributed to the use of thresholding [1–6]. This is because thresholding is particularly effective as the main approach of extracting the region of interest (lungs) from the body region due to the difference in contrast between the two regions. This attribute is due to content of air in the lungs that causes less attenuation of than body regions [6]. Other examples of methods include graph cut [7], texture [8] and active contours [9]. Thus the objective and quantitative evaluation of segmentation is important especially with the array of segmentation methods available today [10].

Limited lung slices or levels can be used to view particular diseases in a particular patient [11]. These levels correspond to different anatomic landmarks and height of the lung as well. There are possible categorisations of levels to three or five level slice viewing. For five levels, the levels can be represented as L1: aortic arch, L2: trachea carina, L3: pulmonary hilar, L4: pulmonary venous confluence and L5: 1–2 cm above right hemi-diaphragm [12].

Segmentation quality can be found using several measures. There are many quality measures used such as volume overlap error, relative volume difference, average symmetric surface difference [13]. For this study, the regression curve which is often overlooked is highlighted to show its capability to exhibit the segmentation quality across levels of the lung. Regression analysis is a powerful tool to show the segmentation quality over a large sample base.

## 2 Data Collection

High Resolution Computed Tomography (HRCT) scans for 15 normal patients and 81 diseased patients were collected retrospectively from Hospital Kuala Lumpur. The images were obtained using a Siemens SomatomPlus4 CT scanner, and observed by a senior radiologist using SyngoFastView version VX57G27. Each slice was obtained at 10 mm intervals. Patients were at a supine position will full suspended inspiration when scans were taken. Images taken were in the size of  $512 \times 512$  pixels.

A senior radiologist was tasked to select the five predetermined levels of the lungs to be processed with the segmentation algorithm. The five predetermined levels correspond to anatomical landmarks were selected by senior radiologist as Level 1 (L1) - aortic arch, Level 2 (L2) - trachea carina, Level 3 (L3) - pulmonary hilar, Level 4 (L4) - pulmonary venous confluence, and Level 5 (L5) - 1–2 cm above right hemi-diaphragm.

Since the gold standard or ground truth of segmentation is also done by a human expert, in this study, the manual tracings of the ground truth were done by a trained lung image expert who plotted the borders of the lung using a software interface. The points plotted were saved in a (x,y) coordinate format.

## 3 Methodology

### 3.1 Segmentation Algorithm

The segmentation algorithm was developed and tested firstly for the normal patients HRCT scans and was applied to the diseased patients for all five predetermined levels. The segmentation approach used here was able to segment across all levels and in a fast manner. The segmentation algorithm combines known techniques such as thresholding and morphology [12]. The flowchart of the segmentation algorithm is shown in Fig. 1.

The HRCT scan image first undergoes an automatic global thresholding known as Otsu thresholding. This enables the separation the body region from

the surrounding region. Secondly objects less than 3000 pixels in area are removed. Next, the body region is applied an empirical threshold of -324HU. Morphology operations of dilation and erosion with a square structure element with a size of  $3 \times 3$  pixels are then executed to smoothen lung boundaries. A connected component analysis is done to extract the two lung regions. Besides the body and surrounding there should be two other regions that can be seen to have 8-connectivity. This property is used to detect the presence of both lungs. A check is done to see whether there are two lungs present. If there are not, lung separation operations are executed. Lung regions are then filled and contours are extracted.

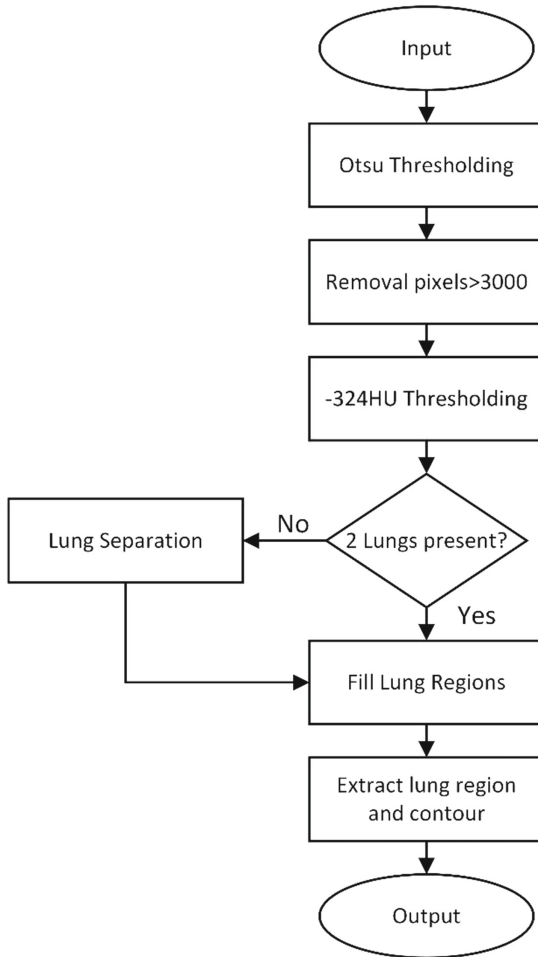


Fig. 1. Flowchart of segmentation algorithm.

### 3.2 Regression Analysis

Regression analysis, using MINITAB, was done to analyze the relationship between two measurements [14]. The first measurement is the lung area enclosed by the automatically segmented boundary using the segmentation algorithm termed as  $y$ . The second measurement is the lung area enclosed by the manually traced lung borders termed as  $x$ . The lung area is calculated by summing the area of pixels enclosed and the area of each pixel can be calculated as Eq. (1).

$$A = h \times l \quad (1)$$

where  $A$  is the area of a pixel in  $\text{mm}^2$ ,  $h$  is the pixel height in mm and  $l$  the pixel length in mm.

Regression analysis firstly plots the trendline that shows the relationship between  $x$  and  $y$ . The trendline can be signified by an equation that models the value of  $y$  when  $x$  changes. The parameters contributing to this model are the slope of the trendline represented by  $\beta$ , and the intersect of the  $y$  axis represented by  $\alpha$ . The model can be represented as Eq. (2).

$$y = \alpha + \beta x \quad (2)$$

In this study, the prediction intervals (PI) with 95% probability are also used. A prediction interval can be defined as the range that is likely to contain the response value of a new value of  $y$  when there are provided specified settings of the predictors in the model Eq. (3).

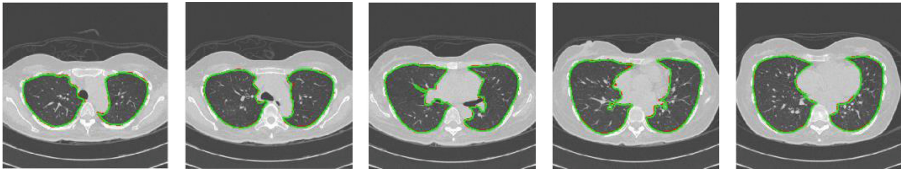
$$y = \alpha + \beta x + \epsilon \quad (3)$$

where  $y$  is the area of the segmented lung,  $x$  is the area of the manually traced lung.  $\alpha$  the intersection of  $y$  axis  $\beta$  is the slope of the trendline, and  $\epsilon$  is the random error term.

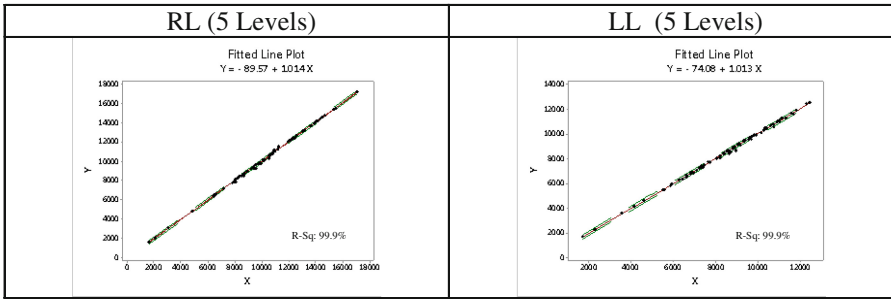
## 4 Results

The Fig. 2 shows the sample of good segmentation that the segmentation algorithm was able to produce. The high quality segmentation is signified by the degree of overlap between two contours noticed by the naked eye. There are two contours present; green signifying the automatic segmented borders and red borders signifying the manually traced lung borders. In some instances such as in Fig. 2 the green totally overlaps the red contours. The overlap signifies the high degree of similarity between the automatic segmentation and ground truth.

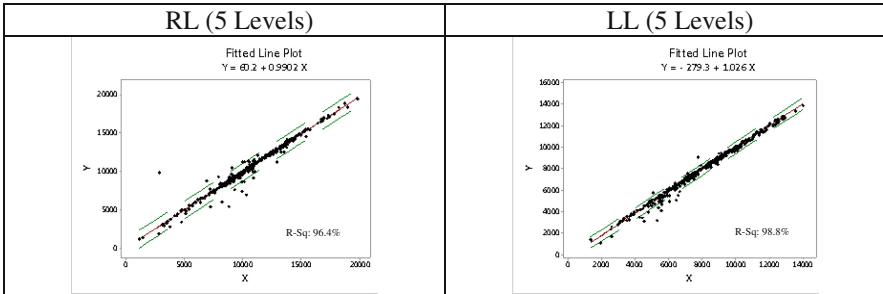
Next the regression curves for the combined five levels are shown for both dis-eased and normal lungs are presented in Figs. 3 and 4. The X axis represents the ground truth traced lung area and the Y axis represents the automatically segmented lung area. The regression equations are inserted as well in the plots. The dashed lines represent the 95% prediction intervals. The solid red line represents the fitted linear regression line.



**Fig. 2.** Samples of high quality segmentation across five levels (Color figure online).

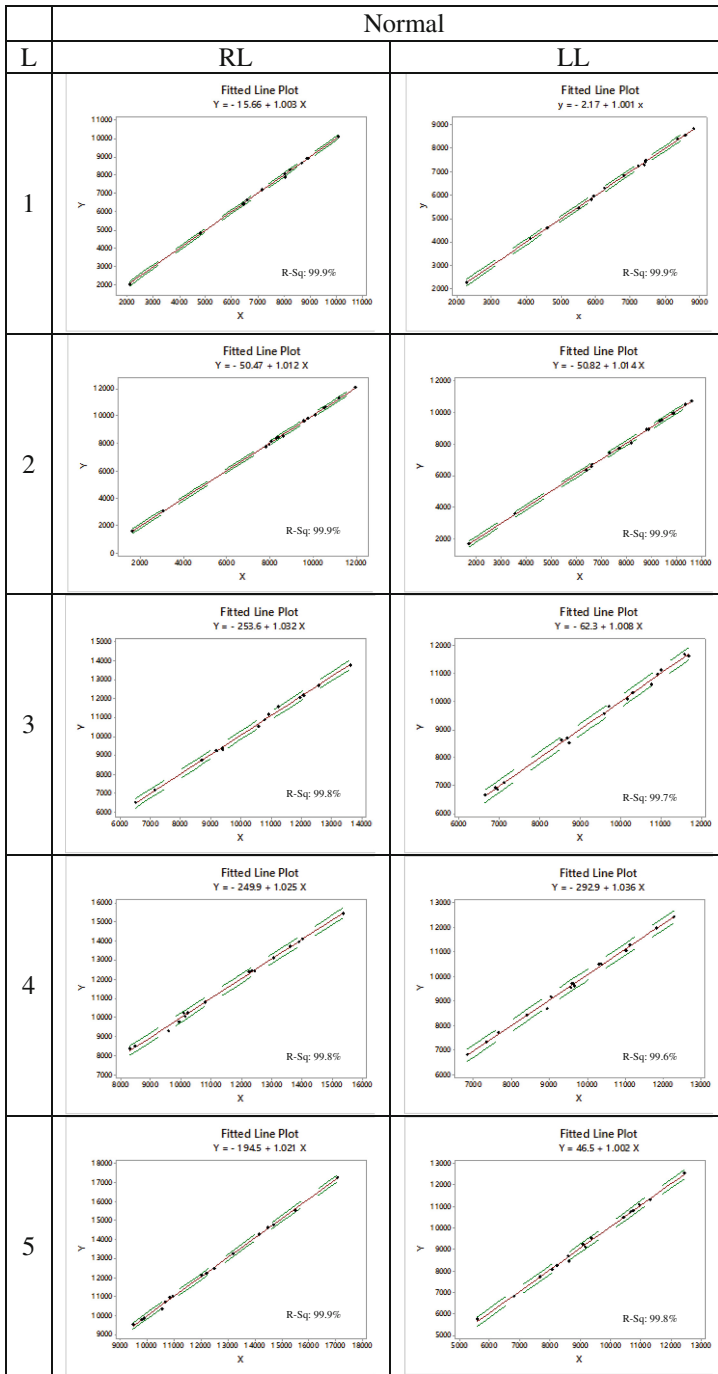


**Fig. 3.** Regression plots for normal patients right (RL) and left (LL) lungs for five levels combined (Color figure online).

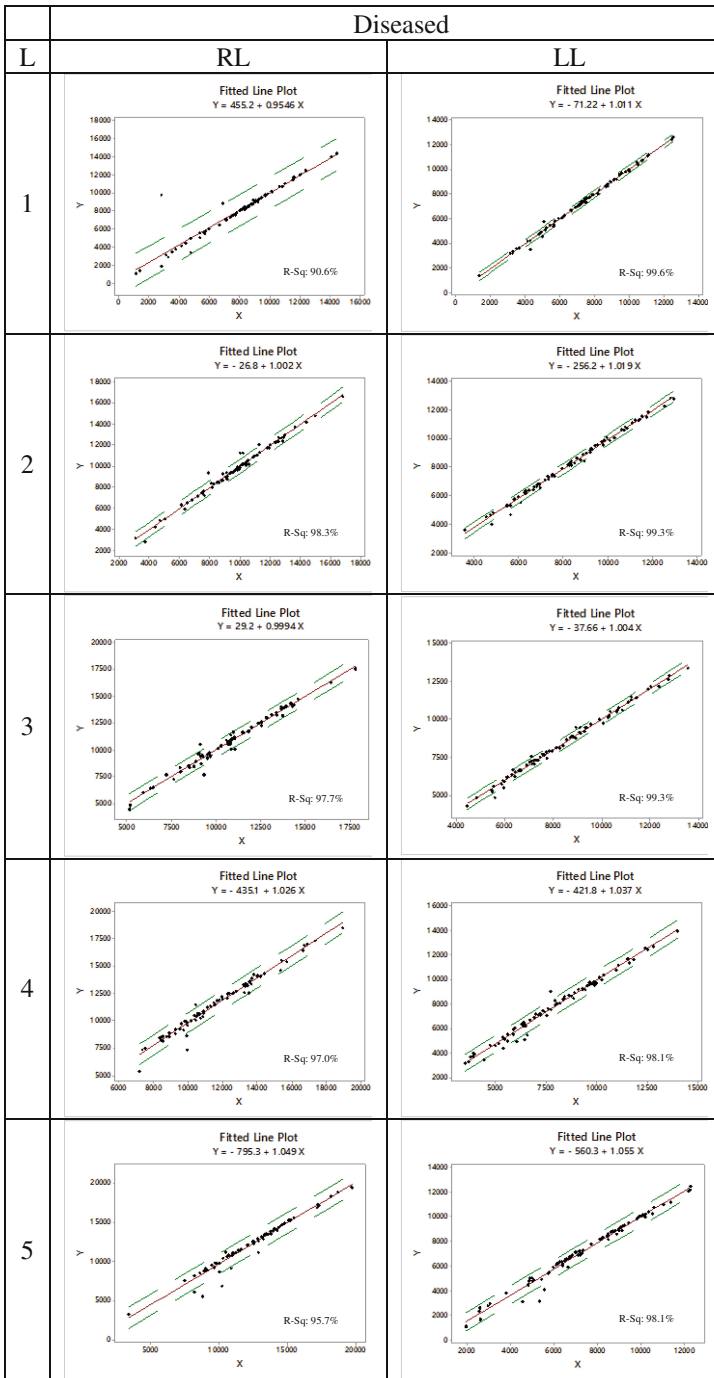


**Fig. 4.** Regression plots for diseased patients right (RL) and left (LL) lungs for five levels combined (Color figure online).

To further see in detail the segmentation quality, the regression curve of separate five levels of right and left lungs for diseased and normal patients are shown in Figs. 5 and 6. The dashed lines represent the 95 % prediction intervals. The solid red line represents the fitted linear regression line. Again, the X axis represents the ground truth traced lung area and the Y axis represents the automatically segmented lung area.



**Fig. 5.** Regression curves for separate five levels for normal right (RL) and left (LL) lungs for normal lungs (Color figure online).



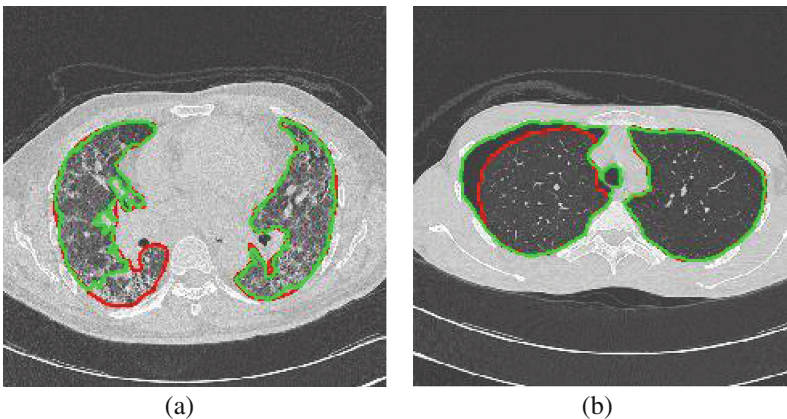
**Fig. 6.** Regression curves for separate five levels for diseased right (RL) and left (LL) lungs for normal lungs (Color figure online).

## 5 Discussion

The segmented area from the proposed segmentation algorithm ( $Y$ ) being highly cor-related to the area obtained by manual tracings (ground truth,  $X$ ) is the basic definition of segmentation quality. The regression plots in Figs. 3, 4, 5 and 6 showing high correlation strongly suggesting that a high segmentation quality was consistently achieved across all five levels. This is illustrated by the  $\beta$  values of all the regression equations being close to one, this situation being the case for normal lungs and more importantly diseased lungs as well. However, non-zero  $\alpha$  values may be considered as deviation of  $Y$  from  $X$ , but their values are relatively small compared to the magnitude of the segmented area. This result strongly suggest that the proposed segmentation algorithm is promising because it shows the segmentation was able to exhibit segmentation quality similar to ground truth provided by a human expert for most diseased lungs.

As a second indicator of segmentation quality, the regression plots displayed should show points within the prediction interval, this being the general case. It is of interest to note that prediction intervals are wider for disease cases, which could suggest a possible method of differentiating disease cases from normal case.

Absence of outliers in the regression plots may also be used as an indicator of segmentation quality. However, when outliers do exist such observation should be separately studied. For example, in Fig. 7(a)-(b), two samples of such outliers are presented. On further investigation, one observation showed the problem of lung hardening whilst the second observation was a patient with a collapse lung. A senior radiologist has confirmed that the lung hardening case has severe reticular pattern and pulmonary consolidation which are features of Interstitial Lung Disease (ILD). Henceforth outliers whilst suggesting lower segmentation quality may provide other useful information.



**Fig. 7.** Samples of outliers from regression curves (a) due to lung hardening and (b) due to collapse lung.



In summary, the simple regression model may be used as an indicator of segmentation quality while outliers provide useful information of the lung condition.

## 6 Conclusion

As a conclusion this study showed that the regression plot has managed to show the segmentation quality in detail when comparing five levels of diseased and normal lungs. The width of the prediction interval of the regression plot and outliers may in turn also provide supplementary information of the lung condition.

**Acknowledgement.** This research is funded by the Universiti Teknologi Malaysia Research University Grant: GUP QK130000.2540.06H35, and Ministry of Higher Education Malaysia and supported by Universiti Malaya.

## References

1. Brown, M.S., Goldin, J.G., McNitt-Gray, M.F., Greaser, L.E., Sapra, A., Li, K.-T., Sayre, J.W., Martin, K., Aberle, D.R.: Knowledge-based segmentation of thoracic computed tomography images for assessment of split lung function. *Med. Phys.* **27**, 592–598 (2000)
2. Brown, M.S., McNitt-Gray, M.F., Mankovich, N.J., Goldin, J.G., Hiller, J., Wilson, L.S., Aberle, D.R.: Method for segmenting chest CT image data using an anatomical model: preliminary results. *IEEE Trans. Med. Imaging*. **16**, 828–839 (1997)
3. Armato, S.G., Sensakovic, W.F.: Automated lung segmentation for thoracic CT impact on computer-aided diagnosis. *Acad. Radiol.* **11**, 1011–1021 (2004)
4. Leader, J.K., Zheng, B., Rogers, R.M., Sciruba, F.C., Perez, A., Chapman, B.E., Patel, S., Fuhrman, C.R., Gur, D.: Automated lung segmentation in X-ray computed tomography: development and evaluation of a heuristic threshold-based scheme. *Acad. Radiol.* **10**, 1224–1236 (2003)
5. Sun, X., Zhang, H., Duan, H.: 3D computerized segmentation of lung volume with computed tomography. *Acad. Radiol.* **13**, 670–677 (2006)
6. Hu, S., Hoffman, E.A., Reinhardt, J.M.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans. Med. Imaging* **20**(6), 490–498 (2001)
7. Massoptier, L., Misra, A., Sowmya, A.: Automatic lung segmentation in HRCT images with diffuse parenchymal lung disease using graph-cut. In: 24th International Conference on Image and Vision Computing New Zealand, IVCNZ09, pp. 266–270 (2009)
8. Korfiatis, P., Kalogeropoulou, C., Karahaliou, A., Kazantzi, A., Skiadopoulos, S., Costaridou, L.: Texture classification-based segmentation of lung affected by interstitial pneumonia in high-resolution CT. *Med. Phys.* **35**, 5290–5302 (2008)
9. Tobata, K., Hospital, K.: A segmentation method of lung areas by using snakes. *Int. J. Innov. Comput. Inf. Control*. **3**, 277–284 (2007)
10. Zhang, Y.: A review of recent evaluation methods for image segmentation. In: Proceedings of the 6th International Symposium on Signal Processing and its Applications, vol. 1, pp. 148–151 (2001)

11. Kazerooni, E.A., Martinez, F.J., Flint, A., Jamadar, D.A., Gross, B.H., Spizarny, D.L., Cascade, P.N., Whyte, R.I., Lynch, J.P., Toews, G., Lynch 3rd, J.P.: Thin-section CT obtained at 10-mm increments versus limited three-level thin-section CT for idiopathic pulmonary fibrosis: correlation with pathologic scoring. *AJR. Am. J. Roentgenol.* **169**, 977–983 (1997)
12. Chia, J., Than, M., Noor, N.M., Rijal, O.M., Yunus, A., Kassim, R.M., Chia, J., Noor, N.M., Rijal, O.M., Yunus, A., Kassim, R.M.: Lung segmentation for HRCT thorax images using radon transform and accumulating pixel width. In: 2014 IEEE Region 10 Symposium, pp. 157–161 (2014)
13. Heimann, T., van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G., Bischof, H., Bornik, A., Cashman, P.M.M., Chi, Y., Cordova, A., Dawant, B.M., Fidrich, M., Furst, J.D., Furukawa, D., Grenacher, L., Hornegger, J., Kainmüller, D., Kitney, R.I., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H.-P., Nemeth, G., Raicu, D.S., Rau, A.-M., van Rikxoort, E.M., Rousson, M., Rusko, L., Saddi, K.A., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J.M., Wimmer, A., Wolf, I.: Comparison and evaluation of methods for liver segmentation from CT datasets. *IEEE Trans. Med. Imaging* **28**(8), 1251–1265 (2009)
14. Montgomery, D.C., Peck, E.A., Vining, G.G.: *Introduction to Linear Regression Analysis*. John Wiley Sons, New York (2012)