Automated CT Image Evaluation of the Lung: A Morphology-Based Concept

R. A. Blechschmidt*, Member, IEEE, R. Werthschützky, and U. Lörcher

Abstract—Computed tomography (CT) provides the most reliable method to detect emphysema in vivo. Commonly used methods only calculate the area of low attenuation [pixel index (PI)], while a radiologist considers the bullous morphology of emphysema. The PI is a good, well-known measure of emphysema. But it is not able to detect emphysema in cases in which emphysema and fibrosis occur at the same time. This is because fibrosis leads to a low number of low-attenuation pixels, while emphysema leads to a high number of pixels. The PI takes the average of both and, consequently, may present a result within the normal range.

Method: The main focus of this paper is to present a new algorithm of thoracic CT image evaluation based on pulmonary morphology of emphysema. The PI is extended, in that it is enabled to differentiate between small, medium, and large bullae (continuous low-attenuation areas). It is not a texture-based algorithm. The bullae are sorted by size into four size classes: class 1 being within the typical size of lung parenchyma; classes 2–4 presenting small, medium, and large bullae. It is calculated how much area the different classes take up of all low-attenuation pixels. The bullae index (BI) is derived from the percentage of areas covered, respectively, by small, medium, and large bullae.

From the relation of the area of bullae belonging to class 4, to that of those belonging to class 2, a measure of the emphysema type (ET) is calculated. It classifies the lung by the type of emphysema in bullous emphysema or small-sized, diffuse emphysema, respectively.

Results: The BI is as reliable as the PI. In cases in which the PI indicates normal values while in fact emphysema is coexisting with fibrosis, the BI, nevertheless, detects the destruction caused by the emphysema. The BI combined with the ET reflects the visual assessment of the radiological expert.

Conclusion: The BI is an objective and reliable index in order to quantify emphysematous destruction, hence, avoiding inter-observer variance. This is particularly interesting for follow-up. The classification of the ET is a helpful and unique approach to achieving an exact diagnosis of emphysema.

Index Terms—Algorithm, classification by size and type, densitometry, emphysema, image evaluation, lung CT, morphology.

I. INTRODUCTION

A NAMNESIS and spirometric tests supply the basic diagnostic tools of lung diseases such as emphysema. Since spirometric tests provide globally measured values, the distribution of pathological lesions cannot be pointed out. Combined effects of restrictive and obstructive problems render the spirometric test evaluation even more difficult. Clinical diagnosis of emphysema is unreliable [14]. Hence, another diagnostic tool is needed for lung-diagnosis.

Computed tomography (CT) is the only method to establish the in vivo diagnosis of pulmonary emphysema by morphology and quantitative analysis [24]. Further, quantitative image analysis is a useful extension of the evaluation of CT scans. Although much research has been done in the field of CT image analysis, there are still some unsettled problems. In Section II, we give a brief review of CT image analysis.

Former algorithms have their origin in describing the distribution of density as a histogram. This has been implemented in scanners by Siemens (Erlangen, Germany), known as the Pulmo Software, and by GE Medical Systems (Waukesha, WI), known as the Density Mask. These software tools provide a semiautomatic segmentation of the lung, and a calculation of the area within a certain threshold. Recently, there has been the extension of the two-dimensional image analysis to a three-dimensional analysis [29], [40].

The classical method of CT image evaluation is the pixel index (PI). It was introduced by Kalender et al. [16]. The PI is the number of pixels (pix) with lower attenuation than the limit value (lim), divided by the number of pixels of the entire segmented lung

\[ PI = \frac{|\{\text{pix} \in \text{CT} : \text{pix} \in [-1000 \ldots \text{lim HU}]\}|}{A_{\text{lung}}} \]  (1)

\[ \text{lim denotes the upper limit, given as being between } -950 \text{ HU and } -900 \text{ HU by other authors. } A_{\text{lung}} \text{ denotes the set of all pixels of the segmented lung} \]

\[ A_{\text{lung}} = |\{\text{pix} \in \text{CT} : \text{pix} \in [-1000 \ldots -200 \text{HU}]\}|. \]  (2)

The PI (1) represents the percentage of lung area with lower attenuation values than the limit value. All pixels below this limit are thought to belong to air-filled lung regions. Thus, this index should describe the amount of air and, hence, detect emphysematous lesions [20].

Other research groups use the mean lung density [12], the full-width at half-maximum of the density histogram [30] or the fifth or tenth percentile of the CT histogram data [11], [37]. All mentioned methods only consider the total area of air-filled lung
regions, and to consider the morphology of emphysema. Values range from 3% to 10%. In daily practice, it has to show which method performs better and can be interpreted more easily by the radiologist. The size classes are defined in Fig. 4. The limits of the size classes using four-neighbor region labeling to avoid voxel trails. Following this, all bullae are sorted by size into four classes, including the classical method. In this paper, we refer to continuous marked low-attenuation areas as “bullae.” We can infer from this to the type of emphysema, which is unique. Third, we focus on the influence of coexisting fibrosis on the diagnosis of emphysema. We were able to show that our method is still expressive in these cases.

B. The New Algorithm in Detail

The purpose of this paper is to introduce the new method and compare it with current techniques. The result of the new algorithm is useful as a parameter of the homogeneity of air sacs in the lung. Usefulness and efficiency of the new approach will be discussed using pulmonary emphysema—a pathological dilation of air sacs with an impaired lung function. This is currently an important field of research in radiology, presented in several reviews [9], [26], [35], [36].

A. New Approach

We developed the new method in order to avoid the averaging effect of the PI and to consider the morphology of emphysema. How can we benefit from the experience of the radiologist for the automatic evaluation? Figs. 1–3 show magnified CT images of normal and emphysematous lungs for different thresholds, varying from −950 HU to −910 HU. The left panel shows a frame of the CT image. The others are threshold images. The white dots refer to regions with lower attenuation values than the threshold. This technique is called voxel highlighting.

A normal subject (Fig. 1) has with increasing thresholds from −950 HU to −910 HU an increase in the number of all pixels of air in the lung. The white dots represent areas with lower attenuation values than −950 HU, −930 HU, and −910 HU, respectively. The calculated PI’s are: $PI_{-950} = 3\%, PI_{-930} = 5\%,$ and $PI_{-910} = 10\%$.

The bullae are marked with the grey arrows. Pixel index: $PI_{-950} = 8\%$, $PI_{-930} = 13\%$, and $PI_{-910} = 21\%$.

The purpose of this paper is to introduce the new method and compare it with current techniques. The result of the new algorithm is useful as a parameter of the homogeneity of air sacs in the lung. Usefulness and efficiency of the new approach will be discussed using pulmonary emphysema—a pathological dilation of air sacs with an impaired lung function. This is currently an important field of research in radiology, presented in several reviews [9], [26], [35], [36].

A. New Approach

We developed the new method in order to avoid the averaging effect of the PI and to consider the morphology of emphysema. How can we benefit from the experience of the radiologist for the automatic evaluation? Figs. 1–3 show magnified CT images of normal and emphysematous lungs for different thresholds, varying from −950 HU to −910 HU. The left panel shows a frame of the CT image. The others are threshold images. The white dots refer to regions with lower attenuation values than the threshold. This technique is called voxel highlighting.

A normal subject (Fig. 1) has with increasing thresholds from −950 HU to −910 HU an increase in the number of all pixels of air in the lung. The white dots represent areas with lower attenuation values than −950 HU, −930 HU, and −910 HU, respectively. The calculated PI’s are: $PI_{-950} = 3\%, PI_{-930} = 5\%,$ and $PI_{-910} = 10\%$.

The bullae are marked with the grey arrows. Pixel index: $PI_{-950} = 8\%$, $PI_{-930} = 13\%$, and $PI_{-910} = 21\%$.

The purpose of this paper is to introduce the new method and compare it with current techniques. The result of the new algorithm is useful as a parameter of the homogeneity of air sacs in the lung. Usefulness and efficiency of the new approach will be discussed using pulmonary emphysema—a pathological dilation of air sacs with an impaired lung function. This is currently an important field of research in radiology, presented in several reviews [9], [26], [35], [36].

A. New Approach

We developed the new method in order to avoid the averaging effect of the PI and to consider the morphology of emphysema. How can we benefit from the experience of the radiologist for the automatic evaluation? Figs. 1–3 show magnified CT images of normal and emphysematous lungs for different thresholds, varying from −950 HU to −910 HU. The left panel shows a frame of the CT image. The others are threshold images. The white dots refer to regions with lower attenuation values than the threshold. This technique is called voxel highlighting.

A normal subject (Fig. 1) has with increasing thresholds from −950 HU to −910 HU an increase in the number of all pixels of air in the lung. The white dots represent areas with lower attenuation values than −950 HU, −930 HU, and −910 HU, respectively. The calculated PI’s are: $PI_{-950} = 3\%, PI_{-930} = 5\%,$ and $PI_{-910} = 10\%$.

The bullae are marked with the grey arrows. Pixel index: $PI_{-950} = 8\%$, $PI_{-930} = 13\%$, and $PI_{-910} = 21\%$.

The purpose of this paper is to introduce the new method and compare it with current techniques. The result of the new algorithm is useful as a parameter of the homogeneity of air sacs in the lung. Usefulness and efficiency of the new approach will be discussed using pulmonary emphysema—a pathological dilation of air sacs with an impaired lung function. This is currently an important field of research in radiology, presented in several reviews [9], [26], [35], [36].

A. New Approach

We developed the new method in order to avoid the averaging effect of the PI and to consider the morphology of emphysema. How can we benefit from the experience of the radiologist for the automatic evaluation? Figs. 1–3 show magnified CT images of normal and emphysematous lungs for different thresholds, varying from −950 HU to −910 HU. The left panel shows a frame of the CT image. The others are threshold images. The white dots refer to regions with lower attenuation values than the threshold. This technique is called voxel highlighting.

A normal subject (Fig. 1) has with increasing thresholds from −950 HU to −910 HU an increase in the number of all pixels of air in the lung. The white dots represent areas with lower attenuation values than −950 HU, −930 HU, and −910 HU, respectively. The calculated PI’s are: $PI_{-950} = 3\%, PI_{-930} = 5\%,$ and $PI_{-910} = 10\%$.

The bullae are marked with the grey arrows. Pixel index: $PI_{-950} = 8\%$, $PI_{-930} = 13\%$, and $PI_{-910} = 21\%$.

The purpose of this paper is to introduce the new method and compare it with current techniques. The result of the new algorithm is useful as a parameter of the homogeneity of air sacs in the lung. Usefulness and efficiency of the new approach will be discussed using pulmonary emphysema—a pathological dilation of air sacs with an impaired lung function. This is currently an important field of research in radiology, presented in several reviews [9], [26], [35], [36].

A. New Approach

We developed the new method in order to avoid the averaging effect of the PI and to consider the morphology of emphysema. How can we benefit from the experience of the radiologist for the automatic evaluation? Figs. 1–3 show magnified CT images of normal and emphysematous lungs for different thresholds, varying from −950 HU to −910 HU. The left panel shows a frame of the CT image. The others are threshold images. The white dots refer to regions with lower attenuation values than the threshold. This technique is called voxel highlighting.

A normal subject (Fig. 1) has with increasing thresholds from −950 HU to −910 HU an increase in the number of all pixels of air in the lung. The white dots represent areas with lower attenuation values than −950 HU, −930 HU, and −910 HU, respectively. The calculated PI’s are: $PI_{-950} = 3\%, PI_{-930} = 5\%,$ and $PI_{-910} = 10\%$.

The bullae are marked with the grey arrows. Pixel index: $PI_{-950} = 8\%$, $PI_{-930} = 13\%$, and $PI_{-910} = 21\%$.
Fig. 4. Definition of the size classes. Shown is a graphical representation of the size in a relative scale.

\[ B_i \] denotes the number of the size classes. \( n_{\text{bull}} \) is defined as follows:

\[ n_{\text{bull}} = \left\{ B_i \in \{ B_0 \ldots B_n \} : |B_i| \in \text{class}_j \right\} \]

(3)

where \( j = 1.4 \) number of size class;
\( B_1 \ldots B_n \) sets of all pixels of one area;
\( \text{class}_1 \ldots \text{class}_4 \) size classes as defined in Fig. 4;
\( n \) numbers of marked areas.

For normal subjects, there are typically multiples of thousand of bullae in the first class and less than a hundred in the other classes. The area these bullae occupy is not normally taken into consideration. In order to improve this, not only the number of bullae in each class but the area these bullae represent should be calculated

\[ AR_{\text{bull}} = \{|\text{pix} \in \text{CT} : \text{pix} \in \{ B_i \in \{ B_0 \ldots B_n \} : |B_i| \in \text{class}_j \}|, \]

(4)

Taking the area of bullae in a class into account, the percentage of area in each class related to the area of the lung \( pAR_{\text{c}j} \) can be derived as

\[ pAR_{\text{c}j} = \frac{AR_{\text{bull}}}{AR_{\text{lung}}} \cdot 100. \]

(5)

This formula is related to the \( PI \) as follows:

\[ PI = \sum_{j=1}^{4} pAR_{\text{c}j}. \]

(6)

Our goal was to develop an expressive index that quantifies inhomogeneous air distribution, based on the opinion of the radiological expert. The expert takes the size of the bullous destruction (severity) and extent into account. With our algorithm, we are able to model these two criteria. The severity is defined by the size of low-attenuation areas (bullae). The extent is calculated by the involved area corresponding to the size classes.

With this knowledge, we defined the \( BI \) as the product of severity and extent. Therefore, we calculated for each severity a partial score \( g_2, g_3 \) and \( g_4 \)

\[ g_2 = \begin{cases} pAR_{\text{c}2} & : 0 \leq pAR_{\text{c}2} \leq 4 \\ 4 & : pAR_{\text{c}2} > 4 \end{cases}, \]

(7)

\[ g_3 = \begin{cases} pAR_{\text{c}3} & : 0 \leq pAR_{\text{c}3} \leq 4 \\ 4 & : pAR_{\text{c}3} > 4 \end{cases}, \]

(8)

\[ g_4 = \begin{cases} pAR_{\text{c}4} & : 0 \leq pAR_{\text{c}4} \leq 4 \\ 4 & : pAR_{\text{c}4} > 4 \end{cases}. \]

(9)

\[ BI = \frac{g_2 + 2 \cdot g_3 + 3 \cdot g_4}{2.4}. \]

(10)

The \( BI \) ranges from zero to ten, corresponding to no emphysema bullae \( (BI \to 0) \) and to many emphysema bullae in all size classes \( (BI \to 10) \), respectively.

Another goal was to classify the emphysematous destruction by their type. As emphysema is divided into two main types: bullous emphysema or nonbullous small-sized bullae emphysema, we defined the classification of the \( ET \) as being the relation of large emphysema bullae \( (pAR_{\text{c}2}) \) to small emphysema bullae \( (pAR_{\text{c}3} \to pAR_{\text{c}4}) \)

\[ ET = \begin{cases} \frac{pAR_{\text{c}2}}{s} & : s > 1\% \\ \text{not calculated} & : s \leq 1\% \end{cases} \]

(11)

with

\[ s = \sum_{j=2}^{4} pAR_{\text{c}j}. \]

(12)

The \( ET \) varies from \(-1 \to +1\), corresponding nonbullous to bullous type, respectively. The classifier \( ET \) is only meaningful, if the area of class 2 to 4 \( (12) \) exceeds a certain amount of the \( PI \). This limit is empirically set to 1%, as this area for nonemphysematous patients is less than 1% \( (11) \). In Fig. 5, we give the new algorithm in pseudocode based on standard pascal.
**C. Pulmological and Radiological Expert Opinion**

In order to proof the $PI$ and the new $BI$, particularly in view of difficulties finding gold standards, all patients were diagnosed and the emphysema was scored. We examined two main diseases of the lung: obstructive and restrictive disorders, represented by pulmonary emphysema and fibrosis, respectively.

The diagnosis (normal, emphysema, emphysema with coexisting fibrosis, fibrosis) was done combining pulmological and radiological expert opinion. In the following, the word disease denotes this diagnose. The normal group was comprised only of subjects which both experts rated as being normal. All other subjects were diagnosed additive (i.e., if one expert diagnosed emphysema and the other had a suspicion on fibrosis then this subject was included in the group of emphysema with coexisting fibrosis). These diagnoses were not collected under blinding conditions.

The scoring of the emphysema was done by the radiological expert under blinding conditions only. In consideration of the emphysema definition, Sakai et al. suggested a score for visual assessment. The score was based on the assessment of two aspects of emphysema: severity and extent. Severity and extent were graded on a four-point scale (see Table I).

The emphysema score ($ES$) of all CT images was rated twice, in different weeks with blinding for the observer. For each image, the severity score was multiplied by its extent. The scores yield an emphysema score ranging from zero to 12 [32] (see Table II). An argument in support of this approach is that the score has been used successfully by other research groups [21], [22], [24]. A second radiological expert opinion was taken from the written findings. Thus, this data is not blinded, but it was only taken to evaluate the interobserver variance.

### Table I
SEVERITY AND EXTENT OF THE EMPHYSEMA SCORE

<table>
<thead>
<tr>
<th>$ES$</th>
<th>severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>no emphysema</td>
</tr>
<tr>
<td>1</td>
<td>$d$ LAA &lt; 5 mm</td>
</tr>
<tr>
<td>2</td>
<td>$d$ LAA &gt; 5 mm with</td>
</tr>
<tr>
<td></td>
<td>intervening normal lung</td>
</tr>
<tr>
<td>3</td>
<td>diffuse LAA without</td>
</tr>
<tr>
<td></td>
<td>intervening normal lung</td>
</tr>
</tbody>
</table>

$d$: diameter, LAA: low attenuation areas, $ES$: points emphysema score.

### Table II
MAPPING OF EMPHYSEMA SCORE WITH GRADING

<table>
<thead>
<tr>
<th>emphysema score ($ES$)</th>
<th>grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>normal</td>
</tr>
<tr>
<td>1 - 3</td>
<td>suspicion on emphysema</td>
</tr>
<tr>
<td>4 - 6</td>
<td>emphysema</td>
</tr>
<tr>
<td>7 - 12</td>
<td>severe bullous emphysema</td>
</tr>
</tbody>
</table>

### Table III
BASIC STATISTIC

<table>
<thead>
<tr>
<th>group of disease</th>
<th>$n$</th>
<th>$PI$ [%]</th>
<th>$BI$ [%p]</th>
<th>$RV$ [%]</th>
<th>FEV$_1$ [%p]</th>
<th>TCO [%p]</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>20</td>
<td>18 ± 5.5</td>
<td>1.7 ± 1.2</td>
<td>92 ± 19</td>
<td>101 ± 20</td>
<td>93 ± 41</td>
</tr>
<tr>
<td>emphysema</td>
<td>12</td>
<td>31 ± 17</td>
<td>5.4 ± 2.9</td>
<td>154 ± 62</td>
<td>59 ± 28</td>
<td>56 ± 47</td>
</tr>
<tr>
<td>emphysema +</td>
<td>15</td>
<td>21 ± 9</td>
<td>3.8 ± 2.7</td>
<td>123 ± 32</td>
<td>70 ± 20</td>
<td>75 ± 32</td>
</tr>
<tr>
<td>fibrosis</td>
<td>30</td>
<td>7.6 ± 4.3</td>
<td>0.7 ± 1.3</td>
<td>103 ± 43</td>
<td>80 ± 22</td>
<td>75 ± 41</td>
</tr>
</tbody>
</table>

$n$ denotes the number of subjects, all other numbers are mean ± standard deviation. Threshold $lim _{HU}$ = 930 HU. %p: percent predicted.

### III. MATERIAL

During a two-year period, we performed CT of the lung in 80 patients with normal lung function and obstructive and/or restrictive lung diseases as emphysema or fibrosis. The scanner Somatom plus 4 (Siemens, Erlangen Germany) was used to obtain the CT data. No spirometric gated CT was performed, because of some extremely dyspnoeic subjects [22]. Image parameters are slice 5 mm, table feed 15 mm, 140 kV, 50 mAs, high-resolution convolution kernel. Spirometric tests and clinical data were obtained for all patients [28]. We chose 60 patients selected by a pulmological expert for CT. Additionally we selected 20 subjects with no emphysema, who had CT performed for other reasons. There are subsets with restrictive lung diseases, i.e., idiopathic lung fibrosis (see Table III for numbers). The sex of all patients was equally distributed and the age ranged closely followed a Gaussian normal distribution.

For each subject, six CT scans with 15-mm table feed were selected. The fourth slice was used as a reference at the carina. After segmenting the lung [6], all 480 images were evaluated with both the classical and the new method.

### IV. RESULTS

First, we calculated the means and standard deviations of the $PI$, $BI$, residual volume (RV), forced expiratory volume (FEV$_1$), and the diffusing capacity (TCO) for the given diseases (see Table III). The spirometric data are given as percent predicted. The normal control has all variables within the normal range. The emphysema group does reflect the typical changes seen by obstructive diseases: a little less than double of the $PI$, the quadruples of the $BI$, increase of the RV up to 150% as predicted, reduced FEV$_1$ and a reduced TCO. It has to be stated that the $PI$ for normal and emphysema with coexisting fibrosis are very close, while the $BI$ is still double. All quantitative predictions were based on a medium threshold of $lim _{HU}$ = 930 HU.

This difference has to be tested for significance, which is done with the Mann–Whitney U-test (see Table IV). It is the nonparametric equivalent to the Student $t$-test. Gaussian normal distribution as a precondition for the $t$-test was not provable. As expected, there was a significant difference for all variables except of the $PI$, discriminating emphysema with coexisting fibrosis from normal. Significance level for all statistics was minimum $p < 0.05$.  

---

**TABLE I**
SEVERITY AND EXTENT OF THE EMPHYSEMA SCORE

<table>
<thead>
<tr>
<th>$ES$</th>
<th>severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>no emphysema</td>
</tr>
<tr>
<td>1</td>
<td>$&lt; 25 %$ lung involvement</td>
</tr>
<tr>
<td>2</td>
<td>25 - 50 %</td>
</tr>
<tr>
<td>3</td>
<td>50 - 75 %</td>
</tr>
<tr>
<td>4</td>
<td>$&gt; 75 %$</td>
</tr>
</tbody>
</table>

**TABLE II**
MAPPING OF EMPHYSEMA SCORE WITH GRADING

<table>
<thead>
<tr>
<th>emphysema score ($ES$)</th>
<th>grading</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>normal</td>
</tr>
<tr>
<td>1 - 3</td>
<td>suspicion on emphysema</td>
</tr>
<tr>
<td>4 - 6</td>
<td>emphysema</td>
</tr>
<tr>
<td>7 - 12</td>
<td>severe bullous emphysema</td>
</tr>
</tbody>
</table>

**TABLE III**
BASIC STATISTIC

<table>
<thead>
<tr>
<th>group of disease</th>
<th>$n$</th>
<th>$PI$ [%]</th>
<th>$BI$ [%p]</th>
<th>$RV$ [%]</th>
<th>FEV$_1$ [%p]</th>
<th>TCO [%p]</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>20</td>
<td>18 ± 5.5</td>
<td>1.7 ± 1.2</td>
<td>92 ± 19</td>
<td>101 ± 20</td>
<td>93 ± 41</td>
</tr>
<tr>
<td>emphysema</td>
<td>12</td>
<td>31 ± 17</td>
<td>5.4 ± 2.9</td>
<td>154 ± 62</td>
<td>59 ± 28</td>
<td>56 ± 47</td>
</tr>
<tr>
<td>emphysema +</td>
<td>15</td>
<td>21 ± 9</td>
<td>3.8 ± 2.7</td>
<td>123 ± 32</td>
<td>70 ± 20</td>
<td>75 ± 32</td>
</tr>
<tr>
<td>fibrosis</td>
<td>30</td>
<td>7.6 ± 4.3</td>
<td>0.7 ± 1.3</td>
<td>103 ± 43</td>
<td>80 ± 22</td>
<td>75 ± 41</td>
</tr>
</tbody>
</table>
TABLE IV
MANN–WHITNEY U-TEST

<table>
<thead>
<tr>
<th>disease</th>
<th>PI</th>
<th>BI</th>
<th>RV</th>
<th>FEV₁</th>
<th>TCO</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>61</td>
<td>29</td>
<td>34</td>
<td>28</td>
<td>65</td>
</tr>
<tr>
<td>emphysema</td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>normal – emphysema</td>
<td>144</td>
<td>73</td>
<td>60</td>
<td>44</td>
<td>90</td>
</tr>
<tr>
<td>with coexisting fibrosis</td>
<td>n.s.</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>*</td>
</tr>
</tbody>
</table>

Numbers are U-values. Significance level is given as **< 0.01, * < 0.05 and n.s. not significant. Threshold: \( U_m = 930 \) HU.

In Section III, we present two correlations. We have chosen the Spearman rank correlation coefficient due to ordinal data. For all of the cases, the values of each of the variables are ranked from smallest to largest, and the Pearson correlation coefficient is computed on the ranks. First, the \( BI \) is correlated with the \emph{emphysema score} (see Fig. 6). The Spearman correlation coefficient is 0.88.

Second, we correlated the \( BI \) with the \emph{PI} labeled by the \emph{disease} (see Fig. 7). One can say that there are some outliers with a low \emph{PI} and a high \emph{BI}. These outliers are diagnosed as emphysema with coexisting fibrosis that have a correct high \emph{BI}, but a false low \emph{PI}.

In the following, we present two charts differentiating the \emph{PI}. The bars are the \emph{PI}, defined as the sum of the percentage of bullae in each class [see (6)]. In Fig. 8, we have divided the sample by the \emph{disease}. The increased percentage of bullae in the upper classes in patients with emphysema is obvious. Although the \emph{PI} for patients with a combination of emphysema and fibrosis is less than normal, there is a significant increase of the bigger bullae.

In Fig. 9, we examined the relation of the size classes to the severity of the emphysema given by the \emph{emphysema score} more closely. We believe that the radiological expert rates the arrangement of low-attenuation areas. Hence, the \emph{BI} should correlate with the \emph{emphysema score}, as this score is based on similar features: severity is corresponding to the size of bullae and the extent corresponds to the area, respectively.

Finally, we wanted to examine the observers variance. In order to do so, we calculated the association of two observers gamma, also called Goodman and Kruskal’s gamma, and the Spearman correlation coefficient (see Tables V and VI). The interobserver variance denotes the agreement of the radiological expert at different reading sessions. The interobserver variance characterizes the agreements between two radiological experts.

V. DISCUSSION

In this section, we propose six hypotheses. These hypotheses will be substantiated by our results and discussed taking account of the relevant literature.

**Hypothesis:** The \emph{PI} is a good, well-known measure of emphysema, but it may provide false results when fibrosis coexists.

The \emph{PI} is the sum of all bullae (continuous low-attenuation areas) per lung. In theory, the \emph{PI} would be expected to be composed of many very small bullae corresponding to parenchyma of alveoli and only a few bigger bullae representing bronchi [4]. Emphysematous destruction causes an increase of bigger bullae.
The $PI$ has been successfully used since 1990. It is commonly accepted. Recent publications still use this index [19], [40]. There are many studies which demonstrate that the $PI$ is reliable, in that it differentiates between normal and emphysematous subjects. The $U$-test showed this for our data, too. Emphysematous patients and the normal reference group do have different means of the $PI$, $BI$, RV, FEV$_{1}$, and TCO ($p < 0.01$). These results are comparable to that of other research groups [8].

The sample of subjects used by most researchers suffers only from emphysema without any other diseases. We found few studies dealing with emphysema and coexisting fibrosis. A coexisting fibrosis, however, may lead to two effects: First, we found in patients with fibrotic disease, areas of the lung with no pixels below the threshold, as if there were air-free areas [see Fig. 10(b)]. If one segment is air-free and a large emphysema is in another segment, the $PI$ takes the average of the low-attenuation areas in these two segments. Hence, it may indicate normal values, even if there is an emphysematous destruction.

Second, there is the partial volume effect to be kept in mind. We believe that if a fibrosis coexists it affects mainly the pixels in the periphery of the bullae. Assuming, that the mean density in the interstitial parenchyma is increased by a fibrosis, a small bulla (nine pixels, size class 1) is compared to a bigger bulla (81 pixels, size class 3). If all periphery pixels are affected by the partial volume effect, then the smaller bulla is reduced to 12% and the bigger bulla is only reduced to 61% (see Fig. 11). In general, the partial volume effect is the stronger the smaller the bullae are. This is due to the relation of the circumference $c$ to the area $a$ ($c^2 \sim a$). Applying our method, we found that in a healthy subject, 95% of the bullae are smaller than 5 mm$^2$ (see Fig. 9). It is these small bullae which are affected by the partial volume effect and which again lead to a reduction of the $PI$. The bigger bullae are less affected. Hence, the $BI$ is less sensitive to fibrosis. All in all, the partial volume effect should not be neglected [17] although does not affect all periphery pixels.

**Hypothesis:** The $BI$ enables the diagnosis of bullous emphysema, even if a fibrosis coexists.

The $PI$ may be falsely low, as stated in the first hypothesis. Why is our approach less dependent on coexisting fibrosis?

If we take the subjects of the first hypothesis [see Fig. 10(b)], there is in one lobe emphysematous destruction while the other is nearly air-free because of fibrosis. The $BI$ registers the emphysematous bullae in the right lobe, by virtue of an increase of area in size class 2–4. In view of the $BI$, it does not matter if the left lobe contains bullae in size class 1 (normal) or does not contain bullae in size class 1 (fibrosis). In other words: the $BI$ rates the bullous destruction based on class 2–4, while fibrotic areas mainly reduce the area of size class 1. Hence, the $BI$ is not reduced in the way the $PI$ is.

Having said this, we have to state that the partial volume effect may nevertheless affect the $BI$. As discussed in the first hypothesis, we claim that the partial volume effect reduces the size of the bullae (see Fig. 11). Although the bigger bullae are less affected, it may happen, that bullae are rated in a lower class than before. Hence, the distribution of bullae may shift to lower classes, thus reducing the $BI$.

The second hypothesis is substantiated by our statistical evaluation. For normal subjects and subjects with emphysema and fibrosis, the U-test showed a significant difference for the $BI$, but not for the $PI$ (see Table III). Thus, the $BI$ detects bullous emphysema even if fibrosis coexists.

The $BI$ covers all cases that are detected by the $PI$. The correlation of the $BI$ with the $PI$ is given in Fig. 7. Besides some outliers, there is a good correlation of both parameters. Examining the outliers more closely, one finds that most of them suffered from a fibrotic lung disease on top of the emphysema. This substantiates our assumption, that the $BI$ is less affected by the coexistence of fibrotic lesions.

Our sample represents a typical set of patients sent in by the pulmonological expert. In 13 out of 27 patients with emphysema, fibrotic aspects accompany the emphysematous destruction. This underlines the importance of introducing new methods into image evaluation. There are several papers depicting effort aimed at improving the $PI$. Saitoh et al. examined the lobar distribution of emphysema. They found that the pulmonary function is significantly different between predominantly upper- and
lower-lobe emphysema groups [31]. Similar results are provided by Gierada et al. [10]. In addition, Gierada et al. introduced a ratio of upper PI and lower PI. Uppaluri et al. applied regional analysis, by dividing the lungs into six equal regions, ventral to dorsal, and analyzing each region separately [39]. Nakano et al. studied the regional distribution of emphysema between the inner and outer segments of the lung. They concluded that low-attenuation areas on CT scans are more often found in the inner segment of the lung than in the outer segment, and that the contribution of the inner segment to pulmonary function tests may be greater than the outer segment [27].

All this goes to show that the global PI has to be replaced by more sensitive methods. As Gurney et al. state: ”Emphysema is an inhomogeneous process. If the entire lung is not studied, sampling errors may occur.” [13] We believe that the BI is an alternative way of looking at the lung more precisely, avoiding sampling errors.

_Hypothesis: The new method classifies emphysema into small-, medium-, and large-sized bullae types._

Emphysema can be divided into two main types: bullous emphysema and nonbullous, diffuse emphysema [7], [15]. The radiologist classifies CTs into these types, due to the visual impression. We suggest a measure of this classification, the ET. It is based on the quotient of areas in the classes 2 ($pAR_{c2}$) and 4 ($pAR_{c4}$). If most of the bullae are bigger than 5 mm² but smaller than 40 mm², then a small-sized bullae type is assumed ($ET = -1$). On the other hand, if most of the bullae are bigger than 80 mm² then a severe bullous emphysema with huge confluent air sacs can be assumed ($ET = +1$). An ET of zero unveils a mixed emphysema, with small, medium, and large-sized bullae next to each other. This is a new classification of emphysema varying from $-1$ to $+1$. The classification (ET) is only meaningful if the area of classes 2–4 exceeds a certain value. In order to provide an impression of the ET's explanatory power, we selected three patients depicted in Fig. 12.

The significance of the ET has to be shown. At the moment, we can state that the validity of the ET is high for bullous emphysema. The correlation of the emphysema score ES and ET (for $ET > 0$) is high: $r = 0.90$. It is difficult to proof the validity of the ET for nonbullous emphysema. This can only be done through correlating the ET with macroscopic morphology of the lung. We believe that the ET covers only a subset of all subjects suffering from diffuse emphysema and that texture recognition may support additional information.

_Hypothesis: The visual assessment varies. The interobserver variance is higher than the intraobserver._

This is a well-known problem in the visual assessment of CT images. Our findings of interobserver and intraobserver variance are similar to those of Bankier et al. [2]. This goes to show that the subjective visual assessment should be replaced. Gevenois et al. accurately observe that “the use of HRCT and objective quantitative techniques should take the place of subjective visual scoring methods” [9]. Archer et al. also emphasise the need for new evaluation methods [1].

The main problem of rating CT images is the high interobserver variance, while the intraobserver variance is sufficiently favorable. In addition, all visual scoring systems are time-intensive [41]. Bankier et al. compared the subjective visual grading with objective quantification, including macroscopic morphometry and thin-section CT densitometry. They concluded that systematic overestimation and moderate interobserver agreement may compromise subjective visual grading of emphysema, which suggests that subjective visual grading should be supplemented with objective methods to achieve precise, reader-independent quantification of emphysema. In our study, we found comparable values of observers’ variance (see Table VI) [2].

_Hypothesis: The radiological expert rates morphological aspects of emphysema in CT images, based on the arrangement of the low-attenuation areas. Automated CT image evaluation has to consider this._

The examples depicted in the Section I illustrate how an expert distinguishes between normal and emphysema. An expert is in the position to distinguish, due to rating morphological aspects. The dilation of air sacs is pathognomonic for a typical form of emphysema. These dilated air sacs appear as bullae in CT images. The expert is rating the distribution or homogeneity of these low-attenuation areas, respectively. It would appear, then, that modern algorithms of image evaluation have to consider both features.

The quantification of these subjective morphological features does not offer an easy solution. It seems to be clear, however, that the typical dilation of air sacs may be a key to the problem. Stern et al. describe these aspects and their influence on diagnosis and quantification [35]. In conclusion, it is obvious that the PI does not reflect these morphological aspects. Uppaluri et al. attempt to solve that problem with a texture-based algorithm using image classification with evolutionary programming [23]. This is a promising approach. We chose to extend the classical PI because of its widespread acceptance. There are more articles published dealing with the PI than with other methods. Our approach is not a texture recognition, but it orientates to morphological aspects of emphysema. Both approaches, while
different, are not mutually exclusive. A comparison with the results of Uppaluri’s working group would be most interesting, in particular whether both methods are complementary to one another. There is much to be said on both sides of this question.

There is a high correlation between the PI and the emphysema score ($r = 0.78$). But we found a higher correlation between the BI and the emphysema score. This underlines the hypothesis that the BI quantifies exactly what the radiologist rates. We would like to emphasise that we did not optimize the BI in order to correlate with the emphysema score, nor did we select our sample. From the high correlation we, thus, conclude that the BI may replace the visual scoring.

**VI. CONCLUSION**

We have developed an automated CT image evaluation of the lung based on morphology of the parenchyma. The main feature of our approach is its emphasis on the size of bullous destruction of emphysema rather than the percentage of low-attenuation areas. One can say the BI extends the classical PI.

The BI is reliable as the PI, but covers additionally cases with coexisting fibrosis. This results in a better sensitivity detecting bullous emphysema compared to the PI.

We invented a new measure of the emphysema type. It is differentiating emphysema in nonbullous and bullous emphysema. The evidence for bullous emphysema is substantiated, but further studies are needed to confirm the validity of the nonbullous emphysema type.

The BI and the emphysema type are objective and expressive indexes avoiding observers variance. This is particularly interesting for followup.

**ACKNOWLEDGMENT**

The authors would like to thank H. Brabant and R. Lowe for reviewing the manuscript.

**REFERENCES**


