Identification of emphysema patterns in high resolution computed tomography images

Musibau A. Ibrahim*1, Oladotun A. Ojo2, Peter A. Oluwafisoye2

1Department of ICT, Osun State University, Osogbo, Nigeria
2Department of Physics, Osun State University, Osogbo, Nigeria

Received: November 1, 2017    Accepted: November 30, 2017    Online Published: December 27, 2017
DOI: 10.5430/jbei.v4n1p16
URL: https://doi.org/10.5430/jbei.v4n1p16

ABSTRACT

Fractal dimension (FD) is a very useful metric for the analysis of image structures with statistically self-similar properties. It has applications in areas such as texture segmentation, shape classification and analysis of medical images. Several approaches can be used for calculating the fractal dimension of digital images; the most popular method is the box-counting method. It is also very challenging and difficult to classify patterns in high resolution computed tomography images (HRCT) using this important descriptor. This paper applied the Holder exponent computation of the local intensity values for detecting the emphysema patterns in HRCT images. The absolute differences between the normal and the abnormal regions in the images are the key for a successful classification of emphysema patterns using the statistical analysis. The results obtained in this paper demonstrated the effectiveness of the predictive power of the features extracted from the Holder exponent in the analysis and classification of HRCT images. The overall classification accuracy achieved in lung tissue layers is greater than 90%, which is an evidence to prove the effectiveness of the methods investigated in this paper.

Key Words: Multi-fractal analysis, Holder exponent, Fractal dimension, Emphysema identification, HRCT images, Classification analysis

1. INTRODUCTION

Automatic identification of emphysema patterns in HRCT images could pose a very big challenge due to the irregular distributions and the complexity of this pattern in the images. Most of the previous work that uses the emphysema database was based on classification of different classes of emphysema patterns, but the quantity and locations of the patterns in an image is another very important challenge that should be investigated.[1–3] Much work has been done on the region of interest (ROI) identification by using the general or classical approaches.[4, 5] In this case, the computation of the alpha values generated by the Holder exponent of the pixel intensity of the original HRCT images would be applied for automatic identification of ROI in the images. The details on the computation of the Holder exponent of the original image with size $M \times N$, which generates the corresponding alpha values that eventually translate to alpha-images, can be found in the previous work.[6–9]

High resolution computed tomography (HRCT) is a very effective scanning method that is more sensitive than chest radiography in measuring the type, extent and distribution of emphysema. Some patients with early emphysema, particularly those with early disease might still have the symptoms and this is the stage where HRCT is most useful.

*Correspondence: Musibau A. Ibrahim; Email: Kunle_ibrahim2001@yahoo.com; Address: Department of ICT, Osun State University, Osogbo, Nigeria.
Emphysema in HRCT is characterized by the presence of areas of abnormally low attenuation, which can be easily contrasted with surrounding normal lung parenchyma\cite{3,10}. Emphysema can be classified into three different classes: (i) centrilobular emphysema (CLE), defined as multiple small low-attenuation areas, (ii) paraseptal emphysema (PSE), defined as multiple low-attenuation areas in a single layer along the pleura often surrounded by interlobular septa that is visible as thin white walls, and (iii) panlobular emphysema (PLE), defined as a low-attenuation lung with fewer and smaller pulmonary vessels. Examples of CLE and PSE, as well as normal tissue (NT) as seen in HRCT scans, are shown in Figure 1.

![Figure 1. HRCT images showing emphysema disease patterns. The images in the first column were obtained from Radiopaedia\cite{11}. Histogram equalization is used to enhance the contrast, and emphysema regions are indicated with a red arrow in the second column.](image)

We have not considered panlobular emphysema (PLE) in our study as PLE is often an advanced stage of a CLE\cite{12}.

Previous studies have shown several ways by which the CT imaging could be used for early detection of emphysema in patients with COPD\cite{12}. HRCT of early centrilobular emphysema (CLE) shows uniformly distributed tiny areas of low attenuation while in panlobular emphysema; HRCT shows either panlobular low attenuation or ill-defined diffuse low attenuation of the lung. Paraseptal emphysema (PSE) is characterized by subpleural well-defined spaces\cite{13}.

The paper by Masashi et al\cite{13} comprehensively described emphysema diseases, different classification of emphysema and their properties. It is also demonstrated that features from HRCT images could form important quantitative measures for identifying different classes of emphysema and the assessment of their severity.

The alpha value computation could be a very suitable measure to detect and identify an emphysema pattern in a particular location as it reveals the local behavior or characteristics of the images using the pixel intensity distributions. It can be used to resolve the irregularities present in biomedical images, which makes the patterns identification approach become easier. An alpha image has a very powerful and important feature that is very useful to uniquely describe the local characteristic properties of the image. It has an index to measure the local complexity of the image tissue, which can be characterized by the parameters like the minimum and the maximum of the alpha values, the interval between the minimum and the maximum, the differences in the mean alpha values plus the corresponding standard deviation. Furthermore, the important characteristic parameters listed could be extracted from the patches and the HRCT images for the classification purposes. Additionally, the average values of the class intervals of the alpha intensity values can be calculated by dividing the alpha values such that the
mean alpha values of each class interval would be measured in terms of the frequency distributions of the alpha belonging to each class for further analysis and investigations of the emphysema patterns.

The applications of fractal dimensions have been previously used to locate and identify ROI in various digital images.\textsuperscript{[14–17]} The complexities of the images within different locations can be varied with different measures of fractal dimensions; the regions or sections with the emphysema in the CT images, for instance, are expected to be more complex than the other regions without, and would have higher values of fractal dimensions.

Fractal dimension has also been used to diagnose patients with an Alzheimer disease. It was discovered by Gómez et al.,\textsuperscript{[18]} that the Higuchi fractal dimension of the signals from the patients with Alzheimer has lower FD values compared to the control subjects recorded. There is a statistical significant difference between the Higuchi’s dimension values of the patients with an Alzheimer disease and that of the elderly control subjects. In Paramanathan et al.,\textsuperscript{[19]} the Higuchi’s fractal dimension of the waveform was calculated and the criteria for determining the minimum and maximum values of the interval was established in order to improve the computational accuracy and speed.

This paper therefore proposes the implementation of Holder exponent of the image intensity for efficient recognition of the emphysema patterns. In this paper, the proposed system would focus on detecting the emphysema patterns based on the fractal properties of the image textures as highlighted above. The proposed ideas can be achieved in various ways; the absolute differences in the overall mean values of the alpha values between the normal tissue images and the histological images could be used. The regions with the statistically significant absolute mean differences are more likely to contain the emphysema patterns why the regions without significant differences would definitely have no emphysema patterns. The same procedure can be used in the analysis of Higuch’s dimension such that the absolute differences in the dimension between the normal and other classes would reflect the quantity of pattern detections.

The Holder exponent of the local distribution of the intensity measure has been previously implemented and applied for the development of multi-fractal spectrum in CT images.\textsuperscript{[5–8, 20–22]} The major difference between the previous approaches and the proposed idea is that the previous work applied a global descriptor (multi-fractal spectrum) for identifying different areas in biomedical images, but this paper would use the local descriptor (Holder exponent) of the CT images for detecting the emphysema patterns. The computation of a multi-fractal descriptor depends on the calculated Holder exponent, which has been successfully used as a global descriptor for efficient classification in biomedical images, but since we are focusing on identifying the patterns within the images; the Holder exponent could be more appropriate for this task.

2. METHODS FOR IDENTIFYING EMPHYSEMA PATTERNS

An intensity measure $\mu_P(r)$ in the neighborhood of a pixel $P$ can be defined as a function $f$ of intensities within a square window $W$ of size $r$, which is centered at $P$. Using the values of the intensity measures $\mu_P(r)$ described in the studies,\textsuperscript{[6–8, 20–22]} we can explore how they scale with $r$ and establish the power laws satisfied by them.

$$\mu_P(r) = Cr^{\alpha_P}, \alpha > 0$$  \hspace{1cm} (1)

where $C$ is a constant of proportionality and $\alpha_P$ represents the Holder exponent for the power law of the measure under consideration at the point $P$.

![Figure 2](image-url)
As in the case of fractal dimension, the value of $\alpha_P$ is calculated as the slope of the linear regression line of the log-log plot where $\log(r)$ is plotted on the $x$-axis and $\log(\mu)$ along $y$-axis. The linear regression lines and the corresponding values of $\alpha_P$ for the graphs are shown in Figure 2.

Since for every $P$, we get a value $\alpha_P$ of the Holder exponent for the chosen measure, we can form a rectangular array of values $\alpha_P$ having the same size as the original image. This two-dimensional array is called the $\alpha$-image.

The range of $\alpha$ values $[\alpha_{\text{min}}, \alpha_{\text{max}}]$ computed from an image is further subdivided into a set of $n$ discrete intervals $[\alpha_i, \alpha_{i+1}]$, $i = 0, 1, ..., n - 1$, where

$$\alpha_i = \alpha_{\text{min}} + (\frac{\alpha_{\text{max}} - \alpha_{\text{min}}}{n})i$$  \hspace{1cm} (2)

The number of pixels having $\alpha_P$ values in each of these subintervals is counted, and this computational process yields the $\alpha$-histogram of the $\alpha$-image. Further, the subdivision produces different sections of the $\alpha$-image. The pixel belonging to a subinterval is called an $\alpha$-slice. The $\alpha$-slices give a mutually disjoint partitioning of the $\alpha$-image. Since each $\alpha$-slice contains pixels belonging to a narrow range of $\alpha$ values; we can view an $\alpha$-slice as the set of pixels that have similar intensity variations (with the same Holder exponent) within their neighborhoods. If we treat $\alpha_P$ as a local feature representing intensity variations within the neighborhood of $P$, the $\alpha$-histogram provides the relative strengths of each of these features, and is in itself an important feature descriptor.

The summary of the general procedures and statistical analysis for identifying and detecting the emphysema patterns in HRCT images using the Holder exponent approach is presented in Figure 3. The Figure 3 presents the Holder exponent computation of each pixel intensity value of the image based on the methods described in the previous work.[6–9]

![Image](image.png)

**Figure 3.** Overview of emphysema pattern detection of the HRCT images using the Holder exponent computation

There are several ways or methods of extracting the ROI in an image. The general overview of the procedures involving the detection of the emphysema patterns using the absolute differences in the alpha values is presented in Figure 4. The alpha values obtained from the computation are used for further analysis in detecting the patterns. For a classification process to be successful, it is good for the images in the same group to exhibit minimum intra-class variation while the images across the groups should possess large inter-class variation. For this reason, some preliminary statistical tests were carried out in order to determine how significant the differences between the images across different emphysema classes are and how insignificant or how similar the images of the same emphysema classes within the ground truth patches are. The details of the experimental tests would be discussed later in this paper.

Further analysis of the experiments has been demonstrated by generating different models to display a visual representation of the images, the differences in the images from the normal ground truth patches and so on. Further statistical investigations are used to identify the locations of the images with the emphysema patterns, but how this is achieved would
be explained later. The ground truth approach is a terminology used to draw an inference from the information provided after preliminary tests or observation.

The information is extracted from the tested images and compared with the standard or pre-labelled images in order to determine the similarities and differences in their features. The ground truth can be done by measuring the characteristic features of the pixel intensity values compared with image sub-division of the HRCT image. In other words, the ground truth images are those images whose identity and characteristic features are used as the training sets to test other images with unknown identity and features. Additionally, it could be used to establish a confusing matrix, which validates the accuracy of the classifier that would minimize the error level and provide the best results for the datasets. How this approach is used to establish the relationship between the ground truth patches and the HRCT image slice would be discussed in the following sections.

The local characteristic of the alpha intensity values has been used for the computation of analysis of variance (ANOVA) for each alpha image. The experimental results at this stage demonstrate that the images from the same patch class have minimum intra-class variance as the deviations in their group means are not statistically significant using the ANOVA. This process is repeated for the HRCT images of the same emphysema classes and the inter-class across the other classes. The results also show that the images from the same slice class have little or no variations in group means while the images across other classes have significant differences in their group means with about 95% confidence interval and a \( p \)-value < .0001. The multiple comparison tests are also applied to the datasets obtained from the patches and the slices, after the one way analysis of variance, which compare the means of several groups to test the hypothesis that the samples are all the same or against the general alternative that they are not all the same.

The alpha values of the HRCT images computed by the implementation of the Holder exponent have been applied for measuring the differences between the ground truth images and the HRCT image slices. Also, absolute FDs differences across different regions between the normal tissues and histological images are calculated, and applied for the ROI detections. The inferences are drawn based on the results of the statistical analysis and the important components of the slice images are used to validate the performances of the algorithms. The findings from the ground truth images, which stated that the patches are derived from the HRCT images, have been a very useful measure for developing an efficient classification system for emphysema pattern detections.

In this experiment, three patch images were selected randomly from each lung type and three image slices were also selected for comparison using the calculated alpha intensity values. The overall mean of each sub-divided image in the slice is calculated for each emphysema class. Each sub-divided image in the slice was compared with the ground truth patches in order to validate the label on the ground truth images.

3. Results and Discussion

The methods previously discussed have been implemented as illustrated in Figure 4 for accurate detection of emphysema patterns. Each image slice with the pixel intensity of 512 × 512 is subdivided into 64 sub-images. The 64 sub-images are arranged in the form of 8 × 8 matrix, and number from left to right and top to bottom such that the matrix location in the first row, first column is equivalent to the first image while the sub-image at the matrix location in the 8th row, 8th column is equivalent to the sub-image number 64 and so on.

The aim at this stage is to focus on the comprehensive analy-
sis of the sub regions within the images in order to identify the locations of the emphysema patterns. The absolute mean differences of the calculated alpha-values generated using the Holder exponent approach between the normal tissues and the emphysema images, and the results of paired samples t-test were used to determine the locations of the emphysema patterns. The results of the experiments, which show the group means and the corresponding p-values for each experiment that was investigated, are shown in Table 1.

Table 1. Intra-Class variation between the patches and corresponding sub-image slices

<table>
<thead>
<tr>
<th>Patches/Slices</th>
<th>Mean1</th>
<th>Mean2</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT1-NT1</td>
<td>1.9915</td>
<td>1.9949</td>
<td>.2947</td>
</tr>
<tr>
<td>CLE1-CLE1</td>
<td>1.9915</td>
<td>1.9961</td>
<td>.2907</td>
</tr>
<tr>
<td>PSE1-PSE1</td>
<td>1.9925</td>
<td>1.9965</td>
<td>.2919</td>
</tr>
</tbody>
</table>

In each group, it has been established that the values of the group mean between the emphysema class of the ground truth images, and the corresponding sub-images extracted from the slices overlap each other. The corresponding p-values in each test show that there are no significant differences between the images. The numbers in Table 1 indicate that the means of the NT of the ground truth patches minus the mean of the corresponding sub-image slice is estimated to be -0.0033 and a 95% confidence interval for the true difference of the means is [-0.0096 0.0029]. In this test, since the confidence interval contains 0.0, it shows that the difference is not significant at the 0.05 level.

The mean alpha values for some of the sub-divided images in the three classes of emphysema are shown in Table 2, the few selected images in the Table 2 are presented in order to demonstrate those locations with the emphysema patterns and those without. The mean alpha values in each class are also used to generate the linear regression models using the smoothing algorithms as this would help to determine the relationship between the classes, and also to visually display how the normal tissues are deviated from other classes. That is, the absolute mean differences between normal tissues and CLE pair classes and the NT-CLE pair classes are calculated for each sub-divided image in order to determine the regions with the emphysema patterns. For instance, in Table 2, the absolute mean deviations between the normal tissues and the CLE and PSE classes revealed that few regions contain the emphysema patterns while others do not according to the paired samples t-test results. The paired t-test results showed that the differences between the normal and emphysema images are statistically significant with a p-value < .001 and 95% confidence interval on the mean that does not contain 0 only in those regions with the emphysema patterns. However, the p-values are higher than the significant level of .05 in those regions without emphysema patterns.

Table 2. Absolute mean deviations of CLE and PSE from the normal tissues

| Sub-images | NTmean | CLEmean | PSEmean | |NT — CLE| |NT — PSE| |
|------------|--------|---------|---------|----------------|----------------|----------------|
| Image1     | 1.9953 | 1.9949  | 1.9949  | 0.0004         | 0.0004         |
| Image2     | 2.0242 | 2.0491  | 2.0171  | 0.0249         | 0.0071         |
| Image3     | 2.0818 | 2.0783  | 2.0596  | 0.0035         | 0.0222         |
| Image4     | 2.0168 | 2.0117  | 2.0369  | 0.0051         | 0.0201         |
| Image5     | 2.0738 | 2.1114  | 2.131   | 0.0376         | 0.0572         |
| Image6     | 2.0789 | 2.1072  | 2.1265  | 0.0283         | 0.0476         |
| Image7     | 2.0774 | 2.1153  | 2.1283  | 0.0379         | 0.0509         |
| Image8     | 2.0622 | 2.1119  | 2.1202  | 0.0497         | 0.058          |

Take for instance, the Table 2 showed that the p values of the results from image1 to image4 are higher than the significance level and hence not statistically significant, but the absolute mean deviations from emphysema images between image5 and image8 have lower p values (p value < .0001) than the significant level of .05. These findings are also in line with the results in Figure 5 when the generalized regression models of the mean deviation values between the normal tissues and CLE classes are used. The results of the absolute mean deviations of the normal tissues from other emphysema classes against sub-images are presented in Figure 5.

In this case, sub-images in the range of 52 to 56 have higher deviation values compared to other regions even with a p value < .0001 and 95% confidence interval for the mean. This indicates the presence of emphysema patterns since these particular regions have been verified by the paired sample t-test that the differences are statistically significant. These significant differences appear in almost the same regions for both CLE and PSE with just little differences. Other deviations across the regions in the graph do not show significant differences when tested, which implies, these regions do not contain the presence of emphysema patterns.
The presence of emphysema patterns in those regions with higher mean deviations of the Holder exponent, which is the alpha values, determines the intensity distributions of the HRCT images and how complex a particular region is. This complexity in the image tissues is associated with the mean deviations or differences between the normal and the abnormal tissues. In other words, the higher the absolute mean deviations from the normal tissues the larger the presence of emphysema patterns and vice versa. This strategy has really demonstrated to be very effective in terms of identifying the regions of interest in the HRCT images.

The next experiments would be to determine how effective this method could be, and the performance level of the classification accuracy that can be accomplished. This can be achieved by validating this approach using the ground truth as the training sets as we have discussed previously and the sub-regions in each lung tissue type for testing the classification system.

### 4. CLASSIFICATION APPROACH OF THE HRCT IMAGES

The robustness of the algorithm developed in this section was evaluated by randomly selecting 42 images in each emphysema class from the ground truth patches for training purposes. The original patch of size $61 \times 61$ pixels has been re-scaled into the same size with the slice image ($512 \times 512$ pixels). 6 image slices were randomly selected from each of the three classes to test the performances and validity of the algorithms with different classifiers but using the summation intensity measure.

The classification results for the image slice are presented in Table 3. The datasets of the images after the computation of the Holder exponent, which resolve the 16 bit intensity values to the normalized alpha values as the pixel intensity, are used for the classification purposes.

<table>
<thead>
<tr>
<th>Confusion matrix</th>
<th>NT</th>
<th>CLE</th>
<th>PSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT</td>
<td>95</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>CLE</td>
<td>5</td>
<td>91</td>
<td>4</td>
</tr>
<tr>
<td>PSE</td>
<td>5</td>
<td>3</td>
<td>92</td>
</tr>
</tbody>
</table>

The purpose of this section is to check the classification accuracy of the local Holder exponents of the images in terms of identifying the regions with or without the emphysema patterns. The classifiers would determine the level of accuracy of the local features extracted from the calculated alpha images using only the summation intensity measures at this stage. As presented in Table 3, the classification accuracy of over 90% is an evidence to demonstrate the predictive capability of the classifier and the datasets used.

The first comparison is conducted between the confusion matrix in Table 3 achieved by the Holder exponent approach and the feature selection approach for improving the classification accuracy in emphysema CT images in the PhD thesis.[22] The datasets are the combined features extracted from the alpha-histogram and multi-fractal datasets of the
Our methods in this study also extract the features using the local descriptors of the CT emphysema images. Our methods in this study also extract the features using the local descriptors of the CT emphysema images. The comparison of the classification accuracy between the two methods are shown in Table 4.

Table 4. Comparison of classification accuracy between the proposed method and existing results

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Bagged decision tree algorithm</th>
<th>Our Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF without SFS</td>
<td>CF with SFS</td>
<td></td>
</tr>
<tr>
<td>NT</td>
<td>CLE</td>
<td>PSE</td>
</tr>
<tr>
<td>NT</td>
<td>91</td>
<td>0</td>
</tr>
<tr>
<td>CLE</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>PSE</td>
<td>9</td>
<td>17</td>
</tr>
</tbody>
</table>

In Table 4, the binary decision tree (BT) algorithm achieved the best classification results compared with the Naive Bayes as presented in the article. As can be seen in Table 4, the proposed method achieved a classification accuracy of 92.27%, this result outperforms the BT technique, which has the classification accuracy of 86% and 88.66% for feature selection (FS) and without FS respectively. This is a very good accomplishment as the difference in classification accuracy between the existing results and the new results are statistically significant with a p-value < .0001. The same results obtained by the proposed approach are then compared with the state-of-the-art methods by Sorensen et al. as shown in Table 5 since the authors also used the same emphysema database.

Table 5. Comparison of classification results between the proposed methods and recently published work

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Bagged decision tree algorithm</th>
<th>Our Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP2</td>
<td>GFB2</td>
<td></td>
</tr>
<tr>
<td>NT</td>
<td>CLE</td>
<td>PSE</td>
</tr>
<tr>
<td>NT</td>
<td>93</td>
<td>0</td>
</tr>
<tr>
<td>CLE</td>
<td>2</td>
<td>98</td>
</tr>
<tr>
<td>PSE</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5 shows that our methods are comparable with the best two approaches, although the classification accuracy in LBP2 is slightly greater than our methods but our result provides about 95% classification accuracy for the normal tissue, which is a good thing compared to the 93% that they have got in LBP2 and GFB2 methods.

5. CONCLUSIONS

The research work in this paper has presented some of the applications of Holder exponents in CT emphysema images. The alpha values obtained from the Holder exponent computation showed the effectiveness of the local intensity distributions for identification and detection of ROI in HRCT images. The useful local features extracted from the image textures showed excellent performance. The methods demonstrated to be a very good measure and accurate enough to separate the regions with emphysema patterns from the regions without emphysema patterns. The classification result of over 90% accuracy in the lung tissue types is an evidence to prove the effectiveness and reliability of the developed systems. The results achieved so far have shown that both local features from the images and the global features from the calculated Higuchi fractal dimension could be very effective in the classification and analysis of the regions of interest.

Further research work could be done in this area by applying better methods in terms of accuracy and efficiency for calculating the fractal dimension of the CT images or improving the computational accuracy and time complexities of the algorithms. This research approach can also be further extended by applying Higuchi’s method in a multi-fractal framework, such that the multi-fractal spectra of digital images can be calculated using the Higuchi’s method and this can probably improve the recognition accuracy.

ACKNOWLEDGEMENTS

The authors would like to sincerely thank Associate Professor Mukundan, Department of Computer Science and Software Engineering at the University of Canterbury for his substantial contributions towards this study.
REFERENCES


