

# Fractal analysis in histology classification of non-small cell lung cancer

Citation for published version (APA):

Patil, R., Srinidhi Bhat, G. M., Wee, L., & Dekker, A. (2019). Fractal analysis in histology classification of non-small cell lung cancer. In *Medical Imaging Artificial Intelligence, Image Recognition, and Machine Learning Techniques* (pp. 62-73). CRC Press, Taylor & Frances Group.

## Document status and date:

Published: 01/01/2019

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# 4

## *Fractal Analysis in Histology Classification of Non-Small Cell Lung Cancer*

Ravindra Patil, Geetha M, Srinidhi Bhat, Dinesh M.S,  
Leonard Wee, and Andre Dekker

### CONTENTS

4.1 Introduction.....	63
4.2 Methodology .....	65
4.2.1 Image Analysis.....	65
4.2.2 Computation of Fractal Dimension.....	66
4.2.3 Extraction of Radiomics Features.....	67
4.2.4 Classification.....	68
4.2.5 Results .....	69
4.3 Conclusion .....	71
References.....	72

### 4.1 Introduction

In both sexes combined, lung cancer is the most commonly diagnosed cancer (11.6% of the total cases) and the leading cause of cancer death. The total number of lung cancer cases in 2018 alone amounted to 2,093,876, the number of deaths with lung cancer being 1,761,007. Non-small cell lung cancer (NSCLC) accounts for 85% of all the lung cancers [1]. The cause of illness and the survival of NSCLC subjects vary across age, genetic profile, size of tumor, and histopathology of tumor. There are various studies that have established a correlation between the subtypes of NSCLC (squamous cell carcinoma, large cell carcinoma, adenocarcinoma, and “not otherwise specified”) to the patient’s survival. Also, it was studied that the prognosis for adenocarcinoma is poor compared to those for non-adenocarcinoma [2]. It was also concluded that surgical management should be different for each sub-category of NSCLC [3]. The current approach of subtype detection is performed using a biopsy procedure, where the tissue under observation is biopsied to determine the subtype, which is invasive in nature. The invasive approach is painful, costly, and not devoid of complications [4]. In recent

times, several studies have been undertaken to identify the sub-categories of NSCLC non-invasively using radiomics, wherein large amount of quantitative features are mined and decision support models are built to achieve the desired objective [5]. Lately, radiomics has been applied to several medical problems such as tumors of lung, breast, and prostate, and also to images extracted from different medical imaging techniques (computed tomography (CT), magnetic resonance (MR), and positron emission tomography (PET)) [6–9, 10–13], showing promising results in each case.

Also, there has been lot of interest in the application of fractals in the oncology domain. Fractals are mathematical objects that have a non-integer dimension. These objects manifest a repeating pattern at different size scales; this property is quantified by a parameter named fractal dimension, which measures the self-similarity grade of the structure under analysis [14]. Such mathematical objects can be self-similar and resemble the repeated pattern within itself. These patterns have been studied in the oncology domain to differentiate between malignant and benign tumors in case of breast cancer. There has been comprehensive reviews on the use of fractal dimensions in various medical research areas, such as pathology, as shown by the literature [15, 16]. Recent trends show fractals to be a useful measure of the pathologies of the vascular architecture, tumor/parenchymal border, and cellular/nuclear morphology. A procedure that combines fractal and segmentation analyses has been proposed to investigate heterogeneity in cancer cells on MR images, similar to the approach described by Szigeti et al. for studying lung tumor heterogeneity on mice CT scans. More details on the fractals and their applications in oncology can be found in Baish and Rakesh [15].

Other major studies include different fractal measures, such as the power-law behavior of the Fourier spectrum of gray-scale images, which was used by Heymans et al. [17] to characterize the microvasculature in cutaneous melanoma. Fractal dimension quantified the degree of randomness to the vascular distribution, a characteristic that cannot be easily captured by the vascular density. Another study by Cusumano et al. aimed to use a fractal-based radiomic approach to predict the complete pathological response after chemotherapy in locally advanced rectal cancer (LARC). Fractal played an important role, giving information not only about the gross tumor volume (GTV) structure, but also about the inner populations in the GTV [18].

In this study, we investigate the role of fractals in the classification of NSCLC histology. We follow the box-counting approach, which aims to overlay boxes of various sizes over the area of interest and to optimize this quantity with the size of each box. Overall, our technical contributions are:

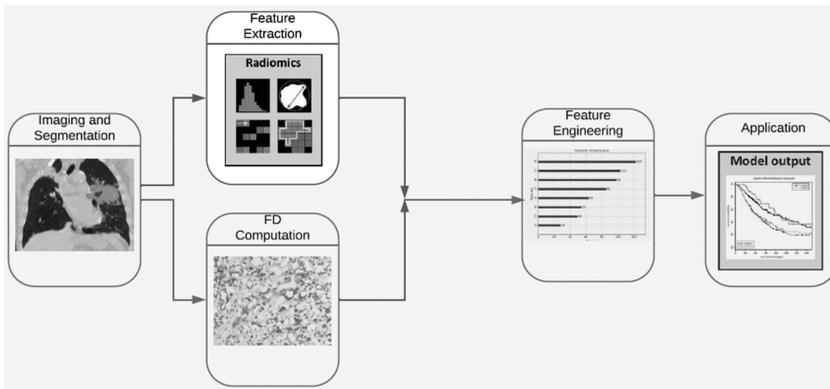
1. Building an algorithm which can compute the fractal dimension of a 2D region of interest in a given volume
2. Analyzing the extent to which fractal dimension aids in classification of NSCLC subtypes with the presence of other radiomic features

## 4.2 Methodology

The methodology adapted in this study consists of acquisition of NSCLC images with varied histology subtypes ( i.e. squamous cell carcinoma, large cell carcinoma, adenocarcinoma, and “not otherwise specified”), through segmentation of the GTV, computation of fractals, extraction of radiomic features and validation of the model to identify the histology of the underlying tumor. The pictorial depiction of the workflow is shown in Figure 4.1

### 4.2.1 Image Analysis

The following experiment involved computerized tomography (CT) images of NSCLC. Images from 317 patients were used in this study, and the dataset was obtained from the collection of NSCLC radiomics in <http://www.cancerimagingarchive.net/> [9]. The images were in digital imaging and communications in medicine (DICOM) format, and the CT image of each patient had the corresponding structural report file (RTSTRUCT), which consists of GTV delineation performed by a team of expert radiologists. The distribution of the demographic of data is depicted in Table 4.1.



**FIGURE 4.1**  
Adapted workflow and approach.

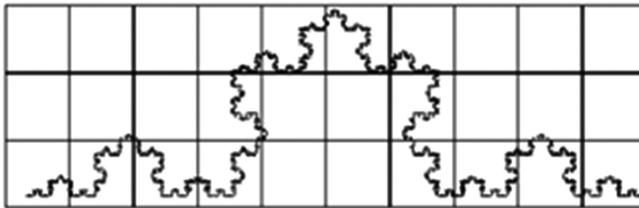
**TABLE 4.1**  
Subject Demographics

Subject Characteristics	Adenocarcinoma	Large Cell Carcinoma	Squamous Cell Carcinoma	NOS
Number of subjects	40	108	110	59
Male	20	65	70	41
Female	20	43	40	18
Mean Age (years)	67.2	66.9	70.2	65.6

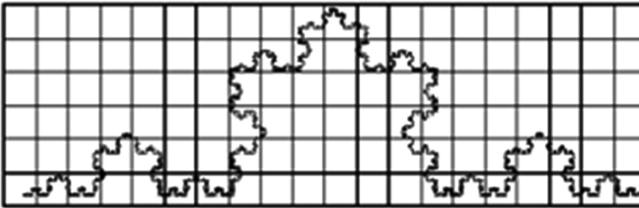
Accessing the DICOM tags, which contained the contour information of the GTV, using the information in the RTSTRUCT file, we created a mask depicting the GTV. The extracted mask was superimposed on the actual image to delineate the region of interest. Further, the minmax normalization of the images was performed to minimize the effects of spike pixels.

#### 4.2.2 Computation of Fractal Dimension

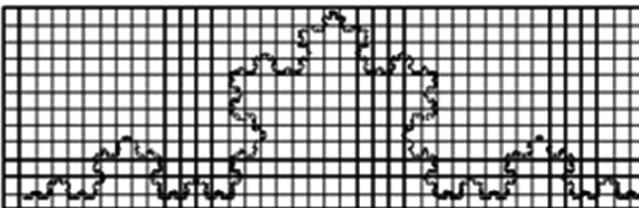
The fractal dimension was computed based on the box counting approach and was further optimized to deduce the fractal dimension for each region of interest. In our approach, we extracted every surface contour of the GTV, and a box-counting algorithm was applied to compute the fractal dimension (FD) of each of the slices containing the tumor volume. The pictorial representation of the approach can be seen in Figure 4.2 with varied grid sizes [19].



The Koch curve with unit 1 grid size, with 18 containing the curve.



The Koch curve with unit 1/2 grid size, with 41 containing the curve.



The Koch curve with unit 1/4 grid size, with 105 containing the curve.

FIGURE 4.2

Sample representation of box-counting approach with varied  $N$  [19].

In this study, the FD was computed on the impinging GTV on the CT slices. The equation for the computation is mentioned in (4.1).

$$FD = \frac{\log(N)}{\log\left(\frac{1}{r}\right)} \quad (4.1)$$

Where:

$FD$  is the fractal dimension

$N$  is the number of boxes needed to cover the region of interest

$r$  is the size of each box

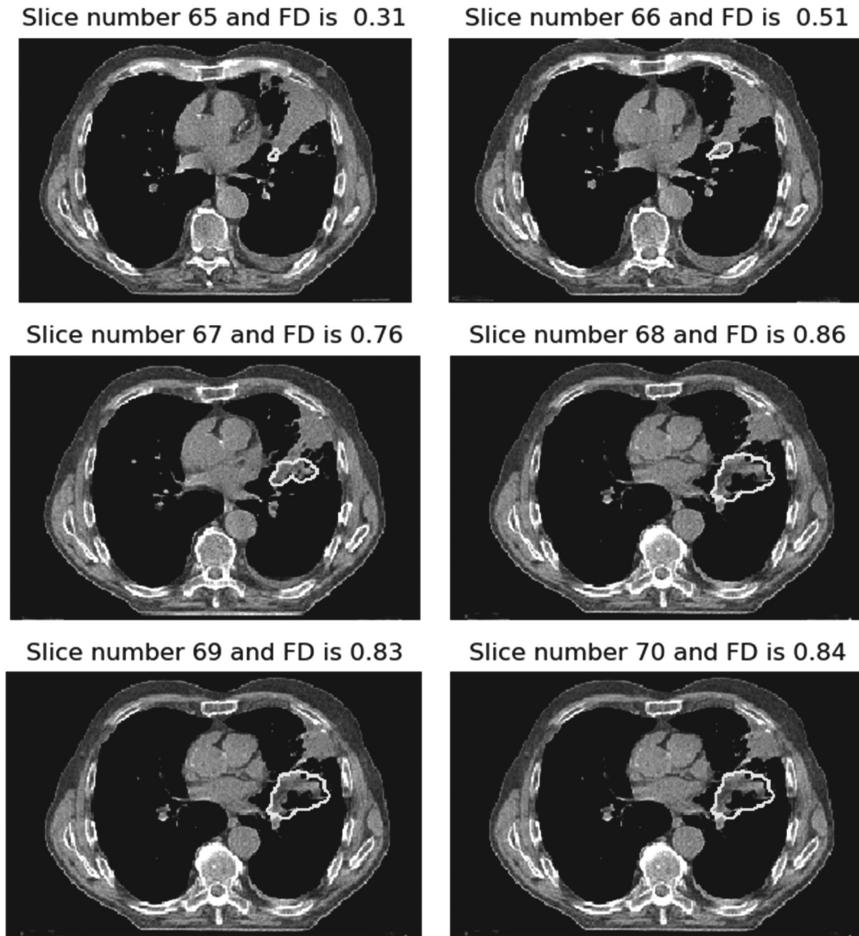
The above equation is recursively applied by varying the size of the box, thereby converting it into a curve-fitting solution. In principle, this corresponds to a line-fitting problem, where points corresponding to  $N$  and  $1/r$  are fit and the slope of line provides the FD. This is one of the reasons we have applied a logarithm to Equation 4.1. This reduces a curve-fitting method to a line-fitting method, which is computationally simpler to solve. Also the logarithm, being a monotonic function in nature, doesn't alter the behavior of the original equation. The FD values for sample subsets are shown in Figure. 4.3, computed based on each of the slices.

### 4.2.3 Extraction of Radiomics Features

The quantitative image features were extracted from the GTV, wherein these imaging features were divided into four sub-categories: (1) first order statistics, (2) textural features, (3) shape- and size-based features, (4) wavelet features. The first order features provide the voxel intensity distribution within GTV. The textural features are computed using gray-level co-occurrence and gray-level run-length texture matrices, which aid in providing the relative position of various gray-level distributions. Shape and size features provide information on how spherical, elongated, or rounded the tumor is, and also about area, tumor compactness, and tumor volume. The wavelet features provide information by decoupling the region into high and low frequencies with GTV as input. In this approach, the original images are decomposed into eight levels of wavelet decomposition ( $X_{LLL}$ ,  $X_{LLH}$ ,  $X_{LHL}$ ,  $X_{LHH}$ ,  $X_{HLL}$ ,  $X_{HLH}$ ,  $X_{HHL}$ , and  $X_{HHH}$ ), where L and H are low pass and high pass. For example,  $X_{LHL}$  is interpreted as the low-pass sub-band resulting from directional filtering in  $\times$  with low pass, high pass along the  $y$ -direction, and low pass in the  $z$ -direction.

$$X_{LLH}(i, j, k) = \sum_{p=1}^{N_H} \sum_{q=1}^{N_L} \sum_{r=1}^{N_H} H(p)L(q)H(r)X(i+p, j+q, k+r) \quad (4.2)$$

where  $N_H$  is the length of filter H and  $N_L$  is the length of filter L.

**FIGURE 4.3**

Subject Lung001 from the NSLC dataset with FD-computed value.

In total, 431 Radiomics features were extracted, and these formed the feature vector for each of the subjects. Further, to this feature vector, max FD, average FD obtained from the fractal dimension computation was augmented, making it 433 feature vector for each subject.

#### 4.2.4 Classification

Two data models were built in this experiment, one with all radiomic features, including extracted fractal features, and the other with only radiomic features, excluding the fractal features. A random forest classifier (RFC) was used to model the multiclass classification problem of predicting the histology of the tumor into one of the following sub-categories: squamous cell

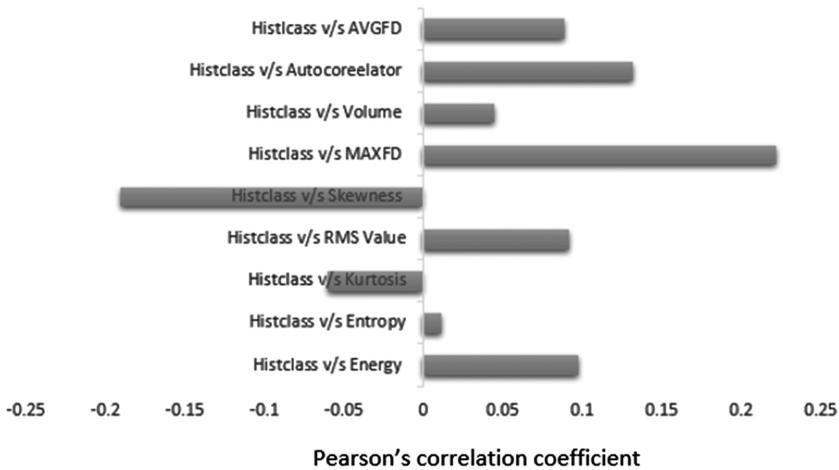
carcinoma, large cell carcinoma, adenocarcinoma, and “not otherwise specified”. The RFC was implemented using the sklearn ensemble package in python and, to obtain the best set of parameters, we defined a grid of hyper-parameter ranges using the Scikit-Learn’s RandomizedSearchCV package by performing 10-fold cross-validation with each combination of values. The hyper-parameters and the values of the RFC tuned in this experiment were:

- `max_features`: This parameter describes the maximum number of features random forest is allowed to try in an individual tree. The value in this experiment was chosen as “auto”, which will take all the features in every tree. Though this option decreases the speed of the algorithm, it provides a high number of options to be considered at each node. Moreover, in this experiment we first chose the top 15 contributing features and then used them for the classification task, hence the sample space is reduced compared to the original 433.
- `n_estimators`: This parameter describes the number of trees to be built before taking the maximum voting of the predictions. Empirically, in this experiment, the optimal value for this parameter was 10.
- `criterion`: This parameter describes the criteria of split. In this experiment we have used the Gini index impurity measure.
- `max_depth`: This parameter represents the depth of each tree in the forest. The deeper the tree, the more splits it has, and the more information it can capture regarding the data. In this case, the optimal depth value was found to be 15.
- `min_samples_leaf`: This describes the minimum samples required to be at the leaf node. In this experiment, the optimal value of the parameter was found to be 3.

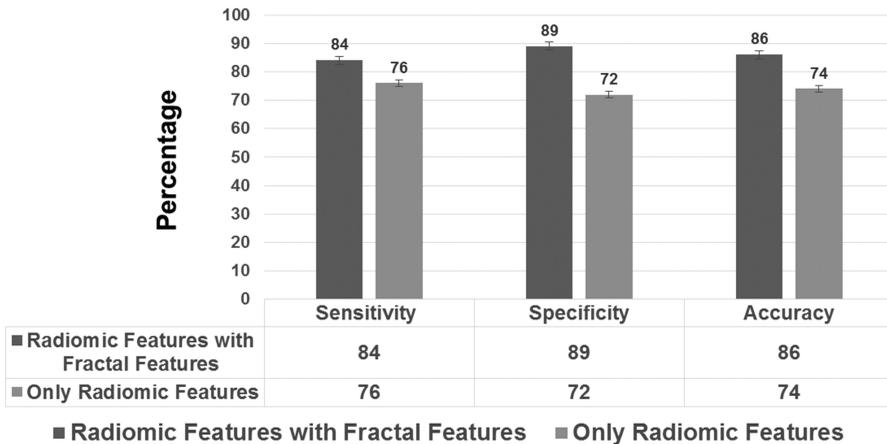
#### 4.2.5 Results

A total of 317 subjects were considered for the histology classification, out of which 40% of subjects were female, the average age being 68 years. Pearson’s correlation analysis was performed to understand the relationship of first order radiomics and FD features with the histology class. It can be observed that max FD has maximum correlation with respect to histology class compared to other first order features. Further, it can also be seen that tumor volume as an independent feature ranks much lower than the FD in Figure 4.4. This is in line with the understanding that the morphology of the tumor provides a more vital distinction than does its volume in histology differentiation.

The classification accuracy considering the radiomics features accounted for 74%; however, with the inclusion of FD features, it increased to 86% ( $P < 0.001$ ). Also, there is an improvement of 8% and 17% in terms for sensitivity and specificity respectively by considering the FD features (Figure 4.5).

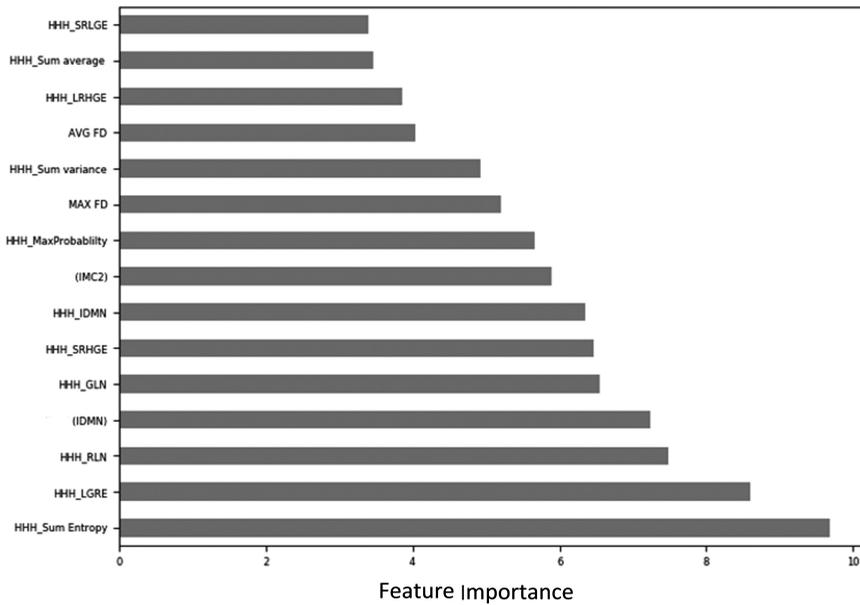


**FIGURE 4.4**  
Correlation of histology with various first order radiomics features and FD features.



**FIGURE 4.5**  
Classification metric comparison between with-FD and without-FD features inclusion.

Further, the features were ranked based on the random forest classifier on the priority of importance for classification of histology, as shown in Figure 4.6. The top features contributing toward histological classification using radiomic features include HHH\_Sum Entropy, HHH\_LRGE, HHH\_RLN, inverse difference moment normalized (IDMN), HHH\_GLN, HHH\_SRHGE, HHH\_IDMN, informational measure of correlation 2 (IMC2), maximum FD, average FD, HHH\_Sum Variance. Also, it's interesting to note



**FIGURE 4.6**  
Feature ranking based on importance.

that the FD-derived, max FD, and avg FD features were ranked among the top 15 features that aid in histology classification of NSCLC. In essence, wavelet-based features and FD parameters dominate as top contributing features for histology classification.

### 4.3 Conclusion

In this study, we established that FD features play an important role in histology classification of NSCLC. Applying fractal analysis on a 2D contour region can provide valuable information and reflect the idea of overall tumor aggressiveness. However, our study has a limitation: the fractal computation was applied on 2D contour GTV regions. This study could be extended to a 3D fractal analysis algorithm using the mesh approach, so that the region of GTV would be better delineated for computation FD. Also, for a tumor contour that is very small, there might not be enough points in the number of boxes vs the size of each box plane ( $\log(N)$  vs  $\log(1/r)$ ) to describe the tumor contour. In this case, the best fit line will be only an approximation, and this will be reflected in the FD value as well.

## References

1. F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancer in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 0, pp. 1–31, 2018.
2. L. H. Ma, G. Li, H. W. Zhang, Z. Y. Wang, J. Dang, S. Zhang, and L. Yao, "The effect of nonsmall cell lung cancer histology on survival as measured by the graded prognostic assessment in patients with brain metastases treated by hypofractionated stereotactic radiotherapy," *Radiation Oncology*, vol. 11, p. 92, 2016.
3. M. Yano, J. Yoshida, T. Koike, K. Kameyama, A. Shimamoto, W. Nishio, K. Yoshimoto, T. Utsumi, Y. Matsumura, S. Moriyama, and Y. Fujii, "The outcomes of a limited resection for non small cell lung cancer based on differences in pathology," *World Journal of Surgery*, 40(11):2688–2697, 2016.
4. C. C. Wu, M. M. Maher, and J. A. Shepard, "Complications of CT-guided percutaneous needle biopsy of the chest: prevention and management," *American Journal of Roentgenology*, vol. 196, no. 6, pp. W678–W682, 2011.
5. R. Patil, G. Mahadevaiah, and A. Dekker, "An approach toward automatic classification of tumor histopathology of non-small cell," *Tomography*, 2(4):374–377, 2016.
6. R. Gillies, P. Kinahan, H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, p. 2, 2016.
7. P. Lambin, E. Rios-Velazquez, R. Leijenaar *et al.* "Radiomics: extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, pp. 441–446, 2012.
8. V. Kumar, Y. Gu, S. Basu, *et al.* "Radiomics: the process and the challenges," *Magnetic Resonance Imaging*, vol. 30, no. 9, pp. 1234–1248, 2012.
9. H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, *et al.* "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, p. 40, 2014.
10. K. C. Santosh and Sameer K. Antani, "Automated chest X-ray screening: can lung region symmetry help detect pulmonary abnormalities?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 5, pp. 1168–1177, 2018.
11. S. Vajda, A. Karargyris, S. Jäger, K. C. Santosh, S. Candemir, Z. Xue, S. K. Antani, and G. R. Thoma, "Feature selection for automatic tuberculosis screening in frontal chest radiographs," *Journal of Medical Systems*, vol. 42, no. 8, pp. 146:1–146:11, 2018.
12. F. T. Zohora and K. C. Santosh, "Foreign circular element detection in chest X-rays for effective automated pulmonary abnormality screening," *International Journal of Computer Vision and Image Processing*, vol. 7, no. 2, pp. 36–49, 2017.
13. K. C. Santosh, Szilárd Vajda, Sameer K. Antani, and George R. Thoma, "Edge map analysis in chest X-rays for automatic pulmonary abnormality screening," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 9, pp. 1637–1646, 2016.
14. B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York, 1982.
15. J. W. Baish and R. K. Jain, "Fractals and cancer," *Perspectives in Cancer Research*, vol. 60, no. 14, 2000.

16. I. Sokolov, "Fractals: a possible new path to diagnose and cure cancer?" *Future Oncology*, vol. 11, no. 22, pp. 3049–30513, 2015.
17. O. Heymans, S. Blacher, F. Brouers, and G. E. Pierard, "Fractal quantification of microvasculature heterogeneity in cutaneous melanoma," *Dermatology*, vol. 198, pp. 212–217, 1999.
18. D. Cusumano *et al.* *Development and Validation of New Radiomic Features Based on Fractal Analysis*.
19. Fractal Explorer. "Calculating fractal dimension." Available at: [https://www.wahl.org/fe/HTML\\_version/link/FE4W/c4.htm](https://www.wahl.org/fe/HTML_version/link/FE4W/c4.htm), last accessed and verified on November 4, 2018.

