# A Rapid Segmentation-Insensitive "Digital Biopsy" Method for Radiomic Feature Extraction: Method and Pilot Study Using CT Images of Non–Small Cell Lung Cancer

Sebastian Echegaray[1], Viswam Nair[2,3,4], Michael Kadoch[2], Ann Leung[2], Daniel Rubin[2,3], Olivier Gevaert[3], and Sandy Napel[2]

[1]Department of Electrical Engineering, Stanford University, Stanford, California; [2]Department of Radiology, Stanford University School of Medicine, Stanford, California; [3]Department of Medicine, Stanford University School of Medicine, Stanford, California; and [4]Canary Center for Cancer Early Detection, Stanford University, Stanford, California

**Corresponding Author:**
Sebastian Echegaray
S323 Campus Drive, S344 Stanford,
CA 94305-5450;
E-mail: sechegaray@gmail.com

**ABSTRACT**

Quantitative imaging approaches compute features within images' regions of interest. Segmentation is rarely completely automatic, requiring time-consuming editing by experts. We propose a new paradigm, called "digital biopsy," that allows for the collection of intensity- and texture-based features from these regions at least 1 order of magnitude faster than the current manual or semiautomated methods. A radiologist reviewed automated segmentations of lung nodules from 100 preoperative volume computed tomography scans of patients with non–small cell lung cancer, and manually adjusted the nodule boundaries in each section, to be used as a reference standard, requiring up to 45 minutes per nodule. We also asked a different expert to generate a digital biopsy for each patient using a paintbrush tool to paint a contiguous region of each tumor over multiple cross-sections, a procedure that required an average of <3 minutes per nodule. We simulated additional digital biopsies using morphological procedures. Finally, we compared the features extracted from these digital biopsies with our reference standard using intraclass correlation coefficient (ICC) to characterize robustness. Comparing the reference standard segmentations to our digital biopsies, we found that 84/94 features had an ICC >0.7; comparing erosions and dilations, using a sphere of 1.5-mm radius, of our digital biopsies to the reference standard segmentations resulted in 41/94 and 53/94 features, respectively, with ICCs >0.7. We conclude that many intensity- and texture-based features remain consistent between the reference standard and our method while substantially reducing the amount of operator time required.

## INTRODUCTION

Quantitative features computed from medical images have been investigated for use in computer-aided diagnosis (1, 2), computer-aided detection (3, 4), and radiomics (5-7). Although these image features have the potential to provide consistent descriptors of the object being analyzed, segmentation of the volume of interest (VOI) is the necessary first step for obtaining these values. Manual segmentation of tumors in 2-dimensional (2D) computed tomography (CT)-images is labor-intensive and time-consuming (8, 9), and even more onerous when segmenting 3-dimensional (3D) volumes. The literature contains several automatic and semiautomatic image segmentation algorithms (10-18). Although these algorithms reduce the time taken

to segment a tumor (19, 20), they do not always achieve accurate or consistent results (21); therefore, all segmentations must be reviewed and possibly edited, which, in turn, requires additional time and introduces variability.

The typical approach to using the radiomics features to build predictive models involves computing a large number of image features from within a VOI. Because many of these computed features are correlated, one of the several machine learning methods (22, 23) can be used to select a subset of nonredundant informative features that can be combined in the model (24-26). We hypothesized that a set of features required for a robust predictive model could be extracted without requiring accurate or precise segmentation of the tumor boundary. The

**Figure 1.** Cross-section through part-solid nodule in the right upper lobe (A), and its intersection with the reference standard 3-dimensional (3D) segmentation (B), and the 3D digital biopsy (C).
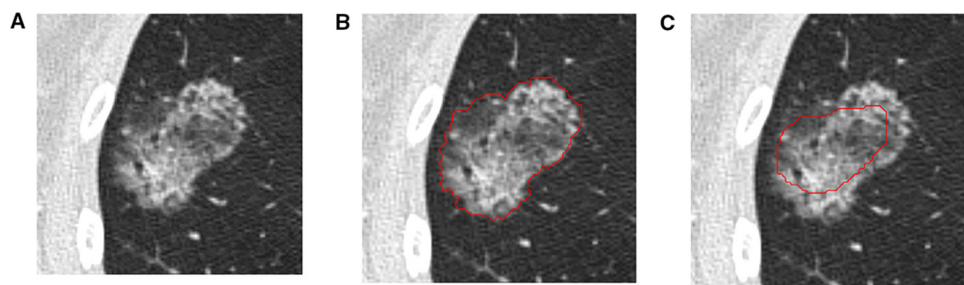
first step in testing this hypothesis involves determining which features extracted from such a segmentation are consistent with a segmentation that attempts to capture the tumor margin. Therefore, we developed a new method, which we call "digital biopsy," in which a human annotator is asked to capture the heterogeneity of the tumor without carefully segmenting the tumor boundary, by "painting" the inside of the tumor using a suitable tool. Then, we analyzed the stability of the intensity and texture features of segmentations of the entire tumor computed from these digital biopsies compared with those computed semi-automatically by calculating their intraclass correlation coefficient (ICC) (27, 28). Because we did not expect the tumor shape or margin to be captured by these digital biopsies, we did not compute ICCs for such features.

In doing so, we introduced a new segmentation paradigm, in which the expert focuses on capturing the heterogeneity of the lesion instead of the tumor boundary. In a preliminary study of 100 patients with preoperative CT scans of proven non–small cell lung cancer, we show that multiple image features that characterize the intensity and texture of lung nodules are consistent with the same features that could be obtained using detailed segmentations, while obtaining an order of magnitude reduction in human operator time.

## MATERIALS AND METHODS
### Data
Following institutional review board approval by Stanford University's Research Compliance Office, that waived the requirement for written informed consent, we selected the first 100 subjects (male, 75; mean age, 70 years) who were part of a radiogenomics cohort obtained from a larger database of patients referred to Stanford University Medical Center with diagnosed non–small cell lung cancer and had CT scans before surgical resection. Scanners (GE Medical Systems, Waukesha, WI; Siemens, Erlangen, Germany; Toshiba, Otawara, Japan; and Phillips, Andover, MA, with 120 kVp), with current and section thickness ranging from 25 to 697 mA and 0.625 to 3 mm, respectively, were used to acquire scans. The tumors ranged in size from 0.37 to 306 cm$^3$ and in attenuation from ground glass to solid. All scans were performed between April 2008 and October 2014.

### Semiautomated—Manual Segmentations
For this study, we used an in-house segmentation algorithm (21) under supervision of a fellowship-trained thoracic radiologist with 23 years of experience. For each subject, the radiologist

first identified the tumor in the CT scan volume and chose a small "seed circle" to initialize the segmentation algorithm, which then derived a 3D VOI containing the voxels within the tumor, which was then superimposed on the CT slices using ePAD, an open-source annotation tool (29, 30). Finally, the radiologist spent between 5 and 45 minutes manually editing each VOI using ePAD's paintbrush tool on 2D cross-sections to correct what he perceived as local over- and/or under-segmentation.

### Digital Biopsies
A nonradiologist human reader with 30 years of experience with medical image processing and analysis (10 years specifically with CT scans of patients with lung cancer) independently viewed each series and created a digital biopsy using the same paintbrush tool used by the radiologist on 2D cross-sections of the tumor. Using the reference standard for each of the 100 subjects as a guide, this reader marked a contiguous section of the tumor volume for inclusion as he scrolled through the 2D cross-sections in the image series. The reader was instructed to capture the gray-scale heterogeneity of the tumor and to not be concerned with capturing the detail of the boundary. We gave no additional instructions about how to segment the sections or which section to select. We tracked the time required to create each digital biopsy and stored each of them to compare with the radiologist's reference standard. Figure 1 shows a central section through one of the nodules and the reference standard 3D segmentations and 3D digital biopsy.

### Simulating Multiple Readers
As the creation of digital biopsies is subjective and therefore prone to intra- and inter-reader variation, we simulated additional digital biopsies using the morphological procedures of erosion and dilation (31). These were achieved using multiple spherical structuring elements with radii ranging from 0.5 to 1.5 mm inclusive, with an interval of 0.5 mm between the radii, for a total of 3 erosions and 3 dilations. The erosions simulated more conservative additional readers, while the dilations simulated more aggressive additional readers. Dilations were not constrained to stay within any region and therefore may have led to the digital biopsy going beyond the tumor borders. To avoid splitting the tumor into multiple portions during erosion, we followed each with a morphological closing procedure. Figure 2 shows an example of the erosions and dilations from the same nodule and digital biopsy shown in Figure 1.

|  | 0.5 mm | 1.0 mm | 1.5 mm |
|---|---|---|---|
| Erosion | | | |
| Dilation | | | |
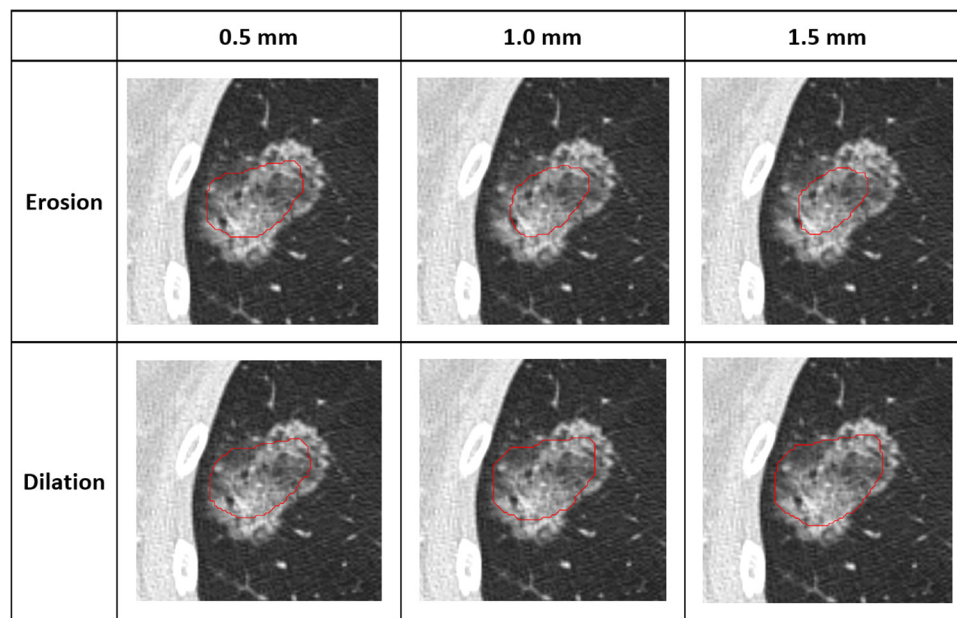
**Figure 2.** Cross-sections of digital biopsies obtained by applying morphological operations to the manual digital biopsy shown in Figure 1. The first and second rows show erosions and dilations, respectively.

### 3D Image Features

There are many algorithms in the literature that extract quantitative features from VOIs (32-34) within a CT scan. These algorithms can be categorized as measuring intensity, shape, and margin or textures' characteristics. Intensity features express statistics of the pixel values within a VOI. Shape features describe the boundary of the VOI. Margin features characterize the transition between the intensity values inside the VOI and the values surrounding it. Finally, texture features measure the spatial distribution of pixel intensities inside the VOI. For this study, we only computed intensity and texture features (a total of 94 features) because our digital biopsies, by definition, do not attempt to capture the shape or margins of the tumors. Appendix 1 provides a list of the specific intensity and texture features that we computed, with references to the literature as needed.

### Metrics

*Overlap.* To analyze the agreement between *i* VOIs in a given series, k, we define overlap as the ratio between their intersections and their unions as follows:

$$O_k = \frac{\cap_i VOI_i}{\cup_i VOI_i}$$

### Feature Consistency

We used the ICC to measure the consistency of the features extracted for each segmentation looking across patients, readers, cores, and sections. The ICC describes how members from the same group resemble each other and has often been used to quantify the consistency of measurements made by different experts (35, 36). A high ICC shows that a feature is consistent across multiple measurements. There are multiple algorithms in the literature to calculate ICC (37); for this study, we used the A-1 method, also known as criterion-referenced reliability, which is expressed as follows:

$$ICC = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where $MS_R$ is the mean square for rows (observations), $MS_E$ is the mean square error, and $MS_C$ is the mean square for columns (segmentations). n and k represent the total number of rows and columns, respectively. In our study, rows represent each study where features were extracted. Columns represent the different segmentations, which can be provided by the reference standard segmentations, digital biopsies, or the morphological operations. We used this method, as it measures the degree of absolute agreement taking into consideration the systematic variations between methods.

## RESULTS

### Digital Biopsies: Time to Obtain and Volume Overlap with Reference Standard

Using ePAD's paintbrush tool, the reader averaged 171 seconds, with a median of 132 seconds, a standard deviation of 152 seconds, and a range 25 to 900 seconds, to create a digital biopsy. Table 1 shows the mean and standard deviation, median, minimum, and maximum of the volume overlap of the reference standard with the digital biopsies with the 3 erosions and the 3 dilations. One can see that the volume overlap decreases with erosion of the digital biopsies, as expected, and that it can also decrease with the dilation of the digital biopsies as the dilated volumes grow bigger than the digital biopsies. Figure 3 and Figure 4 show the distribution of these overlaps as a function of erosion and dilation, respectively, of the digital biopsies. Erosions caused the distribution to shift to the left (lower overlap), while dilations caused an initial shift to the right (higher overlap) at 0.5 mm of dilation, with a shift to the left as larger dilations expanded the volumes to exceed the nodule borders in many cases.

**Table 1.** Overlap Between Digital Biopsies and the Reference Standard Segmentation

| Method | Size | Mean (%) | SD (%) | Median (%) | Minimum (%) | Maximum (%) |
|---|---|---|---|---|---|---|
| Original digital biopsies | None | 74.04 | 10.77 | 76.65 | 25.26 | 85.23 |
| Erosions | 0.5 mm | 64.46 | 11.27 | 67.11 | 20.94 | 82.81 |
| | 1.0 mm | 53.04 | 12.03 | 54.87 | 13.63 | 75.52 |
| | 1.5 mm | 42.33 | 13.11 | 42.61 | 8.47 | 69.61 |
| Dilations | 0.5 mm | 77.73 | 10.38 | 79.49 | 27.11 | 88.58 |
| | 1.0 mm | 78.21 | 9.66 | 80.36 | 28.22 | 90.94 |
| | 1.5 mm | 71.43 | 10.22 | 73.34 | 28.55 | 89.17 |

Abbreviation: SD, Standard deviation.

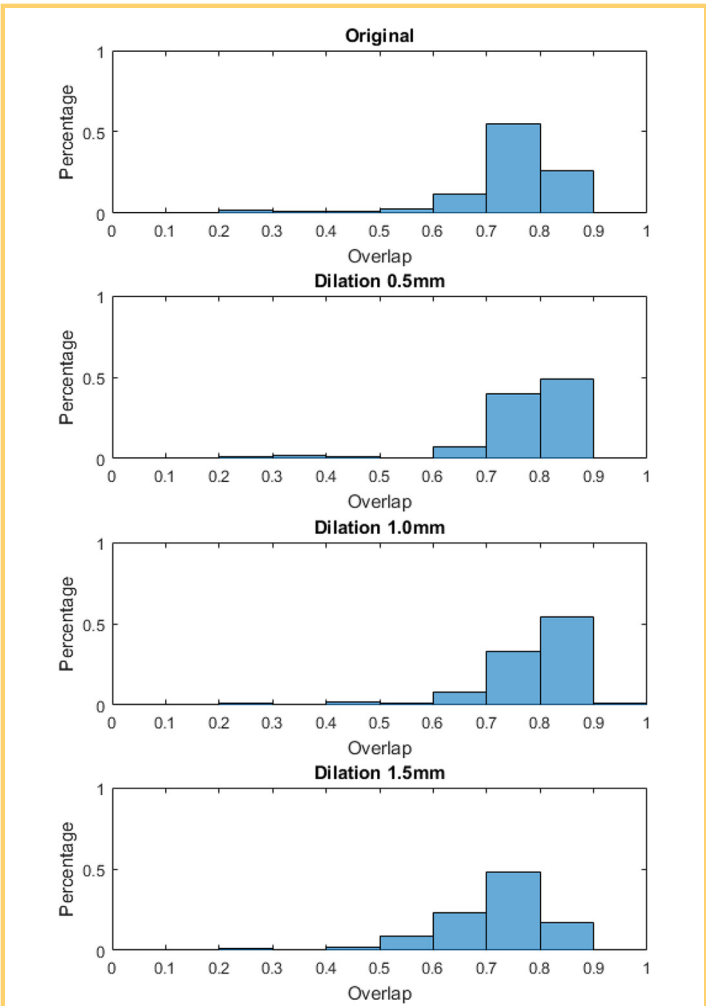## Digital Biopsies: Agreement of Features With Those Obtained Using the Reference Standard Segmentations

We obtained the ICC score for each of the features extracted from the digital biopsies and their erosions and dilations com-pared with the features extracted from the reference standard. Figure 5 and Figure 6 plot the ICC scores, with features ordered from the highest ICC on the left to the lowest ICC on the right for
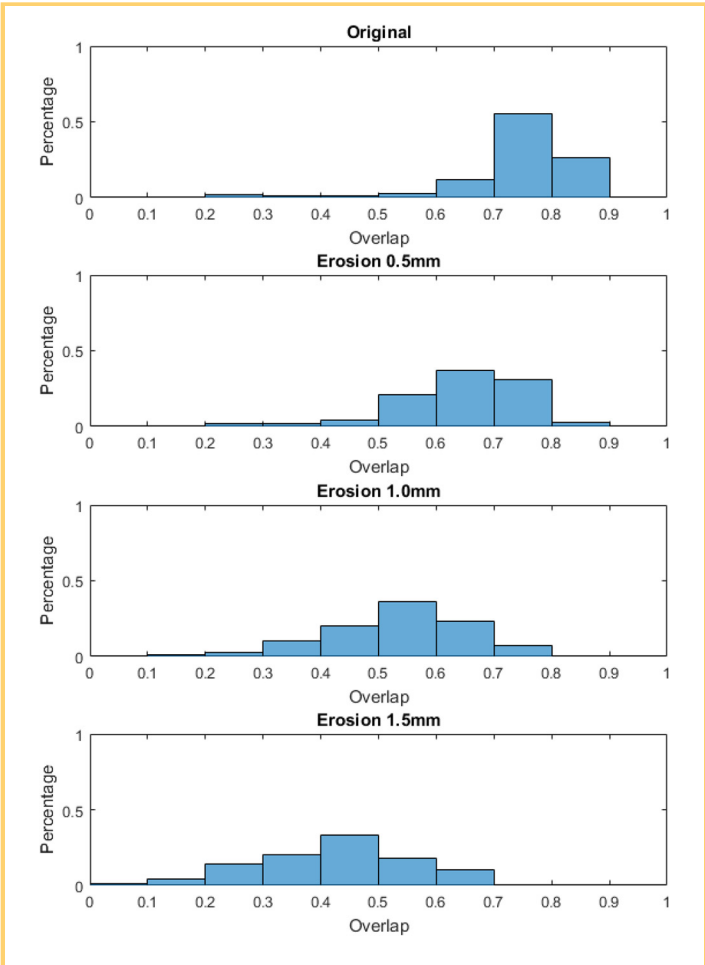


**Figure 3.** Distribution of the overlap of the reference standard segmentation and in order from top to bottom: the original biopsy, 0.5-mm erosion, 1.0-mm erosion, and 1.5-mm erosion. Statistics regarding the distributions are shown in Table 1.
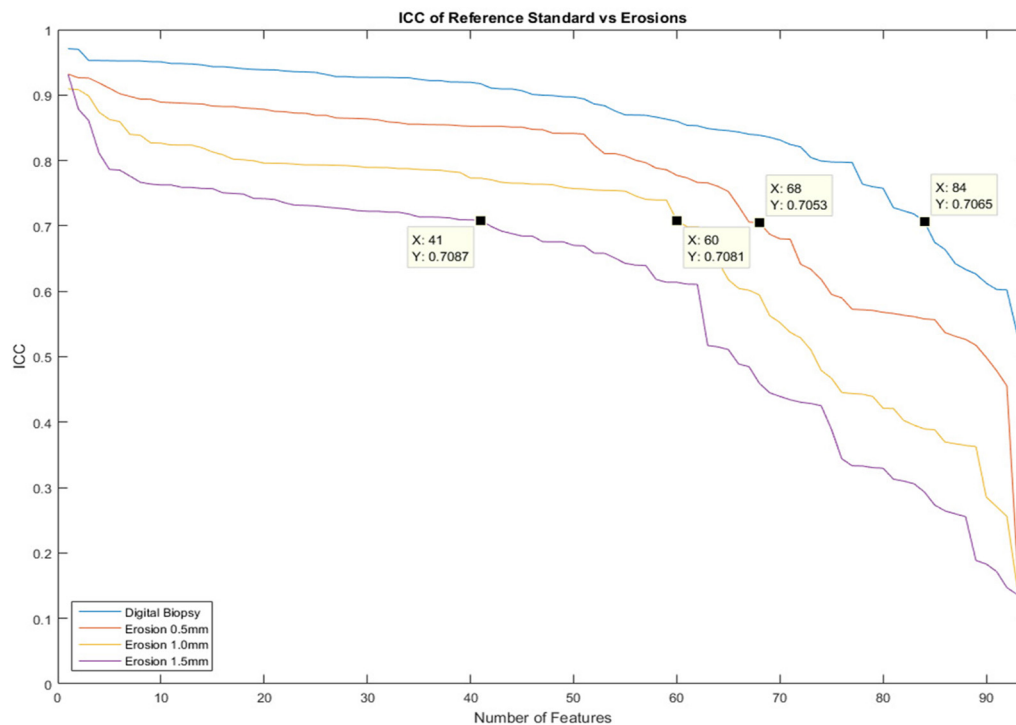


**Figure 4.** Distribution of the overlap of the reference standard segmentation and in order from top to bottom: the original biopsy, 0.5-mm dilation, 1.0-mm dilation, and 1.5-mm dilation. Statistics regarding the distributions are shown in Table 1.
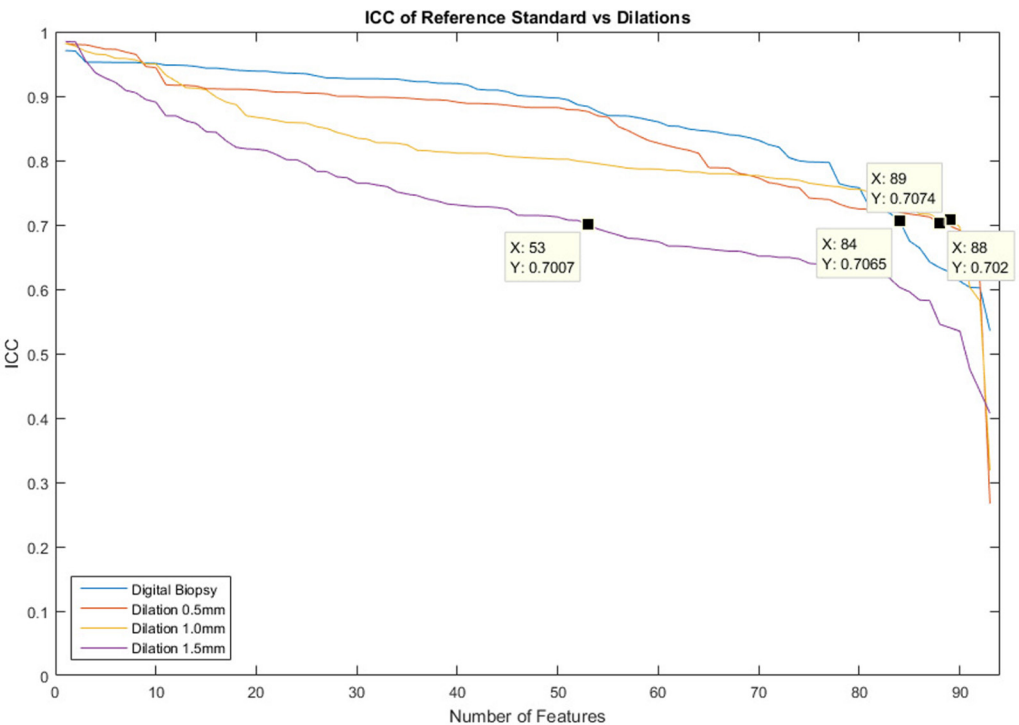
**Figure 5.** The intraclass correlation coefficient (ICC) curves for the features extracted from the digital biopsies and each of the morphological erosions compared with their reference standard segmentation. The features are organized in the descending order by their ICC value. Each line has been marked to indicate the number of features, with ICC > 0.7.

each curve separately, for the erosions and dilations of the digital biopsies, respectively. These figures show that 84/94 features computed using the original digital biopsies have excellent agreement (ICC > 0.7) (37-39), with the same features computed using the reference standard segmentations. Moreover, eroding the digital biopsies continued to produce many features showing excellent agreement with those computed using the reference standard segmentations (68/94, 60/94, and 41/94 for 0.5 mm, 1.0 mm, and 1.5 mm erosions, respectively). Similarly, dilation resulted in many features showing excellent agreement with those computed using the reference standard segmentations (88/94, 89/94, and 53/94 for 0.5 mm, 1.0 mm, and 1.5 mm dilations, respectively). Table 2 shows the number of features above several other thresholds of agreement (ICCs of 0.6 through 0.9), revealing that many features are insensitive to the exact borders of the segmentation. Appendix 2 contains a ranked list showing the most robust features across all 7 experiments.

Figure 7 and Figure 8 show the distribution of ICC scores for intensity and texture features, respectively, and how these distributions change as a function of erosion and dilation. Figure 7 shows that intensity features maintain high ICCs under dilation, but fall off under erosion. Figure 8 shows that the ICCs of texture features decrease under both procedures, although, as was the case for intensity features, more strongly under erosion.

## DISCUSSION

The radiomics methods, in widespread development, compute many (sometimes hundreds of) image features from VOIs or regions of interest (ROIs) in radiological images and link these features to clinical data, such as response to treatment, survival, and gene expression (5, 15, 25, 26, 40, 41). Once these features are computed, they can be used to build predictive models. Several studies have shown that the collection of images' features computed are highly correlated (25, 41, 42), and the resulting models usually are based on a small subset of the robust and independent computed features. Most often, ROIs are selected using segmentation algorithms and/or manual delineation. Although several software packages have been developed by academic institutions (10-15, 21, 43-48) and commercial vendors that offer automated segmentation of lung nodules in chest CT scans, in our experience, none is foolproof; on the contrary, each segmentation must be reviewed for quality control and perhaps edited. Thus, we wondered if a set of features could be computed from an easier-to-obtain ROI that shows consistency with features from full quality-controlled segmentation. The median time and maximum time required to obtain a digital biopsy were 132 and 900 seconds, respectively. This is an order of magnitude that is faster than our reference standard procedure, wherein a trained radiologist had to inspect and modify the semiautomated segmentation to trace the tumor boundary, which required anywhere from 600 to 2700 seconds

**Figure 6.** The ICC curves for the features extracted from the digital biopsies and each of the morphological dilations compared with their reference standard segmentation. The features are organized in the descending order by their ICC. The features are organized in the descending order by their ICC. Each line has been marked to indicate the number of features, with ICC > 0.7.

(5–45 minutes) per case. Further, despite the lack of precise overlap of our digital biopsies to the reference standard, we have shown many features that remain highly consistent. Table A2 (Appendix 2) lists the many robust features we have found that are important descriptors of tumoral heterogeneity, and that have been used in many radiomics studies in several cancer imaging scenarios (5-7, 15, 22, 41, 49).

Although our digital biopsies were much faster to obtain, we acknowledge that for this technique to be used in practice, it would have to be even faster. We have identified multiple addressable factors that limited our speed in obtaining them.

First, this was the first time this expert used this tool for this purpose. It is likely that familiarity would result in an increase in speed. Second, although ePAD is a powerful Digital Imaging and Communications in Medicine (DICOM) viewer and annotation tool, it was not created with this use in mind; painting was accomplished 1 transaxial section at a time, and the reader had to navigate through multiple menus to change the ePAD's brush size to gain efficiency as the cross-sectional area on each section changed. We are confident that a tool that works on multiple planes and has streamlined operations could reduce the required time by another order of magni-

**Table 2.** Features Above Thresholds of Agreement

| ICC | Original | Erosion | | | Dilation | | |
|---|---|---|---|---|---|---|---|
| | | 0.5 mm | 1.0 mm | 1.5 mm | 0.5 mm | 1.0 mm | 1.5 mm |
| >0.9 | 47 | 6 | 2 | 1 | 27 | 15 | 8 |
| >0.8 | 74 | 56 | 18 | 4 | 64 | 51 | 24 |
| >0.7 | 84 | 68 | 60 | 41 | 88 | 89 | 53 |
| >0.6 | 93 | 74 | 67 | 62 | 92 | 91 | 84 |

The number of features presented in the table are out of 94 that presented ICC > 0.9, > 0.8, > 0.7, and > 0.6 in each of the digital biopsies and its morphological modifications when compared with the features extracted from the reference standard segmentation. Appendix 2 names and ranks the individual features with the highest ICCs across all 7 digital biopsy variations.
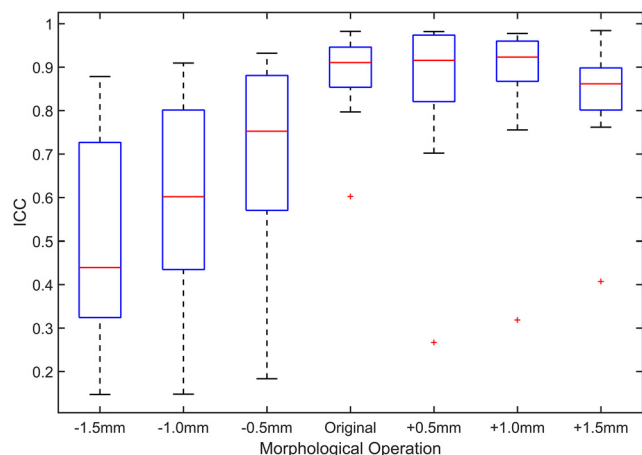
**Figure 7.** Boxplot of ICC of intensity features for original digital biopsies and their erosions/dilations compared with the reference standard segmentations. The Y-axis shows the ICC score and the X-axis is the morphological operation, with "−" and "+" representing that the segmentation underwent erosion and dilation, respectively.
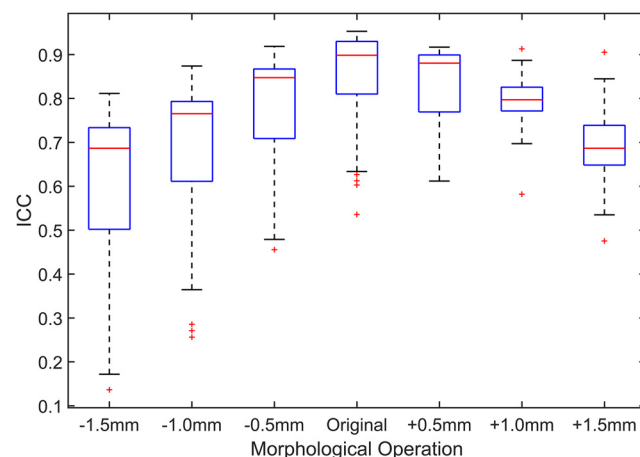


**Figure 8.** Boxplot of ICC of texture features for original digital biopsies and their erosions/dilations compared with the reference standard segmentations. The Y-axis shows the ICC score and the X-axis is the morphological operation, with "−" and "+" representing that the segmentation underwent erosion and dilation, respectively.

tude, thus reducing the median time to under 1 minute per tumor.

Obtaining fast digital biopsies is only useful if the features that are being extracted are consistent among operators and methods. As shown in Table 2, 89% (84/94) of the features remain highly consistent (ICC > 0.7) between the reference standard segmentation and the original digital biopsies, even while the digital biopsy only segments 74% of the original tumor (Table 1). This is promising, in that it shows that we are able to capture the same information using a less precise but faster segmentation. This stability remains mainly present as we dilate and erode our digital biopsies with a spherical structuring element to up to 1.5 mm in radius, representing readers that are more or less conservative with their segmentations, as shown in Figures 5 and 6. These erosions and dilations, simulating additional readers, also produced a large number of features with ICC > 0.7. The discrepancies between erosions and the original digital biopsies are mainly caused by the change of point statistics (eg, extrema), and the loss of higher wavelength textures. In contrast, discrepancies between dilations and the original biopsies can be additionally caused by inclusion of voxels beyond the periphery of the tumor, which can result in some strong texture responses (eg, at the boundary between the tumor and the air in the lung or an adjacent chest wall). We can observe these discrepancies in both Figures 7 and 8, where we see that the intensity features remain mostly stable under dilation, while stability is reduced by erosion, and that texture features' consistency is reduced under both procedures. However, in all circumstances, the majority of features remain consistent with those obtained by our reference standard segmentation.

## Limitations

One limitation is that we only used one manual digital biopsy per subject; different readers would not necessarily result in the same VOI. We attempted to overcome this limitation by generating multiple simulated biopsies by morphologically altering our reader's segmentations. As we have shown, multiple texture and intensity features remain highly stable across the simulated segmentations; therefore, we expect similar results if we were to include multiple readers, as long as their mean volume overlap with the tumor itself is at least 75%. However, future studies with multiple users are required to validate that intra- and inter-reader variation of ICCs of features derived from these digital biopsies is acceptable. Other painting strategies may also be useful, such as creating multiple small digital biopsies per tumor, and remain to be investigated.

Second, to obtain a preliminary comparison of the segmentation time improvement, we compared the time required by a radiologist to carefully adjust segmentation boundaries to that required by a nonradiologist to acquire the digital biopsies, resulting in an order of magnitude speed advantage for the digital biopsies. Although comparing a radiologist with a nonradiologist could confound the comparison, it is not clear that either participant would have a speed advantage over the other given the different tasks. Future studies should compare multiple radiologists and trained nonradiologists (who ideally would be preferred for this task on the basis of cost).

Third, as part of our experimental design, we have eliminated all boundary and shape features from our study. The literature has reports that boundary and shape are important markers in characterization of certain cancers (50-52). Our proposed method for segmenting does not capture these features,

and it remains to be shown that the intensity and texture features that we do capture are sufficient to build strong predictive models. It is also possible that complementing the features obtained via digital biopsy with a small set of easy-to-obtain semantic features (eg, "spiculated," "lobulated," "pleural attachment," and "poorly defined margins") could strengthen the model at little cost; this too, remains, for future evaluation.

Fourth, we acknowledge that the reader who provided the original digital biopsies used the reference standard as a guide to locate each tumor, information that would not be available in general. It is therefore possible that performance with unguided digital biopsies would be different than that reported here. Although our experiments simulating additional readers with morphological operations provide some reassurance, performance with unguided digital biopsies by trained readers should be evaluated in the future.

Fifth, in this study, we have only focused on the stability of the features and have not proven, nor intended to prove, that these features correlate with any clinical or outcomes data. As Aerts (24) and others have shown, a first step in building these models is to find features that are robust, that is, insensitive to segmentation. We have shown that our digital biopsy technique is robust in the context of CT for lung cancer, and it remains to show this in other cancer types and imaging scenarios. Discovering relationships and building models linking these features to disease is the next step. Rapidly obtained digital biopsies may prove quite useful in this endeavor.

Finally, our study only addresses the stability of the features to changes in tumor segmentation. However, it is well known that CT acquisition and reconstruction parameters and conditions can also affect quantitative feature values (53-57). Future studies that compare quantitative image features acquired at different points in time must be aware of this possibility and control for this effect.

In conclusion, we have proposed a new paradigm for selecting a VOI for the radiomics analysis that captures the heterogeneity of a given lesion in 3D. This method is 1 or 2 orders of magnitude faster than semiautomated segmentations, which has remained the dominant strategy because completely automatic segmentation has not been shown. We have shown that the texture- and intensity-based features extracted in this way are robust to morphological transformations and remain highly correlated with those from curated segmentations and, therefore, we think that the use of digital biopsies will accelerate the potential of researchers to develop, and for clinicians to use, quantitative imaging methods to characterize cancer in medical images.

## APPENDIX 1
### Intensity Features
To quantify the intensity characteristics of the volume, we extracted all the classical statistical descriptors such as mean, variance, kurtosis, skewness, entropy, maximum, and minimum from inside the VOI. These features summarize the global information of the intensity distribution inside the VOI without considering local differences.

**Table A1.** List of Haralick Features Extracted from the GLCM

| Features | References |
|---|---|
| Energy | (59, 61) |
| Contrast | (59, 61) |
| Sum of Means | (61) |
| Cluster Tendency | (59) |
| Entropy | (59) |
| Homogeneity | (59) |
| Inertia | (58, 59) |
| Max Probability | (59) |
| Correlation | (59, 61) |
| Variance | (61) |
| Cluster Shade | (59) |
| Inverse Variance | (58, 59) |

Abbreviation: GLCM, gray-level co-occurrence matrix.
Details of the implementation of each of these features can be found in the references mentioned alongside the features.

### Texture Features
To characterize local changes in intensity, we computed the Haralick features. The Haralick features are measurements taken from a gray-level co-occurrence matrix (GLCM) (58-60). A GLCM is defined as the distribution of co-occurring values in an image at a fixed offset. This matrix is defined by the following equation:

$$C(i, j, \Delta x, \Delta y, \Delta z) = \sum_{p=1}^{N} \sum_{q=1}^{M} \sum_{r=1}^{O} \begin{cases} 1, & \text{if } I(p, q, r) = i \text{ and} \\ & I(p + \Delta x, q + +\Delta y, r + +\Delta x) = j \\ 0, & \text{otherwise} \end{cases}$$

where i and j are the row and columns indices of the GLCM, respectively. $\Delta x$, $\Delta y$, and $\Delta z$ are the fixed offsets in the three axes of the volume. I(p,q,r) is the gray level at point (p,q,r) and N,M,O, are the sizes of the volume in each dimension. A set of statistics, shown with their references in Table A1, is then computed from the GLCM. To obtain rotation-invariant features, we calculate the Haralick features in 13 directions, aggregate them, and report their mean and standard deviation.

## APPENDIX 2
In Table A2, we count how many times each feature was over a certain ICC value over each of the 7 experiments. (Reference Standard vs Digital Biopsy and 0.5 mm, 1.0 mm, 1.5 mm erosions and dilations). Features that did not have an ICC > 0.7 in any of the experiments are obviated. Features that consistently have high ICC across all experiments indicate to be more robust to changes in their contour than features with rapidly changing ICC.

**Table A2.** Features of Important Descriptors of Tumoral Heterogeneity

| Feature Name | ICC > 0.7 | ICC > 0.8 | ICC > 0.9 |
|---|---|---|---|
| "Haralick D=3mm std sum of means" | 7 | 7 | 3 |
| "Intensity Entropy" | 7 | 6 | 5 |
| "Intensity Mean" | 7 | 6 | 3 |
| "Intensity Median" | 7 | 6 | 3 |
| "Intensity Trimmed Mean (25%)" | 7 | 6 | 3 |
| "Haralick D=1mm std energy" | 7 | 6 | 1 |
| "Haralick D=2mm std entropy" | 7 | 6 | 1 |
| "Haralick D=1mm std max probability" | 7 | 5 | 4 |
| "Haralick D=1mm std entropy" | 7 | 5 | 2 |
| "Haralick D=1mm mean cluster tendency" | 7 | 5 | 1 |
| "Haralick D=1mm std sum of means" | 7 | 5 | 1 |
| "Haralick D=2mm std sum of means" | 7 | 5 | 1 |
| "Haralick D=3mm std cluster shade" | 7 | 5 | 1 |
| "Haralick D=2mm std max probability" | 7 | 5 | 0 |
| "Haralick D=2mm mean cluster tendency" | 7 | 4 | 1 |
| "Haralick D=3mm mean cluster tendency" | 7 | 4 | 1 |
| "Haralick D=2mm std cluster shade" | 7 | 4 | 1 |
| "Haralick D=3mm std contrast" | 7 | 4 | 1 |
| "Haralick D=3mm std inertia" | 7 | 4 | 1 |
| "Haralick D=2mm std energy" | 7 | 4 | 0 |
| "Haralick D=1mm std variance" | 7 | 3 | 0 |
| "Haralick D=2mm std variance" | 7 | 1 | 0 |
| "Intensity Under −291 HU Percentage" | 6 | 5 | 3 |
| "Intensity Over −291 HU Percentage" | 6 | 5 | 3 |
| "Haralick D=1mm mean variance" | 6 | 5 | 2 |
| "Haralick D=2mm mean variance" | 6 | 5 | 2 |
| "Haralick D=3mm mean variance" | 6 | 5 | 2 |
| "Haralick D=3mm std max probability" | 6 | 5 | 0 |
| "Haralick D=2mm std contrast" | 6 | 4 | 2 |
| "Haralick D=2mm std inertia" | 6 | 4 | 2 |
| "Haralick D=1mm mean cluster shade" | 6 | 4 | 1 |
| "Haralick D=2mm mean cluster shade" | 6 | 4 | 1 |
| "Haralick D=1mm std cluster shade" | 6 | 4 | 1 |
| "Haralick D=2mm std cluster tendency" | 6 | 4 | 1 |
| "Haralick D=3mm std cluster tendency" | 6 | 4 | 1 |
| "Haralick D=3mm mean cluster shade" | 6 | 4 | 0 |
| "Haralick D=1mm std contrast" | 6 | 3 | 2 |
| "Haralick D=1mm std inertia" | 6 | 3 | 2 |
| "Haralick D=1mm mean entropy" | 6 | 3 | 0 |
| "Haralick D=2mm mean entropy" | 6 | 3 | 0 |
| "Haralick D=3mm std variance" | 6 | 3 | 0 |
| "Haralick D=3mm mean energy" | 6 | 2 | 0 |
| "Haralick D=3mm mean max probability" | 6 | 2 | 0 |
| "Haralick D=1mm mean energy" | 6 | 1 | 0 |
| "Haralick D=2mm mean energy" | 6 | 1 | 0 |

(Continued)

**Table A2.** Continued

| Feature Name | ICC > 0.7 | ICC > 0.8 | ICC > 0.9 |
|---|---|---|---|
| "Haralick D=2mm mean max probability" | 6 | 1 | 0 |
| "Haralick D=3mm std entropy" | 5 | 5 | 1 |
| "Intensity Skewness" | 5 | 4 | 2 |
| "Intensity Min" | 5 | 4 | 2 |
| "Haralick D=1mm mean contrast" | 5 | 4 | 2 |
| "Haralick D=1mm mean inertia" | 5 | 4 | 2 |
| "Haralick D=1mm std cluster tendency" | 5 | 4 | 1 |
| "Haralick D=2mm mean contrast" | 5 | 3 | 2 |
| "Haralick D=2mm mean inertia" | 5 | 3 | 2 |
| "Haralick D=3mm mean contrast" | 5 | 3 | 1 |
| "Haralick D=3mm mean inertia" | 5 | 3 | 1 |
| "Haralick D=1mm std homogeneity" | 5 | 3 | 1 |
| "Haralick D=2mm std homogeneity" | 5 | 3 | 1 |
| "Haralick D=3mm mean entropy" | 5 | 3 | 0 |
| "Haralick D=3mm std homogeneity" | 5 | 3 | 0 |
| "Haralick D=1mm mean max probability" | 5 | 1 | 0 |
| "Intensity Mean Absolute Difference" | 4 | 4 | 1 |
| "Intensity Standard Deviation" | 4 | 4 | 0 |
| "Intensity Interquartile Difference " | 4 | 2 | 1 |
| "Intensity Range" | 4 | 2 | 0 |
| "Intensity Max" | 4 | 2 | 0 |
| "Haralick D=3mm std energy" | 4 | 1 | 0 |
| "Haralick D=1mm mean sum of means" | 4 | 0 | 0 |
| "Haralick D=2mm mean sum of means" | 4 | 0 | 0 |
| "Haralick D=3mm mean sum of means" | 4 | 0 | 0 |
| "Haralick D=1mm mean correlation" | 3 | 1 | 0 |
| "Intensity Kurtosis" | 3 | 0 | 0 |
| "Haralick D=1mm mean homogeneity" | 3 | 0 | 0 |
| "Haralick D=1mm mean inverse variance" | 3 | 0 | 0 |
| "Haralick D=2mm mean homogeneity" | 3 | 0 | 0 |
| "Haralick D=2mm mean inverse variance" | 3 | 0 | 0 |
| "Haralick D=3mm mean homogeneity" | 3 | 0 | 0 |
| "Haralick D=3mm mean inverse variance" | 3 | 0 | 0 |
| "Haralick D=2mm std correlation" | 3 | 0 | 0 |
| "Haralick D=1mm std correlation" | 2 | 1 | 0 |
| "Haralick D=3mm std inverse variance" | 2 | 1 | 0 |
| "Haralick D=2mm mean correlation" | 2 | 0 | 0 |
| "Haralick D=3mm mean correlation" | 2 | 0 | 0 |
| "Haralick D=2mm std inverse variance" | 2 | 0 | 0 |
| "Intensity Harmonic Mean" | 1 | 1 | 1 |
| "Intensity Mode" | 1 | 1 | 0 |
| "Haralick D=3mm std correlation" | 1 | 0 | 0 |

Abbreviation: ICC, Intra-class correlation.
The features mentioned were computed for each of the 7 digital biopsies (original, 3 erosions, and 3 dilations) and the number of times their ICC compared with the reference standard was higher than 0.9, 0.8, and 0.7, ranked by ICC. Features that never scored >0.7 are not shown.

## REFERENCES

1. Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM. Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography. Med Phys. 1987;14(4):538–548.
2. Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Comput Med Imaging Graph. 2007;31(4-5):198–211.
3. Zhang J, Liu Y. Cervical cancer detection using SVM based feature screening. Med Image Comput Assist Interv. 2004;3217:873–880.
4. Yu H, Barriga S, Agurto C, Echegaray S, Pattichis M, Zamora G, Bauman W, Soliz P. Fast localization of optic disc and fovea in retinal images for eye disease screening. IEEE Trans Inf Technol Biomed. 2012;16(4):644–657.
5. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJWL, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA, Gillies RJ. Radiomics: the process and the challenges. Magn Reson Imaging. 2012;30(9):1234–1248.
6. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, Aerts HJWL. Radiomics: Extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48(4):441–446.
7. Napel S, Giger M. Special Section Guest Editorial:Radiomics and Imaging Genomics: Quantitative Imaging for Precision Medicine. J Med Imaging (Bellingham). 2015;2(4):041001.
8. Chen W, Giger ML, Bick U. A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. Acad Radiol. 2006;13(1):63–72.
9. Stammberger T, Eckstein F, Michaelis M, Englmeier K-H, Reiser M. Interobserver reproducibility of quantitative cartilage measurements: comparison of B-spline snakes and manual segmentation. Magn Reson Imaging. 1999;17(7):1033–1042.
10. Armato SG 3rd, Sensakovic WF. Automated lung segmentation for thoracic CT. Acad Radiol. 2004;11(9):1011–1021.
11. Opfer R, Wiemker R. A new general tumor segmentation framework based on radial basis function energy minimization with a validation study on LIDC lung nodules. Proc. SPIE 6512, Medical Imaging 2007: Image Processing, 651217 (3 March 2007).
12. Brown MS, McNitt-Gray MF, Mankovich NJ, Goldin JG, Hiller J, Wilson LS, Aberle DR. Method for segmenting chest CT image data using an anatomical model: preliminary results. IEEE Trans Med Imaging. 1997;16(6):828–839.
13. Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, Lewis JH, De Ruysscher D, Kikinis R, Lambin P, Aerts HJWL. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. Sci Rep. 2013;3:3529.
14. Gu Y, Kumar V, Hall LO, Goldgof DB, Li C-Y, Korn R, Bendtsen C, Velazquez ER, Dekker A, Aerts H, Lambin P, Li X, Tian J, Gatenby RA, Gillies RJ. Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. Pattern Recognit. 2013;46(3):692–702.
15. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, Mitra S, Shankar BU, Kikinis R, Haibe-Kains B, Lambin P, Aerts HJ. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One. 2014;9(7):e102107.
16. Xuan J, Adali T, Wang Y. Segmentation of magnetic resonance brain image: integrating region growing and edge detection. In: Image Processing, 1995 Proceedings, International Conference on Washington, DC, 1995. p. 544–547.
17. Horsch K, Giger ML, Venta LA, Vyborny CJ. Automatic segmentation of breast lesions on ultrasound. Med Phys. 2001;28(8):1652–1659.
18. Pham DL, Xu C, Prince JL. Current methods in medical image segmentation. Annu Rev Biomed Eng. 2000;2(1):315–337.
19. Hermoye L, Laamari-Azjal I, Cao Z, Annet L, Lerut J, Dawant BM, Van Beers BE. Liver segmentation in living liver transplant donors: comparison of semiautomatic and manual methods 1. Radiology. 2005;234(1):171–178.
20. Zhang X, Tian J, Xiang D, Li X, Deng K. Interactive liver tumor segmentation from CT scans using support vector classification with watershed. Conf Proc IEEE Eng Med Biol Soc. 2011;2011:6005–6008.

21. Kalpathy-Cramer J, Zhao B, Goldgof D, Gu Y, Wang X, Yang H, Tan Y, Gillies R, Napel S. A comparison of lung nodule segmentation algorithms: methods and results from a multi-institutional study. J Digit Imaging. 20163;29(4):476–487.
22. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. machine learning methods for quantitative radiomic biomarkers. Sci Rep. 2015;5:13087.
23. Liyang Wei, Yongyi Yang, Nishikawa RM, Yulei Jiang. A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications. IEEE Trans Med Imaging. 2005;24(3):371–380.
24. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.
25. Itakura H, Achrol AS, Mitchell LA, Loya JJ, Liu T, Westbroek EM, Feroze AH, Rodriguez S, Echegaray S, Azad TD, Yeom KW, Napel S, Rubin DL, Chang SD, Harsh GR, Gevaert O. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. Sci Transl Med. 2015;7(303):303ra138.
26. Echegaray S, Zamora G, Yu H, Luo W, Soliz P, Kardon R. Automated analysis of optic nerve images for detection and staging of papilledema. Invest Ophthalmol Vis Sci. 2011;52(10):7470–7478.
27. Koch Gary G. Intraclass correlation coefficient. In Samuel Kotz and Norman L. Johnson. Encyclopedia of Statistical Sciences. Vol. 4. New York: John Wiley & Sons. pp. 213–217; 1982.
28. Caceres A, Hall DL, Zelaya FO, Williams SCR, Mehta MA. Measuring fMRI reliability with the intra-class correlation coefficient. Neuroimage. 2009;45(3):758–768.
29. ePAD web-based platform for quantitative imaging in the clinical workflow. Available from: http://epad.stanford.edu
30. Rubin DL, Willrett D, O'Connor MJ, Hage C, Kurtz C, Moreira DA. Automated tracking of quantitative assessments of tumor burden in clinical trials. Transl Oncol. 2014 Feb;7(1):23–35.
31. Höhne KH, Hanson WA. Interactive 3D segmentation of MRI and CT volumes using morphological operations. J Comput Assist Tomogr. 1992;16(2):285–294.
32. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognit. 1996;29(1):51–59.
33. Chen W, Giger ML, Li H, Bick U, Newstead GM. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. Magn Reson Med. 2007;58(3):562–571.
34. Sato Y, Westin C, Bhalerao A, Nakajima S, Shiraga N, Tamura S, Kikinis R. Tissue classification based on 3D local intensity structures for volume rendering. IEEE Trans Vis Comput Graph. 2000;6(2):160–180.
35. Lu L, Shara N. Reliability analysis: calculate and compare intra-class correlation coefficients (ICC) in SAS. Northeast SAS Users Gr. 2007;14.
36. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–428.
37. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30–46.
38. Tukey JW. Comparing individual means in the analysis of variance. Biometrics. 1949;5(2):99–114.
39. Dancey CP, Reidy J. Statistics Without Maths for Psychology: Using SPSS for Windows. Upper Saddle River, NJ: Prentice-Hall, Inc.; 2004.
40. Doyle S, Agner S, Madabhushi A, Feldman M, Tomaszewski J. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In: Biomedical Imaging: From Nano to Macro, 2008 ISBI 2008 5th IEEE International Symposium on. 2008. p. 496–499.
41. Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, Cheshier SH, Napel S, Zaharchuk G, Plevritis SK. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. Radiology. 2014; 273(1):168–174.
42. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, Rubin DL, Napel S, Plevritis SK. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data–methods and preliminary results. Radiology. 2012;264(2):387–396.

43. Way TW, Hadjiiski LM, Sahiner B, Chan H-P, Cascade PN, Kazerooni EA, Bogot N, Zhou C. Computer-aided diagnosis of pulmonary nodules on CT scans: Segmentation and classification using 3D active contours. Med Phys. 2006;33(7): 2323–2337.

44. Kostis WJ, Reeves AP, Yankelevitz DF, Henschke CI. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical ct images. IEEE Trans Med Imaging. 2003;22(10):1259–1274.

45. Armato SG 3rd, Li F, Giger ML, MacMahon H, Sone S, Doi K. Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program. Radiology. 2002;225(3):685–692.

46. Gurcan MN, Sahiner B, Petrick N, Chan H-P, Kazerooni EA, Cascade PN, Hadjiiski L. Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system. Med Phys. 2002;29(11): 2552–2558.

47. Zhao B, Schwartz LH, Moskowitz CS, Ginsberg MS, Rizvi NA, Kris MG. Lung cancer: computerized quantification of tumor response–initial results. Radiology. 2006;241(3):892–898.

48. Zhao B. Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. J Appl Clin Med Phys. 2003;4(3):248–260.

49. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

50. Zacharaki EI, Wang S, Chawla S, Soo Yoo D, Wolf R, Melhem ER, Davatzikos C. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magn Reson Med. 2009;62(6):1609–1618.

51. Schwartz LH Colville JA, Ginsberg MS, Wang L, Mazumdar M, Kalaigian J, Hricak H, Ilson D, Schwartz GK. Measuring tumor response and shape change on CT: esophageal cancer as a paradigm. Ann Oncol. 2006;17(6):1018–1023.

52. Bricault I, Kikinis R, Morrison PR, VanSonnenberg E, Tuncali K, Silverman SG. Liver metastases: 3D shape-based analysis of CT scans for detection of local recurrence after radiofrequency ablation. Radiology. 2006;241(1):243–250.

53. Lo P, Young S, Kim HJ, Brown MS, McNitt-Gray MF. Variability in CT lung-nodule quantification: Effects of dose reduction and reconstruction methods on density and texture based features. Med Phys. 2016;43(8):4854.

54. Solomon J, Mileto A, Nelson RC, Roy Choudhury K, Samei E. Quantitative features of liver lesions, lung nodules, and renal stones at multi-detector row CT examinations: dependency on radiation dose and reconstruction algorithm. Radiology. 2016;279(1):185–194.

55. Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA, Qin Y, Riely GJ, Kris MG, Schwartz LH. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. Radiology. 2009;252(1):263–272.

56. Fletcher JG, Leng S, Yu L, McCollough CH. Dealing with uncertainty in CT images. Radiology. 2016;279(1):5–10.

57. Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, Goldgof DB, Hall LO, Korn R, Zhao B, Schwartz LH, Basu S, Eschrich S, Gatenby RA, Gillies RJ. Test-retest reproducibility analysis of lung CT image features. J Digit Imaging. 2014; 27:805–823.

58. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Trans Syst Man Cybern. 1973;3(6):610–621.

59. Soh L-K, Tsatsoulis C. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. IEEE Trans Geosci Remote Sens. 1999;37(2):780–795.

60. Clausi DA. An analysis of co-occurrence texture statistics as a function of grey level quantization. Can J Remote Sens. 2002;28(1):45–62.

61. Haralick R, Shanmugan K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern Syst. 1973;3(6):610–621.