

HOW RELEVANT ARE THE ESTIMATIONS OF THE FRACTAL DIMENSION OF THE TEXTURE AND CONTOUR TO DISCRIMINATE BETWEEN MALIGNANT AND BENIGNANT SINUS TUMORS – A STATISTICAL STUDY

Ion ANGHEL¹, Raluca Roxana PURNICHESCU PURTAN², Andreea UDREA³

The goal of the paper is to develop methods and algorithms based on the chaotic dynamical systems theory (fractal dimension analysis) and specific nonparametric statistics methods (bootstrap and bias-corrected and accelerated percentile intervals) and to apply them to study CT images of sinus tumors. The statistical analysis shows that the use of the fractal dimension of the mass texture discriminate best between malignant and benignant sinus tumor, therefore leads to an improvement of the diagnosis.

Scopul acestui articol este de a dezvolta algoritmi având ca punct de plecare metode din teoria sistemelor dinamice haotice (analiza dimensiunii fractale) și metode statistice nonparametrice (bootstrap și determinarea intervalelor de încredere) și de a-i aplica la studiul imaginilor CT reprezentând tumori maligne sau benigne ale zonei para-sinusale. Studiul statistic a arătat că dimensiunea fractală a texturii discriminează mai bine între tumorile maligne și benigne, ducând astfel la îmbunătățirea posibilităților de diagnostic.

Key words: Fractal dimension, box-counting, nonparametric bootstrap, bias-corrected and accelerated percentile intervals, paranasal sinus tumors.

1. Introduction

Fractal analysis and nonlinear time series analysis are branches of chaos theory. They provide useful methods for the characterization of single and multi variable signals (images and time series).

Usually, fractal analysis refers to a collection of methods for the description and quantization of geometric features of irregular forms and patterns. It was largely applied for the study of biological systems and subsystems at microscopic and macroscopic scale, image enhancement and compression,

¹ Reader, UMF "Carol Davila", Clinical Hospital ORL "Coltea", Bucharest, Romania

² PhD Student, University POLITEHNICA of Bucharest, Romania, e-mail:

raluca.purtan@gmail.com

³ PhD, University POLITEHNICA of Bucharest, Romania

fracture analysis, texture classification because of their fractal-like structure of the manipulated objects ([3], [7], [8]).

Its most known measure is the fractal dimension used to provide information on the irregularity of an object contour or self similarities in a texture.

Fractal analysis was largely applied for distinction between normal and modified tissue and between benignant and malignant tissue ([3], [7], [8]).

The case study considered in this paper targets the classification of masses into benignant and malignant. Benignant and malignant tumors of the paranasal sinuses may be asymptomatic for a long time and represents 3% of head and neck carcinomas ([6], [1]). When symptoms arise they often mimic the clinical picture of chronic rhinosinusitis, which is by far more frequent. For these reasons tumor diagnosis is often delayed ([9], [6]). As there is a myriad of different tumor entities, it may be beneficial for the clinician to distinguish between epithelial tumors (papilloma, pleomorphic adenoma), mesenchymal tumors (fibromatosis, osteoma), neural-related tumors (schwannoma) and tumor like lesions (giant cell granuloma, mucocoele) ([1]). Imaging is indicated whenever there is clinical suspicion for a neoplastic process. Computed tomography [CT] and magnetic resonance imaging [MRI] are complimentary in the pretreatment evaluation of patients with sinonasal tumors.

In this paper we present a study based on the fractal dimension of the analyzed masses formed in the paranasal sinuses region.

We consider a series (a number of 34) of CT images taken with contrast substance containing the tissue that has fractal – like structure. The 34 CT images consist in 17 malignant tumors and 17 benignant (two samples of same size). Distinction between malignant and benignant masses is made by calculating:

1. the fractal dimension of the mass's texture
2. the fractal dimension of the mass's contour.

A nonparametric bootstrap analysis with a jackknife correction for errors ([5], [4]) is run to assess the accuracy of our primary statistical results. We applied this technique to obtain more reliable statistical estimators (mean, standard deviation and confidence intervals for the mean) in order to make statistical inference, that is, to decide the power of discrimination between malignant and benignant masses, in both measurements (the fractal dimension of the mass's texture and the fractal dimension of the mass's contour).

We conclude that a more accurate discrimination is obtained when using the F_d of texture.

2. Methods

In the last 2 decades fractal analysis become a powerful tool for analyzing the form, pattern and growth in biological systems. This kind of details are

captured, best, by using the fractal dimension (F_d) which provides us a measure of non/smoothness of a contour or a texture. The fractal dimension is calculated by applying the „box-counting” estimation algorithm because, in comparison to other methods, it offers two major advantages: it is easy to and can be applied for images no matter how complex.

The statistical inference is based on a nonparametric bootstrap technique. The general bootstrap method introduced by Efron (1979) is a resampling procedure for estimating the distributions of statistics based on independent observations. The bootstrap method is shown to be successful in many situations, which is being accepted as an alternative to the asymptotic methods. In fact, it is better than some other asymptotic methods, such as the traditional normal approximation and the Edgeworth expansion ([2]). The *nonparametric bootstrap* is suitable when the population distribution function is completely unknown and when the sample size is small (less than 30).

Box-Counting Fractal Dimension Estimation Method

The „box-counting” fractal dimension, derived from the *Hausdorff* coverage dimension is given by the following approximation:

$$Fd \approx \frac{\log(N(s))}{\log(1/s)} \quad (1)$$

where: - $N(s)$ is the number of squares that contain information when grid covers the image (parts of the extracted contour/texture)

- s squares' side length.

It is expected, that for a smaller s value, the above approximation should be better:

$$Fd = \lim_{s \rightarrow 0} \log \frac{N(s)}{\log(1/s)} \quad (2)$$

If this limit exists, it is called the „box-counting” dimension of the measured object. In practice, this limit converges slowly, that is why the following expression is used:

$$\log(N(s)) = Fd \cdot \log\left(\frac{1}{s}\right) \quad (3)$$

This is the equation of a straight line of slope F_d , the „log-log” curve described by the points of $(\log(N(s)), \log(1/s))$ for different values of the square's side s . Through linear regression (least squares method) the slope of the line that approximates the points' distribution is determined; this is the „box-counting” fractal dimension.

The „box-counting” algorithm assumes to determine the fractal dimension in accordance with the dependence of the object contour or texture upon the used scale factor.

It consists in :

1. to transform the grey shades CT image into a binary image by applying a certain threshold and all the pixels with grey levels over that threshold are made black and the rest white.
2. to cover successivly the image by applying grids of squares with equal sides s of dimensions powers of two: 2, 4, 8, ... and counting every time the squares that contain at least one black pixel - some part of the analyzed object ($N(s)$).
3. the points of coordinates $(\log(N(s)), \log(1/s))$, where s is the common side of the coverage squares, and $N(s)$ the number of squares that contain information, are positioned approximately in a line and its slope will be the fractal dimension in "box-counting" context/perspective.

We'll apply the "box-counting" algorithm, described above, for different scale values s , using software – MorfoFractal, a graphical application developed by the authors.

We have used images that contain those parts of the masses that are not obstructed by nearby bones, such that we can obtain information on the free evolution of the mass.

The "box-counting" fractal dimension is estimated for:

1. the masses (tumors) textures and
2. the masses (tumors) contours.

In the second case, the contour of the mass is extracted by using following procedure:

- search for all the pixels that are black and are not part of the object, their neighbors are all white (or, in some cases, have a limited number of black neighbors – one or two) and make them white.
- search for all the pixels that are black and have the upper, lower, right and left neighbors black and make them white. (Fig.1).

By employing this procedure we will obtain the interconnected pixels in the image that delimitate the object from the background.

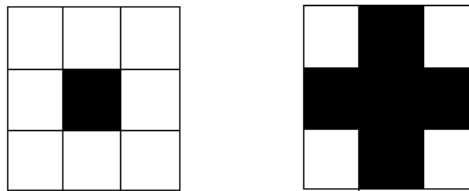


Fig. 1. a) Black pixel with all the neighbors of white color – becomes white ; b) Black pixel with upper, lower and lateral neighbors of black color – becomes white

Statistical Methods

A traditional approach to statistical inference is to make assumptions about the structure of the population (e.g., an assumption of normality), and, along with the stipulation of random sampling, to use these assumptions to derive the sampling distribution, on which classical inference is based. This familiar approach has two potentially important deficiencies: if the assumptions about the population are wrong, then the corresponding sampling distribution of the statistic may be seriously inaccurate. On the other hand, if asymptotic results are relied upon, these may not hold to the required level of accuracy in a relatively small sample.

In contrast, the *nonparametric bootstrap* allows us to estimate the sampling distribution of a statistic empirically without making assumptions about the form of the population, and without deriving the sampling distribution explicitly.

The general nonparametric bootstrap procedure is based on five steps as follows (refer to Efron and Tibshirani [5] for detailed discussions):

Consider the case where a random sample of size n is drawn from an unspecified probability distribution, F , with the observations x_1, x_2, \dots, x_n viewed as realizations of independent random variables with common distribution function F . We denote the interest estimate parameter as $\hat{\theta}$. The basic steps in the nonparametric bootstrap procedure are:

Step 1: Construct an empirical discrete probability distribution, F_n , that gives probability of $1/n$ to each observed value x_1, x_2, \dots, x_n . This is the nonparametric maximum likelihood estimate of the population distribution, F .

Step 2: Draw the resample, that is, a random sample of size n , with replacement, from the empirical distribution F_n . Calculate the statistic of interest, $\hat{\theta}$, for this resample, denoted $\hat{\theta}^*$.

Step 3: Repeat Step 2 " B " times, by Monte Carlo resampling and compute for each resample the statistic $\hat{\theta}^*$ (B is at least equal to 1000 when an estimate of confidence interval around $\hat{\theta}$ is required).

The empirical distribution of the resulting values $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$ is an approximation to the distribution function of $\hat{\theta}$. The bootstrap statistic $\hat{\theta}$ is approximated by the average of $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$ values.

Step 4: Calculate the confidence intervals for the statistic of interest, $\hat{\theta}$ using the *bootstrap bias-corrected and accelerated (BCa) percentile intervals*.

There are several approaches to constructing bootstrap confidence intervals. One of the most widely used is the *bootstrap percentile interval*: the endpoints of a

0.05 α -level confidence interval for θ (the population parameter estimated by $\hat{\theta}$) would be the ordered bootstrap replicates of the statistic values of $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$ at the 2.5th and the 97.5th quantiles of the bootstrap distribution. Although they do not artificially assume normality, percentile confidence intervals often do not perform well in terms of bias and skewness in the bootstrap distribution ([4]). The *bootstrap bias-corrected and accelerated* (BCa) *percentile intervals* method adjusts for both bias and skewness in the bootstrap distribution ([4]). This approach is accurate, has reasonable computation requirements, and does not produce excessively wide intervals.

To find the BC_a interval for the interest parameter θ , we calculate:

$$z = \Phi^{-1} \left[\frac{\sum_{i=1}^B \mathbb{I}(\theta_i^* \leq \hat{\theta})}{B+1} \right]$$

where $\Phi^{-1}(\cdot)$ is the standard normal quantile function and $\sum_{i=1}^B \mathbb{I}(\theta_i^* \leq \hat{\theta})$ is the adjusted proportion of bootstrap replicates at or below the original sample estimate $\hat{\theta}$ of θ . If the bootstrap sampling distribution is symmetric, and if $\hat{\theta}$ is unbiased, this proportion will be close to 0.5 and the correction factor z will be close to 0.

Let $\theta_{(-i)}$ represent the value of $\hat{\theta}$ produced when the i^{th} observation is deleted from the original sample (the 1-deleted jackknife); there are n of this quantities. Let $\hat{\theta}_m$ represent the average of the $\theta_{(-i)}$ values. We calculate another correction factor:

$$a = \frac{\sum_{j=1}^n (\theta_{(-j)} - \hat{\theta}_m)^3}{6 \left(\sum_{j=1}^n (\theta_{(-j)} - \hat{\theta}_m)^2 \right)^{3/2}}$$

With the correction factors z and a , compute:

$$a_1 = \Phi \left[z + \frac{z - z_{1-\alpha/2}}{1 - a(z - z_{1-\alpha/2})} \right] \text{ and } a_2 = \Phi \left[z + \frac{z + z_{1-\alpha/2}}{1 - a(z + z_{1-\alpha/2})} \right]$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The values a_1 and a_2 are used to locate the endpoints of the corrected percentile confidence interval:

$$\theta_{[B \cdot a_1]}^* < \theta < \theta_{[B \cdot a_2]}^*$$

When the correction factors z and a are both 0, $a_1 = \Phi(-z_{1-\alpha/2}) = \Phi(z_{\alpha/2}) = \alpha/2$, and $a_2 = \Phi(z_{1-\alpha/2}) = 1 - \alpha/2$, which corresponds to the (uncorrected) percentile interval.

The above algorithm and the random generator of the bootstrap samples was implemented as a MatLab routine. The bootstrap analysis was conducted for $B=2000$, for both samples of $n=17$ and in each sample, for both types of measurements (the fractal dimension of the mass's texture and the fractal dimension of the mass's contour). In each of the four resulted data sets, the statistics of interest were the mean and standard deviation. The confidence interval was calculated for the mean with the BC_a method.

3. Results

Two sets of CT images representing sinus tumors are analyzed: 17 malignant (Fig.2) and 17 benignant (Fig.3). The fractal dimension values of the contours and textures of all this images were estimated using the procedures described in Chapter 2. Some results are displayed in Table 1.

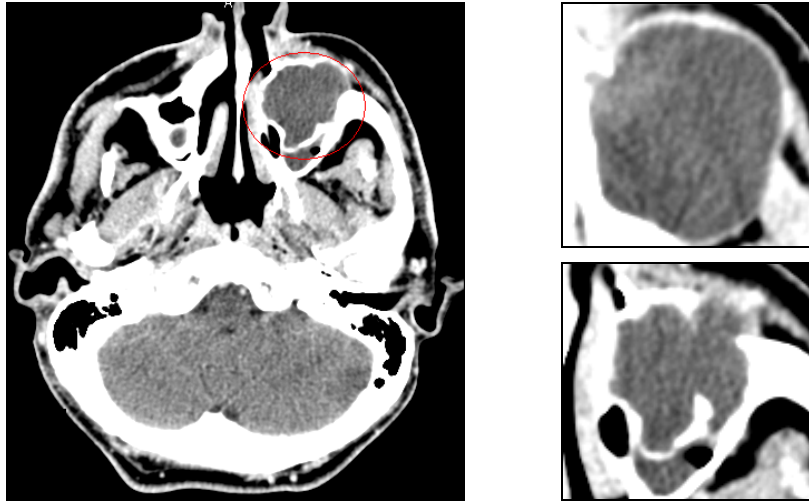


Fig 2. CT slice containing an analyzed malignant maxillary tumor (left) and another two regions from other CT slices also presenting malignant tumors (upper and lower right).

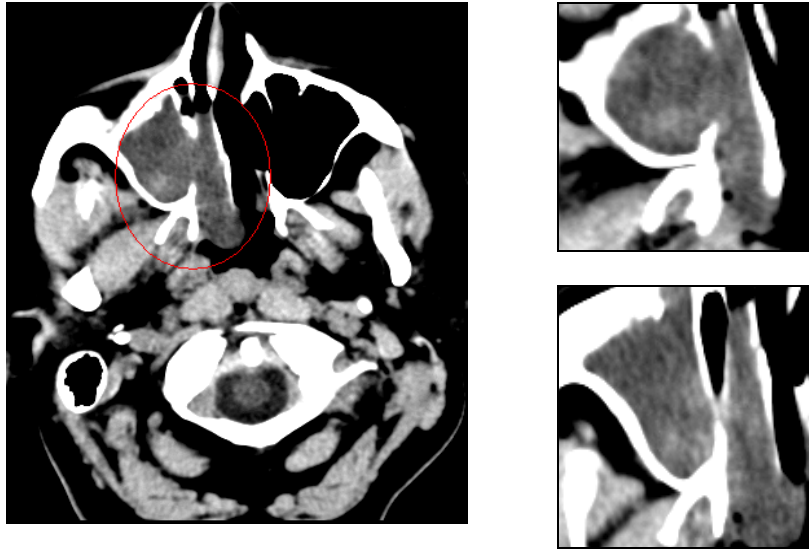


Fig 3. CT slice containing an analyzed polipoide formations (benignant sinus tumor) - left and another two regions from other CT slices also presenting benignant tumors - upper and lower right.

Table 1.

F_d of the textures and contours of some of the analyzed tumors (masses)

Group 1 – maxillary tumors (malign)		Group 2 – polipoide formations (benign)	
Fractal dimension of the mass texture	Fractal dimension of the mass contour	Fractal dimension of the mass texture	Fractal dimension of the mass contour
1.83	1.65	1.60	1.40
1.78	1.60	1.70	1.53
1.75	1.58	1.66	1.48
1.80	1.70	1.72	1.55
1.73	1.59	1.65	1.4
1.70	1.60	1.75	1.64

The statistical results obtained for malign and benign tumors by calculating the masses texture F_d and the masses contours F_d , with the nonparametric bootstrap method described in Chapter 2 are presented in Table 2.

Table 2.

Statistical estimations for malign and benign tumors (mean, SD, CI)

	Fractal dimension of the mass texture		Fractal dimension of the mass contour	
	Estimated mean (SD)	95% Confidence Interval (estimated mean)	Estimated mean (SD)	95% Confidence Interval (estimated mean)
Group 1 (malign)	1.776 (0.43)	(1.736 , 1.815)	1.617 (0.41)	(1.571 , 1.668)
Group 2 (benign)	1.665 (0.51)	(1.623 , 1.715)	1.497 (0.36)	(1.402 , 1.597)

The values obtained for the estimated mean of mass texture and contour discriminate both between the malignant and benign groups (Kolmogorov-Smirnov test, $p=0.016$, respective $p=0.025$). Because of the overlapping 95% confidence interval in the case of fractal dimension of the mass contour (values between 1.571 and 1.597 can be misclassified), we conclude that the fractal dimension of the mass texture is a more accurate discrimination measurement between benignant and malignant masses formed in the paranasal sinuses region.

4. Conclusions

Because of the asymptomatic or confounding symptoms of the paranasal sinuses tumors, any improvement in early diagnosis is beneficial for the rhinologist practitioner. The goals of patient evaluation for suspected sinonasal cancers include a precise diagnosis, determining the extend of disease and developing a treatment plan. The two powerful algorithms involved in the study – the box-counting and the nonparametric bootstrap with BCa revealed that computing the fractal dimension of the mass texture is an accurate and valuable method for a precise diagnosis.

The F_d for masses texture leads to a better statistical discrimination because it takes into consideration all the details of the analyzed image sections.

By improving the CT noninvasive diagnosis methods, the presence of a sinonasal tumor can be more accurate predicted, leading to a more efficient medical evaluation and treatment.

R E F E R E N C E S

- [1] *I. Anghel, A. Anghel*, (2011). "Inverted papilloma", *Practica Medicală*, Bucharest, vol 6, nr.1: p.30-35;
- [2] *A.C.Davison, D.V. Hinkley*, (1997). "Bootstrap Methods and their Application". Cambridge: Cambridge University Press.
- [3] *R.Dobrescu, C. Vasilescu* (2004). "Interdisciplinary Applications Of Fractal And Chaos Theory", *Academia Română*, Bucuresti,
- [4] *B. Efron*, , (1987). "Better Bootstrap Confidence Intervals". *Journal of the American Statistical Association* 82 (397): 171–185.
- [5] *B. Efron*, , *R.J. Tibshirani*, (1993). "An introduction to the bootstrap". Chapman and Hall, London.
- [6] *G. Galzada, R. Weber*, (2009). "Malignant Tumors of the Nose and Paranasal Cavity", *Rhinology and Facial and Plastic Surgery*, Springer: p.388-394;
- [7] *G. Landini* (1998). „Complexity in tumor growth patterns” in *Fractals in Biology and Medicine*, ed. G.A. Losa, Birkhauser Verlag : p.268-284

- [8] *M. Olteanu, A. Udrea*, (2010). "Image Analysis based on the study of the attractor of a time series", U.P.B. Sci. Bull. Series A, Vol. 72, Iss. 4, ISSN 1223-7027 p.11-18
- [9] *B.Schik, J. Długaiczek*, (2009). "*Benign Tumors of the Nasal Cavity and Paranasal Sinuses*", Rhinology and Facial Plastic Surgery, Springer: p.378-385.