

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Tumor Malignancy Characterization in Clinical Environments: An Approach Using the FYC-Index of Spiculation and Artificial Intelligence

*Fernando Yepes-Calderón, Flor M. Medina,  
Nolan D. Rea and José Abella*

## Abstract

According to the World Health Organization, cancer is the second leading cause of death in the world. The myriad of variations, paths of development, and mutations make this abnormality challenging to treat. With the advent of medical imaging, complex qualitative information is collected with the aim of characterizing the pathology; however, the uncomfortable and time-consuming histology remains the state of care within hospitals. This manuscript presents a strategy to extract quantifiable features from the images. The method captures shape perturbation as variations in reference to a perfect circle that is used in a standardized dimensional space. A multifeatured scheme is created when the quantification is applied in all slices produced by imaging modalities such as computed tomography, magnetic resonance imaging, and tomosynthesis. Later, the numbers obtained by the introduced algorithm are used in an artificial intelligence pipeline that correlates spicularity with aggressiveness using the histology as supervising factor.

**Keywords:** medical image analysis, tumor grading, cancer, tumor characterization

## 1. Introduction

Classifying cancer lesions in form and intensity from the images is of interest in radiology units [1–3]. Currently, histology is the gold standard to define cancer type, stage, and grade; nevertheless, histology comes with its associated costs and delays and has been reported to increase morbidity [4, 5]. When diagnosing from the images, the desired classification is accurate and repeatable only if the operator includes the quantitative domain to the set of available tools that are mostly from the qualitative domain. The quantification is accomplished by separating the neo-mass from the anatomical parts in the image employing segmentation.

Regarding segmentation, authors have proposed assisting techniques that partially or fully accomplish the tasks with different levels of accuracy [6–8].

After segmentation, the challenge is finding a repeatable and performant method for all kinds of cancer manifestations. Some quantifying approaches target cancer in specific parts of the body [9, 10], while others focus on particular kinds of cancer [11, 12].

Although technology has invaded the medical facilities, currently assisting tools are not of help in diagnosing cancer. The tasks are still performed by human experts employing purely qualitative judgment. There is a need to quantify and thus abandon the uncertainty produced by human variability.

In practice, qualitative features suggested by X-rads [13, 14] such as roughness and stiffness are difficult to conceptualize with mathematical models; therefore, indexes based on these features are complicated to model [15]. However, the shape of the captured objects is a stable feature in the field of view [16, 17] and, conveniently, has the required sensibility across all cancer manifestations because it captures the core manifestations of the disease, the disordered growth pattern [18, 19]. More importantly, tumor shape is quantified in a feature-enriched scheme to favor further machine-learning implementation. In this document, we employ the FYC-Index of spiculation [20] to assert quantification on the edges of breast tumors imaged with tomosynthesis [21, 22]. The numbers yielded by the FYC-Index strategy are fed to an artificial intelligence classifier that initially differentiates between benign and malign neo-masses, showing a high degree of accuracy in supervised experiments. The presented strategy is equally performant in all imaging techniques that generate volumetric representations by slicing, including MRI, CT, and tomosynthesis.

## 2. Materials and methods

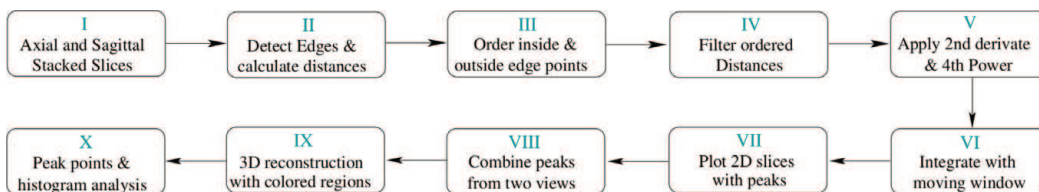
### 2.1 Clinical data

A cohort of 48 breast tomosynthesis images underwent segmentation performed by an expert radiologist. Histology was performed on the 48 masses yielding 29 malignant cases and 19 benign. The resulting masks hold the specifications of the original images regarding the field of view and spatial resolution. Since the algorithm explained in Section 2.2 is immune to resolution changes and the field of view is standardized, records of the images' specifications are not provided in this document.

### 2.2 The FYC-Index of spiculation

The reader is invited to refer to **Figure 1**. Recall that the procedure explained below is used on both views, axial and sagittal.

**Block I.** Images of CT, MRI, or tomosynthesis are suitable for this processing due to their slicing nature. The FYC-Index re-sample all masks to isometric



**Figure 1.**  
Block diagrams for the FYC-Index pipeline.

voxels of 1 mm before selecting the biggest mask through area calculation. Next, the dimensions of the biggest bounding box are used as a dimensional template. Then, the other slices in the study—including those of other tumors in case we are working with a population—are scaled to the dimensions of the biggest bounding box. This process also centers the masks. Distortion in the mask growing process is avoided by using the adaptive supersampling method [23, 24]. After scaling, all the images share the same field of view (FOV) and therefore, the same planar coordinates for the center point.

**Block II.** Then, the edges of the masks are detected using the Canny edge detector [25].

**Block III.** As the Canny detector does not create single-pixel edges, the system detects two paths corresponding to the outer and inner edges. The Euclidean distance from the artificially created center of coordinates to each point in the edge is saved in two arrays, one corresponds to the outer edge and the other to the inner edge. The two arrays are averaged in an ordered array of distances (AoD). The run along the edge that creates the AoD is standardized by starting the distance calculation at the top center of the image and taking the edging points in a clockwise fashion until the starting point is located at a distance of  $\sqrt{(2)}$  mm of the current point or below. Recall that voxels are all set to 1 mm.

**Block IV.** The AoD is Gaussian filtered creating the FAoD. This filtering is intended to eliminate the high-frequency components produced by the digital grid. The filter is implemented in the frequency domain, keeping 80% of the original spectral power. According to [26], maintaining the 80% of the signal spectral power assures that the important content of the signals is kept.

**Block V.** A five-point differentiation is applied to find the regions of rapid change; next, a second five-point differentiation is executed to recover the inflection points. These operations are generalized to each point in FAoD as shown in Eq. 1:

$$p'_n = [(d_n - d_{n-2}) + (d_n - d_{n-1}) + (d_{n+1} - d_n) + (d_{n+2} - d_n)]/4 \quad (1)$$

The second derivative “p”—obtained with the second pass of Eq. (1)—is where peaks are detected. The peak elements on FAoD are exalted, while regions of low dynamics in the same array are diminished when raising FAoD to the fourth power.

**Block VI.** A moving window integration selects peaks in the exponential second derivative. Given points  $p(s) = (s, d)$ , the area ( $A$ ) under the curve section with a width ( $N$ ) is calculated for step ( $s$ ), as it is shown in Eq. (2). If  $A(s) > T$  for a chosen value of  $T$ ,  $s$  is added to this list of locations of peaks:

$$A(s) = \sum_{i=-\frac{N}{2}}^{\frac{N}{2}} p \left[ \left( \frac{sN}{2} \right) - \left( \frac{N}{2} + i \right) \right] \quad (2)$$

**Block VII.** The spiculation quantifying process is executed in axial and sagittal views.

**Block VIII.** The location of the detected peaks is then crossed; we only kept the points that coincide in both views.

**Block IX.** These points when mapped back in the images uplift the regions where the tumor presents a highly disorganized growing pattern.

**Block X.** Each slice in the study contributes to the histogram signature of the tumor. The FYC-Index defines the span of the histograms by using the maximum and minimum amount of points found in the slices when working on a single mass or among all analyzed tumors when working on populations.

Under the FYC-Index domain, while more spiculation, the histogram profiles are more populated in the right side.

### 2.3 Proof of concept on synthetic data

Testing on extremes is a common practice in engineering. Unfortunately, finding extremes on clinical data is cumbersome. The difficulty relies on the nature of the information; in the clinics, where the patients are imaged on the presumption that some abnormality is present, the images often yield moderately spiculated masses, posing a problem overall for the lower extreme reference. Regarding the highly spiculated reference, one can use the histogram signature to pinpoint the slice yielding the most right-filling pattern. However, a sounded proof of concept should comply with the common complexity found in the clinics, where two masses can have similar volumes and have a different nature regarding malignancy; thus, conventional methods are unable to detect differences. To overcome this problem, we have created a synthetic framework where lower references are created by stacking the less spiculated slice among all the data analyzed. A mildly spiculated mass is created by stacking a mildly spiculated slice among the study, and, analogously, the extreme spiculated sample is created by stacking the most spiculated slice found in the study. For the three samples, the stacking is driven in a manner that the masses end by having a similar volume.

### 2.4 Artificial intelligence (AI) implementation

Every column in the histogram signature created by employing the procedure in Section 2.2 is seen as a feature in classification postulate that aims to distinguish between malign and benign samples. This is possible due to the independence of the peaks counting in a slice by slice fashion. In general, the perturbations on the slice  $n$  do not have any correlation with the perturbations on slice  $m$ ; therefore, orthogonality is granted. In addition to the bins counting, the number of bins fulfilled—some bins may end empty—those filled from the middle bin to the right and those filled to the middle bin to the left, is also used in the featuring space.

Every tumor population has a different span in the histogram signature; however, the amount of peak-counting-derived features have been set constant by forcing seven equally spaced bins regardless of the peak-counting range. Thus, the experiments always create an analyzing matrix containing 11 columns, 10 columns for the features, and 1 column to register the supervising factor provided by the histology. The current exercise presents a boolean support vector machine (SVM) classifier, where the machine is trained to provide a benign or malign verdict.

The data matrix is scaled and normalized using Python-Pandas [27, 28]. The classifier estimators are proved by cross correlation where the train and test samples are gathered from the original dataset using (0.7 : 0.3) (train:test) in a fivefold experimental scheme. For the classification experiments, Scikit-learn [29] is employed.

Listing 1: Python code use to run the SVM classifier while progressively adding features.

```
def run Features Testing Classification (features, mdata, lbls):
    atregs = []
    ascores = []
    for i in length(features):
        est = feature_selection.SelectKBest (k=i)
        est . fit (mdata, lbls)
        tregs = est . get_support (indices = True)
```

```
ndata = est . transform (mdata)
estsvm = svm . LinearSVC ()
gs = grid_search . GridSearchCV (estsvm, {'C' : np . logspace (-4,3)})
tscore = np . mean (cross_validation . cross_val_score (gs, ndata, lbls, n_jobs=5))
```

As it is shown in Listing 1, the SVM classification is done after progressively adding features which are grabbed from the *mdata* matrix using the indexes saved in the features' array. The accuracy records presented in Section 4 correspond to the experiment that yielded the highest accuracy values per folding.

3. Results

3.1 FYC-Index extraction

Figure 2 shows how the algorithm yields two different outcomes based on the tortuosity of the two analyzed shapes. The small shape refers to a mostly rounded region of interest (ROI), therefore, does not present abrupt changes in the distances from the edging points to the center of FOV. In contrast, the same measure yields rapidly changing distances in the big ROI. Those rapidly changing distances are captured by the first derivate and framed in their inflection points by the second derivate. Later, those points are amplified and made all positive by the fourth power function, while the same fourth power function diminishes changes in which the derivate yielded values in the range  $(-1, 1)$ . As the moving window adds up all values encountered in its domain, the regions of rapid change represented by large values compute to higher numbers within the domain of the moving window, and that is where the enhanced points appear in the plot. As all the points are mapped with their original coordinates, a crossing of 3D positions among the selected points

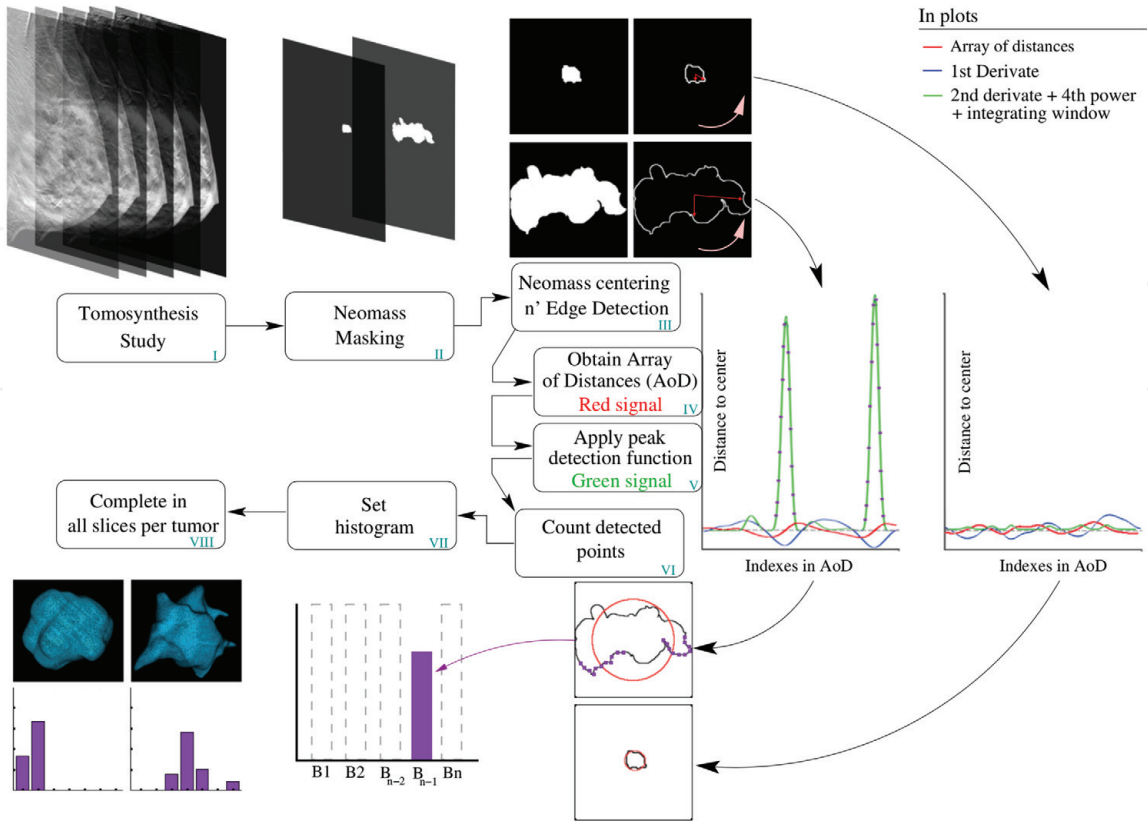
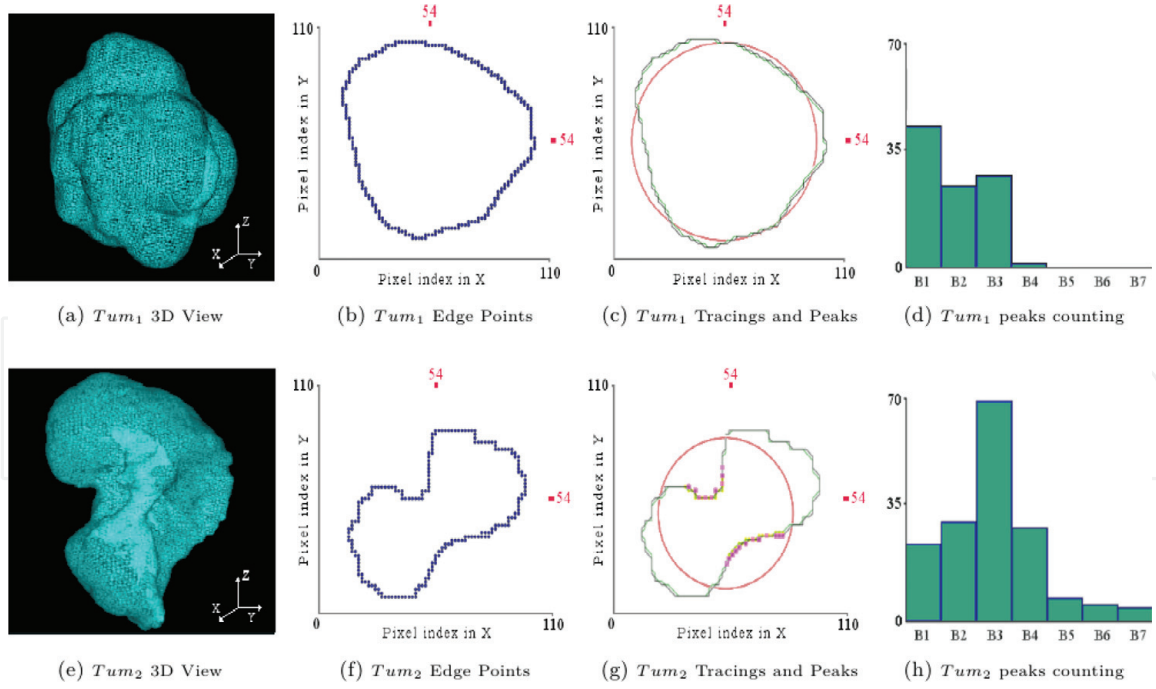


Figure 2. FYC-Index extraction. The inner loop is the detailed block diagram similar to the one shown in Figure 1 but specific for tomosynthesis images. The outer loop shows sampling images on each block.



**Figure 3.**

*Histogram signature of the FYC-Index on two tumors and intermediate steps (b, c, d, g) of processing. The circles in frames in (c) and (g) correspond to perfectly rounded regions where the area is equal to the one of the mask.*

in two image views filters out positions erroneously selected. Finally, the presented procedure allocates an item of frequency in a histogram where the bins contain ranges of point counting. Naturally, highly spiculated slices contribute mostly to the right bins of the histogram. When all slices in a tumor have been analyzed, the operator could be sure that the histogram is descriptive of the degree of homogeneity of the mass which is also associated with aggressiveness (see **Figure 3**).

A sample of the process where the 3D reconstruction of the masses together with the respective normalized FYC-Index histogram is presented in **Figure 4**.

### 3.2 Analysis of synthetic data

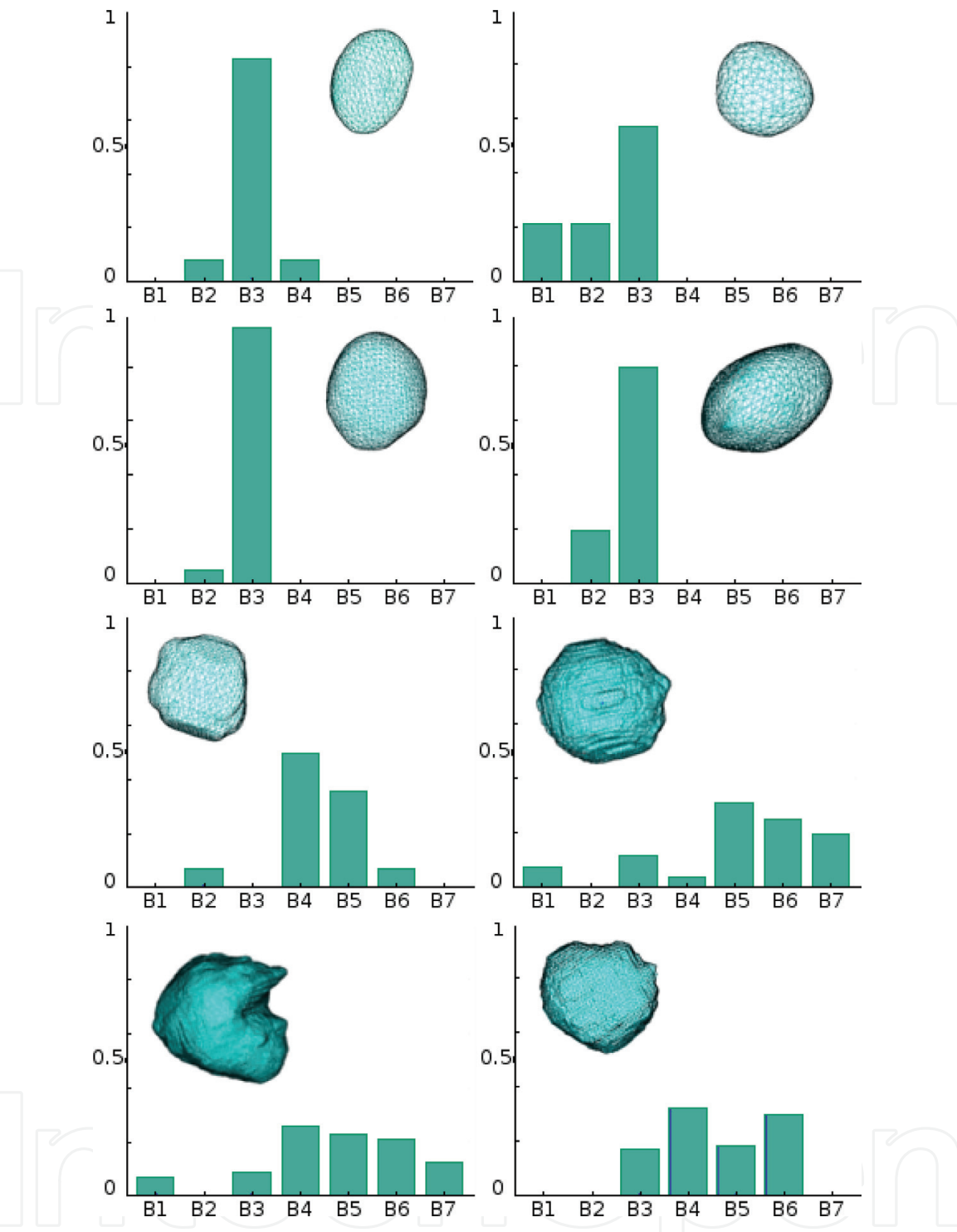
As explained in Section 2.3, extreme references are created to demonstrate the span of the method and the capacity to deliver a representation of easy interpretation. The synthetic creations are shown in **Figure 5**.

The results obtained on synthetic data corroborates that the FYC-Index is sensible to the changes in the edges that distinguish between malign and benign masses. In contrast, commonly used geometrical indexes are not sensitive to changes. In this exercise, we have isolated the spiculation by equalizing the volumes of the studied software objects. A complete set of 3D geometrical functions are applied on the clinical data in use, with the aim of comparing the performance of standard of care tools in the clinics, and the FYC-Index is shown in **Figure 6**.

## 4. Verdicts dictated by (AI) implementation

The fivefolding SVM exercise proposed in Section 2.4 was executed using a Python-Pandas dataframe and Scikit-Learn SVM. The results are registered in **Table 1**.

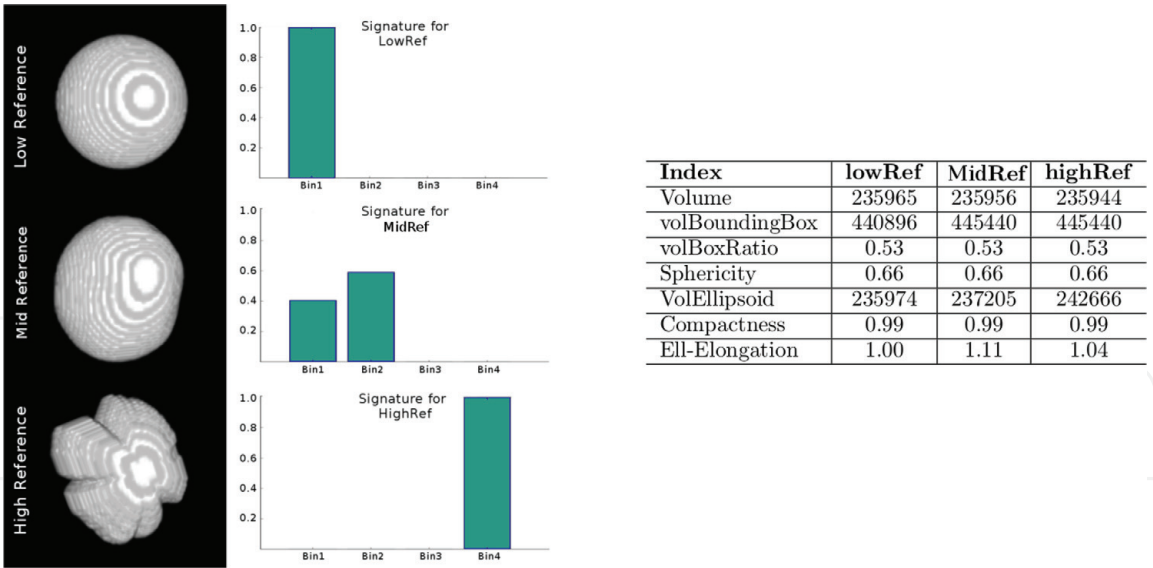
The strong-force algorithm presented in Section 2.4 executed the supervised classification with a high degree of accuracy. The design of the experiments turns the classification into the capacity to differentiate whether a mass is benign or malign.



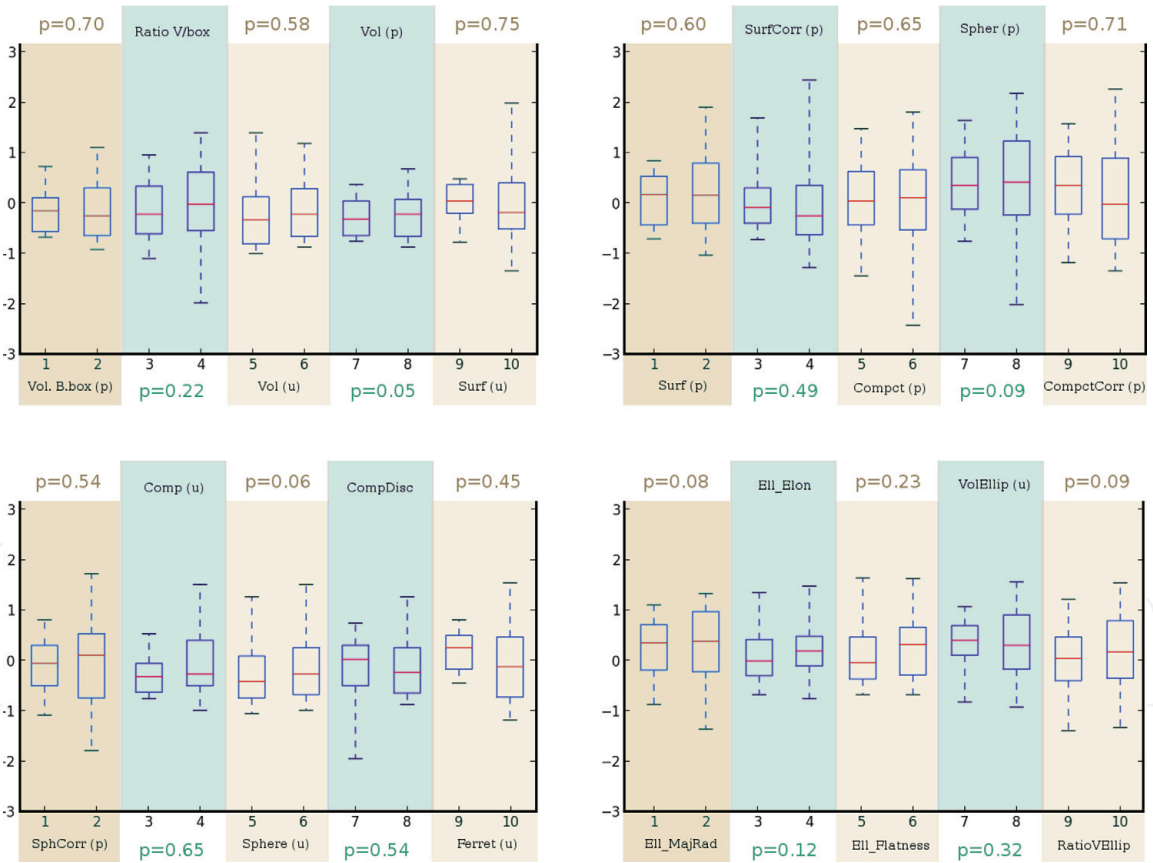
**Figure 4.**  
*A sample of the processed tumors and their FYC-Index signatures.*

## 5. Discussion

The proposed method is sensitive to slight changes in the edges of the masses that are characteristically malignant. The same method includes a stage of quantification that has proven to be descriptive at a simple glance even for nonspecialized operators. Since the procedure has been automated, it is compliant with the confidentiality regulations and, therefore, can be easily implemented in hospitals and clinics. The FYC-Index is a flexible method equally performant when analyzing masses in individuals and populations. The method presents a signature which results in a measure of lobularity. This strategy works regardless of factors such as size and spatial resolution. Moreover, the results are direct and easy to interpret.



**Figure 5.** Performance of the FYC-Index in software-created references. On the right, a table with records of often used 3D geometrical indexes. Note that these indexes are not sensible within the characteristics that require to be quantified.



**Figure 6.** The two boxes per colored column correspond to the clinical data detailed in Section 2.1. Normality was discarded by Kolmogorov test [30] done in the two groups separately. As normality was not met, the nonparametric Kruskal-Wallis test [31] was employed. The p-values are mapped back and forward in the chi-square distribution.

The specifications of the FYC-Index make it suitable to analyze all sort of cancer manifestations, regardless of localization or pathogenic roots. The presented strategy uses a machine-learning classifier to rapidly characterize the malignancy of a mass. However, the real challenge consists of defining malignancy together with aggressiveness. Such an approach requires more rounds of training/testing sessions

Folding	Accuracy (%)	Sensibility (%)	Specificity (%)
1	93.4	92.1	89.1
2	91.3	92.4	90.4
3	90.3	90.2	89.6
4	92.7	90.1	89.3
5	89.9	90.0	89.6

*Results for the fivefolding experiments on histograms acquired with the FYC-Index of spiculation.*

**Table 1.**  
*SVM-supervised classification.*

with sufficient samples in all grading range. This multilevel classification should be designed to follow the classification directives presented in the X-RADS standards; thus, the existing automatic tools can also provide insights for selecting more accurate treatments. To the best of our knowledge, no other authors are integrating the tools as we have proposed. The use of the features we have proposed is a novel view of the solution; therefore, we do not include in this report a comparison with other methods.

6. Conclusion

Cancer is the second most threatening disease which humanity has not been able to neutralize. Other diseases that were considered pandemics in the past, costing millions of human lives, have been eradicated through vaccination. Rapidly mutating diseases such as AIDS have been downgraded from mortal to chronic. Maladies like high blood pressure, stroke, or cirrhosis among several other chronic afflictions have been associated with race, genetics, habits, or exposition factors, providing a way to reduce the probability of acquiring them or a path of development where scientists still have space to explore. Cancer instead affects all humans regardless of any factor. The only aspect that increases the surviving expectations, without a doubt, is early detection, and it is here where the method presented in this manuscript gains relevance. Detection from the images is possible, and automatic diagnosis not only avoids the painful and uncomfortable biopsy, but it also contributes to faster and more accurate verdicts.

IntechOpen

### **Author details**

Fernando Yepes-Calderón<sup>1\*</sup>, Flor M. Medina<sup>2</sup>, Nolan D. Rea<sup>1</sup> and José Abella<sup>2</sup>

<sup>1</sup> Children's Hospital Los Angeles, Los Angeles, CA, USA

<sup>2</sup> Fundación Valle del Lili, Cali, Colombia

\*Address all correspondence to: fernandoyepesc@gmail.com

### **IntechOpen**

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Catalano OA, Samir AE, Sahani V, Hahn PF. Pixel distribution analysis: Can it be used to distinguish clear cell carcinomas from angiomyolipomas with minimal fat? *Genitourinary Imaging*. 2007;**247**(3):11-38. URL: [www.rsna.org/rsnarights](http://www.rsna.org/rsnarights)
- [2] Chaudhry HS, Davenport MS, Nieman CM, Ho LM, Neville AM. Histogram analysis of small solid renal masses: Differentiating minimal fat angiomyolipoma from renal cell carcinoma. *Genitourinary Imaging*. 2011;**1**:1-3. DOI: 10.2214/AJR.11.6887. URL: [www.ajronline.org](http://www.ajronline.org)
- [3] Arakeri MP, Reddy RM. A novel CBIR approach to differential diagnosis of liver tumor on computed tomography images. *Procedia Engineering*. 2012;**38**: 528-536. DOI: 10.1016/j.proeng.2012.06.066
- [4] Stigliano R, Marelli L, Yu D, Davies N, Patch D, Burroughs AK. Seeding following percutaneous diagnostic and therapeutic approaches for hepatocellular carcinoma. What is the risk and the outcome? Seeding risk for percutaneous approach of HCC. *Cancer Treatment Reviews*. 2007;**33**(5): 437-447. DOI: S0305-7372(07)00041-2 [pii] 10.1016/j.ctrv.2007.04.001. URL: <http://www.ncbi.nlm.nih.gov/pubmed/17512669>
- [5] Bedossa P. Liver biopsy. *Gastroentérologie Clinique et Biologique*. 2008;**32**:4-7
- [6] dos Santos DP, Kloeckner R, Wunder K, Bornemann L, Düber C, Mildenerberger P. Effect of kernels used for the reconstruction of MDCT datasets on the semi-automated segmentation and volumetry of liver lesions. *Technique and Medical Physics*. 2013; **19**:780-784. DOI: 10.1055/s-0033-1356178
- [7] Guo Y, Feng Y, Sun J, Zhang N, Lin W, Sa Y, et al. Automatic lung tumor segmentation on PET/CT images using fuzzy Markov random field model. *Computational and Mathematical Methods in Medicine*. 2014;**2014**:8-10. DOI: 10.1155/2014/401201
- [8] Schulz J, Skrivseth SO, Tømmerås VK, Marienhagen K, Godtliebsen F. A semiautomatic tool for prostate segmentation in radiotherapy treatment planning. *BMC Medical Imaging*. 2014; **12**(1):1. DOI: 10.1186/1471-2342-12-25. URL: [BMC Medical Imaging](http://www.biomedcentral.com/BMCMedicalImaging)
- [9] Fütterer J, Heijmink S, Scheenen T, Veltman J, Huisman HJ. Prostate cancer localization with dynamic contrast-enhanced MR imaging and proton MR spectroscopic imaging. *RSNA Radiology*. Nov 2006;**241**(2):449-458
- [10] Oto A, Kayhan A, Jiang Y, Tretiakova M, Yang C, Antic T. Prostate cancer: Differentiation of central gland cancer from Benign prostatic hyperplasia by using diffusion-weighted and dynamic contrast-enhanced MR imaging. *RSNA Radiology*. 2010;**257**
- [11] Mendez A, Tahoces P, Lado M, Souto M, Correa J, Vidal J. Automatic detection of breast border and nipple in digital mammograms. *Computer Methods and Programs in Biomedicine*; May 1996;**49**(3):253-262
- [12] Doyle S, Rodriguez C. Detecting prostatic adenocarcinoma from digitized histology using a multi-scale hierarchical classification approach. In: 2006 International Conference of the IEEE EMB, New York, NY; 2006. pp. 4759-4752. DOI: 10.1109/IEMBS.2006.260188
- [13] Dorsi EME, Sickles C, Morris E. *ACR BI-RADS Atlas, Breast Imaging*

Reporting and Data System. American College of Radiology; 2013

[14] Weinreb J, Barentsz J, Choyke P, Cornud F, Haider M, Macura K, et al. PI-RADS prostate imaging-reporting and data system: 2015 version 2. *European Urology*; Jan 2016;**69**(1): 16-40

[15] Yu H, Wilson SR. Liver masses with acoustic radiation force impulse technique. *Ultrasound Quarterly*. 2011; **27**:217-223. URL: [www.ultrasound-quarterly.com](http://www.ultrasound-quarterly.com)

[16] Rangayyan RM, Mudigonda NR, Desautels JEL. Boundary modeling and shape analysis methods for classification of mammographic masses. *Medical & Biological Engineering & Computing*. 2000;**38**(5):487-496

[17] Guliato D, Rangayyan RM, Carvalho JD, Santiago SA. Polygonal modeling of contours of breast tumors with the preservation of spicules. *IEEE Transactions on Biomedical Engineering*; **55**(1). DOI: 10.1109/TBME.2007.899310

[18] Brown JM, Giaccia AJ. The unique physiology of solid tumors: Opportunities (and problems) for cancer therapy. *Cancer Research*. 1998; **7**:1408-1416

[19] Erasmus JJ, Connolly JE, McAdams HP, Roggli VL. Solitary pulmonary nodules: Part I. Morphologic evaluation for differentiation of benign and malignant lesions. *Radiographics*. 2000; **20**(1):43-58

[20] Yepes-Calderon F, Johnson R, Lao Y, Hwang D, Coloigner J, Yap F, et al. The 3D edge runner pipeline: A novel shape-based analysis for neoplasms characterization. In: *SPIE 9788, Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging*; 97882N; 29 March 2016. DOI: 10.1117/12.2217238

[21] Abdel R, Bakry RE. Breast tomosynthesis: A diagnostic addition to screening digital mammography. *The Egyptian Journal of Radiology and Nuclear Medicine*. 2018;**49**(2):529-535

[22] Eghtedari M, Tsai C, Robles J, Blair SL, Ojeda-Fournier H. Tomosynthesis in breast cancer imaging: How does it fit into preoperative evaluation and surveillance? *Surgical Oncology Clinics of North America*. 2018;**27**(1):33-49

[23] Whitted T. An improved illumination model for shaded display. *Communications of the ACM*. 1980;**4**(4): 343-349

[24] Rigau J, Feixas M, Sbert M. New contrast measurements for pixel supersampling. In: *Proceedings of CGI02*. July 2002:439-451

[25] Canny J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*; **PAMI-8**. 1986:679-698. DOI: 10.1109/TPAMI.1986.4767851

[26] Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal*. 1948;**27**:379-423, 623-656

[27] McKinney W. Data structures for statistical computing in python. In: van der Walt S, Millman J, editors. *Proceedings of the 9th Python in Science Conference*; 2010. pp. 51-56

[28] McKinney W. *Pandas: A Foundational Python Library for Data Analysis and Statistics*

[29] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. *Scikit-Learn: Machine learning in Python*. *The Journal of Machine Learning Research*. 2011;**12**:2825-2830

[30] Massey FJ. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*;

1951;**46**(253):68-78. DOI: 10.1080/  
01621459.1951.10500769

[31] Weaver KF, Morales V, Dunn SL,  
Godde K, Weaver PF. An Introduction  
to Statistical Analysis in Research: With  
Applications in the Biological and Life  
Sciences, Chapter 8. USA: John Wiley &  
Sons; 2017. DOI: 10.1002/  
9781119454205.ch8