Multi-scaled morphological features for the characterization of mammographic masses using statistical classification schemes

Harris Georgiou[1], Michael Mavroforakis[1], Nikos Dimitropoulos[2], Dionisis Cavouras[3], Sergios Theodoridis[1]

[1] University of Athens, Informatics Department, TYPA buildings, University Campus, 15771, Athens, Greece

[2] Medical Imaging Department, EUROMEDICA Medical Center, 2 Mesogeion avenue, Athens, Greece

[3] Medical Imaging Technologies Department, TEI-Athens, 12210, Athens, Greece

Corresponding Author:

Harris Georgiou (Mr), 11 Vas. Dipla str, P.O. 11745, Athens, Greece.

Tel: +30 210 9313361, Fax: +30 210 9313631, Email: xgeorgio@di.uoa.gr

Summary

Objective: A comprehensive signal analysis approach on the mammographic mass boundary morphology is presented in this article. The purpose of this study is to identify efficient sets of simple yet effective shape features, employed in the original and multi-scaled spectral representations of the boundary, for the characterization of the mammographic mass. These new methods of mass boundary representation and processing in more than one domain greatly improve the information content of the base data that is used for pattern classification purposes, introducing comprehensive spectral and multi-scale wavelet versions of the original boundary signals. The evaluation is conducted against morphological and diagnostic characterization of the mass, using statistical methods, fractal dimension analysis and a wide range of classifier architectures.

Methods and material: This study consists of: (a) the investigation of the original radial distance measurements under the complete spectrum of signal analysis, (b) the application of curve feature extractors of morphological characteristics and the evaluation of the discriminative power of each one of them, by means of statistical significance analysis and dataset fractal dimension, and (c) the application of a wide range of classifier architectures on these morphological datasets, in order to conduct a comparative evaluation of the efficiency and effectiveness of all architectures, for mammographic mass characterization. Radial distance signal was exploited using the discrete Fourier transform (DFT) and the discrete wavelet transform (DWT) as additional carrier signals. Seven uniresolution feature functions were applied over these carrier signals and multiple shape descriptors were created. Classification was conducted against mass shape type and clinical diagnosis, using a wide range of linear and non-linear classifiers, including linear discriminant analysis (LDA), least-squares minimum distance (LSMD), k nearest neighbor (k-NN), radial basis function (RBF) and multi-layered perceptron (MLP) neural networks (NN), and support vector machines (SVM). Fractal analysis was employed as a dataset analysis tool in the feature selection phase. The discriminative power of the features produced by this composite analysis is subsequently analyzed by means of multivariate analysis of variance (MANOVA) and tested against two distinct classification targets, namely (a) the morphological shape type of the mass and (b) the histologically verified clinical diagnosis for each mammogram.

Results: Statistical analysis and classification results have shown that the discrimination value of the features extracted from the DWT components and especially the DFT spectrum, are of great importance. Furthermore, much of the information content of the curve features in the case of DFT and DWT datasets is directly related to the texture and fine-scale details of the corresponding envelope signal of the spectral components. Neural classifiers outperformed all other methods (SVM not used because they are mainly two-class classifiers) with overall success rate of 72,3% for shape type identification, while SVM achieved the overall highest 91,54% for clinical diagnosis. Receiver operating characteristic (ROC) analysis has been employed to present the sensitivity and specificity of the results of this study.

Keywords

Mammography, morphological analysis, multi-scaled analysis, fractal dimension, medical diagnostics.

1. Introduction

In localized breast cancers, the morphology and shape characteristics of the detected masses have been established as factors of great importance in the discrimination between fibroadenomas, cysts and carcinomas, due to inherent anatomical differences directly related to malignancy [1-3]. Approximately 80% to 85% of diagnostic information is retrieved from the mammographic appearance of the mass [4-7]. However, the cognitive task of tissue characterization and classification of a mass as probable benign or probable malignant is extremely complex and includes advanced inference mechanisms [1, 2]. Clinical studies have shown that malignant masses usually infiltrate the surrounding tissue as they expand and, as a result, the mass' appearance exhibits vague and fuzzy boundaries, with linear strands extending irregularly outwards [8, 9]. In the benign pathological cases, masses appear as mass shapes of well-defined boundaries and non-stellate structures [4]. Myelloid breast carcinomas, papillary colloid carcinomas, as well as phylloid cystosarcomas, do not follow the rule and form special groups with exceptional morphology [10, 11].

Based on these morphological and anatomical differences, indicated by the mammographic appearance of the mass, four distinct shape types have been established as the most fundamental clinical cases [12]. These are

round, lobulated, micro-lobulated and stellate (Figure 1). Round shape morphologies relate primarily to benign pathology, while stellate shape morphologies relate to malignancy [4, 13, 14].

Although previous studies have been focused on descriptive measurements for the characterization of each shape type [14-16], little work has been done on the area of spectral and multi-resolution analysis of the shape signal. The introduction of wavelet transformations [17-21], as well as the traditional approach of Fourier analysis [22, 23], offer a promising new area for the study of shape analysis and the extraction of valuable discriminative information. Furthermore, in the area of the shape classification itself with respect to benignancy or malignancy of the corresponding mass, only a limited group of models has been applied in the specific problem, consisting mainly of minimum distance classifiers, utilizing standard Euclidean distance measures [24]. Little evaluation for the same classification task has been conducted for more recent non-linear classifiers, such as support vector machine (SVM) [24]. Furthermore, there are only a few studies available addressing the general issue of investigating the inherent complexity and dimensionality of the extracted morphological datasets used by these classifiers [25].

*(*** PLACEHOLDER FOR: **FIGURE-1** ***)*

This study is focused on three main areas of interest:

(1) The investigation of the original radial distance measurements under the complete spectrum of signal analysis. Morphological shape analysis has been conducted, not only to the radial distance signal itself, but also to the envelope of the discrete Fourier transform (DFT) [22] spectrum and the discrete wavelet transform (DWT) [18, 19] decomposed sub-band signals for multiple wavelet function selections.

(2) The application of seven uniresolution quantitative measurements of morphological characteristics to all the extracted signals and the evaluation of their respective discriminative power, by means of multivariate analysis of variance (MANOVA) [26, 27]. Additionally, a of fractal feature space dimension analysis is employed in order to establish a preliminary estimation of the complexity and intrinsic dimensionality of the training datasets [25], as a tool for feature selection.

(3) The application of a wide range of classifier architectures, namely linear discriminant analysis (LDA), least-squares minimum distance (LSMD), k nearest neighbor (k-NN), radial basis function (RBF) and multi-layered

perceptron (MLP) neural networks (NN), and SVM classifiers, on the same morphological datasets, in order to conduct a comparative evaluation of the efficiency and effectiveness of all architectures, for the specific problem of mammographic mass characterization.

The discriminative power of the features produced by this composite analysis is subsequently tested against two distinct classification targets, namely **(a)** the morphological shape type of the mass and **(b)** the histologically verified clinical diagnosis for each mammogram. Both targets are examined using the same morphological datasets and classifier architectures. For the purposes of this study, an original mammogram database was studied in the context of verified clinical results. The database contained detailed qualitative information for each mammogram, including a thorough and compact list of clinical findings provided by human experts [12]. Results [4] have verified the statistical significance of the quantitative morphological properties of mammographic masses, in particular the benignancy or malignancy indicated by the shape type of the mass. In the current study a set of morphological feature extractors was applied on the mammogram database in order to construct complete datasets of morphological "signatures" for benign and malignant cases of masses. MANOVA was subsequently used for the construction of additional subsets of statistically independent morphological features. Both the original and the reduced datasets were analyzed using statistical analysis and fractal dimension techniques [25], in order to establish a lower bound for the inherent dimensionality of the input space and how it is affected when using feature selection methods like MANOVA.

The material of this study is organized in sections as follows. Section 2 contains the details regarding the base dataset and the methodologies used throughout the study. Section 3 presents the results of the morphological features evaluation, the dataset fractal analysis and the classification tests. Next, section 4 discusses the choices and performance considerations over the various morphological dataset types and classifier architectures. Finally, section 5 summarizes the results and consequences of various aspects of using morphological features, the fractal-based analysis of the datasets and the various classifier architectures, for establishing a robust and complete framework to analyze the shape and morphology of mammographic masses.

2. Materials

Two distinct resource materials were employed in this study: (a) a prototype mammographic image database, annotated by human experts, and (b) the corresponding dataset of qualitative features, also provided by an expert physician.

The requirement for patient's clinical history and positive histological verification of the benignancy or malignancy of each case was assessed as one of extreme importance for the quality and validity of the subsequent results. Thus, a special-purpose image set was assembled, using cases of mammographic masses with complete radiological evaluation and histological diagnosis, already used in other studies in the past successfully [28]. The initial set contained 1350 cases of women that proceeded for the evaluation of a clinical breast problem, each containing a full set of 4 mammograms, i.e., cranio-caudal (CC) and medio-lateral oblique (MLO) projections per breast, and it was used as a base for the final selection of mass cases with positive clinical verification by surgical biopsy and histological examination. For the construction of the raw material for this study, a set of 130 MLO mammograms containing a mass were selected by an expert physician for digitization, including a total of 46 benign mass cases and 84 mass malignancies of various types. The selection was made on the basis of unbiased statistical distribution of the underlying textural features and the completeness of the dataset with respect to various clinical findings, always focused in pathological cases that included presence of at least one suspicious mass. In the set of 130 mammograms, almost two-thirds of the cases are malignant, because many different types of malignancies had to be represented in the database with an adequate statistical sample. For each mammogram, a complete list of qualitative information was provided by the attendant physician [12], containing details about the age of the patient, number of present masses, presence of microcalcifications, density of the mass, percentage of fat inside the mass, mass boundary vagueness, mass homogeneity, mass shape type and clinical diagnosis. All these qualitative clinical descriptions, as well as their logging range of values, are shown in Table 1. The various qualitative details were included according to the expert's recommendations and proposals, as explicit information related to various types of malignant mammogram abnormalities, including architectural distortion, microcalcification clusters, micro-lobulated or stellate masses [29-31].

The film mammograms where converted into a digital format using a digital scanner (Epson Perfection 1640SU with film adapter) at the default optical density (OD) of 3.1, while the scanning specifications for its grayscale range and optical resolution were 14 bits and 1600 dpi respectively. The digitized mammograms were

subsequently transformed digitally to a resolution of 63 μm (400 dpi) at 8-bit gray level, which is consistent with other typical image databases of digitized mammograms that are used as a reference in similar studies [32]. The final set of 130 mammograms was used in all cases, with no further reduction in spatial resolution or gray level depth.

Advanced algorithms for automated mammographic lesion detection have been proposed, however their level of sensitivity and specificity is still under investigation [33-36]. Furthermore, their fine-scale accordance to the corresponding expert's detailed description of mass boundaries, especially in cases of non-trivial stellate distortions, exhibits many practical problems [34]. In order to obtain mass boundaries of high quality and detail, a manual segmentation was applied. Specifically, each mass was manually described by the radiologists at the maximum available spatial resolution, using a high-resolution digitizer device and stored as an embedded boundary descriptor via alpha channel data. These mass descriptions were subsequently used as a fine-scale tissue inclusion/exclusion mask for all subsequent extraction and analysis of textural features.

*(*** PLACEHOLDER FOR: **TABLE-1** ***)*

3. Methods

Four categories of methods and algorithms were employed in this study: (a) multi-scaled signal processing and spectral decomposition, (b) a set of 1-D morphological feature functions, (c) dataset fractal dimension evaluation, and (d) linear, k-NN, NN (MLP and RBF) and SVM classifier architectures.

3.1 Multi-scaled spectral decomposition

Seven traditional uniresolution morphological features were used to extract quantitative information from a one-dimensional signal. These one-dimensional features have been widely used in the past, as invariant and unbiased descriptors for the characterization of the properties and morphology of mammographic masses [4, 13, 14]. The one-dimensional signal normally consists of the radial distance function, sampled over a well-defined mass borderline that is defined by a closed curve of arbitrary shape. Additionally, the envelope of the DFT spectrum and the DWT components of the previously described one-dimensional signal, produced from each mass, were also used as carrier signals.

187

188 3.1.1 Radial Distance Signal

189 In general, each radial distance function is computed in two steps: (1) the centroid of the mass is computed, using
190 cumulative distributions of the projections in both x and y axes and (2) the acquired centroid of the mass is identified
191 as the center of a polar coordinate system and the mass borderline is sampled over fixed intervals to extract the radial
192 distance sequence, starting from angle 0:

193

194 $$d(i) = \sqrt{(x(i) - X_0)^2 + (y(i) - Y_0)^2}\ , i = 1,2,...,N \qquad (1)$$

195

196 where the point (X_0, Y_0) is the acquired centroid of the mass and N is the number borderline samples taken.
197 As the mass boundary can have arbitrary (e.g., non-convex) shape and displacement in relation to the mass centroid,
198 many abnormal cases, especially those regarding stellate masses, were considered too complex to be analyzed by a
199 direct centroid-to-boundary measurement. In other words, the parameterized form of radial distance versus radial
200 angle theta (θ) does not define a valid function $r(\theta)$, due to the possibility of multiple $r(\theta)$ values for a fixed θ value,
201 in cases of arbitrary mass shapes.

202 In this study, the boundary of each mass was acquired using a prototype edge-following method.
203 Specifically, the starting point on the mass boundary was acquired at an arbitrary angle and was then registered in
204 full resolution, using a line-following scheme, before calculating radial distances for all the boundary pixels that
205 were detected (Figure 2). This means that, instead of taking radial distance samples from the mass center towards the
206 boundary over fixed angular intervals, the boundary itself is traced at pixel-level and all radial distances are
207 calculated backwards, i.e., from the boundary pixels towards the mass center. Essentially, the new parameterized
208 form of radial distance $r(t)$ includes a spatial transition (t) on the boundary, indicating the transition between two
209 adjacent boundary pixels. Thus, the total number N of borderline samples is exactly the length (in pixels) of the mass
210 perimeter in full resolution. The same concept, in its generalized form, is best manifested by the various Fourier
211 Descriptor functionals [22, 24]. This type of radial distance functionals $r(t)$ are directly related to the (variable) total
212 length of the mass' boundary outline, instead of the (fixed) range of angular intervals as in the case of $r(\theta)$. Thus, the

initial resolution and fine-scale details, as well as the intrinsic morphological properties of the boundary, are fully preserved.

*(*** PLACEHOLDER FOR: **FIGURE-2** ***)*

In order to avoid typical problems with edge-following techniques, especially in cases of sharp edges, a pre-processing stage was applied before acquiring the borderline curve. Specifically, a dilation filter [22] of pre-determined mask size was used to remove localized noise effects on the borderline, usually attributed to operator instability during the manual segmentation of the mass boundaries. A minimum dilation mask of 3x3 has been experimentally proved adequate for all the digitized mammograms of the dataset. No significant alternation due to dilation was observed in the morphological characteristics of any of the identified masses, while the variation in the length of the borderline was less than ±1% at maximum and ±0,1% on average. Thus, it was determined that the dilation pre-processing stage introduced no immediate side-effect on the quality and the informational content of the borderline curve in all cases. Comparative results revealed considerable increase in quality and resolution of the extracted radial measurements in all cases and thus this technique was applied throughout this study.

All radial distance signals were normalized using the minimum and maximum values of radial distances in the dataset, in order to avoid excessive differences in magnitude and total energy between masses of different sizes, using the typical formula:

$$\hat{d}(i) = \frac{d(i) - d_{\min}}{d_{\max} - d_{\min}}, \; i = 1, 2, ..., N \qquad (2)$$

3.1.2 Fourier Analysis – Discrete Spectrum

Preliminary results and spectral analysis proved the significance of properties of the frequency spectrum as important discriminating factors between the various morphological shape types of the masses and the final diagnostic

conclusion. As the normalized radial distance signal represents the shape of each mass' perimeter, abnormalities in the initial radial distance signal should be evident as explicit properties of the discrete spectrum of the signal.

The uniresolution features were applied as descriptive curve features for the DFT spectral envelope in each case. The qualitative significance of the uniresolution features is similar to the case of the radial distance signal, as they can positively describe attributes of the underlying one-dimensional signal, which in this case is the spectrum of the signal, instead of the original signal itself.

3.1.3 Wavelet Decomposition

L.M.Bruce and R.R.Adhami [4] have shown the significance and importance of multi-resolution features that are based on wavelet transform using modulus-maxima method. Specifically, they have shown that DWT could be used in one or more scales iteratively to transform the original signal into one or more decomposed forms that could be used for detailed sub-band analysis.

In this study, instead of creating a modulus-maxima representation of the signal produced by the wavelet decomposition, each detail and approximation component was treated as a separate signal by itself. In relation to the original radial distance signal, each decomposition level should capture different aspects of the qualitative and quantitative properties of the mass boundary and thus produce detailed discriminative information. Specifically, at small scales (i.e., the first decomposition levels) the DWT reveals information content related to localized sharp variations of the signal, while at large scales (i.e., deeper decomposition levels) it reveals global signal variations [4]. The uniresolution features were, again, used as curve descriptors for each wavelet component separately, producing a different set of features for each one of them.

In order to estimate the effect of wavelet function selection over the discriminative power of the curve features calculated in each case, several wavelet functions were applied in the wavelet decomposition process. Specifically, five wavelet functions [37] were examined: (i) Haar, (ii) sym2, (iii) coiflet-1, (iv) biorthogonal-1.5 and (v) discrete Meyer. These selections were based on the property that on each scale the DWT should reveal information related to sharp edges and peaks on the original radial distance signal, as they would represent abnormalities on the perimeter and the boundaries of the mass itself [4, 38]. Furthermore, experimental studies have shown that for shape features based on the zero-crossing property, a particular class of wavelet functions, based on

the derivatives of a smoothing filter function, provides the means to identify the signal's sharp variation points and different scales [39]. Therefore, only Gaussian-like wavelet kernels were considered as appropriate candidates. Features extracted separately at each decomposition level could also reveal the quality and value of the discriminative information contained on that level. The first four DWT decomposition levels have been used for extracting multi-scaled wavelet component of the radial distance signal, after evaluating cases of applying wavelet decomposition methods [38] in mass perimeter signal and according to the minimum and maximum signal lengths available in the current database.

3.2 Curve feature extraction

The creation of robust and yet simple representation of the four shape type categories of mammographic masses, a small set of uniresolution 1-D curve functions were used. The selection of these particular shape descriptor functions was based on four main factors: (a) simplicity in implementation and physical meaning, (b) invariance in terms of translation, scale and rotation, (c) well-suited for application in all base signal types, namely in spatial, DFT and DWT domains, and (d) similar physical meaning in all transformed signal domains.

Numerous shape features have been proposed in the past for various applications [21, 40]. In general, they can be grouped in two major categories [21]: contour-based, i.e., exploiting only the boundary shape information of an object, and region-based, i.e., all the pixels within a shape region are used to obtain the shape representation. In both cases, shape representation can be global, describing a property of the complete shape, or structural, were the shape is broken down into local boundary segments called primitives [21]. For contour-based shape features, the normalized centroid distance or normalized radial length measurement has been proposed as one of the most content-rich shape representations for closed-contour boundaries of objects [21], especially when spectral shape features are to be used [23]. Based on this spatial representation of a shape profile, various feature extractors have been proposed for exploiting gradient transitions, statistical variances and moments [40-43]. Based upon the way they are implemented in practice, various versions of these spatial features can be defined, yet many of them are overlapping in terms of the information content they capture [40]. However, most spatial shape descriptors, especially the gradient-based, suffer from noise sensitivity due to their inherent heavy dependency to boundary variations.

Furthermore, some statistical properties, such as higher-order moments, have no specific physical meaning regarding the original boundary shape information [21].

To overcome the problem of noise sensitivity, transformation to spectral domain has been proposed for the shape descriptor signal. Specifically, Fourier descriptors (FD) [22, 44] and wavelet descriptors (WD) [22, 45] have been used to capture shape properties that emerge from the spectral and multi-scale decomposition of the original boundary information of an object. Although these techniques are extremely powerful in describing arbitrary shape types and identifying invariant shape properties, their application is restricted by some limitations. While FD features are well-suited for describing spectral properties of the object's boundary, the size of the convolution kernel that defines the locality of the transformation is a free parameter, essentially defining short-time Fourier descriptors (SFD) and conventional or global FD, each suited for different type of applications [46]. The problem of scaling is resolved very efficiently in the case of WD, where the boundary shape signal is decomposed at several scales. However, the much larger volume of data (compared to the FD) and, in many cases, the dependency of the robustness of the WD features to the exact choices of wavelet functions, often makes the used of WD shape features very problem-specific. Furthermore, most shape-matching algorithms, such as the modulus-maxima representation [38] and the zero-crossing points scheme [45], exploit only the dominant attributes and features of the DWT decomposed signal, instead of the complete signal components in all scales. As a result, the design of most WD features and shape matching algorithms, proposed until now, require complex similarity measurements that are usually shift-variant and dependant on the complexity of the shape [21].

In this study, a small well-studied set of seven 1-D uniresolution curve features have been selected as the most prominent for application in all three signal domains, namely (a) the radial distance signal, i.e., spatial domain, (b) the envelope of its DFT, i.e., the Fourier domain, and (c) the envelope of its DWT components, i.e., the wavelet domain. These are [4, 14]: (i) radial distance mean, (ii) radial distance standard deviation, (iii) mass circularity, (iv) entropy of the radial distance histogram, (v) area ratio parameter, (vi) zero-crossing count and (vii) mass boundary roughness.

The radial distance mean m was calculated by the following formula:

$$m = \frac{1}{N}\sum_{i=1}^{N} d(i) \qquad (3)$$

The radial distance standard deviation s was computed by:

$$s = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (d(i) - m)^2} \qquad (4)$$

Mass circularity C was defined as:

$$C = \frac{P^2}{A} \qquad (5)$$

where A is the area and P is the perimeter of the specified mass. The area was calculated as the integral of the normalized radial distance signal at full resolution against the sampling angle, while the perimeter was extracted by summing the differences of all samples of the same signal sequentially along the complete angle range.

The entropy E of the radial distance formula:

$$E = \sum_{k=1}^{N} P_k \log(P_k) \qquad (6)$$

was applied on the normalized histogram of radial distances with N=100 bins. P_k is the probability that the radial distance measure will be between: 1 and: 1+1/N times the current value. The number of times the radial distance passed through each bin was summed and the sums were divided by the total number of radial distance samples available. The number of bins was determined experimentally and has been confirmed as of adequate histogram resolution.

The area ratio parameter A_R was estimated by the formula:

$$A_R = \frac{1}{m \cdot N} \sum_{i=1}^{N} (d(i) - m) \qquad (7)$$

where

$$A_R = 0 \,,\ \forall d(i) \le m \qquad (8)$$

The zero-crossing count ZC was calculated as a count of the number of times the radial distance signal crossed the average radial distance value.

Finally, the mass boundary roughness R was calculated by dividing the radial distance signals into small segments of equal length and then estimating a roughness index for each one of them according to the formula:

$$R(j) = \sum_{i=j}^{L+j} \left| d(i) - d(i+1) \right| \quad ,\ j = 1,\ldots,\left\lceil \frac{N}{L} \right\rceil \qquad (9)$$

$$R = \frac{\sum_{j=1}^{k} R(j) + \frac{L_{k+1}}{L} \cdot R(k+1)}{k + \frac{L_{k+1}}{L}} \quad ,\ k = \left\lceil \frac{N}{L} \right\rceil - 1 \qquad (10)$$

where R(j) is the roughness index for the j-th segment, L is the number of boundary points in each segment, k denotes the last segment of length L and N is the total number of boundary points available. The rather complicated form of the averaging formula for calculating R is due to the fact that the length L_{k+1} of the last segment is less than or equal to L, thus it has to be weighted accordingly. In the current study, N varies according to the size of the mass perimeter, but a fixed value of L=16 was experimentally determined as adequate for the specific spatial resolution of 63 μm, i.e., the roughness index was calculated over segments of about 1mm.

The same seven uniresolution features were employed as curve descriptors in all three domain spaces (spatial, DFT, DWT). In terms of shape feature categorization, mass circularity (A_R) is a region-based global feature,

while the roughness index (R) is a contour-based structural feature. The remaining five features are dependant solely on the boundary shape signal and thus can be considered as pure contour-based global features. Since they are calculated in the normalized centroid-centered space and no directional boundary indexes are used, all seven features are translation-, scale-, and rotation-invariant. Although these uniresolution features have been successfully applied in similar cases using radial distance measurements as the base signal [14], or in modulus-maxima representation of wavelet domains [4], they have never been used comparatively in all three spatial, DFT and DWT domains, using the full versions of the transformed signals and testing their efficiency and robustness in all three domains simultaneously for mammographic mass characterization.

For the physical meaning of these features in the spatial, DFT spectral and DWT multi-scaled decomposed domains, see [21, 22]. In short, the radial distance measurement can be considered as a measure of energy distribution: in spatial domain this is the signal energy as a function of time/space, in DFT as a function of frequency and in DWT as a function of the multi-scale wavelet components. Thus, the statistical properties of the shape signal essentially capture more or less the same attributes of the original boundary shape, but in different representation domains. Furthermore, the choice of invariant and noise-resilient shape features that can be applied in the same way in all three domains, i.e., upon the spatial signal curve and the envelope of the DFT/DWT spectral components, essentially validates their use as efficient and robust energy-related descriptors of the original boundary shape. Thus, the initial dimension of the training dataset is 217 per sampled mass, containing 7x1 features from the spatial domain, 7x5 features from various versions (initial, abs, log) of the DFT spectrum envelope, and 7x5 features from 4-level DWT decomposition (4 detail + 1 approx) using 5 different wavelet functions.

3.3 Fractal feature analysis

In order to establish a preliminary estimation of the complexity and intrinsic dimensionality of the training datasets, fractal feature analysis was applied to the initial training datasets. Fractal feature analysis, specifically the calculation of intrinsic fractal dimension of the input datasets, provides the quantitative means of investigating the non-linearity and the correlation between the available features by means of dimensionality of the resulting feature space [47, 48]. Fractal dimension has also been used as an alternative way of characterizing the discriminative power of each feature separately, thus providing a non-statistical way of ranking them in terms of importance for the classification task

[25]. The fractal analysis of datasets has been used successfully in previous studies [28] and it has been proved very valuable as a tool for comparing arbitrary datasets of extracted features with the qualitative clinical properties that an experience physician uses to characterize a mammographic image.

In short, the two most commonly used methods of calculating the fractal dimension of a dataset are the pair-count (PC) and the box-counting (BC) algorithms [17, 48, 49]. In the pair-count algorithm, all Euclidean distances between the samples of the dataset are calculated and a closure measure is then used to cluster the resulting distances space into groups, according to various ranges (r), i.e., the maximum allowable distance within samples of the same group. The PC value is calculated for various sizes of (r) and it has been proved that PC(r) can be approximated by:

$$PC(r) = K \cdot r^{D} \qquad (11)$$

where K is a constant and D is called pair-count exponent. The PC(r) plot is then a plot of: log(PC(r)) versus log(r), i.e., D is the slope of the linear part of the PC(r) plot over a specific range of distances (r). The exponent D is called correlation fractal dimension of the dataset, or D_2. The box-counting approach calculates the exponent D in a slightly different way in order to accommodate case of large datasets with size in the order of thousands; however it essentially calculates an approximation of that same correlation fractal dimension value, i.e., D_2.

In both cases, the D_2 value of the initial dataset characterizes the intrinsic or "true" dimension of the input space [48]. In other words, D_2 would be the minimum dimension of the dataset if only "perfect" features were allowed, i.e., totally uncorrelated and with the best discriminative power available within the specific set of features. In this study, fractal feature analysis was applied to both the initial set of qualitative characteristics, provided by the expert physician, as well as the constructed datasets of morphological features. In all cases, the pair-count algorithm employing Euclidean distances was used, due to the relatively small number of samples available, as well as the better stability and accuracy for D_2 against the box-counting approach [49].

In order to calculate the slope at the linear part of the PC(r) plot, a parametric sigmoid function was used for fitting between the sample points of the plot. In the parametric sigmoid function:

$$y = Y_0 + C_y \cdot \left[\frac{1}{1 + \exp\left(-C_x \cdot (x - X_0)\right)} \right] \qquad (12)$$

the (X_0,Y_0) identifies the transposition of the axes, while C_x and C_y identify the appropriate scaling factors. Specifically, the value of C_x affects the steepness of the central part of the curve, while C_y specifies the Y-axis width of the sigmoid curve. Then, the slope of the linear part around the central curvature point, i.e. the value of D_2, is:

$$\frac{\partial^2 y(X_0)}{\partial x^2} = 0 \quad \Rightarrow \quad D_2 = \frac{\partial y(X_0)}{\partial x} = \frac{C_x \cdot C_y}{4} \qquad (13)$$

3.4 Feature selection, classification and testing

Classification was conducted for two separate configurations. The first one contained the morphological shape type as the classification target, while the second one contained the clinical diagnosis as the classification target. Both configurations were combined with all the available input datasets, including uniresolution curve feature values extracted from (a) the radial distance signal, (b) the DFT spectrum signal and (c) the DWT multi-scale decomposed signals. Features extracted from each case were rated by applying MANOVA significance analysis and reduced subsets of sizes up to 20 features, contained in the MANOVA top-20 features list, were employed in all classification setups. Several classifiers were used to evaluate each one of the available feature combinations for optimized subsets selection. Combined datasets, using features extracted from the original radial distance signal, the DFT spectrum and the DWT decomposed signals, were also examined and classified using multivariate analysis, producing sorted feature lists and optimal feature set selections.

Several classifier architectures were applied during the classification phase. A LDA model was used in the form of linear classifier [50]. A LSMD classifier [24] was employed, using Mahalanobis distance measure and least-squares dataset pre-processing on the input [24]. These two models were also used for feature set optimization by applying exhaustive search throughout all the available feature combinations. A k-NN model was also used, including estimation of an optimal value k for the size of the neighborhood set and sub-optimal feature set [24].

Two different types of NN architectures were employed: a RBF NN with Gaussian activation function and linear output function [51], and a MLP NN with hyperbolic tangent internal activation and softmax output function [52], both implemented with topology optimally adapted to each configuration and dataset. In the case of the MLP, the output function was optimally selected between linear, softmax and logistic for each case separately. All topologies used one hidden layer of optimized size.

For more advanced investigation of the feature set, typical SVM models were applied in relation to the final diagnosis. Specifically, the C-support vector classification (C-SVC) model was used in combination with standard RBF kernel functions, optimizing the penalty factor (C) and the Gaussian spread parameter (σ) during training [53, 54]. SVM classifiers that employ various non-linear kernels, combined with new algorithms for fast training of these architectures, are considered state-of-the-art in Pattern Recognition today [24, 54, 55] and they can be regarded as a realistic upper limit in the performance of automated systems in similar applications in practice.

For consistent and realistic estimation of each classifier's generalization ability, k-fold cross validation techniques are usually employed, specifically leave-one-out and leave-k-out methods [24, 56, 57]. In this study, all configurations used the leave-one-out method for dataset manipulation during training and testing phases, combined with optimal feature set selection for the linear classifiers, or the complete selected (optimal) feature sets for the neural networks. SVM classifiers employed limited feature set optimizations, using iterative runs of enlarging inclusions of several features, available on the feature ranking lists created by MANOVA significance analysis. One of the reasons that full feature set optimization (i.e., exhaustive feature combinations) was not applied with NN or SVM architectures, is that the training phase, combined with the optimization of the size of the hidden layer in the case of NN, becomes computationally too expensive. Furthermore, it also relies on the fact that trained NN architectures apply optimal weight values at the input layer, thus they can be examined during a post-training pruning phase to optimally reduce the dimensionality of the input set if necessary [58]. SVM classifiers have also proven exceptionally efficient in classification tasks of high dimensionality [24, 53, 54]. In all cases, classifiers were re-trained for every new dataset that was produced after the extraction of one training sample, according to the leave-one-out method, and then classified this sample treating it as unknown input. Thus, the quality and generalization of the classification results depended solely on the quality and unbiased distribution of the training samples in the complete dataset for each case.

4. Results

Preliminary studies on the original mammographic database confirmed the significance of morphological shape characterization with the clinical diagnosis [12]. Specifically, it has been statistically verified that round and lobulated cases together exhibit only 12% malignancy, while the combined micro-lobulated and stellate cases exhibit 96% malignancy (Table 2). This result shows that utilizing the shape characterization alone as discrimination measure between benignancy and malignancy can establish a success rate over 93% [12].

*(*** PLACEHOLDER FOR: **TABLE-2** ***)*

All seven uniresolution feature functions were used for all base signals, including the original radial distance signal, its DFT spectrum envelope and its wavelet decomposition signals up to level four. For the wavelet decomposition, five wavelet functions were selected and applied. Finally, absolute value and logarithmic versions of the DFT spectrum signals were used in order to determine the value of pre-processing transformations of the spectrum data. A total of 217 features were collected for all the available mass descriptions and all training samples were grouped into separate datasets according to their base signal, along with one dataset containing a combination of all of them.

4.1 Fractal analysis of morphological datasets

Figure 3 illustrates the PC(r) plot and the calculation of D_2 for the combined dataset of all 217 morphological features. Table 2 summarizes the results of D_2 value, calculated over the qualitative features set and the constructed morphological features datasets.

*(*** PLACEHOLDER FOR: **FIGURE-3** ***)*

As the Table 3 shows, the combined dataset of all 217 morphological features exhibits roughly the same D_2 value with the qualitative properties identified by the expert physician. In other words, the combined morphological

features dataset employs at least the same or better descriptive power as the qualitative dataset does. On the other hand, the corresponding D_2 value when using only the best 20 morphological features, as identified by MANOVA for statistical independence, should not be characterized as conclusive for describing the complete dataset of dimension 217, although the D_2 value indicates an intrinsic dimension lower than 5. In any case, the difference of 1,4 between the corresponding D_2 values (i.e., for full morphological feature set and for MANOVA top-20 set) can be considered as safe and expected, provided that these best 20 morphological features exhibit most of the descriptive power of the original set, with only 9,2% of all the available features. As a result, the MANOVA analysis for the selection of the 20 highly uncorrelated morphological features was considered safe and consistent for constructing training datasets of minimal dimensionality.

*(*** PLACEHOLDER FOR: **TABLE-3** ***)*

4.2 Shape Type Classification Mode

Statistical analysis and classification results have shown that the discriminating power of features extracted from the DFT spectrum and especially the DWT components is of great importance. Specifically, 24 out of the 34 combined best (unique) MANOVA-selected features are extracted from the DWT-decomposed signal, 8 from the DFT spectrum and only 2 from the spatial signal of the mass boundary. Although the exact choice of wavelet function is of some importance, it is worth noticing that in 22 out of the 24 cases these wavelet features refer to the “detail” components rather than the “approximation” components in the DWT domain [17-20]. Furthermore, in the corresponding DWT components refer to decomposition at levels 2 or more, i.e., not at the small scales at level 1 that usually refer to noise artifacts [20, 59].

Using all the available classifiers, optimized for each setup and leave-one-out testing, the success rate for classification, based only on features extracted from the original radial distance signal, ranged between 46,9% and 57,7%. The same results, for feature sets based on DWT component signals, ranged from 45,4% to 67,7% over all wavelet function selections, while, for feature sets based on DFT spectrum success rates, ranged from 57,7% to 72,3%. SVM classifiers are inherently not well-suited for multi-class (i.e., not binary) classification problems. Although there are many ways of combining multiple individual SVM classifiers into one compact architecture for

multi-class problems (e.g., optimum decision trees, one-against-all schemes, etc [24]), these methods are not directly comparable to the rest of the classifier architectures used in this study. Therefore, no SVM classification results were acquired for the shape type target.

Using the complete feature set of dimension 217 and applying multivariate statistical significance for selecting the MANOVA top-20 features, optimal or sub-optimal feature sets were selected according to the current classification target. All the resulting feature sets contained morphological features extracted primarily from DFT spectrum and DWT components, while the classification results were similar with the success rates already achieved by each one of these feature groups by itself. The best classification configuration contained the selection of features extracted from the DFT spectrum, combined with an optimized LSMD or a MLP with one hidden layer and softmax output, both resulting in 72,3% success rate for leave-one-out testing mode (Table 4).

*(*** PLACEHOLDER FOR: **TABLE-4** ***)*

For DFT spectrum feature sets, absolute value and logarithmic transformations on pre-processing produced a small positive or negative result in the final success rate according to the selected classifier model, ranging between -3% and +4% in relation to the success rate achieved by using the non-transformed DFT spectrum. Discriminative information content was greatly decreased, about –20% to –26%, when DFT spectrum resolution was reduced to only 128 points, instead of the full-resolution spectrum. Although the overall envelope of the spectrum was still maintained even at this limited resolution, classification results have shown that the quality and the discriminative value of the extracted features was greatly degraded in comparison to the full DFT spectrum resolution mode. This implies that much of the information content of the curve features in the case of DFT spectrum is directly related to the texture and fine detail of DFT the envelope signal, which is in complete accordance with clinical practice and intuition. Decrease in the DFT resolution resulted in decrease in the quality and fine-scale details of the resulting signal, thus producing loss of a significant portion of the discriminative information captured by the curve feature functions at higher DFT resolutions.

For DWT decomposition features, the exact choice of the wavelet function proved to be a factor of some importance in combination to each classifier choice, as some wavelet functions with specific classifier models

produced the best results, while other combinations produced poor overall success rates. However, there is no statistically significant evidence of specific wavelet functions being robust and consistently efficient for all classifier choices when shape type classification target is considered.

In the case of the shape type classification mode, the SVM classifier was chosen for one-against-all classification testing [24], in order to investigate the differences in sensitivity and specificity of the overall best classifier against each of the four shape type categories. Results have shown that differences in performance, when round, lobulated or stellate masses are considered, are not significant. Exception to this is the case of micro-lobulated masses, were the classification problem seems to be much harder. The corresponding success rates for round, lobulated and stellate classes were between 89% and 94%, while for the micro-lobulated class it was only 74%. Receiver operating characteristic (ROC) analysis was employed as an additional evaluation tool for the overall-best classifier. Specifically, the ROC curves have been produced using SVM classifier, which is the most suitable performance-wise for the investigation of the sensitivity and specificity of the selected feature subsets. The SVM was employed for two-class classification tasks, using each one of the shape types against the other three. Figure 4 illustrates the ROC plot for each one of the four shape type classifications, as well as the corresponding area under curve (AUC) or $\underline{A_z}$ index [40] value.

*(*** PLACEHOLDER FOR: **FIGURE-4** ***)*

4.3 Clinical Diagnosis Classification Mode

As in the case of morphological shape type classification, overall results have shown that the discriminating power of features extracted from the DFT spectrum and especially the DWT components is of great value. Specifically, 17 out of the MANOVA top-20 features are extracted from the DWT-decomposed signal of the mass boundary. The exact choice of wavelet function is of some importance, however it is worth noticing that in 14 out of the 17 cases these wavelet features refer to the “detail” components rather than the “approximation” components in the DWT domain [17-20]. Furthermore, in 12 out of these 17 cases these wavelet features refer to decomposition levels 3 and 4, i.e., at larger DWT decomposition scales that are usually noise-resilient [20, 59].

Using all the available classifiers, optimized for each setup and leave-one-out testing, the success rate for classification, based only on features extracted from the radial distance signal, ranged between 68,4% and 73,8%. The same configurations for feature sets based on DWT component signals resulted in 73,9% to 85,4% success rates over all wavelet function selections, roughly the same success rates for feature sets based solely on DFT spectrums. For the combined dataset of 217 features and conducting multivariate statistical significance feature selection, optimal feature subsets were selected for the current classification target. All the created feature subsets contained features extracted primarily from DFT spectrum or DWT components, while the classification results were marginally increased in relation to the success rates already achieved by each one of these feature groups, i.e., DFT or DWT, alone. The best classification configuration was the selection of mixed features extracted from DFT spectrum and DWT components of various wavelet functions. The overall best success rate of 91,54% over the MANOVA-selected top-20 features was achieved by SVM classifier, closely followed by MLP with one hidden layer of optimum size and softmax output resulting in 89,2% success rate, both for leave-one-out testing mode (Table 5).

*(*** PLACEHOLDER FOR: **TABLE-5** ***)*

Discriminative information quality was, again, greatly degraded, about –10% to –27%, when DFT spectrum resolution was reduced to only 128 points. As in the case of shape analysis classification, the classification results on final diagnosis revealed once again the significance of the information loss in texture and fine-scale details of the DFT envelope signal. Although the overall envelope of the spectrum was maintained at this low resolution, the discriminative value of the extracted features was greatly reduced in comparison to the full resolution mode.

For DFT spectrum feature sets, absolute value and logarithmic transformations on pre-processing produced a small positive or negative variation in the final success rate of any given classifier model, ranging between -3% and +2% in relation to the success rate achieved by using the non-transformed DFT spectrum. For DWT decomposition features, the exact choice of the wavelet function was important in certain cases, increasing the success rate up to +7,7%, for specific classifier choices.

Figure 5 illustrates the ROC plot for the diagnosis classification using the SVM classifier, as well as the corresponding area under curve (AUC) or $\underline{A_z}$ index [40] value.

*(*** PLACEHOLDER FOR: **FIGURE-5** ***)*

5. Discussion

5.1 Optimal feature functions selection and evaluation

Comparative results over all the available datasets and feature groupings have shown that the effectiveness of each feature function depends greatly on each specific base signal under investigation, as well as the intended classification target, i.e., mass shape type or diagnosis. Morphological feature functions, which are valuable for either the discrimination of shape types or clinical diagnosis, when a specific domain (i.e., spatial, DFT or DWT) is used, may not be equally efficient for the same classification target, when applied to a different type of base signal. This is expected and justified due to the different nature and physical context of the information extracted by the same feature functions, when applied to the different base signals, i.e., the radial distance signal, the DFT spectrum envelope or the DWT component signals. Thus, there is only some qualitative indication on the preference of one feature function over another when applied in base signals of different informational content, although some of them may be of greater importance within the same dataset. The overall ranking of the seven curve feature functions, in order of preference, is illustrated in Table 6 for the shape type and diagnosis classification targets.

*(*** PLACEHOLDER FOR: **TABLE-6** ***)*

Among the total of 217 features collected from all the carrier signals, i.e., radial distance measurements, DFT spectrum envelope and DWT components, MANOVA analysis was applied to select the ones with maximum independence and outmost statistical significance. For these optimal feature selections, both for the shape type and the clinical diagnosis target cases, specific curve feature functions were favored at some degree, like zero-cross count

(for shape type only), roughness index (for diagnosis only), mean value and area ratio (for both classification targets). On the contrary, some feature functions like circularity seems to be of little or no use in terms of discriminative power. However, combined results over all the optimal classifier configurations showed no strong evidence supporting or rejecting the selection of any specific curve feature function, either for any specific carrier signal or for hybrid signal configurations.

For optimized classifier structures, misclassification analysis has shown that in almost all cases the results were compact and consistent. Misclassified morphological shape types were rarely farther than one class away from the correct one, while completely invalid classifications, including round instead of stellate shape types or vice versa, were zero or near-zero for optimized classifiers using any of the datasets. In the case of clinical diagnosis classification, the false-positive and false-negative were, in general, equally distributed within the overall classification errors. For optimized configurations on final diagnosis classification, the actual error rates were similar in volume and content to the error rates involved in the statistical covariance analysis between mass shape types and clinical diagnosis in the original mammogram database [12]. Thus, the descriptive power of these morphological features was confirmed to be directly related to the inherent shape and morphology of the mass' boundaries.

5.2 Classifier Architectures Evaluation

Classification results for both targets, i.e., morphological shape type and clinical diagnosis, proved the value of non-linear architectures versus linear models. Initial studies and statistical analysis have shown that the class distributions in both cases are too complex to be approximated by linear discriminate functions. With the exception of the case of mass shape type property versus clinical diagnosis, where evidently there is high correlation between them, all the other qualitative properties and morphological characteristics of the mass seem quite independent or at least non-linearly dependent with each other.

The k-NN classifier produced at some extent lower overall success rates, even though the size k of the neighborhood was optimized during training. Between linear classifiers, the LSMD classifier resulted in significant increase in overall success rate against both the k-NN and the LDA classifiers, and in some cases similar to the one produced by various neural network architectures.

For NN, comparison between RBF and MLP architectures presented evidence that RBF networks result in either the same or lower overall success rate. Differences in success rates ranged between 0% and 10%, favoring the choice of MLP architectures over RBF, when optimized topologies were employed for both classifiers. For SVM, the C-SVC classifier architecture has proved to be the best choice over all the other linear and non-linear classifiers when clinical diagnosis is considered.

In almost all cases, the application of NN with optimized topology versus a linear classifier, like LDA or LSMD, resulted in at least the same or better overall results, over the same datasets. The overall average difference in success rate between linear classifiers and NN over various datasets ranged between 0% and 6,4% for the shape type classification case, although LDA and mostly LSMD produced results similar to neural network models when used with DWT-only feature sets for various wavelet function selections. Similar conclusions were drawn for the clinical diagnosis classification case, where in some cases the overall success rates of some neural classifier were matched by the LSMD linear classifier. However, for both classification targets, the best accuracies where achieved by optimized non-linear classifiers, specifically 72,3% for morphological shape type by NN and 91,54% for clinical diagnosis by SVM classifiers, all evaluated using leave-one-out testing method. SVM classifier achieved the overall best success rate for the diagnosis target, marginally higher (+2,3%) from the best NN classifier.

It should be noted that, although receiver operational characteristic (ROC) analysis of classifiers is common in medical applications, it was used only as a supplemental performance evaluation method in this study. The main property of ROC is their qualitative presentation of a classifier's sensitivity and specificity levels for various decision thresholds [60]. However, their contribution becomes obscure and cumbersome in cases where clear quantitative and comparative results are needed, when more than one classifier are employed comparatively [61, 62] or when there are more than two output classes [63-65] or when the generalization capabilities of classifiers is the main concern [66-67]. Furthermore, their transformation to subjective measurements for comparison purposes is usually limited to calculations involving the area under curve (AUC) or A_z index values [61, 66, 67]. In terms of a ROC plot, for a specific misclassification cost ratio between the classes (e.g., equal), the best success rate always coincides with the point on the ROC curve that exhibits the minimum distance from the ideal classifier response, which is the point (0,1) in the ROC space.

Since this study is focused on investigating the discriminative power of morphological features and their relative efficiency when used with various classifier architectures, the optimal performance was measured as the single value of their best accuracy with equal class misclassification costs, namely the success rate or hit ratio [24], which is calculated directly from the corresponding confusion matrix, instead of a ROC curve. Furthermore, the success rate, calculated by means of k-fold cross-validation methods such as the leave-one-out method employed in this study, is the most commonly accepted practice for robust evaluation of classifier performance in the discipline of Pattern Recognition [24].

6. Conclusions

Morphological shape type of masses in mammographic screening has been established as one of the leading factors for the conclusive clinical diagnosis. Radial distance measurements have been used as the basis of extensive signal analysis, including DFT spectrum analysis and DWT decomposition for multi-resolution analysis. Seven uniresolution measurements have been used as feature functions for the compact description of the properties and attributes of each of the base signals. These feature functions have been applied in all forms of base signals, producing a large set of descriptive measurements concerning the characteristics of the original shape signal.

Using various feature sub-sets with optimized linear and neural classifier models, it has been established that the DFT spectrum and the DWT components capture discriminative information of significant importance, in relation to both morphological shape type and clinical diagnosis. Furthermore, NN and SVM classifiers outperformed linear ones in all cases, while least-squares based classifiers exhibited the highest accuracies over the linear architectures.

The use of spectral properties and wavelet components of the original radial distance signal, in conjunction with optimized NN and SVM classifiers, produced significant increase in the overall success rate of the diagnostic system.

References

[1] R.L. Egan, Breast imaging: Diagnosis and morphology of breast diseases (Saunders, Philadelphia, 1988).

[2] S.L. Robbins, M. Angell, V. Kumar, Basic Pathology (Saunders, Philadelphia, 1981).

[3] L. Bocchi, G. Coppini, R. De Dominicis, G. Valli, Tissue characterization from X-ray images, Med. Eng. Phys., 19(4) (1997) 336-342.

[4] L.M. Bruce, R.R. Adhami, Classifying mammographic mass shapes using the wavelet transform modulus-maxima method, IEEE Trans. Med. Im., 18 (12) (1999) 1170-1177.

[5] G.M. Newstead, P.B. Baute, H.K. Toth, Invasive lobular and ductal carcinoma: mammographic findings and stage at diagnosis, Radiology, 184 (1992) 623-627.

[6] J.E. Meyer, E. Amin, K.K. Lindfors, J.C. Lipman, P.C. Stomper, D. Genest, Medulary carcinoma of the breast: mammographic and US appearance, Radiology, 170 (1989) 79-82.

[7] E.A. Sickles, Breast masses: mammographic evaluation, Radiology, 173 (1989) 297-303.

[8] J.L. Encarnacao, H.-O. Peitgen, G. Sakas, G. Englert, Fractal Geometry and Computer Graphics (Springer-Verlag, Berlin, 1992).

[9] C.B. Caldwell, S.J. Stapleton, D.W. Holdsworth, R.A. Jong, W.J. Weiser, G. Cooke, M.J. Yaffe, Characterization of mammographic parenchymal pattern by fractal dimension, Phys. Med. Biol., 35 (1990) 235-247.

[10] M.J. Homer, Imaging features and management of characteristically benign and probably benign lesions, Radiol. Clin. of N. Amer., 25(5) (1987) 939-951.

[11] M.J. Homer, Breast imaging: Pitfalls, controversies and some practical thoughts, Radiol. Clin. of N. Amer., 23(3) (1985) 459-472.

[12] M. Mavroforakis, H. Georgiou, N. Dimitropoulos, D. Cavouras, S. Theodoridis, Significance Analysis of Qualitative Mammographic Features, Using Linear Classifiers, Neural Networks and Support Vector Machines, Europ. J. Radiol., 54 (2005) 80-89.

[13] H. Georgiou, D. Cavouras, N. Dimitropoulos, S. Theodoridis, Mammographic mass shape characterization using neural networks, in: Proceedings of the 2nd European Symposium On Biomedical Engineering and Medical Physics, Patras, Greece, (2000) BME 13.

[14] J. Kilday, F. Palmieri, M.D. Fox, Classifying mammographic lesions using computerized image analysis, IEEE Trans. Med. Im., 12 (4) (1993) 664-669.

[15] J. Kilday, F. Palmieri, M.D. Fox, Linear discriminant based mammographic tumor classification using shape descriptors, in: Proceedings of the 12th Annual International Conference of IEEE Engineering in Medicine and Biology 1 (IEEE, New Jersey, 1990) 434.

[16] E.M. Petersen, D. Ridder, H. Handels, Image processing with neural networks – A review, Pat. Rec., 35 (2002) 2279-2301.

[17] P.R. Massopust, Fractal Functions, Fractal Surfaces and Wavelets (Academic Press, San Diego, 1994).

[18] C.S. Burrus, R.A. Gopinath, H. Guo, Introduction to wavelets and wavelet transforms – A primer (Prentice Hall, New Jersey, 1998).

[19] I. Daubechies, Ten lectures on wavelets, 2nd edition, CBMS-NSF regional conference series in applied mathematics 61 (SIAM, Philadelphia, 1992).

[20] M. Vetterli, C. Herley, Wavelets and filter banks: theory and design, IEEE Trans. Sig. Proc., 40(9) (1992), 2207-2232.

[21] D. Zhang, G. Lu, Review of shape representation and description techniques, Pat.Rec., 37 (2004) 1-19.

[22] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 3rd edition (Prentice-Hall, New Jersey, 2006).

[23] D.S. Zhang, G. Lu, A comparative study of Fourier descriptors for shape representation and retrieval, in: Proceedings of the 5th Asian Conference on Computer Vision (ACCV02), Melbourne, Australia, Jan. 22-25, (2002) 646-651.

[24] S. Theodoridis, K. Koutroumbas, Pattern Recognition, 3rd edition (Academic Press, San Diego, USA, 2006).

[25] C.Jr. Traina, A.J.M. Traina, L. Wu, C. Faloutsos, Fast feature selection using fractal dimension, in: XV Brazilian Database Symposium, Joao Pessoa, PA Brazil, 1 (2000) 158–171.

[26] W.W. Cooley, P.R. Lohnes, Multivariate data analysis (John Willey & Sons, New York, 1971).

[27] R. Leeden, K. Vrijburg, J. Leeuw, A review of two different approaches for the analysis of growth data using longitudinal mixed linear models: Comparing hierarchical linear regression (ML3, HLM) and repeated measures designs with structured covariance matrices (BMDP5V), The Statistical Software Newsletter, Computational Statistics and Data Analysis (SSNinCSDA), 21 (5) (Elsevier Science, Amsterdam, 1996) 583-605.

[28] M. Mavroforakis, H. Georgiou, N. Dimitropoulos, D. Cavouras, S. Theodoridis, Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers, Artificial Intelligence in Medicine, 37 (2) (2006) 145-162.

[29] C.J. D'Orsi, D.J. Getty, J.A. Swets, R.M. Pickett, S.E. Seltzer, B.J. McNeil, Reading and decision aids for improved accuracy and standardization of mammographic diagnosis, Radiology, 184 (1992) 619-622.

[30] Y. Wu, M.L. Giger, K. Doi, C.J. Vyborny, R.A. Schmidt, C.E. Metz, Artificial neural networks in mammography: Application to decision making in the diagnosis of breast cancer, Radiology, 187 (1993) 81-87.

[31] L.V. Ackerman, A.N. Mucciardi, E.E. Gose, F.S. Alcorn, Classification of benign and malignant breast tumors on

the basis of 36 radiographic properties, Cancer, 31 (1973) 342-352.

[32] J. Suckling, J. Parker, D.R. Dance, S. Astley, I. Hutt, C.R.M. Doggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S.L. Kok, P. Taylor, D. Betal, J. Savage, The Mammographic Image Analysis Society Digital Mammogram Database, Exerpta Medica, International Congress Series, 1069 (1994) 375-378.

[33] H. Li, Y. Wang, K.J.R. Liu, S.-C.B. Lo, M.T. Freedman, Computerized radiographic mass detection – Part I: Lesion site selection by morphological enhancement and contextual segmentation, IEEE Trans. Med. Im., 20 (4) (2001) 289-301.

[34] B. Sahiner, N. Petrick, H.P. Chan, L.M. Hadjiiski, C. Paramagul, M.A. Helvie, M.N. Gurcan, Computer-Aided Characterization of Mammographic Masses: Accuracy of Mass Segmentation and Its Effects on Characterization, IEEE Trans. Med. Im., 20 (12) (2001) 1275-1284.

[35] Y. Hatanaka, T. Hara, H. Fujita, S. Kasai, T. Endo, T. Iwase, Development of an Automated Method for Detecting Mammographic Masses with a Partial Loss of Region, IEEE Trans. Med. Im., 20 (12) (2001) 1209-1214.

[36] H. D. Cheng, Muyi Cui, Mass lesion detection with fuzzy neural network, Pat. Rec., 37 (2004) 1189–1900.

[37] MathWorks Inc., Matlab 7.0 documentation: Signal processing toolbox (MathWorks, Massachusetts, 2004).

[38] L.M. Bruce, R.R. Adhami, J.W. Bruce, Appropriate scales when using wavelets for feature extraction, in: Dagli et al, eds., Proceedings of the Artificial Neural Networks in Engineering ’96, St.Louis, MO, (1996) 507-512.

[39] S.G. Mallat, Zero-crossings of a wavelet transform, IEEE Trans. Information Theory, 37 (4) (1991) 1019-1033.

[40] H.D. Cheng, X.J. Shi, R. Min, L.M. Hu, X.P. Cai, H.N. Du, Approaches for automated detection and classification of masses in mammograms, Pat. Rec., 39 (2006) 646-668.

[41] N. Petrick, H.P. Chan, B. Sahiner, M.A. Helvic, Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms, Med. Phys., 26 (1999) 1642-1654.

[42] B. Sahiner, H.P. Chan, N. Petrick, M.A. Helvic, L.M. Hadjiiski, Improvement of mammographic mass characterization using spiculation measures and morphological features, Med. Phys. 28 (7) (2001) 1455-1465.

[43] N.M. El-Faramawy, R.M. Rangayyan, J.E.L. Desautels, O.A. Alim, Shape factors for analysis of breast tumors in mammograms, Canadian Conference on Electrical and Computer Engineering ’96, Calgary, Canada (Elsevier Science, New Jersey, 1996) 355-358.

[44] P.J. van Otterloo, A contour-oriented approach to shape analysis (Prentice-Hall, New Jersey, 1991).

[45] Q.M. Tieng, W.W. Boles, Recognition of 2D object contours using wavelet transform zero-crossing representation, IEEE Trans. PAMI, 19 (8) (1997) 910-916.

[46] D.S. Zhang, G. Lu, A comparison of shape retrieval using Fourier descriptors and short-time Fourier descriptors, in:

H.-Y. Shum, M. Liao, S.-F. Chang, eds., Proceedings of the 2nd IEEE Pacific-Rim Conference on Multimedia (PCM01), Beijing, China, October 24-26 (Lecture Notes in Computer Science vol.2195, Springer, 2001) 855-860.

[47] C.C. Chen, J.S. Daponte, M.D. Fox, Fractal Feature Analysis and Classification in Medical Imaging, IEEE Trans. Med. Im., 8 (2) (1989) 133-142.

[47] A.W. Moore, M.S. Lee, Efficient algorithm for minimizing cross validation error, in: Proceedings of the 11th.International Conference on Machine Learning (ML-94), 10-13 July 1994, New Brunswick, New Jersey (Morgan Kaufmann, San Francisco, 1994).

[48] B. Abrahao, L. Barbosa, Characterizing Datasets Using Fractal Methods, Technical report, Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, MG Brazil, 2003.

[49] S. Pierre, J.F. Rivest, On the Validity of Fractal Dimension Measurements in Image Analysis, J. Vis. Com. Im. Proc., 7 (3) (1996) 217-229.

[50] L. Devroye, L. Gyorfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition (Springer-Verlag Inc., New York, 1996).

[51] T. Poggio, F. Girosi, Networks for approximation and learning, in: Proceedings of the IEEE, 78 (9) (1990) 1481-1497.

[52] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd Ed, (Prentice Hall, New Jersey, 1999).

[53] N. Christianini, J. Shawe-Taylor, An Introduction to Support Vector Machines (Cambridge University Press, UK, 2000).

[54] V.N. Vapnik, Statistical Learning Theory (John Wiley & Sons, New York, 1998).

[55] M. Mavroforakis, S. Theodoridis, Reduced Convex Hulls: A Geometric Approach to Support Vector Machines, IEEE Sig. Proc. Mag., 24 (3) (2007) 120-123.

[56] M. Stone, Cross-validatory choice and assessment of statistical predictions, J. R. Stat. Soc., 21 (1) (1974) 111-147.

[57] P.A. Lachenbruch, R.M. Mickey, Estimation of error rates in discriminant analysis, Technometrics, 10 (1968) 1-11.

[58] R. Reed, Pruning algorithms – A survey, IEEE Trans. N.N., 4(5) (1998) 740-747.

[59] C.R. Jung, J. Scharcanski, Adaptive image denoising and edge enhancement in scale-space using the wavelet transform, Pat. Rec. Letters, 24 (2003) 965-971.

[60] J.A. Swets, R.M. Pickett, Evaluation of Diagnostic Systems: Methods from Signal Detection Theory (Academic Press, New York, 1982).

[61] T. Fawcett, An introduction to ROC analysis, Pat. Rec. Lett., 27 (2006) 861-874.

[62] N.M. Adams, D.J. Hand, Comparing classifiers when the misallocation costs are uncertain, Pat. Rec., 32 (1999)

1139-1147.

[63] J.E. Fieldsend, R.M. Everson, Formulation and comparison of multi-class ROC surfaces, in: Proceedings of ROCML 2005, part of the 22nd International Conference on Machine Learning (ICML 2005), 41-48.

[64] A. Srinivasan, Note on the location of optimal classifiers in n-dimensional ROC space, Technical Report PRG-TR-2-99, Oxford University Computing Laboratory (Oxford University, UK, 1999).

[65] T. Fawcett, ROC Graphs: Notes and Practical Considerations for Data Mining Researchers, Intelligent Enterprise Technologies Laboratory, HP Laboratories Palo Alto, HPL-2003-4.

[66] S. Agarwal, T. Graepel, S. Har-Peled, R. Herbrich, D. Roth, Generalization Bounds for the Area Under the ROC Curve, J. Machine Learning Research, 6 (2005) 393-425.

[67] C. Cortes, M. Mohri, AUC Optimization vs. Error Rate Minimization, Technical Report, AT&T Labs Research (AT&T, New Jersey, 2003).

708

Multi-scaled morphological features for the characterization of mammographic masses using statistical classification schemes

Harris Georgiou[1], Michael Mavroforakis[1], Nikos Dimitropoulos[2], Dionisis Cavouras[3], Sergios Theodoridis[1]

[1] University of Athens, Informatics Department, TYPA buildings, University Campus, 15771, Athens, Greece

[2] Medical Imaging Department, EUROMEDICA Medical Center, 2 Mesogeion avenue, Athens, Greece

[3] Medical Imaging Technologies Department, TEI-Athens, 12210, Athens, Greece

Corresponding Author:

Harris Georgiou (Mr), 11 Vas. Dipla str, P.O. 11745, Athens, Greece.

Tel: +30 210 9313361, Fax: +30 210 9313631, Email: xgeorgio@di.uoa.gr

739 **Table 1** – Complete list and quantification details for all the qualitative properties, used by the expert physician when annotating
740 the images in the mammogram database
741

Qualitative feature	**Range**
Patient's age	Years (integer)
Mass existence	Yes / no
Microcalcifications existence	Yes / no
Fat percentage	0%,…,100%
Boundary sharpness	0%,…,100%
Mass density	L (hypo) / M (iso) / H (hyper)
Mass homogeneity	1,…,10
Mass shape type	1 (round) / 2 (lobulated) / 3 (micro-lobulated) / 4 (stellate)
Verified diagnosis	B (benign) / M (malignant)

742
743
744

Table 2 – Distribution of the four morphological shape types against diagnosis in the image database. Percentages are calculated per column (shape type).

	Round	**Lobulated**	**Micro-lobulated**	**Stellate**
Benign	**25**	**18**	**2**	**1**
	83%	95%	5%	3%
Malignant	**5**	**1**	**41**	**37**
	17%	5%	95%	97%

749 **Table 3** – Correlation fractal dimension (D_2) value for qualitative and combined morphological feature datasets, for all 217 and

750 for top-20 MANOVA selected morphological features.

751

D_2 value	**complete set**	**MANOVA top-20**
Qualitative properties set	3,18	–
Combined morphological features set	**4,11**	**2,71**

752

Table 4 – Highest success rates of all classifiers against shape type prediction using the best 20 features, as selected by MANOVA from the complete set of 217 morphological features and for classification against all four shape type classes (* = not used).

Classifier model	Success rate (%)
LDA	66,90%
K-NN	67,70%
LSMD	**72,30%**
MLP	**72,30%**
RBF	67,70%
C-SVC/RBF*	–

Table 5 – Highest success rates of all classifiers against diagnosis prediction using the best 20 features, as selected by MANOVA from the complete set of 217 morphological features.

Classifier model	Success rate (%)
LDA	89,23%
k-NN	85,38%
LSMD	88,46%
MLP	89,23%
RBF	83,85%
C-SVC/RBF	**91,54%**

Table 6 – Overall ranking of the seven curve feature functions, in order of preference, for the shape type and diagnosis classification targets.

Curve feature function ranking	
Shape type	**Diagnosis**
ZC count	Mean
Mean	Area ratio
Area ratio	Roughness
Entropy	Stdev
Roughness	Entropy
Stdev	Circularity
Circularity	ZC count

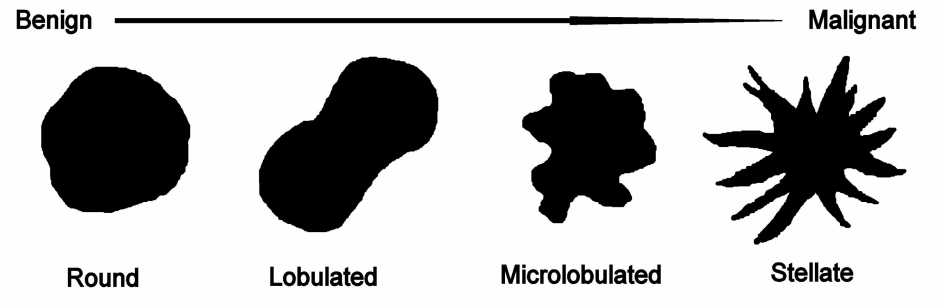

Figure 1 – Morphological shape type representations of mammographic tumors.

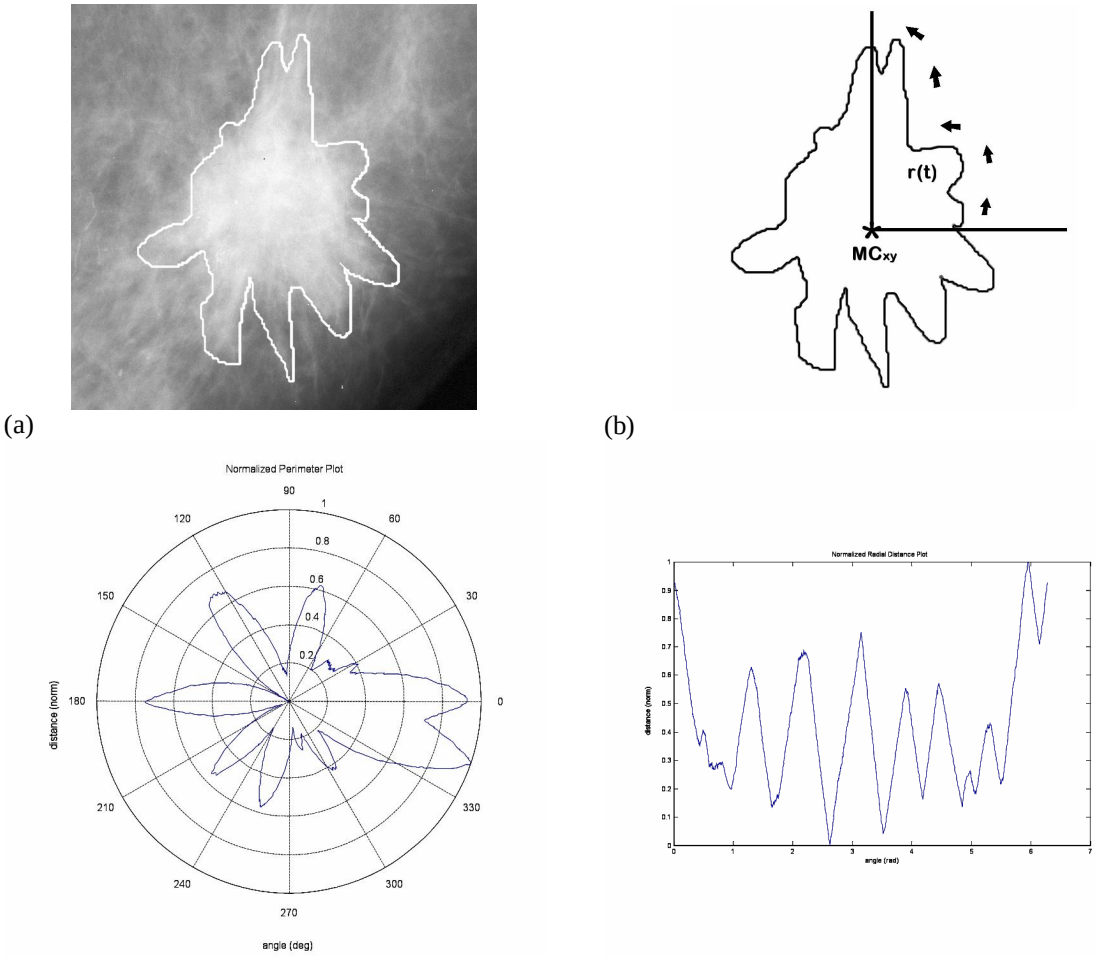

Figure 2 – Radial distance signal extraction: (a) mass boundaries segmentation, (b) mass perimeter radial sampling, (c) normalized mass perimeter, (d) normalized radial distance signal.

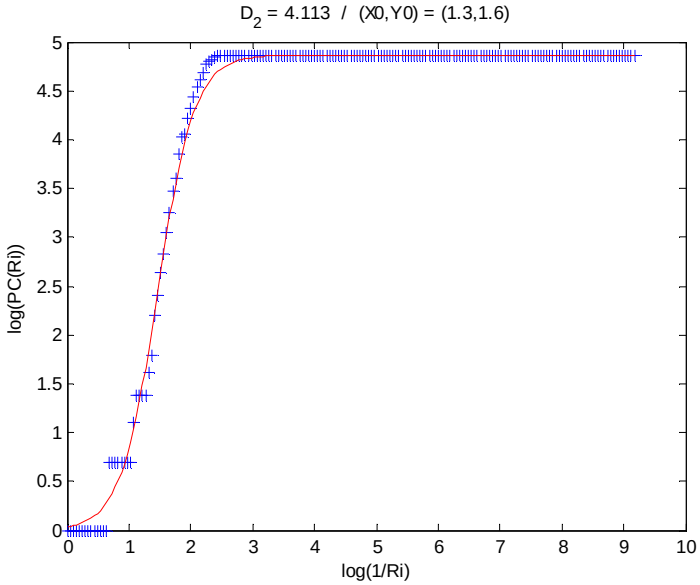

Figure 3 – PC(r) plot and sigmoid fitness function for the combined dataset of all 217 morphological features. For X-axis, (1/r) was used instead of (r) for correct (+) sign for the slope value. The Y-axis represents the log of the pair-count value for parameter (r).

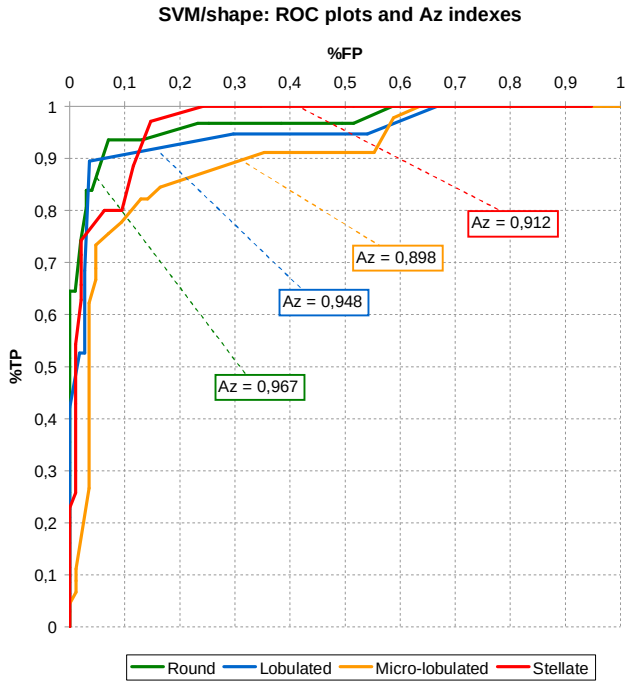

Figure 4 – ROC plot for all four shape type classifications with SVM classifiers, using each one of the shape types against the other three, as well as the corresponding AUC or $\underline{A_z}$ index value for each classification setup.

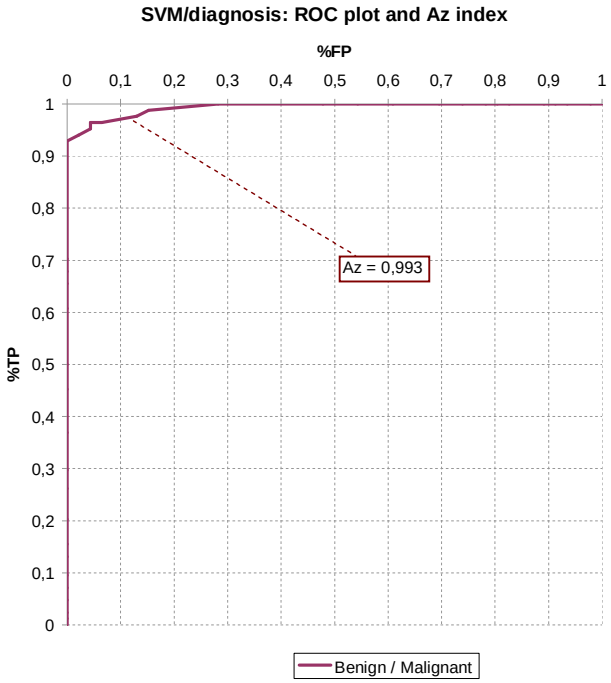

Figure 5 – ROC plot for the diagnosis classification using SVM classifier and the corresponding AUC or $\underline{A_z}$ index value.