Multi-resolution Morphological Analysis and Classification of Mammographic Masses Using Shape, Spectral and Wavelet Features

H. V. GEORGIOU[1], M. E. MAVROFORAKIS[1], D. CAVOURAS[2], N. DIMITROPOULOS[3], S. THEODORIDIS[1]

[1] Univ. of Athens, Informatics Dept., TYPA buildings, Univ. Campus, 15771, Athens, Greece
[2] Medical Imaging Technologies Dept., TEI-Athens, 12210, Athens, Greece
[3] Medical Imaging Dept., EUROMEDICA Medical Center, 2 Mesogeion ave, Athens, Greece

Contact: H. Georgiou (Mr), 11 Vas.Dipla str, 11745, Athens, Greece – mailto:xgeorgio@middle-earth.gr
M. Mavroforakis (Mr), 43 Knossou str, 16561, Athens, Greece – mailto:memav@nbg.gr

Abstract – This study constitutes a comprehensive signal analysis approach to the morphological characterization of mammographic mass shape. Three distinct areas of shape morphology were exploited for feature extraction. Specifically, the radial distance signal, the DFT spectrum envelope and the DWT decomposition with multiple wavelet function choices, were analyzed by seven curve feature functions, as carriers of significant discriminating information. Classification was conducted against the morphological shape type identification, as well as the verified clinical diagnosis, using optimized feature set selections and combinations by multivariate statistical significance analysis. All available datasets and configurations were applied to a wide range of linear and neural classifiers, including linear discriminant analysis, least-squares minimum distance, K-nearest neighbor, RBF and MLP neural networks. Neural classifiers outperformed linear equivalents in all cases, producing an overall accuracy of 72.3% for morphological shape type identification and 89.2% for clinical diagnosis identification.

I. INTRODUCTION

In localized breast cancers, the morphology and shape characteristics of the detected tumors have been established as factors of outmost significance in the discrimination between fibroadenomas, cysts and carcinomas, due to inherent anatomical differences directly related to malignancy [1]-[2]. Approximately 80% to 85% of diagnostic information is retrieved from the mammographic appearance of the tumor [3]. Clinical studies have shown that malignant masses often infiltrate the surrounding tissue as they expand. As a result the tumor's boundaries appear vague and fuzzy, with linear strands extending irregularly outwards, while benign pathological cases appear as mass shapes of well-defined boundaries and non-stellate structures [3].

Based on these morphological differences evidenced by mammographic appearance, four distinct shape types have been established. These are round, lobulated, nodular or micro-lobulated, and stellate or spiculated. Round shape morphologies relate primarily to benign pathology, while stellate shape morphologies relate to malignancy [3]-[5].

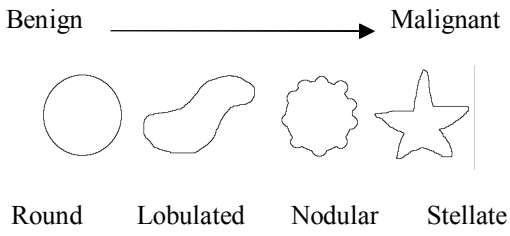

Figure-1. Morphologic shape types of mammographic masses

This study is focused on three main areas of interest. First, the original radial distance measurements were studied under the complete spectrum of signal analysis. Morphological shape analysis is conducted, not only to the radial distance signal itself, but also to the envelope of the DFT spectrum and the wavelet decomposed sub-band signals for multiple wavelet function selections. Second, seven uniresolution shape descriptor measurements were applied on all the extracted signals and the discrimination value of each one of them was evaluated by multivariate analysis (MANOVA). Third, a wide range of classifier architectures were applied on the same datasets in order to conduct a comparative evaluation of the efficiency and effectiveness of each architecture, for the specific problem of mammogram tumor classification.

The three areas of study were conducted in relation to two different classification goals, either the shape type of the tumor or the verified clinical diagnosis for each mammogram. The two cases were treated as separate problems, although dataset processing and classification models were essentially the same.

II. MAMMOGRAM DATABASE

For this study, a subset of 163 mammograms where selected for digitization by an expert. The selection was made on the basis of unbiased statistical distribution and the completeness of the dataset. All cases were positively verified clinically by biopsy and further diagnostic tests. For each mammogram, a complete list of qualitative information was filled by the expert, including details about tumor shape morphology and clinical diagnosis.

The mammograms where digitized at a resolution of 63 μm (400 dpi) at 8-bit gray level and some post-processing was applied in order to enhance the sharpness of the images. From the initial set of 163 mammograms, a total of 33 were not used in this study, either because of the absence of distinct tumors or due

to ambiguity on the exact shape type. The final set of 130 mammograms was used in all cases with no reduction in spatial resolution or gray level depth.

Using the original mammographic database of the 130 images, mass boundaries were manually described by the expert and used later for the extraction of the perimeter and radial distance measurements for each tumor.

III. CURVE FEATURE EXTRACTION

Seven traditional uniresolution shape features were used to extract quantitative information from a one-dimensional signal. These shape features have been widely used in the past to characterize signals extracted as the radial distance function samples over a well-defined mass borderline [3]-[5].

A. Radial Distance Signal

In general, each radial distance function is computed by first detecting the centroid of the mass using cumulative distributions of the projections in both x and y axes. Then the acquired centroid of the mass is defined as the center of a polar coordinate system and, for a specific angular resolution N, the mass borderline is sampled over $2\pi/N$ intervals to extract the radial distance sequence starting from angle 0.

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

Instead of using direct centroid-to-boundary measurements, the boundary of each mass was acquired at an arbitrary angle and then registered in full resolution using a line-following scheme.

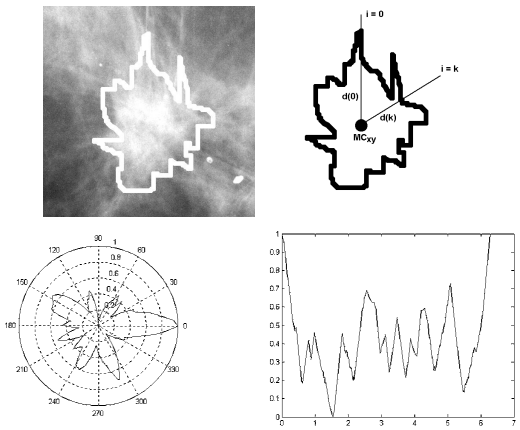

Figure-2. Radial distance signal extraction: (a) mass boundaries detection, (b) mass perimeter radial sampling, (c) normalized mass perimeter, (d) normalized radial distance signal.

All radial distance signals were normalized using min-max values in order to avoid excessive differences in magnitude and energy between masses of different sizes.

B. Curve Features

The seven uniresolution curve features used in this study were: (i) radial distance mean, (ii) radial distance standard deviation, (iii) tumor circularity, (iv) entropy of the radial distance histogram, (v) area ratio parameter, (vi) zero-crossing count and (vii) tumor boundary roughness. The area was calculated as the integral of the normalized radial distance signal, while the perimeter was extracted by summing the differences of all samples sequentially. The entropy E of the radial distance was calculated on the normalized histogram of radial distances with N=100 bins. Finally, for the tumor boundary roughness estimator, a fixed value of segments size L=16 was experimentally determined.

C. Fourier Analysis – Discrete Spectrum

Preliminary results and spectral analysis proved the significance of properties of the spectrum as discriminating factors between the various shape types of the masses and the final diagnostic conclusion. As the normalized radial distance signal represents the shape and the texture of each tumor's perimeter, abnormalities in the initial radial distance signal should be evident as explicit properties of the discrete spectrum of the signal. Thus, instead of using the initial radial distance function samples as the base for the descriptive curve features, the DFT spectrum envelope was also included in the study.

3.4 Wavelet Decomposition

L.M.Bruce and R.R.Adhami [3] have shown the significance and positive value of multiresolution features that are based on wavelet transform using modulus-maxima method. In this study, instead of creating a modulus-maxima representation of the signal produced by the wavelet decomposition, each 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 tumor boundary and thus produce detailed discriminating information.

In order to estimate the effect of wavelet function selection over discriminating power of the curve features calculated in each case, several wavelet functions were applied in the wavelet decomposition process. Specifically, five wavelet functions were selected: (i) Haar, (ii) sym2, (iii) coiflet-1, (iv) biorthogonal-1.5 and (v) discrete Meyer. Wavelet components at small scales reveal abnormalities in localized level, while large scales reveal more general abnormalities for the whole shape. Features extracted separately at each decomposition level could also reveal the quality and value of the discrimination information contained on that level. The maximum decomposition level was established at 4 after evaluating cases of applying wavelet decomposition methods [6] in tumor perimeter signal and according to the minimum and maximum signal lengths available in the current database.

IV. CLASSIFICATION AND TESTING

Classification was conducted for two separate configurations. The first one contained the shape type as the classification target, while the second one contained the clinical diagnosis as the classification target.

Several classifier architectures were applied during the classification phase. A LDA model was used in the form of linear classifier. A least-squares minimum distance classifier (LSMD) was employed, using Mahalanobis distance measure and least-squares dataset pre-processing on the input. A K-nearest neighbor model was also used, including estimation of an optimal value K for the size of the neighborhood set.

Two different types of neural network architectures were employed: a RBF network with Gaussian activation function and linear output function, and a MLP network with hyperbolic tangent internal activation and softmax output function, both implemented with topology adapted in each configuration and dataset. All topologies included one hidden layer with optimized size. All configurations used 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.

V. RESULTS AND DISCUSSION

Preliminary studies on the original mammographic database have confirmed the significance of morphological shape characterization with the clinical diagnostic result. Specifically, it was statistically verified that round and lobulated cases exhibit only 5% to 16% malignancy, while nodular and stellate cases exhibit 90% to 97% malignancy.

	Round	Lobulated	Nodular	Stellate
Benign	25	18	1	1
	83%	95%	2%	3%
Malignant	5	1	42	37
	17%	5%	98%	97%

Table-1. Training patterns true distribution against morphological shape types and clinical diagnosis.

Derived from the base DFT spectrum dataset, absolute value and logarithmic versions of the 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 patterns were grouped into separate datasets according to their base signal, along with one dataset containing all of them.

A. Shape Type Classification Mode

Statistical analysis and classification results have shown that the discrimination value of features extracted from the DWT components and especially the DFT spectrum is of great importance. Using all the available classifiers optimized for each setup and leave-out-testing, the success rate for classification based only on features extracted from the original radial distance signal ranged between 46.1% and 57.7%, while the same results for feature sets based on DWT component signals range from 53.0% to 67.7% over all wavelet function selections, and for feature sets based on DFT spectrum success rates range from 60% to 72%. Wavelet function choice proved to be significant according to the classifier choice, as some wavelet functions with specific classifier models produced the best results, while other combinations produced poor success rates. For DFT spectrum feature sets, absolute value and logarithmic transformations on pre-processing produced a positive or negative result in the final success rate according to the selected classifier model, ranging between 1% and 9% for a specific classifier choice. Discriminating information content was greatly decreased, about –20% to –26%, when DFT spectrum resolution was reduced to 128 points, implying that the texture and fine detail of the spectrum was of great importance in relation to the quality and discrimination power of the extracted features.

Using the complete feature set of dimension 217 and applying multivariate statistical significance feature selection, optimum feature sets were selected according to the current classification target. All created feature sets contained features extracted primarily from DFT spectrum or 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 was the selection of features extracted from the DFT spectrum, combined with an optimized least-squares minimum distance (LSMD) or a multi-layer perceptron (MLP) with one hidden layer and softmax output, both resulting in 72.3% success rate for leave-one-out testing mode.

B. 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 DWT components and especially the DFT spectrum is of great value. Using all the available classifiers optimized for each setup and leave-out-testing, the success rate for classification based only on features extracted from the radial distance signal ranged between 66.9% and 73.8%, while the same results for feature sets based on DWT component signals ranged from 73.8% to 85.4% over all wavelet function selections, roughly the same success rates for feature sets based on DFT spectrums. Wavelet function choice was significant in certain cases for an increase in success rate up to +7.7% for each classifier choice. For DFT spectrum feature sets, absolute value and logarithmic transformations on pre-processing produced little to no significant variance in the final success rate according to the selected classifier model, ranging between 0% and 4.6% for specific classifier choices. Discriminating information content was, again, greatly decreased, about –10% to –27%, when DFT spectrum resolution was reduced to 128 points, implying that the texture and fine detail of the spectrum if of great importance in relation to the quality and discriminating power of the extracted features.

Using the complete feature set of dimension 217 and conducting multivariate statistical significance feature selection, optimum feature sets were selected according to the current classification target. All created feature sets 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 by itself. The best classification configuration was the selection of mixed features extracted from DFT spectrum and DWT components of various wavelet functions, with a multi-layer perceptron (MLP) with one hidden layer of optimum size and softmax output, resulting in 89.2% success rate for leave-one-out testing mode.

C. Classifier Architectures Evaluation

Classification results both for morphological shape type and clinical diagnosis configurations 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.

Between linear classifiers, LSMD resulted in significant increase in overall success rate, similar in some cases to the one produced by various neural network architectures.

For neural networks, comparison between RBF and MLP architectures presented evidence that RBF networks result in somewhat lower overall accuracy. Differences in success rates ranged between 0% and 10%, favoring the choice of MLP architectures over all optimized topology sizes.

In almost all cases, the application of neural networks 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 accuracy between linear and neural classifiers 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 an LSMD linear classifier. However, for both classification targets, the best accuracies where achieved by optimized MLP classifiers, specifically 72.3% for morphological shape type and 89.2% for clinical diagnosis, all evaluated using leave-one-out testing method.

D. Optimal feature selection and evaluation

Results over all the available datasets and groupings have shown that feature selection both during statistical significance analysis and set optimization depends greatly on the dataset. There is no clear preference of one feature function over another when applied in base signals of different information content, although some of them may be of greater discrimination importance within the same dataset.

For optimized classifier structures, misclassification analysis have shown that in almost all cases the results were compact and consistent. Misclassified morphological shape types rarely differed one class away from the correct one, while completely invalid classifications including round instead of stellate types or vice versa, were essentially zero for optimized classifiers using any of the datasets. In the case of clinical diagnosis classification, the false-positive and false-negative results were similar in volume and content to the error rate involved in the statistical covariance analysis between tumor shape types and clinical diagnosis in the original mammogram database.

VI. CONCLUSION

Morphological shape type of tumors 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 the base signal. These feature functions have been applied in all forms of base signals, producing a large set of descriptive measurements about 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 discriminating information of significant importance, in relation to both morphological shape type and clinical diagnosis. Furthermore, neural classifiers outperformed linear equivalents in all cases, while least-squares based classifiers exhibited the highest accuracies over the linear classifiers. The use of spectral properties and wavelet components of the original radial distance signal, in conjunction with an optimized neural classifier, produced significant increase in the overall accuracy of the diagnostic system.

REFERENCES

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

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

[3] L.M.Bruce, R.R.Adhami, Classifying mammographic mass shapes using the wavelet transform modulus-maxima method, IEEE Trans.Med.Im., Vol.18, No.12, Dec.1999, 1170-1177.

[4] H.Georgiou, D.Cavouras, N.Dimitropoulos, S.Theodoridis, Mammographic mass shape characterization using neural networks, 2nd European Symp. On Biom.Eng. and Med.Phys., Patras, Greece, 2000, BME 13.

[5] J.Kilday, F.Palmieri, M.D.Fox, Classifying mammographic lesions using computerized image analysis, IEEE Trans.Med.Im., Vol.12, No.4, Dec.1993, 664-669.

[6] L.M.Bruce, R.R.Adhami, J.W.Bruce, Appropriate scales when using wavelets for feature extraction, Proc.ANN.Eng.'96, St.Louis, MO, Nov.1996.

[7] L.V.Ackerman, Computer classification of radiographs and xerograms of the breast, Ph.D. dissertation, Univ.Illinois, 1971.

[8] H.S.Gallager, J.E.Martin, "The study of breast carcinoma by correlated mammography and subserial whole organ sectioning: Early observations", Cancer, Vol.23, 1969, 855-873.

[9] J.E.Martin, H.S.Gallager, Mammographic diagnosis of minimal breast cancer, Cancer, Vol.28, 1971, 1519-1526.

[10] X.M.Luo, An expert system for mammogram analysis, M.S. thesis, Univ.Tennessee, Knoxville, TN, 1989.

[11] J.Kilday, F.Palmieri, M.D.Fox, Linear discriminant based mammographic tumor classification using shape descriptors, 12th Ann.Int.Conf. IEEE Eng.Med.Biol., Vol.1, 1990, 434-434.