

Diagnostic Classification of Digital Mammograms by Wavelet-Based Spectral Tools: A Comparative Study

Erin K. Hamilton*, Seonghye Jeon*, Pepa Ramírez Cobo[†], Kichun Sky Lee[‡] and Brani Vidakovic*

*Georgia Institute of Technology, Atlanta, Georgia

[†]University of Seville, Seville, Spain

[‡]Emory University, Atlanta, Georgia

Abstract—The aim of this paper is to present results from a comparative investigation into the diagnostic performance of several wavelet-based estimators of scaling, some from published literature and some newly proposed. These estimators are evaluated based on their ability to classify digitized mammogram images from a clinical database, for which the true disease status is known by biopsy. We found that Abry-Veitch and modified weighted Theil-type estimators provided the best classification rates, while the standard wavelet-based OLS estimator performed worst. The results are robust with respect to choice of wavelets (Haar wavelet being an exception) and are of potential clinical value. The diagnostic is based on the properties of image backgrounds (which is an unused diagnostic modality in Mammograms) and the best correct classification rates achieve 90%, varying slightly with the choice of basis, levels used, and size of training set.

Keywords—Wavelets; Scaling; fBm; Mammogram analysis; Breast cancer diagnostic

I. INTRODUCTION

Breast cancer is one of the most common forms of cancer among women in the United States, second only to non-melanoma skin cancer. It has been estimated that 1 in 8 women born today will be diagnosed with breast cancer during her lifetime [1]. One of the most important tools toward reducing breast cancer deaths is advanced precision of screening technologies. Early detection is still the best strategy for improving prognosis and also leads to less invasive options for both specific diagnosis and treatment.

Mammography is currently the best method for detecting a breast cancer early, before the malignant tissue is substantial enough to feel or cause symptoms. However, the radiological interpretation of mammogram images is a difficult task since the appearance of even normal tissue is highly variable and complex, and signs of early disease are often small or indistinct. Reading a mammogram image is a skill that physicians develop over time, and confidently stating whether findings are cancerous or not is often difficult. Suspicious findings are commonly clarified by follow-up images, ultrasound, or MRI. On the other hand, it has been estimated that 10–30% of cancers which could have been detected are missed [2]. Thus, improving both the specificity and the sensitivity of mammographic diagnoses is an important goal in improving prognoses while also reducing the number of unnecessary procedures or surgical operations.

In high frequency and irregular data collected in real-life settings (both naturally occurring and human-made), a commonly occurring phenomenon is that of regular scaling. Examples of this have been found in a variety of systems and

processes including economics (stock market, exchange rate fluctuations), telecommunications (internet data), physics (hydrology, turbulence), geosciences (wind and rainfall patterns), and several applications in biology and medicine (DNA sequences, heart rate variability, auditory nerve-spike trains). The irregular behaviors of these complex structures are difficult or impossible to quantify by standard modeling techniques; but when observations are inspected at different scales, there is a regular relationship between the behavior at each scale. This phenomenon has been demonstrated in many medical images, leading to the diagnostic use of tools capable of quantifying statistical similarity of data patterns at various scales.

The standard measure of regular scaling is the Hurst exponent. This measure can also be connected to measures of long memory, dimension, and fractality in signals and images and is viewed as an informative summary. Many techniques for estimating the Hurst exponent exist, and assessing the accuracy of these estimations can be complicated. In this paper, we compare several scaling estimators, based on various estimation techniques. But rather than focusing on the modeling capabilities of these measures, we focus on their ability to differentiate between cancerous versus normal tissue in the backgrounds of mammogram images. Note that this diagnostic use of information contained in the background tissue of images is novel, since most of the references found in literature dealing with breast cancer detection methods are based on microcalcifications [3], [4], [5]. Only recently has information contained in the background come into consideration [6]. This classifying measure based on background tissue would be a new tool to be used in combination with existing clinical diagnostic tools, thus improving the power of non-invasive diagnostic techniques.

All measures used in this analysis are based on wavelet theory, which continues to grow in its importance for image processing techniques [7], [8], [6]. In this context wavelet transforms are powerful tools because of their innate ability to model statistical similarity of signals and images at different scales.

The paper is organized as follows. Section II briefly describes wavelet transforms and scaling, after which we present the scaling assessment tools (classifiers) in Section III. In Section IV we provide the comparative performance of these classifiers. In Section V we provide discussion and make recommendations for practical use of the methodology. Technical details concerning newly introduced robust measures discussed in Section III are deferred to the Appendix.

II. BACKGROUND

A. The Discrete Wavelet Transform

The advancement of computer and imaging technology in modern medicine has resulted in enormous amounts of medical imaging information available in digital formats. Many techniques have been developed to make image information more accessible and manageable for analysis. Among these techniques, wavelet transforms have been shown to be particularly useful. The wavelet transform decomposes a signal into many different frequency bands, which works very well with the statistical properties of most images, thus lending itself to many useful applications in modeling, analysis, and handling of image data. We now give a brief overview of the discrete wavelet transform, and its extension into the 2-D case.

The discrete wavelet transform (DWT) of a function $\{X(t), t \in \mathbb{Z}\}$ represents this function in terms of shifted and dilated versions of a wavelet (or *mother*) function $\psi(t)$ and shifted versions of a scaling (or *father*) function $\phi(t)$. For specific choices of the scaling functions and wavelets, an orthonormal basis can be formed from the atoms

$$\begin{aligned}\psi_{j,k}(t) &= 2^{j/2} \psi(2^j t - k) \\ \phi_{j,k}(t) &= 2^{j/2} \phi(2^j t - k), \quad \forall j, k.\end{aligned}$$

Then $X(t)$ can be represented by wavelets as

$$X(t) = \sum_k c_{J_0,k} \phi_{J_0,k}(t) + \sum_{j=J_0}^{\infty} \sum_k d_{j,k} \psi_{j,k}(t), \quad (1)$$

where

$$d_{j,k} = \int X(t) \psi_{j,k}(t) dt \quad \text{and} \quad c_{j,k} = \int X(t) \phi_{j,k}(t) dt$$

are detail and scaling coefficients, respectively. Here, J_0 is the coarsest scale or lowest resolution of the transform, and larger values of j correspond to higher resolutions. For a detailed introduction to wavelet theory, the reader is referred to [9], [10], or [11].

Data sets can be easily and quickly transformed by the DWT through coding the data by the wavelet coefficients. When dealing with functions that are given by their sampled values, it is customary to set the sampled values to be “smooth” coefficients at the highest resolution level $j = J$. The subsequent “detail” levels obtained through DWT are denoted by d_j , corresponding to $j = J - 1, J - 2, \dots, J_0$.

Many signals arising in practical applications (astronomy, geophysics, economics, etc.) are multidimensional. The DWT is easily generalized to the multidimensional case. Since we are interested in the wavelet transforms of medical images, the generalization shown here is for the 2-D case. The 2-D wavelet basis functions are constructed via translations and dilations of a tensor product of univariate wavelet and scaling functions:

$$\begin{aligned}\phi(t_1, t_2) &= \phi(t_1)\phi(t_2) & \psi^h(t_1, t_2) &= \phi(t_1)\psi(t_2) \\ \psi^v(t_1, t_2) &= \psi(t_1)\phi(t_2) & \psi^d(t_1, t_2) &= \psi(t_1)\psi(t_2).\end{aligned} \quad (2)$$

The symbols h, v, d in (2) stand for horizontal, vertical and diagonal directions, respectively, since the atoms capture image features in the corresponding directions.

Consider the wavelet atoms

$$\begin{aligned}\phi_{j,\mathbf{k}}(\mathbf{t}) &= 2^j \phi(2^j t_1 - k_1, 2^j t_2 - k_2) \\ \psi_{j,\mathbf{k}}^i(\mathbf{t}) &= 2^j \psi^i(2^j t_1 - k_1, 2^j t_2 - k_2),\end{aligned}$$

for $i = h, v, d$, where $j \in \mathbb{Z}$, $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$, and $\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$. Then, any function $X \in \mathcal{L}_2(\mathbb{R}^2)$ can be represented as

$$X(\mathbf{t}) = \sum_{\mathbf{k}} c_{J_0,\mathbf{k}} \phi_{J_0,\mathbf{k}}(\mathbf{t}) + \sum_{j \geq J_0} \sum_{\mathbf{k}} \sum_i d_{j,\mathbf{k}}^i \psi_{j,\mathbf{k}}^i(\mathbf{t}),$$

where the detail coefficients are given by

$$d_{j,\mathbf{k}}^i = 2^j \int X(\mathbf{t}) \psi^i(2^j \mathbf{t} - \mathbf{k}) dt.$$

Since this transformation is linear, a fast DWT can be achieved by matrix multiplication, similar to a Fast Fourier transform. We direct the reader to [11] (pp 115-116, 153-159) for the construction of these wavelet matrices, both in the 1-D and 2-D case.

B. Scaling and wavelet-based spectra

The methodology used to analyze scaling is based on the analysis of autocovariances, or correlations between observations as a function of the time separation between them. The variance of a signal in its original domain corresponds to its “energy” in the frequency domain. The term “energy” is an informal name for the squared coefficients in frequency-domain representations of signals and images such as (1). Thus, the correlation between time-separated observations in the original domain corresponds to the scaling of energy in the frequency/scale domains. This introduces the idea of energy spectra as a tool for characterizing the behavior of data, and the Hurst exponent (H) as the standard measure of regular scaling. We now describe how this spectra can be represented using wavelet-based methods, and then extend these methods into the 2-D case.

A stochastic process $\{X(t), t \in \mathbb{R}\}$ is self-similar with scaling exponent H if, for any $\lambda \in \mathbb{R}^+$,

$$X(\lambda t) \stackrel{d}{=} \lambda^H X(t), \quad (3)$$

where $\stackrel{d}{=}$ denotes equality of all joint finite-dimensional distributions, throughout this paper. For a fixed level j , it can be shown that under \mathcal{L}_2 normalization,

$$d_{jk} \stackrel{d}{=} 2^{-j(H+1/2)} d_{0,k}.$$

If, in addition, $X(t)$ has stationary increments, then $E(d_{0k}) = 0$ and $E(d_{0k}^2) = E(d_{00}^2)$. Therefore,

$$E(d_{jk}^2) \propto 2^{-j(2H+1)}. \quad (4)$$

By taking logarithms on both sides of (4), we obtain the basis for estimating H , the wavelet spectrum, which is defined as

$$S(j) = \log_2 (E d_{jk}^2) = -(2H + 1)j + C. \quad (5)$$

For a more rigorous description of the wavelet spectra in one dimension, we refer the reader to [12].

Nicolis et al. [6] generalized the definition of traditional wavelet spectra to two-dimensions, which is presented after a brief definition of the fractional Brownian motion.

C. 2-D spectra

A popular standard model for images that scale is the *fractional Brownian motion* (fBm) in two dimensions. A 2-D fBm, $B_H(\mathbf{t})$, for $\mathbf{t} \in [0, 1] \times [0, 1]$ and $H \in (0, 1)$, is a random process with stationary zero-mean Gaussian increments, for which (3) reads $B_H(a\mathbf{t}) \stackrel{d}{=} a^H B_H(\mathbf{t})$.

In Nicolis et al.'s generalization of traditional wavelet spectra to 2-D, three different hierarchies $i = \{h, v, d\}$ (horizontal, vertical, and diagonal directions) constitute the detail spaces. The natural definition of the wavelet spectra then involves power spectrum corresponding to each of these three hierarchies. For the fBm in two dimensions, the expected value of the corresponding detail coefficients will verify that

$$E \left[\left| d_{j,\mathbf{k}}^i \right|^2 \right] = c_i 2^{-(2H+2)j}, \quad (6)$$

for some constant c_i depending on the wavelets ψ^i in (2), but not on the scale j . By taking logarithms on both sides of (6), we obtain the 2-D wavelet-based spectra

$$S^i(j) = \log_2 E \left[\left| d_{j,\mathbf{k}}^i \right|^2 \right] = -(2H + 2)j + C_i, \quad (7)$$

by which H for 2-D fBm is estimated.

While (5) and (7) give the basis for estimating H in 1-D and 2-D, respectively, specific methods for this estimation continue to be investigated and improved upon. This is the motivation behind the comparative analysis of scaling estimators presented in this paper. We now present analysis comparing the discriminatory power of six different wavelet-based scaling estimators, applied to the context of assessing the presence of cancer in mammogram images.

III. METHODS

A. Existing scaling estimators

1) *Ordinary least squares regression (OLS)*: Using ordinary least squares regression (OLS), directional Hurst exponents (diagonal (H_d), horizontal (H_h), and vertical (H_v)) can be estimated from the slopes of the linear equations in (7). The empirical counterpart to this is an OLS regression defined on pairs

$$\left(j, \log_2 \left| d_{j,\mathbf{k}}^i \right|^2 \right), \quad i = h, v, d. \quad (8)$$

where $\left| d_{j,\mathbf{k}}^i \right|^2$ is an empirical counterpart of $E \left[\left| d_{j,\mathbf{k}}^i \right|^2 \right]$. The slope of the regression would estimate H , i.e., $H = -(slope + 2)/2$. This method is in prevalent use for both estimation of H 's and classification by H 's, and as we will see later is suboptimal in the context of classification.

2) *Abry-Veitch weighted regression (AV)*: Veitch and Abry [12] improved the method shown in (III-A1) to a weighted linear regression, because the problem is heteroscedastic. Since the variances of the $\log_2 \left| d_{j,\mathbf{k}}^i \right|^2$ vary with j , this method weights each level by the inverse of the variance of that level. H 's are then estimated from the slopes of these weighted regressions. In addition, a bias correction term is introduced because of the passing with expectation through a logarithm.

B. Newly defined robust scaling estimators

A new, more robust class of Theil-type [13] estimators is introduced. Such estimators use a weighted average of all pairwise slopes, thus are robust with respect to possible outlier levels and free of any distributional assumptions. In our context, the slopes of the linear equations in (7) are assessed as a weighted average of all pairwise slopes, $s_{i,j}$ between levels i and j . The estimator of the overall slope then is $\sum_{i,j} w_{ij} s_{ij} / \sum_{i,j} w_{ij}$. The weights for each pair are designed to reduce the undue influence that outliers can have on estimates. Four new weights of this nature are investigated.

1) *Harmonic average weighted slopes (HA)*: Theoretically optimal weights for slope s_{ij} between levels i and j are proportional to

$$w_{ij} \propto (i - j)^2 \times HA(2^{2i}, 2^{2j}),$$

where HA is the harmonic average. Derivation of this weight is given in the appendix, but the rationale is that each pairwise slope is weighted by an inverse of variance of the estimated slope for that pair. We will denote this estimation method as HA.

2) *Level-enhanced ordinary least squares (EOLS)*: It is well known that the OLS estimator has previously been extended to a Theil-type representation with weights $w_{ij} \propto (i - j)^2$, e.g., [14]. After extensive simulation analysis we found that slopes based on the finer levels in wavelet decompositions are more critical for classification purposes. Thus we modified the weights to further favor the differentiating finer levels as

$$w_{ij} \propto ij(i - j)^2.$$

We will denote this estimation method as EOLS.

3) *Modified HA and EOLS*: Estimation of H and classification by H are two different tasks, and optimal estimators in one context may not perform well in the other. This is especially true for real-life images for which theoretical models are just an approximation. As previously stated, finer levels in wavelet decomposition are critical for correct classification. Utilizing the property that these levels also have more coefficients, we multiplied HA and EOLS weights by a product of the number of coefficients involved, $2^{(i+j)}$. The estimators with such weights emphasize heavily the fine detail levels and we denoted them as MHA and MEOLS.

IV. RESULTS

A. Description of the Data

The collection of digitized mammograms we analyzed was obtained from the University of South Florida's Digital Database for Screening Mammography (DDSM) [15]. The DDSM is described in detail in [16]. Images from this database containing suspicious areas are accompanied by pixel-level "ground truth" information relating locations of suspicious regions to what was assessed and verified through biopsy. We selected 105 normal cases (controls) from volumes normal-01, and 72 cancer cases from volumes cancer-01 and cancer-02. Each case study contains four mammograms (two for each breast: the craniocaudal (CC) and mediolateral oblique (MLO) projections) from a

screening exam. We considered only the CC projections, using the right breast image for all normal controls, and the cancerous breast (right or left) image for cancer cases. A subimage of size 1024×1024 was taken from each mammogram image for analysis. An example of an image and its subimage is provided in Fig. 1.

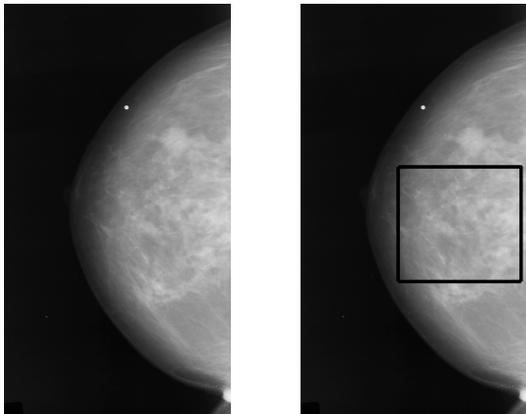


Fig. 1. *Left panel*: right CC mammogram corresponding to a cancer case. *Right panel*: subimage of size 1024×1024 considered for the analysis.

B. Classification of digitized mammograms

For every subimage, we performed the DWT using four different wavelet filters (Haar, Daubechies 4 tap, Symmlet 8 tap, Coiflet 6 tap), ensuring the filter choice did not favor any estimator over the others. Results are comparable except for the Haar wavelet for which the separation of cancer cases and controls was inadequate. We also tried three different level ranges: 2 to 6, 2 to 8, and 5 to 8. We provide here detailed results for the slopes that involve levels 5 to 8 using Daubechies 4, since this basis is the most local. After each transform, we used each of the six estimation methods to compute the estimated directional Hurst exponents (H_d , H_h , and H_v). All estimators are calculated after adjusting for AV bias [12]. Multiple classification methods were then used, to inform the tradeoffs between model simplicity versus power. Performance of each estimator was compared in terms of sensitivity, specificity, and overall misclassification rate.

The estimated density of H 's obtained from diagonal wavelet spectra alone (H_d) are shown in Fig. 2. The most parsimonious classifying model, without significant sacrifice in correct classification rates, includes only H_d as the predictor. Fig. 3 is the logistic regression curve (in red) fitted over scores $b_0 + b_1 H_d$. Green dots represent cancer cases at level 1, and controls at level 0.

Fig. 4 shows an ROC curve of H_d in differentiating between controls and cancer cases. The diagonal line represents a test with a sensitivity of 50% and a specificity of 50%. This shows the ROC curve lying significantly to the left of the diagonal, where the combination of sensitivity and specificity are highest. The area under the ROC curve, which is proportional to the diagnostic accuracy of the test, is 0.8820. Table I summarizes the results of the classification based only on H_d , for each estimation method. In the first

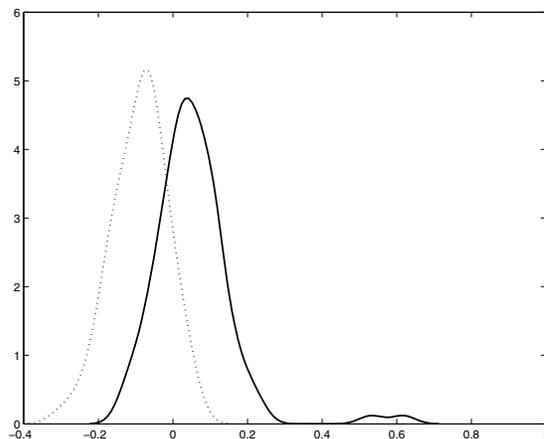


Fig. 2. Estimated density of H_d obtained from 105 controls (*solid line*) and 72 cancer cases (*dotted line*). The estimated H 's are empirical and flat spectra can cause H to be negative.

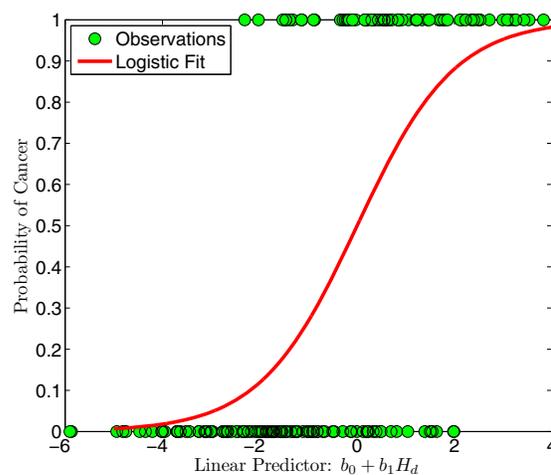


Fig. 3. Logistic regression: $\text{logit}(p) = -0.8927 - 22.7722 \cdot H_d$, where H_d is the Abry-Veitch estimator.

two columns, we provide the area under the ROC curve (AUC) and H_d^* , which is the value for H_d on the ROC curve where the maximum distance from the diagonal is achieved (i.e., maximum Youden index). The last three columns provide $1 - \text{Sensitivity}$, $1 - \text{Specificity}$, and Error (misclassification) rate achieved at the threshold H_d^* .

TABLE I
RESULTS OF CLASSIFICATION BY LOGISTIC REGRESSION USING H_d .

Method	AUC	H_d^*	1-Se	1-Sp	Error
OLS	0.5907	0.0212	0.6250	0.2286	0.3898
AV	0.8820	-0.0240	0.1528	0.2095	0.1864
HA	0.8073	-0.0005	0.2361	0.2667	0.2542
MHA	0.9321	-0.0624	0.1806	0.0952	0.1299
EOLS	0.6251	0.0093	0.5833	0.2191	0.3672
MEOLS	0.7783	-0.0093	0.3472	0.2286	0.2768

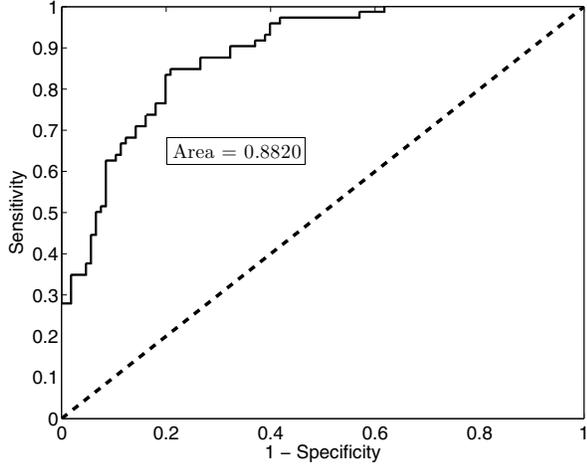


Fig. 4. ROC curve for the logistic regression: $\text{logit}(p) = -0.8927 - 22.7722 \cdot H_d$, where the most distant point from the diagonal (Youden index) is achieved at $H_d^* = -0.0240$ for which Sensitivity was 84.7% and Specificity 79%.

Both linear and quadratic classification methods were then implemented using pairs (H_d, H_h) . The classifier based on (H_d, H_v) is comparable in performance. But the remaining combination (H_h, H_v) turns out to be suboptimal and we dropped it from consideration. We randomly selected 66% of the data as a training set to fit the classifier and used the remaining 34% of the data to test performance. The random selection of training and testing data was repeated 10,000 times, so the reported prediction errors are averaged over 10,000 runs.

Fig. 5 shows a scatter plot by H_h versus H_d , illustrating the differentiation between controls and cancer cases. Table II summarizes the results of linear and quadratic classifications based on pairs (H_d, H_h) , for each estimation method. Results are given in terms of $1 - \text{Sensitivity}$, $1 - \text{Specificity}$, and Error (misclassification) rate. The performance of linear and quadratic classifiers was comparable, and both were slightly better than classification by H_d alone.

Finally, to assess the value of using all three directional Hurst exponents $(H_d, H_h, \text{ and } H_v)$, we also conducted a classification analysis involving Support Vector Machines (SVM) and several modeling scenarios (three exponents, as well as nine exponents obtained for different choices of levels). The classification rates improve, but not significantly. The relative performance of the six estimation methods was consistent regardless of the classification method used.

V. DISCUSSION AND CONCLUSIONS

In this paper we presented a comparative investigation into the diagnostic performance of six wavelet-based estimators of scaling. The HA and EOLS are newly defined robust estimators, while their modifications MHA and MEOLS are motivated by the specific application of diagnostic mammography. We found that MHA, MEOLS, and Abyr-Veitch estimators provided the best classification rates, for a range of wavelets and level choices. The standard wavelet-based

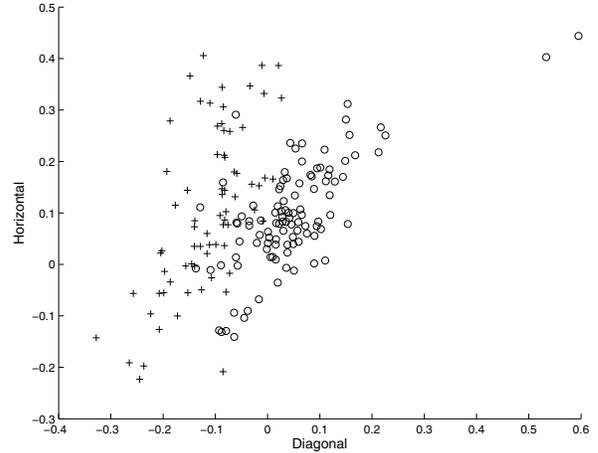


Fig. 5. Scatter plot of H_h versus H_d . Circles denote controls, and crosses denote cancer cases.

TABLE II
RESULTS OF LINEAR AND QUADRATIC CLASSIFICATION BASED ON PAIR (H_d, H_h) .

	Method	1-Se	1- Sp	Error
Linear	OLS	0.3301	0.3430	0.3377
	AV	0.1152	0.1114	0.1130
	HA	0.1496	0.1802	0.1678
	MHA	0.0910	0.0913	0.0912
	EOLS	0.2931	0.3200	0.3090
	MEOLS	0.1658	0.2014	0.1869
Quadratic	OLS	0.2477	0.4144	0.3466
	AV	0.1282	0.1216	0.1243
	HA	0.1511	0.1982	0.1791
	MHA	0.1042	0.0853	0.0930
	EOLS	0.2310	0.3799	0.3194
	MEOLS	0.1590	0.2296	0.2009

OLS estimator did not perform well and our recommendation is that this estimator should not be used in tasks of classification. The overall misclassification rate of the new weights proposed in this paper (HA, MHA, EOLS and MEOLS) was lower than the ordinary least squares estimate in all settings. It should be noted that these are not necessarily global phenomena, rather specific observations in mammography image classification.

Diagonal spectra (H_d) was found to be the most discriminatory and little power is lost if only this spectra is used. But, although H_d itself is strongly discriminatory and the most parsimonious classifying model, the use of H_d in combination with H_h (or H_v) does perform better than H_d alone. Further, the results of SVM classification using all three spectra (H_d, H_h, H_v) did perform slightly better than the linear or quadratic classifications. This implies that each wavelet spectra has some level of power to differentiate between normal and malignant cases.

A meaningful implication of this research is the improve-

ment of both sensitivity and specificity of current clinical diagnostic tests for breast cancer. The ambiguities involved in current diagnostic methods often result in extra costs, painful additional procedures, or missed cancers. With this tool, reasonable misclassification errors are achieved, and a promising new indicator may be added to current screening techniques.

APPENDIX: DERIVATION OF HA METHOD

In this appendix we provide a skeleton of the derivation of HA. For the detailed derivation see <http://www.prism.gatech.edu/~gte925z/papers/BIBM2011Appendix.pdf>.

Let $d_j = d_{j\mathbf{k}}$ be an arbitrary (wrt \mathbf{k}) wavelet coefficient from the j th level of the decomposition of the m -D fBm $B_H(\omega, \mathbf{t})$, $\mathbf{t} \in \mathbb{R}^m$,

$$d_j = \int_{\mathbb{R}^m} B_H(\omega, \mathbf{t}) \psi_{j\mathbf{k}}^*(\mathbf{t}) d\mathbf{t},$$

for some fixed $\mathbf{k} = (k_1, \dots, k_m)$. Here $\psi_{j\mathbf{k}}^*(\mathbf{t}) = \prod_{i=1}^m \psi_{jk_i}^*(t_i)$ where ψ^* is either ψ or ϕ , but in the product there is at least one ψ . It is well known that

$$d_j \stackrel{d}{=} 2^{-(H+m/2)j} d_0,$$

where d_0 is a coefficient from the level $j = 0$, and $\stackrel{d}{=}$ means equality in distributions [6]. Coefficient d_j is a random variable with

$$\mathbb{E} d_j = 0, \quad \text{Var} d_j = \mathbb{E} d_j^2 = 2^{-(2H+m)j} \sigma^2,$$

where $\sigma^2 = \text{Var} d_0$.

The rescaled “energy”

$$\frac{2^{(2H+m)j}}{\sigma^2} d_j^2 \sim \chi_1^2,$$

while assuming the independence of $d_{j\mathbf{k}}$ ’s (as is standard),

$$\frac{2^{(2H+m)j}}{\sigma^2} \sum_{\mathbf{k} \in j\text{th level}} d_{j\mathbf{k}}^2 = \frac{2^{(2H+2m)j}}{\sigma^2} \overline{d_j^2}$$

has $\chi_{2^m}^2$ distribution. Here, $\overline{d_j^2}$ is the average energy in the j th level.

Thus,

$$\overline{d_j^2} \stackrel{d}{=} 2^{-(2H+2m)j} \sigma^2 \chi_{2^m}^2.$$

From this,

$$\mathbb{E} \overline{d_j^2} = \sigma^2 2^{-(2H+2m)j} \mathbb{E} \chi_{2^m}^2 = 2^{-(2H+m)j} \sigma^2$$

and

$$\text{Var} \overline{d_j^2} = \sigma^4 2^{-(4H+4m)j} \times 2 \cdot 2^m = 2^{-4Hj-3mj+1} \sigma^4.$$

Recall that if X has $\mathbb{E} X$ and $\text{Var} X$ finite and φ is a function with finite second derivative at $\mathbb{E} X$, then

$$\mathbb{E} \varphi(X) \approx \varphi(\mathbb{E} X) + \frac{1}{2} \varphi''(\mathbb{E} X) \cdot \text{Var} X.$$

and

$$\text{Var} \varphi(X) \approx (\varphi'(\mathbb{E} X))^2 \text{Var} X.$$

When φ is logarithm for base 2, then

$$\mathbb{E} \log_2 \overline{d_j^2} = -(2H+m)j - \frac{1}{2^m j \log 2} + \log_2 \sigma^2.$$

Note that $-\frac{1}{2^m j \log 2}$ is the Abry-Veitch bias term, and it is free of H and σ^2 . This bias is a second order approximation. Also,

$$\text{Var} \log_2 \overline{d_j^2} = \frac{2}{2^m j (\log 2)^2}.$$

Finally,

$$\text{Var} \left(\frac{\log_2 \overline{d_j^2} - \log_2 \overline{d_i^2}}{j-i} \right) = \frac{2}{(\log 2)^2} \cdot \frac{1/2^{mj} + 1/2^{mi}}{(j-i)^2}.$$

Since weights w_{ij} are inverse-proportional to the variance, then

$$w_{ij} \propto (i-j)^2 \times HA(2^{mi}, 2^{mj}).$$

where HA is the harmonic average.

REFERENCES

- [1] S. Altekruse, C. Kosary, M. Krapcho, N. Neyman, R. Aminou, and W. Waldron, “Seer cancer statistics review, 1975-2007,” Bethesda, MD, 2010.
- [2] J. Martin, M. Moskowitz, and J. Milbrath, “Breast cancer missed by mammography,” *American Journal of Roentgenology*, vol. 37, no. 2, pp. 142–162.
- [3] T. Wang and N. Karayiannis, “Detection of microcalcifications in digital mammograms using wavelets,” *IEEE Transactions on medical imaging*, vol. 17, no. 4, pp. 498–509, 1998.
- [4] T. Netsch and H. Peitgen, “Scale-space signatures for the detection of clustered microcalcifications in digital mammograms,” *IEEE Transactions on medical imaging*, vol. 18, no. 9, pp. 774–786, 1999.
- [5] P. Kestener, J. Lina, P. Saint-Jean, and A. Arneodo, “Wavelet-based multifractal formalism to assist in diagnosis in digitized mammograms,” *Image analysis and stereology*, vol. 20, no. 3, pp. 169–175, 2001.
- [6] O. Nocolis, P. Ramirez-Cobo, and B. Vidakovic, “2d wavelet-based spectra with applications,” *Computational Statistics & Data Analysis*, vol. 55, no. 1, pp. 738–751, 2011.
- [7] I. A. Eckley, G. P. Nason, and R. L. Treloar, “Locally stationary wavelet fields with application to the modelling and analysis of image texture,” *Journal of the Royal Statistical Society Series C-Applied Statistics*, vol. 59, pp. 595–616, 2010.
- [8] H. Wendt, S. G. Rouxl, S. Jaffard, and P. Abry, “Wavelet leaders and bootstrap for multifractal analysis of images,” *Signal Processing*, vol. 89, no. 6, pp. 1100–1114, 2009.
- [9] I. Daubechies, *Ten lectures on wavelets*, ser. CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.
- [10] S. G. Mallat, *A wavelet tour of signal processing*. San Diego, CA: Academic Press, 1998.
- [11] B. Vidakovic, *Statistical Modeling by Wavelets*. New York: John Wiley & Sons, 1999.
- [12] D. Veitch and P. Abry, “A wavelet-based joint estimator of the parameters of long-range dependence,” *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 878–897, 1999.
- [13] H. Theil, “A rank-invariant method of linear and polynomial regression analysis,” *Indagationes Mathematicae*, vol. 12, pp. 85–91, 1950.
- [14] J. Nevitt and H. P. Tam, “A comparison of robust and nonparametric estimators under the simple linear regression model,” *Multiple Linear Regression Viewpoints*, vol. 25, pp. 54–69, 1998.
- [15] “Digital database for screening mammography,” University of South Florida. [Online]. Available: <http://marathon.csee.usf.edu/Mammography/Database.html>
- [16] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, *The Digital Database for Screening Mammography*, ser. 5th International Workshop on Digital Mammography, Toronto, Canada. Madison, WI: Medical Physics Publishing, 2000.