Application of Parametric Statistical Weights in CAD Imaging Systems

Michael Galperin[*a]
[a]Almen Laboratories, Inc., 1672 Gil Way, Vista, CA 92084

Abstract

PURPOSE: To propose a method for Parametric Statistical Weights (PSW) estimations and analyze its statistical impact in Computer-Aided Diagnosis Imaging Systems based on a Relative Similarity (CADIRS) classification approach.
MATERIALS AND METHODS: A Multifactor statistical method was developed and applied for Parametric Statistical Weights calculations in CADIRS. The implemented PSW method was used for statistical estimations of PSW impact when applied to a clinically validated breast ultrasound digital database of 332 patients' cases with biopsy proven findings. The method is based on the assumption that each parameter used in Relative Similarity (RS) classifier contributes to the deviation of the diagnostic prediction proportionally to the normalized value of its coefficient of multiple regression. The calculated by CADIRS Relative Similarity values with and without PSW were statistically estimated, compared and analyzed (on subset of cases) using classic Receiver Operator Characteristic (ROC) analysis methods.
RESULTS: When CADIRS classification scheme was augmented with PSW the Relative Similarity the calculated values were 2-5% higher in average. Numeric estimations of PSW allowed decomposition of statistical significance for each component (factor) and its impact on similarity to the diagnostic results (biopsy proven).
CONCLUSION: Parametric Statistical Weights in Computer-Aided Diagnosis Imaging Systems based on a Relative Similarity classification approach can be successfully applied in an effort to enhance overall classification (including scoring) outcomes. For the analyzed cohort of 332 cases the application of PSW increased Relative Similarity to the retrieved templates with known findings by 2-5% in average.

Keywords: breast ultrasound, cancer, image analysis, segmentation, level of suspicion scoring, diagnostic imaging, ROC, computer-aided, statistical analysis, similarity

1. Introduction

During the last four years of cooperation, Almen Laboratories, with the UCSD School of Medicine and San Diego VA Healthcare System have developed a sophisticated protocol for some clinical imaging based applications. The protocol is based on a software imaging and retrieval software system that provides extensive tools to quantify and extract medically defined objects and their features of interest and analyze the information content. The software developed by Almen Laboratories can store, retrieve and compare different objects and images based on this information for classification purposes including Level of Suspicion (LOS) scoring. A Relative Similarity[8-10] classifier was developed, tested and clinically validated with breast ultrasound imaging data set of 332 Institutional Review Board (IRB) approved cases. The developed system was used in more than 20 biomedical studies and applications and is based on patented image retrieval by the content methods[1]. The in-depth software application was tested and validated in clinical environments for the specific problem of breast ultrasound LOS for cancer by computer-aided scoring (under control of practitioner) and thereby reduction of unnecessary biopsies. This initial effort was aimed at developing tools that accurately (target - 95% and above) identify and confirm masses with lower levels of suspicion, rather than increasing the accuracy of diagnosis of cancers, which requires a biopsy anyway.

The overall breast ultrasound CAD Almen Labs/VA study hypothesis was that the accuracy of decision making of breast ultrasound can be significantly improved by following implementation of a structured method for breast lesions. Such method for description and reading interpretation can be achieved through application of a computer-aided imaging system based on Breast Image Reporting and Documentation System (BIRADS) with FDA approved and accepted lexicon guidelines. The protocol also provides a score for Level of Suspicion (LOS) for cancer using these same guidelines. Our approach applies a database of verified known findings to which an unknown may be compared and evaluated. Storage and retrieval of these images is accomplished through identification of key information content of the breast masses themselves ("case-based reasoning"). We suggested that lesions of lower suspicion level such as complex

cystic masses may be ruled out as candidates for biopsy with a higher degree of confidence when the image interpretation is made with support of this computer-aided imaging system. The ROC analysis was used and a "cut-off values" standard method applied when the cut-off values are tested against their corresponding true positive and false positive rates. Ground truth in this study is the surgical or needle biopsy results or two-year follow up in the case of benign findings. The initial breast cancer study was conducted and validated with PSW being equal for all selected parameters in optimized set [6,8]. In the study diagnostic breast ultrasound (US) image files for 332 women were retrieved chronologically (not randomly) from the image library of one of our hospitals under institutional review board approval (VA, UCSD Thornton Hospital). ROC analysis [10,18,20] was used to validate accuracy and impact on suspicion scoring of the implemented approach and computer-aided methodology. CADIRS approach was validated to be successful in its goal of aiding the reduction of biopsies of benign masses and was ready for further enhancement of its RS classifier through applying PSW during the classification process. Sensitivity achieved by CADIRS on the full set of 332 patients and without PSW used is 87.8% with Specificity 97.3%, Positive Predictive Value of 90.3% and Negative Predictive Value of 96.5% with CAD Efficiency 95.2%. In order to increase even further the Sensitivity and Specificity of the developed CAD system we came with the hypothesis that the Specificity and accuracy of the outcomes produced by CAD can be improved by using statistical weights derived from quantifications of lesion templates with known findings.

The goal of this report is limited to a small sub-study with the goal to propose a method for Parametric Statistical Weights (PSW) estimations and analyze its statistical impact on Relative Similarity numeric estimation in Computer-Aided Diagnosis Imaging Systems.

2. Materials and Methods

Data: IRB approved 332 breast ultrasound cases were acquired as direct digital files from the PACS archives of the participating institutions and made anonymous for development, testing and training of the system. Data will be collected in chronological order for patients who underwent biopsy or have a two-year negative follow up. Images are acquired from the same type of ultrasound scanner at each institution to minimize machine-dependent variables. Routine quality control was performed on all systems including the US workstation monitors. Three subsets of cases were defined by known findings (cystic, solid benign, solid malignant), which were further divided randomly into two groups, a development set and test set of approximately 160 cases each. The lead radiologist selected a minimum of two images for each mass and potentially more than one mass per patient. Various ultrasound image artifacts are also an important source of potential error for an automated scoring process. The system will be used with the direct participation of a human observer (radiologist), who may readily ignore interference of image artifacts on the segmentation algorithms. As is the current clinical practice, images with large interfering artifacts likely were not analyzed.

Segmentation: Segmentation was accomplished by several alternating methods (including multi-level pixel thresholding, region growing and radial gradient edge detection)[2-6] with criterion to yield the highest accuracy results. The accuracy of segmentation was evaluated using the now standard method of False Positive/False Negative (FP, FN) and True Positive/True Negative (TP, TN) pixel summary calculation[6,7]. Negative Predictive Value (NPV) and Positive Predictive Value (PPV) were also estimated[6,7].

Feature Selection: The categories of features[6,8] (a few parameters in each) listed in Table 1 were developed to correspond to the image criteria of the ultrasound reporting lexicons known as BIRADS[11,12].

Table 1. Default classes of image parameters for lesion classification

Image Criteria (Qualitative Categories)	**Sample of Associated Categories of Parameters**
Spherical/ovoid vs. irregular shape	Formfactor, Equivalent circular diameter/Form factor Perimeter/Area, Perimeter/Equivalent circular diameter, Aspect ratio
Linear margin vs. poorly defined margin	Edge Gradient
Homogeneous texture vs. internal echoes	Homogeneity (multiple texture parameters)
Isoechoic/anechoic vs. echoic	Relief (Contrast), Optical Density, Integrated Density
Calcifications	Scatterer density, scatterer size, 2^{nd}, 3^{rd}, 4^{th} moments of inertia
Edge shadowing vs. Central shadowing Distal enhancement	Density measures of a Distal ROI defined by X- and Y-Ferret coordinates
Parallel to skin vs. irregular	X-Ferret/Y-Ferret, Aspect ratio, Relative angle

Several texture features including 1) those measured from the pixel histogram of a defined region of interest, considered first order statistics, 2) second order parameters that involve spatial distribution and relationships features of co-occurrence matrices, 3) probability distributions including measures of the angular second and third moments, sum and difference entropy, sum and difference variance, correlation, contrast, etc. [2-5,13-16]. At the initial phase of development a large number of parameters was included in order to determine which have a promising degree of association with mass characterization, particularly for breast ultrasound - those lesions with a lower LOS score[9,10,12,14]. Overall about 30 imaging parameters of the 332 patients cases were calculated [6,8-10] and prepared for correlation and regression analyses. A Statistical Software System *Data Companion*® developed by Almen Laboratories, Inc. was used for computer calculations.

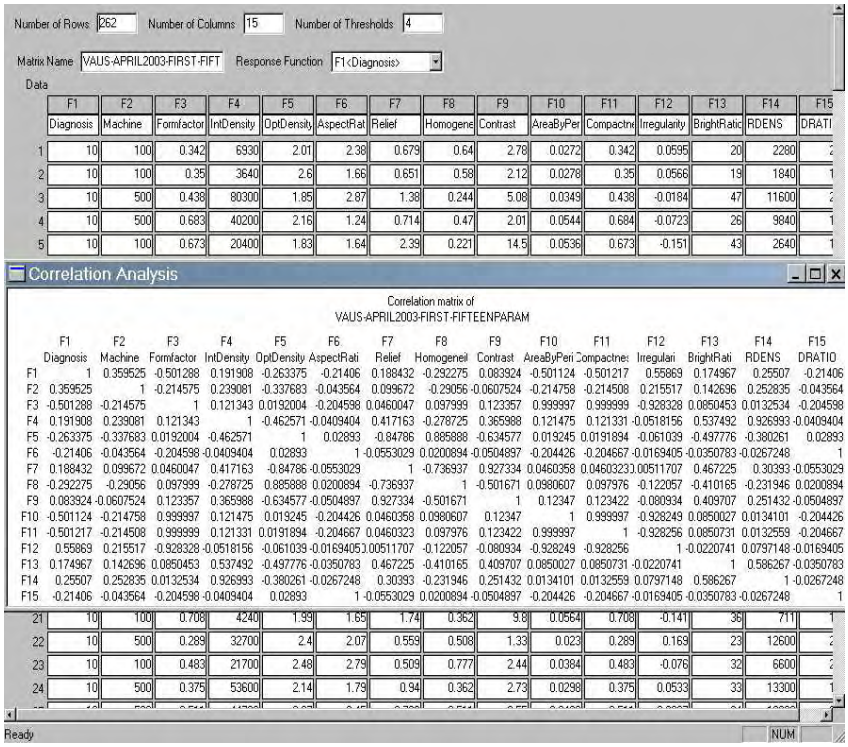
Optimization and reduction of the Parameters Set: The original 30 parameters were reduced to 15 and their colinearity was checked with correlation analysis. We used standard multivariate statistical approaches for decomposing a correlation matrix into linear combinations of variables[16]. The linear combinations were chosen so that the first combination has the largest possible variance, the second combination has the next largest variance, subject to being uncorrelated with the first, the third has the largest possible variance, subject to being uncorrelated with the first and second, and so forth[16]. The Condition Index (CI) we used is a simple function of the eigenvalues, namely,

$$CI_i = \sqrt{\frac{\lambda_{\max}}{\lambda_i}} \ , \qquad (1)$$

where λ is the conventional symbol for an eigenvalue.

Then informative parameters were identified and comprehensive subsets were defined using our Multiple Determination weighting method[13]. In this method parameters that contribute most to deviation from the response factor (which is the diagnostic finding in our case) are assigned the highest statistical scores by the computer. The weight used for a given parameter in the comparison process may thus be derived from the values of the parameter vectors associated with the detected objects in the image database. In using this method a system is represented as a totality of factors.

When reduced to 15 parameters (Figure 1) we used the same method to determine the final subset of significant parameters.


Number of Rows 262 Number of Columns 15 Number of Thresholds 4

Matrix Name VAUS-APRIL2003-FIRST-FIFT Response Function F1<Diagnosis>

Data

	F1 Diagnosis	F2 Machine	F3 Formfactor	F4 IntDensity	F5 OptDensity	F6 AspectRati	F7 Relief	F8 Homogenei	F9 Contrast	F10 AreaByPeri	F11 Compactne	F12 Irregulari	F13 BrightRati	F14 RDENS	F15 DRATI
1	10	100	0.342	6930	2.01	2.38	0.679	0.64	2.78	0.0272	0.342	0.0595	20	2280	2
2	10	100	0.35	3640	2.6	1.66	0.651	0.58	2.12	0.0278	0.35	0.0566	19	1840	1
3	10	500	0.438	80300	1.85	2.87	1.38	0.244	5.08	0.0349	0.438	-0.0184	47	11600	2
4	10	500	0.683	40200	2.16	1.24	0.714	0.47	2.01	0.0544	0.684	-0.0723	26	9840	1
5	10	100	0.673	20400	1.83	1.64	2.39	0.221	14.5	0.0536	0.673	-0.151	43	2640	1

Correlation Analysis

Correlation matrix of
VAUS-APRIL2003-FIRST-FIFTEENPARAM

	F1 Diagnosis	F2 Machine	F3 Formfactor	F4 IntDensity	F5 OptDensity	F6 AspectRati	F7 Relief	F8 Homogeneit	F9 Contrast	F10 AreaByPeri	F11 Compactnes	F12 Irregulari	F13 BrightRati	F14 RDENS	F15 DRATIO
F1	1	0.359525	-0.501288	0.191908	-0.263375	-0.21406	0.188432	-0.292275	0.083924	-0.501124	-0.501217	0.55869	0.174967	0.25507	-0.21406
F2	0.359525	1	-0.214575	0.239081	-0.337683	-0.043564	0.099672	-0.29056	-0.0607524	-0.214758	-0.214508	0.215517	0.142696	0.252835	-0.043564
F3	-0.501288	-0.214575	1	0.121343	0.0192004	-0.204598	0.0460047	0.097999	0.123357	0.999997	0.999999	-0.928328	0.0850453	0.0132534	-0.204598
F4	0.191908	0.239081	0.121343	1	-0.462571	-0.0409404	0.417163	-0.278725	0.365988	0.121475	0.121331	-0.0518156	0.537492	0.926993	-0.0409404
F5	-0.263375	-0.337683	0.0192004	-0.462571	1	0.02893	-0.84786	0.885888	-0.634577	0.019245	0.0191894	-0.061039	-0.497776	-0.380261	0.02893
F6	-0.21406	-0.043564	-0.204598	-0.0409404	0.02893	1	-0.0553029	0.0200894	-0.0504897	-0.204426	-0.204667	-0.0169405	-0.0350783	-0.0267248	1
F7	0.188432	0.099672	0.0460047	0.417163	-0.84786	-0.0553029	1	-0.736937	0.927334	0.0460358	0.0460323	0.00511707	0.467225	0.30393	-0.0553029
F8	-0.292275	-0.29056	0.097999	-0.278725	0.885888	0.0200894	-0.736937	1	-0.501671	0.0980607	0.097976	-0.122057	-0.410165	-0.231946	0.0200894
F9	0.083924	-0.0607524	0.123357	0.365988	-0.634577	-0.0504897	0.927334	-0.501671	1	0.12347	0.123422	-0.080934	0.409707	0.251432	-0.0504897
F10	-0.501124	-0.214758	0.999997	0.121475	0.019245	-0.204426	0.0460358	0.0980607	0.12347	1	0.999997	-0.928249	0.0850027	0.0134101	-0.204426
F11	-0.501217	-0.214508	0.999999	0.121331	0.0191894	-0.204667	0.0460323	0.097976	0.123422	0.999997	1	-0.928256	0.0850731	0.0132559	-0.204667
F12	0.55869	0.215517	-0.928328	-0.0518156	-0.061039	-0.0169405	0.00511707	-0.122057	-0.080934	-0.928249	-0.928256	1	-0.0220741	0.0797148	-0.0169405
F13	0.174967	0.142696	0.0850453	0.537492	-0.497776	-0.0350783	0.467225	-0.410165	0.409707	0.0850027	0.0850731	-0.0220741	1	0.586267	-0.0350783
F14	0.25507	0.252835	0.0132534	0.926993	-0.380261	-0.0267248	0.30393	-0.231946	0.251432	0.0134101	0.0132559	0.0797148	0.586267	1	-0.0267248
F15	-0.21406	-0.043564	-0.204598	-0.0409404	0.02893	1	-0.0553029	0.0200894	-0.0504897	-0.204426	-0.204667	-0.0169405	-0.0350783	-0.0267248	1

21	10	100	0.708	4240	1.99	1.65	1.74	0.362	9.8	0.0564	0.708	-0.141	36	711	1
22	10	500	0.289	32700	2.4	2.07	0.559	0.508	1.33	0.023	0.289	0.169	23	12600	2
23	10	100	0.483	21700	2.48	2.79	0.509	0.777	2.44	0.0384	0.483	-0.076	32	6600	2
24	10	500	0.375	53600	2.14	1.79	0.94	0.362	2.73	0.0298	0.375	0.0533	33	13300	1

Ready NUM

Figure 1. Statistical analysis for the first round of elimination. Original 30 parameters are reduced to 15 and their colinearity was checked with correlation analysis. We used standard multivariate statistical approaches for decomposing correlation matrix into linear combinations of variables. The linear combinations are chosen so that the first combination has the largest possible variance, the second combination has the next largest variance, subject to being uncorrelated with the first, the third has the largest possible variance, subject to being uncorrelated with the first and second, and so forth.

Statistical method for Parametric Weights Calculations: The statistical experiment simulation tools are correlation, regression, and multifactor analyses, where the coefficients of pairwise and Multiple Correlation coefficient (CMC) are computed and linear and non-linear regressions are obtained[13,16]. The data for a specific model experiment are represented as a matrix whose columns stand for factors describing the system and the rows for the experiments (values of these factors). The first factor (response function, in our case – biopsy finding), for which the regression is obtained, is referred to as the system response. The coefficients of the regression equation and the covariances help to "redistribute" the Multiple Determination coefficient (CMD) among the factors; in other words the "impact" of every factor to response variations is determined. The specific impact indicator of the factor is the fraction to which a response depending on a totality of factors in the model changes due to this factor. This specific impact indicator may then be used as the appropriate weight estimation to assign to that factor (i.e. parameter of the parameter set associated with the objects). The impact of a specific factor is described by a specific impact indicator which is computed by the following algorithm:

$$\gamma_j = \alpha * [\, b_j * c_{0j} \,], \qquad j = 1,2,..., k \ , \qquad (2)$$

where γ is the specific impact indicator of the j-th factor; k is the number of factors studied simultaneously; b_j is the j-th multiple regression coefficient; c_{0j} – covariance coefficient, and α - is fraction of Multiple Determination related to the impact of the factor and can be computed as:

$$\alpha = \Theta^2 / \left[\sum \, | b_j * c_{0j} | \, \right] , \qquad (3)$$

where Θ is the coefficient of CMD. The specific contribution indicator is obtained mainly from the coefficient of CMD. The method[13] implies that the specific impact of the j-th factor on R depends only on the ratio of addends in the formula for final multiple regression and its derivatives. That also implies that the addend whose magnitude is the largest is associated with the largest specific impact. Since the regression coefficients may have different signs their magnitudes have to be taken in the total. For this reason coefficients γ_j of the specific impact are bound to be positive while their sign indicates the direction of the impact. The influence of background factors that were not included in the statistical experiment is computed by the formula:

$$\gamma_b = \quad 1 - \sum \gamma_j \qquad (4)$$

The statistical significance of each factor influence (through parametric representation in Relative Similarity) can be computed from Fisher criterion which can be defined by the formula:

$$F_j = \gamma_j * \left(n - k - 1 \right) / \gamma_b \ , \qquad (5)$$

where *n* – is total number tests (in our case total number of patients cases processed).

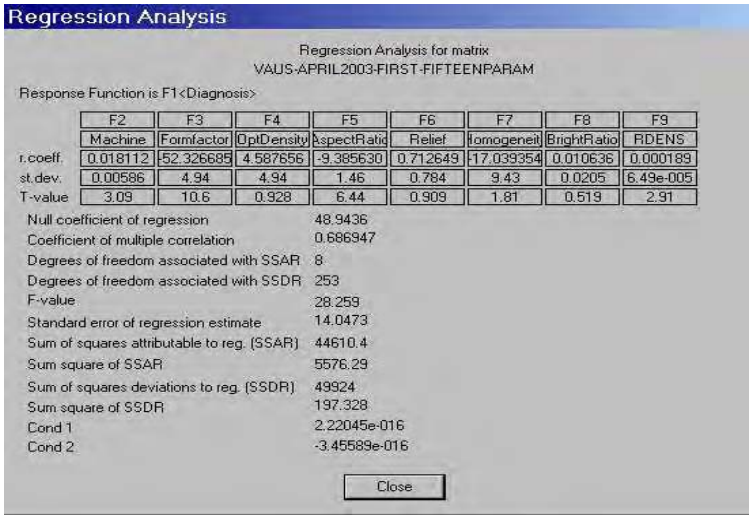


Figure 2. This figure illustrates one of the examples of the second step of elimination in the proposed method. After the original set of 30 parameters are reduced to 15 based on their colinearity, the further reduction is made based on combination of correlation and regression analyses. Multi-factor analysis was then based on subset of 9 significantly contributing parameters. The variance of each of these linear combinations is represented by *eigenvalue*[16].

Classification based on Relative Similarity: The classification method of the development was based on a Relative Similarity (RS) approach that included two stage procedures. The combinations of features to be derived from a parameter optimization method[9,10,13] may be represented by an **L**-dimensional vector **P** used to calculate the "Similarity," **R**, of one lesion to another [8-10,13,16,17].

$$R = \left(\sum_{k=1}^{L} (p_k^t - P_k^t)^s * \gamma_k \right)^{1/s} \quad (6)$$

It was our hypothesis that appropriate weighting factors or also known as Parametric Statistical Weights, γ, may be applied to these results to enhance the classification; and ***s*** – dimensionality of individual feature vector. A new case with an "unknown" finding is compared directly to the database of stored images and a measure of **R** is computed for different benign and malignant lesions. Similarity is calculated for a particular lesion $\mathbf{P}_{it}$ (the index of this "template" object) compared to the other lesions, $\mathbf{P}_k$ (k=1,…L) where **L** is the number of objects, although other measures of distance may be also tested. Then, during the second stage of the calculating procedures, the "Similarity," **R**, was normalized using three different methods and the method yielding the highest accuracy was implemented in CADIRS for RS estimations. The term "Relative Similarity" basically means that, i.e. in breast cancer study, the detected lesion is compared not to a hypothetical "golden template" of the disease but to the digital imaging database of previously analyzed patients and the imaging cases most "similar" to this suspicious mass that are automatically retrieved and displayed.

Estimation of PSW impact on RS numeric values using ROC: To answer a question if there is significant impact of PSW and if the answer is "yes" to numerically assess impact of PSW in RS estimations standard ROC analysis[18-20] was used. We used the following technique to prepare ROC input matrix. We assumed that "diagnostic" test is positive (value 1) if the use of PSW increased RS value in comparison to the same calculations without PSW for "lesion in questions" by at least 1%. If use of PSW decreased RS value or was less than 1% we set the outcome to 0 ("negative finding").

3. Results

Overall statistics: To illustrate in this report PSW impact on RS evaluation we randomly selected a small subset of 15 patient cases out of 296 used for the feature selection and parameter set optimization. The CMD achieved for the processed subset of images was 0.51 which means that 51% of deviation in calculated LOS to cancer can be attributed to mainly contributing factors. The correlation based colinearity threshold was determined by CADIRS as 0.8. The matrix of reduced data was "well statistically conditioned" with CI at a negligible level. When we tested threshold at 0.9 level the CMD increased to 66%. The regression was found statistically credible at F value equal to 28.3.

PSW estimations: We used formula (2) for the numeric estimations of PSW. The numeric estimations for the illustration subset are presented in Table 2.

Table 2. PSW (γ_j) numeric estimations (without normalization)

Formfactor	Optical Density	Aspect Ratio	Relief (Contrast)	Homogeneity (3d moments)	Brightness Ratio	Relative Integrated Density
-0.227	0.03	-0.057	0.014	-0.048	0.005	0.039

Assessment of impact on Relative Similarity: In the next phase of the study we compared RS numeric estimations with and without PSW. An example of the output for two example cases is illustrated by Table 3. In the table we present RS estimations for two patient cases. The "lesions in question" (LIN) were segmented, quantified and then retrospectively compared with a digital database of disease templates with known findings using RS method with and without PSW being present in formulae. The first two parts (with and without PSW) of Table 3 illustrate comparative results for a complex cyst and the second two parts – for a carcinoma.

Table 3. Example of PSW impact on the outcome of RS classification (numbers are rounded to the third digit after the decimal point). Numeric values of PSW presented in Table 2 were used to calculate Relative Similarity with and without PSW. Please note that use of PSW not only increased the value of calculated RS but also but also changed the retrieved composition of the closest matches from the digital database of templates with known findings.

Complex Cysts without PSW

Image ID with closest match to LIN	Similarity	Formfactor	Optical Density	Aspect Ratio	Relief (Contrast)	Homogeneity (3d moments)	Brightness Ratio	Relative Integrated Density
u359asf5	85.80%	0.482	0.823	1.8	6.26	0.002	211	0.965
u487asf3	85.50%	0.628	0.65	1.73	5.97	0.002	229	0.97
u384asf1	84.70%	0.607	0.715	1.52	6.82	0.002	196	0.97
u20asm1	82.60%	0.517	0.748	1.47	5.74	0.001	200	0.955
u453asf10	82.20%	0.686	0.714	1.87	6.22	0.001	208	0.973
u235bsf4	81.90%	0.722	0.691	1.59	6.88	0.000	235	0.97

Complex Cysts with PSW (for numeric values used see Table 2)

Image ID with closest match to LIN	Similarity	Formfactor	Optical Density	Aspect Ratio	Relief (Contrast)	Homogeneity (3d moments)	Brightness Ratio	Relative Integrated Density
u539asf16	87.30%	0.544	0.978	1.77	5.56	0.004	166	0.948
u587asm3	85.20%	0.544	0.769	2.04	6.58	0.002	149	0.938
u149ahf3	84.60%	0.573	0.712	2.66	8.91	0.002	178	0.921
u43asf3	84.40%	0.587	0.861	1.79	5.07	0.004	139	0.937
u384asf1	83.70%	0.607	0.715	1.52	6.82	0.002	196	0.97
u20asm1	83.00%	0.517	0.748	1.47	5.74	0.001	200	0.955

Carcinoma without PSW

Image ID with closest match to LIN	Similarity	Formfactor	Optical Density	Aspect Ratio	Relief (Contrast)	Homogeneity (3d moments)	Brightness Ratio	Relative Integrated Density
u384asf1	91.90%	0.607	0.715	1.52	6.82	0.002	196	0.97
u359asf5	91.80%	0.482	0.823	1.8	6.26	0.002	211	0.965
u20asm1	91.50%	0.517	0.748	1.47	5.74	0.001	200	0.955
u487asf3	89.50%	0.628	0.65	1.73	5.97	0.002	229	0.97
u154asm9	86.40%	0.675	0.783	1.52	6.72	0.001	187	0.964
u427bsm1	86.10%	0.621	0.934	1.58	5.39	0.001	190	0.963

Carcinoma with PSW (for numeric values used see Table 2)

Image ID with closest match to LIN	Similarity	Formfactor	Optical Density	Aspect Ratio	Relief (Contrast)	Homogeneity (3d moments)	Brightness Ratio	Relative Integrated Density
u20asm1	93.60%	0.517	0.748	1.47	5.74	0.001	200	0.955
u12asm6	89.10%	0.558	1.01	1.16	5.06	0.000	182	0.96
u359asf5	87.30%	0.482	0.823	1.8	6.26	0.001	211	0.965
u539asf16	87.20%	0.544	0.978	1.77	5.56	0.004	166	0.948
u427bsm3	86.40%	0.497	0.958	1.59	4.58	0.001	165	0.964
u587asm1	86.20%	0.482	0.757	1.28	5.73	0.001	155	0.968

For the illustration subset of 24 cases (144 "tests" - 24 "lesions in question" with 6 closest template matches based on RS values) we present partial ROC analysis. As mentioned above the "diagnostic" test value was set to 0 if no statistically significant increase (threshold 1%, that means if value of RS with PSW greater than value of RS without PSW by more than 1% the "diagnostic test" was considered "normal" (1) – otherwise "abnormal" (0)) in RS was observed and 1 when PSW did have statistical impact. The specific hypothesis was that PSW will increase RS calculated values. "No discrimination" meant that PSW had no impact on RS numeric estimations or even decreased the calculated values. The ROC results are illustrated on Figure 3 and Table 4.

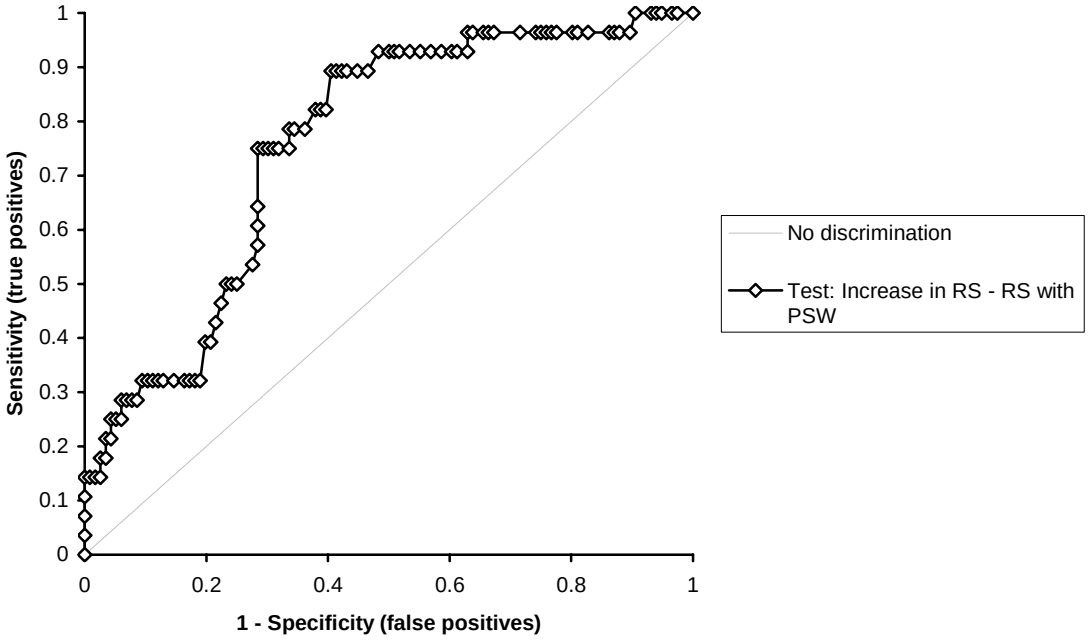


Figure 3. ROC curve (computer generated, no smoothing applied) for PSW impact of RS numeric values, illustration subset of 24 cases. Test is "abnormal" if PSW has not increased the values of the RS for the retrieved templates with known findings.

Table 4. Statistical data for ROC analysis of PSW impact on RS numeric values (based on limited subset of 24 sampled cases)

Curve	A_z	St. error	*p*	95% CI of A_z
Test: Increase in RS - RS with PSW	0.758	0.0464	<0.0001	0.667 to 0.849

Trend of impact on diagnostic scoring: The last step of this sub-study was to check the trend on potential diagnostic outcomes in the developed CADIRS. In order to see the trend we used the same 24 randomly selected cases with equal distribution of malignant and benign cases (such set can not be used for a true CAD study because the cancer occurrence[1] plays significant role in statistical power estimation of the validated clinical data set). When ROC was performed with known diagnostic findings present the accuracy of scoring with RS values calculated with PSW was slightly higher. While Sensitivity of both sets of scores was about the same the Specificity was 8% higher for the scoring set derived from RS values that used PSW numbers. A_z for the subset of 24 cases set with PSW "in play" was also higher – 0.71 vs 0.67 (5.63%)[2].

4. Discussion

In this research, the impact of Parametric Statistical Weights on Relative Similarity classifier in breast cancer CAD system was estimated and analyzed. Large patients' data set was used to statistically define best parameter set that maximized RS values and eliminated colinearity according to the preset threshold level. Statistical weight estimation method was implemented and computer experiment was completed to quantify the impact of PSW use. While overall accuracy of Computer-Aided Diagnosis Imaging system based on Relative Similarity classification approach applied to LOS scoring in breast ultrasound was not the purpose of this sub-study and was reported earlier[6,8,9,10], this research has confirmed a statistically significant impact of PSW on RS numeric estimations. It was determined that for the given data set such impact was in average interval of 2 to 5 percent in absolute RS values with 95% confidence level. Sensitivity of both sets of scores was about the same while the Specificity was 8% higher for the scoring set derived from RS values that used PSW numbers. Also the A_z test value was found to be significantly higher for RS calculations that used PSW. The importance of such improvement should be taken in light of future practical implementation of a breast ultrasound CAD system which will be dealing with thousands of templates with known findings not hundreds like in our development case. Additionally, the reported method allows not only identifying a numeric estimate of a parameter impact but also the direction of such impact. We also discovered that too steep of a threshold drop for a correlation coefficient (0.8) in colinearity elimination caused decline in value Multiple Determination coefficient and therefore more careful selection of parameters elimination thresholding is required. To compensate and increase drop in CMD independent imaging and non-imaging parameters should be introduced to the original data set so that the CMD will stay in 0.7-0.8 range for the final best parameter set evaluation after all eliminations. That can be achieved by using machine dependency, demographic and other patient's data (i.e. mammography results) in conjunction with imaging information extracted from traditional ultrasound imagery as well as volumetric slices rendering estimations.

5. Conclusions

Parametric Statistical Weights in Computer-Aided Diagnosis Imaging Systems based on a Relative Similarity classification approach can be successfully applied in an effort to enhance overall classification (including scoring) outcomes. For the analyzed cohort of 332 cases application of PSW increased Relative Similarity to the retrieved templates with known findings by 2-5% in average with Specificity 6-9% higher without a drop in Sensitivity. The study presented in this article should be taken as further proof of the effectiveness of in-depth computerized classic statistical analyses on the performance of the breast cancer CAD applications for diagnostic use. Enlarged prospective study should be designed and implemented in order to determine the overall dynamics of the optimized set of extracted lesion features and PSW derived from them. In the world of modern CAD that based on three prominent pillars – segmentation accuracy (mass's borders), feature extraction (tissue characterization) and classification (mass's categorization) – statistically

[1] Cancer occurrence rate changes from clinic to clinic but for breast ultrasound it is in average interval of 20-25%.

[2] As mentioned in the introduction overall Sensitivity achieved by CADIRS on the full set of 332 patients and without PSW used is 87.8% with Specificity 97.3%, Positive Predictive Value of 90.3% and Negative Predictive Value of 96.5%, CAD Efficiency 95.2%. These results are being reported separately.

significant improvements within each such component will lead their way to an overall improved usefulness of computer-aided applications in the clinical world and day-to-day practice. In our views Internet deployment of the validated CAD systems for on-line real time use by practitioners should consolidate somewhat dispersed but significant advancements of a few prominent CAD research and development groups.

6. Breakthrough work presented

The assessed CADIRS LOS scoring methods[6,8,9] open a clear path to implementation of a BIRAD™ rule-based lexicon for breast ultrasound findings reporting. The application is based on a CADIRS Relative Similarity and LOS score calculation for a digital database of breast lesions templates with known findings. The developed software and methods facilitated a high accuracy of LOS automated scoring and promised to yield results that will impact the existing clinical practice. The proposed and tested statistical PSW method increased estimated values of RS and also indicated some positive trend in impact on diagnostic scoring classification overall. For that additional prospective studies should be completed with an enlarged original data set.

7. Acknowledgment

The author thanks the members of the breast imaging group of VA and UCSD School of Medicine (Professor L. Olson, Professor M. O'Boyle, Dr. C. Comstock and Dr. K. Richman) for their review of the diagnostic data and help with patient's' digital database of templates with known findings. Our special appreciation and gratitude to Professor Michael P. Andre for his scientific guidance and productive exchange of ideas. We acknowledge Luis and Gustavus Pfeiffer Research Foundation support and commitment to breast ultrasound CAD research.

References

1. Galperin, M. (inventor), "Object based image retrieval", US PTO #09-370,366, PCT International patent # WO01/11489, 2000: 37
2. Parker, J.R. "Algorithms for image processing and computer vision," New York: John Wiley & Sons, 1997.
3. Russ, J, Ed., "The image processing handbook," New York: CBC Press, 1995.
4. Rosenfeld, A, Kak, AC., "Digital picture processing," New York: Academic Press, 1982.
5. Young, I.T.; Gerbrands, J.G.; van Vliet, L.J.. Fundamentals of Image Processing, Project 2987, the ASCI Post Doc program, and the Rolling Grants program of the Foundation for Fundamental Research in Matter (FOM), 1998.
6. André M.P., Galperin, M., Olson L.K., et al, "Preliminary investigation of a method to assess breast ultrasound level of suspicion," SPIE Medical Imaging 4322:507-512, 2001.
7. Stavros, A.T., et al., "Solid Breast Nodules: Use of Sonography to Distinguish between Benign and Malignant Lesions," Radiology 196, pp. 123-134, 1995.
8. André M.P., Galperin M., Olson L.K., et al.: "Improving the accuracy of diagnostic breast ultrasound," Acoustical Imaging 26, 2002. (In Press)
9. Galperin, M., et al., "Statistical Analysis to Assess Automated Level of Suspicion Scoring Methods in Breast Ultrasound", SPIE Medical Imaging 5034, 2003.
10. Galperin, M., et al., "ROC analysis of lesion descriptors in breast ultrasound images," SPIE Medical Imaging 5034, 2003.
11. American College of Radiology (ACR). "Breast Imaging Reporting and Data System (BI-RADS™)". Third Edition. Reston, VA: American College of Radiology; 1998.
12. J.A. Baker, P.J. Kornguth, M.S. Soo, et al., "Sonography of solid breast lesions: Observer variability of lesion description and assessment," Amer J Roentgenol 172, 1621-1625, 1999.
13. Galperin, M. "Simulation Modeling in Marine Ecology", 2nd ed., New York: John Wiley & Sons, 1990.
14. Kupinski, M.A. and Giger, M. L., "Feature selection with limited dataset," Med. Phys. 26(10), pp. 2176-2182, 1999.
15. R.L. Lieber, "Statistical significance and statistical power in hypothesis testing," J Orthopedic Res 8(2), 304-309, 1989.
16. Everitt, B., "Cluster Analysis", Second Edition, McGraw, NY, 1980
17. Vicario, E., " Image Description And Retrieval", Plenum Press, NY, 1998

18. Metz, C. (1978) "Basic principles of ROC analysis" -Seminars in Nuclear Medicine 8, 283-298.
19. Swets, J., "The relative operating characteristic in psychology" (1973) Science 182, 990-1000.
20. Rockette H.E, Li W., Brown M.L., Britton C.A., Towers J.T., Gur D., “Statistical test to assess rank-order imaging studies” Acad Radiol., Jan: 8(1):24-30, 2001

---

* mgalperin@almenlabs.com; phone: 1-760-806-0040; http://www.almenlabs.com