BMC Bioinformatics

METHODOLOGY ARTICLE

Open Access

Supervised Regularized Canonical Correlation Analysis: integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery

Abhishek Golugula[1], George Lee[2], Stephen R Master[3], Michael D Feldman[3], John E Tomaszewski[3], David W Speicher[4] and Anant Madabhushi[2*]

Abstract

Background: Multimodal data, especially imaging and non-imaging data, is being routinely acquired in the context of disease diagnostics; however, computational challenges have limited the ability to quantitatively integrate imaging and non-imaging data channels with different dimensionalities and scales. To the best of our knowledge relatively few attempts have been made to quantitatively fuse such data to construct classifiers and none have attempted to quantitatively combine histology (imaging) and proteomic (non-imaging) measurements for making diagnostic and prognostic predictions. The objective of this work is to create a common subspace to simultaneously accommodate both the imaging and non-imaging data (and hence data corresponding to different scales and dimensionalities), called a metaspace. This metaspace can be used to build a meta-classifier that produces better classification results than a classifier that is based on a single modality alone. Canonical Correlation Analysis (CCA) and Regularized CCA (RCCA) are statistical techniques that extract correlations between two modes of data to construct a homogeneous, uniform representation of heterogeneous data channels. In this paper, we present a novel modification to CCA and RCCA, Supervised Regularized Canonical Correlation Analysis (SRCCA), that (1) enables the quantitative integration of data from multiple modalities using a feature selection scheme, (2) is regularized, and (3) is computationally cheap. We leverage this SRCCA framework towards the fusion of proteomic and histologic image signatures for identifying prostate cancer patients at the risk of 5 year biochemical recurrence following radical prostatectomy.

Results: A cohort of 19 grade, stage matched prostate cancer patients, all of whom had radical prostatectomy, including 10 of whom had biochemical recurrence within 5 years of surgery and 9 of whom did not, were considered in this study. The aim was to construct a lower fused dimensional metaspace comprising both the histological and proteomic measurements obtained from the site of the dominant nodule on the surgical specimen. In conjunction with SRCCA, a random forest classifier was able to identify prostate cancer patients, who developed biochemical recurrence within 5 years, with a maximum classification accuracy of 93%.

Conclusions: The classifier performance in the SRCCA space was found to be statistically significantly higher compared to the fused data representations obtained, not only from CCA and RCCA, but also two other statistical techniques called Principal Component Analysis and Partial Least Squares Regression. These results suggest that SRCCA is a computationally efficient and a highly accurate scheme for representing multimodal (histologic and proteomic) data in a metaspace and that it could be used to construct fused biomarkers for predicting disease recurrence and prognosis.

* Correspondence: anantm@rci.rutgers.edu
[2]Department of Biomedical Engineering, Rutgers University, Piscataway, New Jersey, USA
Full list of author information is available at the end of the article

Background

With the plentitude of multi-scale, multi-modal, disease pertinent data being routinely acquired for diseases such as breast and prostate cancer, there is an emerging need for powerful data fusion (DF) methods to integrate the multiple orthogonal data streams for the purpose of building diagnostic and prognostic meta-classifiers for disease characterization [1]. Combining data derived from multiple sources has the potential to significantly increase classification performance relative to performance trained on any one modality alone [2]. A major limitation in constructing integrated meta-classifiers that can leverage imaging (histology, MRI) and non-imaging (proteomics, genomics) data streams is having to deal with data representations spread across different scales and dimensionalities [3].

For instance, consider two different data streams $F_A(x)$ and $F_B(x)$ describing the same object x. If $F_A(x)$ and $F_B(x)$ correspond to the same scale or resolution and also have the same dimensionality, then one can envision, concatenating the two data vectors into a single unified vector $[F_A(x), F_B(x)]$ which could then be used to train a classifier. However when $F_A(x)$ and $F_B(x)$ correspond to different scales, resolutions, and dimensionalities, it is not immediately obvious as to how one would go about combining the different types of measurements to build integrated classifiers to make predictions about the class label of x. For instance, directly aggregating data from very different sources without accounting for differences in the number of features and relative scaling, can not only lead to the *curse of dimensionality* (too many features and not enough corresponding samples [4]), but can lead to classifier bias towards the modality with more attributes. A possible solution is to first project the data streams into a space where the scale and dimensionality differences are removed; a *meta-space* allowing for a homogeneous, fused, multi-modal data representation.

DF methods try to overcome these obstacles by creating such a metaspace, on which a proper meta-classifier can be constructed. Methods leveraging embedding techniques have been proposed to try and fuse such heterogeneous data for the purpose of classification and prediction [2,3,5-7]. However, all of these DF techniques have their own weaknesses in creating an appropriate representation space that can simultaneously accommodate multiple imaging and non-imaging modalities. Generalized Embedding Concatenation [5] is a DF scheme that relies on dimensionality reduction (DR) methods to first eliminate the differences in scales and dimensionalities between the modalities before fusing them. However, these DR methods face the risk of extracting noisy features which degrade the metaspace [8]. Other

variants of the embedding fusion idea, including Consensus embedding [6] and Boosted embedding [3] have yielded promising results, but come at a high computational cost. Consensus embedding attempts to combine multiple low dimensional data projections via a majority voting scheme while the Boosted embedding scheme leverages the Adaboost classifier [9] to combine multiple weak embeddings. In the case of weighted multi-kernel embedding using graph embedding [7] and support vector machine classifiers [2], insufficient training data can lead to overfitting and inaccurate weights to the various kernels, which can lower the performance of the meta-classifier [10].

CCA is a statistical DF technique that extracts linear correlations, by using cross-covariance matrices, between 2 data sources, X and Y. It capitalizes on the knowledge that the different modalities represent different sets of descriptors for characterizing the same object. For this reason, the mutual information that is most correlated between the two modalities will provide the most meaningful transformation into a metaspace. In recent years, CCA has been used to fuse heterogeneous data such as pixel values of images and the text attached between these images [11], assets and liabilities in banks [12], and audio and face images of speakers [13].

Regularized CCA (RCCA) is an improved version of CCA which in the presence of insufficient training data prevents overfitting by using a ridge regression optimization scheme [14]. Denote p and q as the number of features in X and Y, and n as the sample size. When $n << p$ or $n << q$, the features in X and Y tend to be highly collinear. This leads to ill-conditioned matrices C_{xx} and C_{yy}, which denote the covariance matrix of X with itself and Y with itself, such that their inverses are no longer reliable resulting in an invalid computation of CCA and an unreliable metaspace [15]. The condition placed on the data to guarantee that C_{xx} and C_{yy} will be invertible is $n \geq p + q + 1$ [16]. However, that condition is usually not met in the bioinformatics domain, where samples (n) are usually limited, and modern technology has enabled very high dimensional data streams to be routinely acquired resulting in very high dimensional feature sets (p and q). This creates a need for regularization, which works by adding small positive quantities to the diagonals of C_{xx} and C_{yy} to guarantee their invertibility [17]. RCCA has been used to study expressions of genes measured in liver cells and compare them with concentrations of hepatic fatty acids in mice [18]. However, the regularization process required by RCCA is computationally very expensive. Both CCA and RCCA also fail to take complete advantage of class label information, when available [19].

In this paper, we present a novel efficient Supervised Regularized Canonical Correlation Analysis (SRCCA) DF algorithm that is able to incorporate a supervised feature selection scheme to perform regularization. Mainly, it makes better use of labeled information that in turn allows for significantly better stratification of the data in the metaspace. While SRCCA is more expensive than the overfitting-prone CCA, it provides the needed regularization while also being computationally cheaper than RCCA. SRCCA first produces an embedding of the most correlated data in both modalities via a low dimensional metaspace. This representation is then used in conjunction with a classifier (K-Nearest Neighbor [20] and Random Forest [21] are used in this study) to create a highly accurate meta-classifier.

Along with CCA and RCCA, SRCCA is compared with 2 other low dimensional data representation techniques: Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR). PCA [22] is a linear DR method that reduces high dimensional data to dominant orthogonal eigenvectors that try to represent the maximal amount of variance in the data. PLSR [23] is a DR method that uses one modality as a set of predictors to try to predict the other modality. Tiwari et al. [24] employed PCA in conjunction with a wavelet based representation of different MRI protocols to build a fused classifier to detect prostate cancer in vivo. PLSR has been used with heterogeneous multivariate signaling data collected from HT-29 human colon carcinoma cells stimulated to undergo programmed cell death to uncover aspects of biological cue-signal-response systems [25].

In this work, we apply SRCCA to the problem of predicting biochemical recurrence in prostate cancer (CaP) patients, following radical prostatectomy, by fusing histologic imaging and proteomic signatures. Biochemical recurrence is commonly defined as a detectable elevation of Prostate Specific Antigen (PSA), a key biomarker for CaP [26-28]. However, the nonspecificity of PSA leads to over-treatment of CaP, resulting in many unnecessary treatments, which are both stressful and costly [29-33]. Even the most widely used prognostic markers such as pathologist assigned Gleason grade [34], which attempts to capture the morphometric and architectural appearance of CaP on histopathology, has been found to be a less than perfect predictor of biochemical recurrence [35]. Additionally, Gleason grade has been found to be subject to inter-, and intra-observer variability [36-38]. While some researchers have proposed quantitative, computerized image analysis approaches [1,39,40] for modeling and predicting Gleason grade (a number that goes from 1 to 5 based on morphologic appearance of CaP on histopathology), it is still not clear that an accurate, reproducible grade predictor from histology

will also be accurate in predicting biochemical recurrence and long term patient outcome [41].

Recent studies have shown that proteomic markers can be used to predict aggressive CaP [42,43]. Techniques such as mass spectrometry hold promise in their ability to identify protein expression profiles that might be able to distinguish more aggressive from less aggressive CaP and identify candidates for biochemical recurrence [44-46]. However, more and more, it is becoming apparent that a single prognostic marker may not possess sufficient discriminability to predict patient outcome which suggests that the solution might lie in an integrated fusion of multiple markers [47]. This then begs the question as to what approaches need to be leveraged to quantitatively fuse imaging and non-imaging measurements to build an integrated prognostic marker for CaP recurrence. The overarching goal of this study is to leverage SRCCA to construct a fused quantitative histologic, proteomic marker, and a subsequent meta-classifier, for predicting 5 year biochemical recurrence in CaP patients following surgery.

Our main contributions in this paper are:

- A novel data fusion algorithm, SRCCA, that builds an accurate metaspace representation that can simultaneously represent and accommodate two heterogeneous imaging and non-imaging modalities.
- Leveraging SRCCA to build a meta-classifier to predict risk of 5 year biochemical recurrence in prostate cancer patients following radical prostatectomy by integrating histological image and proteomic features.

The organization of the rest of the paper is as follows: In the methods section, we first review the 4 statistical methods, PCA, PLSR, CCA and RCCA. Next, we introduce our novel algorithm, Supervised Regularized Canonical Correlation Analysis (SRCCA). We then discuss the DF algorithm for metaspace creation and the computational complexities for CCA, RCCA and SRCCA. In the Experimental Design section, we briefly discuss the prostate cancer dataset considered in this study and the subsequent proteomic and histologic feature extraction schemes before moving on to the experiments performed on the dataset where we try to determine the ability of PCA, PLSR, CCA, RCCA and SRCCA to identify patients at risk for biochemical recurrence following surgery. The results are discussed in the subsequent section and the concluding remarks are presented at the end of the paper.

Methods
Review of PCA and PLSR
Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR) are common statistical

methods used to analyze multi-modal data and they are briefly discussed in the following sections. However, further information, explaining how these two methods can be viewed as special cases of the generalized eigen-problem, can be found in [48].

Principal Component Analysis (PCA)

PCA [22] constructs a low dimensional subspace of the data by finding a series of linear orthogonal bases called principal components. Each component seeks to explain the maximal amount of variance in the dataset. Denote two multidimensional variables, $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$, where p and q are the number of features in X and Y and n the number of overall samples. PCA is usually performed on the data matrix, $Z \in \mathbb{R}^{n \times (p+q)}$, obtained by concatenating the individual modalities such that: $Z = [X \ Y]$ [24]. $\bar{Z} \in \mathbb{R}^{n \times (p+q)}$ is then obtained by subtracting the means of all features for a certain sample from its original feature value in Z so that the resultant $\bar{Z}$ has rows with a 0 mean. $\bar{Z}$ is further broken using singular value decomposition into [22]:

$$\bar{Z} = UEV^T \tag{1}$$

where $E \in \mathbb{R}^{n \times n}$ is a diagonal matrix containing the eigenvalues of the eigenvectors which are stored in $U \in \mathbb{R}^{p \times p}$, and $V^T \in \mathbb{R}^{m \times n}$. The eigenvalues stored in E explain how much variance of the original $\bar{Z}$ is stored in the corresponding eigenvector, or principal component. Using these eigenvalues as a rank, the top d embedding components can be chosen to best represent the original data in a lower dimensional subspace.

Partial Least Squares Regression(PLSR)

PLSR [49] is a statistical technique that generalizes PCA and multiple regression. The general underlying model behind PLSR is [23]:

$$X = TP^T + E \tag{2}$$

$$Y = TC^T + F \tag{3}$$

where $T \in \mathbb{R}^{n \times l}$ is a score matrix, $P \in \mathbb{R}^{p \times l}$ and $C \in \mathbb{R}^{q \times l}$ are loading matrices for X and Y, and $E \in \mathbb{R}^{n \times p}$ and $F \in \mathbb{R}^{n \times p}$ are the error terms. PLSR is an iterative process and works by continually approximating, and improving the approximation of the matrices T, P and C [50].

Review of CCA and RCCA

Canonical Correlation Analysis (CCA)

CCA [51] is a way of using cross-covariance matrices to obtain a linear relationship between the two multidimensional variables, $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$. CCA obtains two directional vectors $w_x \in \mathbb{R}^{p \times 1}$ and $w_y \in \mathbb{R}^{q \times 1}$ such that Xw_x and Yw_y will be maximally

correlated. It is defined as the optimization problem [11]:

$$\rho = \max_{w_x, w_y} \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}} \tag{4}$$

where $C_{xy} \in \mathbb{R}^{p \times q}$ is the covariance matrix of the matrices X and Y, $C_{xx} \in \mathbb{R}^{p \times p}$ is the covariance matrix of the matrix X with itself and $C_{yy} \in \mathbb{R}^{q \times q}$ is the covariance matrix of the matrix Y with itself. The solution to CCA reduces to the solution of the following two generalized eigenvalue problems [52]:

$$C_{xy} C_{yy}^{-1} C_{yx} = \lambda C_{xx} w_x \tag{5}$$

$$C_{yx} C_{xx}^{-1} C_{xy} = \lambda C_{yy} w_y \tag{6}$$

where λ is the generalized eigenvalue representing the canonical correlation, and w_x and w_y are the corresponding generalized eigenvectors. CCA can further produce exactly min$\{p, q\}$ orthogonal embedding components (sets of $w_x X$ and $w_y Y$) which can be sorted in order of decreasing correlation, λ.

Regularized Canonical Correlation Analysis (RCCA)

RCCA [53,54] corrects for noise in X and Y by first assuming that X and Y are contaminated with noise, $N_x \in \mathbb{R}^{n \times p}$ and $N_Y \in \mathbb{R}^{n \times q}$. We assume that these noise vectors in the p and q columns of N_X and N_Y, respectively, are gaussian, independent and identically distributed. For this reason, all combinations of the covariances of the p columns of N_X and q columns of N_Y will be 0 except the covariance of a particular column vector with itself. This variance of each column of N_X and N_Y is labeled λ_x and λ_y and these labels are called the regularization parameters. The matrix C_{xy} will not be affected but the matrices C_{xx} and C_{yy} become $C_{xx} + \lambda_x I_x$ and $C_{yy} + \lambda_x I_x$. The solution to RCCA now becomes the solution to these generalized eigenvalue problems [52]:

$$C_{xy}(C_{yy} + \lambda_y I_y)^{-1} C_{yx} = \lambda(C_{xx} + \lambda_x I_x) w_x \tag{7}$$

$$C_{yx}(C_{xx} + \lambda_x I_x)^{-1} C_{xy} = \lambda(C_{yy} + \lambda_y I_y) w_y \tag{8}$$

The regularization parameters next have to be chosen. For $i \in \{1, 2, \ldots, n\}$, let w_x^i and w_y^i denote the weights calculated from RCCA when samples X_i and Y_i are removed. λ_x and λ_y are varied in a certain range $\theta_1 \leq \lambda_x$, $\lambda_y \leq \theta_2$ and chosen via a grid search [55] optimization of the following cost function [18]:

$$\max_{\lambda_x, \lambda_y} [corr(\{X_i w_x^i\}_{i=1}^n, \{Y_i w_y^i\}_{i=1}^n)] \tag{9}$$

where *corr* (·, ·) refers to the Pearson's correlation coefficient [56]. The above cost function essentially measures the change in the produced w_x^i and w_y^i when a sample i is omitted and seeks the optimal λ_x and λ_y where this change is minimized. λ_x and λ_y are chosen using the embedding component with the highest λ and then adjusted for the remaining dimensions [18].

Extending RCCA to SRCCA

Supervised Regularized Canonical Correlation Analysis (SRCCA) chooses λ_x and λ_y using a supervised feature selection method (*t*-test, Wilcoxon Rank Sum Test and Wilks Lambda Test are used in this study). Denote $\mathbb{W}_1$ and $\mathbb{W}_2$ as class 1 and class 2 and μ_1 and μ_2, σ_1^2 and σ_2^2, n_1 and n_2 as the means, variances, and sample sizes of $\mathbb{W}_1$ and $\mathbb{W}_2$. The data in the metaspace, Xw_x or Yw_y, can be split using its labels into the n_1 samples that belong to $\mathbb{W}_1$ and the n_2 samples that belong to class $\mathbb{W}_2$, where $n_1 + n_2 = n$. These two partitions can then be used to calculate the discrimination level between the samples of the two classes in the metaspace representation. In this study, we implement RCCA with the *t*-test (SRCCA$_{TT}$), the Wilcoxon Rank Sum Test (SRCCA$_{WRST}$) and the Wilks Lambda Test (SRCCA$_{WLT}$) to try to choose more appropriate regularization parameters, λ_x and λ_y, that can more successfully stratify the samples in the metaspace compared to the parameters chosen by RCCA. Similar to RCCA, for SRCCA, λ_x and λ_y are chosen using the embedding component with the most discriminatory score as chosen by the feature selection schemes below and then adjusted for the remaining dimensions.

SRCCA$_{TT}$

The *t*-test [57] is a parametric test that assumes the distributions of the two samples are normal and tests whether these distributions have the same means. The *t*-score, which measures the number of standard deviations the two means of n_1 samples of $\mathbb{W}_1$ and n_2 samples of $\mathbb{W}_2$ are away from each other, is maximized using a grid search algorithm as:

$$\max_{\lambda_x, \lambda_y} \frac{||\mu_1 - \mu_2||}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}. \tag{10}$$

SRCCA$_{WRST}$

Wilcoxon Rank Sum Test [58] sorts both the samples in order from lowest value to highest value. It then uses their respective ranks within the population to calculate the discriminatory score:

$$\max_{\lambda_x, \lambda_y} \left\{ \left(\sum_{i=1}^{n_2} b_i - \frac{n_2(n_2+1)}{2} \right) , \left(n_1 n_2 - \sum_{i=1}^{n_2} b_i + \frac{n_2(n_2+1)}{2} \right) \right\}, \tag{11}$$

where b_i represents the rank of the sample $i \in \mathbb{W}_2$ with respect to the rest of the samples.

SRCCA$_{W\,LT}$

In an ideal metaspace representation, samples from each class will be grouped together while the samples from different classes will be grouped separately. The WLT [59] capitalizes on this knowledge and calculates the ratio of within class variance of both samples to the total variance of both samples combined. Wilks Lambda (Λ) is minimized using a grid search algorithm as:

$$\min_{\lambda_x, \lambda_y} \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n \sigma^2}. \tag{12}$$

Data Fusion in the context of CCA, RCCA and SRCCA

DF is performed as described in Foster et al. [60]. When the Xw_x and Yw_y are maximally correlated, each modality represents similar information, and thus either Xw_x or Yw_y can be used to represent the original two modalities in the metaspace. Moreover, X and Y are both descriptors of the same object and thus, the most relevant information is the data that exists and is correlated in both modalities. Thus, a high correlation of Xw_x and Yw_y is indicative that meaningful data, measuring the object of interest, is being added to the metaspace.

In order of decreasing λ, the top d embedding components, up to $\phi = \min\{p, q\}$ can be chosen to represent the two modalities in a metaspace. However, the lower embedding components will have a lower λ, and thus a lower correlation between Xw_x and Yw_y which might imply that non-relevant data is being added to the metaspace. To avoid this issue, a threshold, λ_0, can be selected such that only embedding components with $\lambda \geq \lambda_0$ will be included in the metaspace.

Computational Complexity

Given $\phi = \min\{p, q\}$, CCA has a computational complexity of $\phi!$ (based on the source code in [61]). The regularization algorithm requires a grid search process for each ordered pair (λ_x, λ_y). Assume v potential λ_x and λ_y sampled evenly between θ_1 and θ_2. RCCA requires a training/testing cross-validation strategy, at each ordered pair (λ_x, λ_y), to find the optimal λ_x and λ_y. It will require CCA to be performed an order of n times at each of the v intervals leading to a complexity of $vn\phi!$. SRCCA only requires a CCA factorization once at each of the v intervals leading to a complexity of $v\phi!$.

The computational complexities for each of the CCA schemes are summarized in Table 1. Table 1 indicates that SRCCA is an order of n times faster compared to RCCA. However, SRCCA is also more complex compared to CCA and will have a longer execution time.

Table 1 The computational complexities of all 3 DF methods used in this study

Method	Complexity
CCA	$\phi!$
RCCA	$vn\phi!$
SRCCA	$v\phi!$

$\phi = \min\{p, q\}$, which represents the number of features in the lower dimensional modality, n is the sample size and v is the interval spacing over which λ_1 and λ_2 will be chosen in the range $\{\theta_1, \theta_2\}$.

Experimental Design
Data Description
A total of 19 prostate cancer patients at the Hospital at the University of Pennsylvania were considered for this study. All patient identifiers are stripped from the data at the time of acquisition. The data was deemed to be exempt for review by the internal review board at Rutgers University and the protocol was approved by the University of Pennsylvania internal review board. Hence, the data was deemed eligible for use in this study. All of these patients had been found to have prostate cancer on needle core biopsy and subsequently underwent radical prostatectomy. 10 of these patients had biochemical recurrence within 5 years following surgery (BR) and the other 9 did not (NO BR). The 19 patient studies were randomly chosen from a larger cohort of 110 patient studies at the University of Pennsylvania all of whom had been stage and grade matched (Gleason score of 6 or 7) and had undergone gland resection. Of these 110 cases, 55 had experienced biochemical recurrence within 5 years while the other 55 had not. The cost of the mass spectrometry to acquire the proteomic data limited this study to only 19 patient samples. Following gland resection, the gland was sectioned into a series of histological slices with a meat cutter. For each of the 19 patient studies, a representative histology section on which the dominant tumor nodule was observable was identified. Mass Spectrometry was performed at this site to yield a protein expression vector. The representative histologic sections were then digitized at 40 × magnification using a whole slide digital scanner.

In the next two sections, we briefly describe the construction of the proteomic and histologic feature spaces. Subsequently we describe the strategy for combination of quantitative image descriptors from the tumor site on the histological prostatectomy specimen and the corresponding proteomic measurements obtained from the same tumor site, via mass spectrometry. The resultant meta-classifier, constructed in the fused meta-space, is then used to distinguish the patients at 5 year risk of biochemical recurrence following radical prostatectomy from those who are not.

Proteomic Feature Selection
Prostate slides were deparaffinized, and rehydrated essentially as described in [62]. Tumor areas previously defined on a serial H&E section were collected by needle dissection, and formalin cross-links were removed by heating at 99°C. The FASP (Filter-Aided Sample Preparation) method [63] was then used for buffer exchange and tryptic digest. After peptide purification on C-18 StageTips [64] samples were analyzed using nanoflow C-18 reverse phase liquid chromatography/tandem mass spectrometry (nLC-MS/MS) on an LTQ Orbitrap mass spectrometer. A top-5 data-dependent methodology was used for MS/MS acquisition, and data files were processed using the Rosetta Elucidator proteomics package, which is a label-free quantitation package that uses extracted ion chromatograms to calculate protein abundance rather than peptide counts. A high dimensional feature vector was obtained, denoted $\varphi^P \in \mathbb{R}^{19 \times 953}$, characterizing each patient's protein expression profile following surgery. This data underwent quantile normalization, log(2) transformation, and mean and variance normalization on a per-protein basis.

Quantitative Histologic Feature Extraction
In prostate whole-mount histology, denoted $\varphi^H \in \mathbb{R}^{19 \times 151}$ (Figure 1 (a), (f)), the objects of interest are the glands (shown in Figure 1 (b), (g)), whose shape and arrangement are highly correlated with cancer progression [1,39,65,66]. We briefly describe this process below. Prior to extracting image features, we employ an automatic region-growing gland segmentation algorithm presented by Monaco et al. [67]. The boundaries of the interior gland lumen and the centroids of each gland, allow for extraction of 1) morphological and 2) architectural features from histology as described briefly below. More extensive details on these methods are in our other publications [5,39,68].

Glandular Morphology The set of 100 morphological features [1], (denoted $\varphi^M \in \mathbb{R}^{19 \times 100}$), of attributes, consists of the average, median, standard deviation, and min/max ratio for features such as gland area, maximum area, area ratio, and estimated boundary length (See Table 2).

Architectural Feature Extraction 51 architectural image features, which have been shown to be predictors of cancer [69], (denoted $\varphi^A \in \mathbb{R}^{19 \times 51}$), were extracted in order to quantify the arrangement of glands present in the section (See Table 2). Voronoi diagrams, Delaunay Triangulation and Minimum Spanning Trees were constructed on the digital histologic image using the gland centroids as vertices, the gland centroids having previously been identified via the scheme in [68].

Fusing Proteomic, Histologic Features for Predicting Biochemical Recurrence in CaP Patients Post-Surgery
Experiment 1 - Comparing SRCCA with CCA and RCCA
We performed CCA, RCCA, and SRCCA on selected multimodal combinations, φ^P and φ^J, where $J \in \{M, A,$
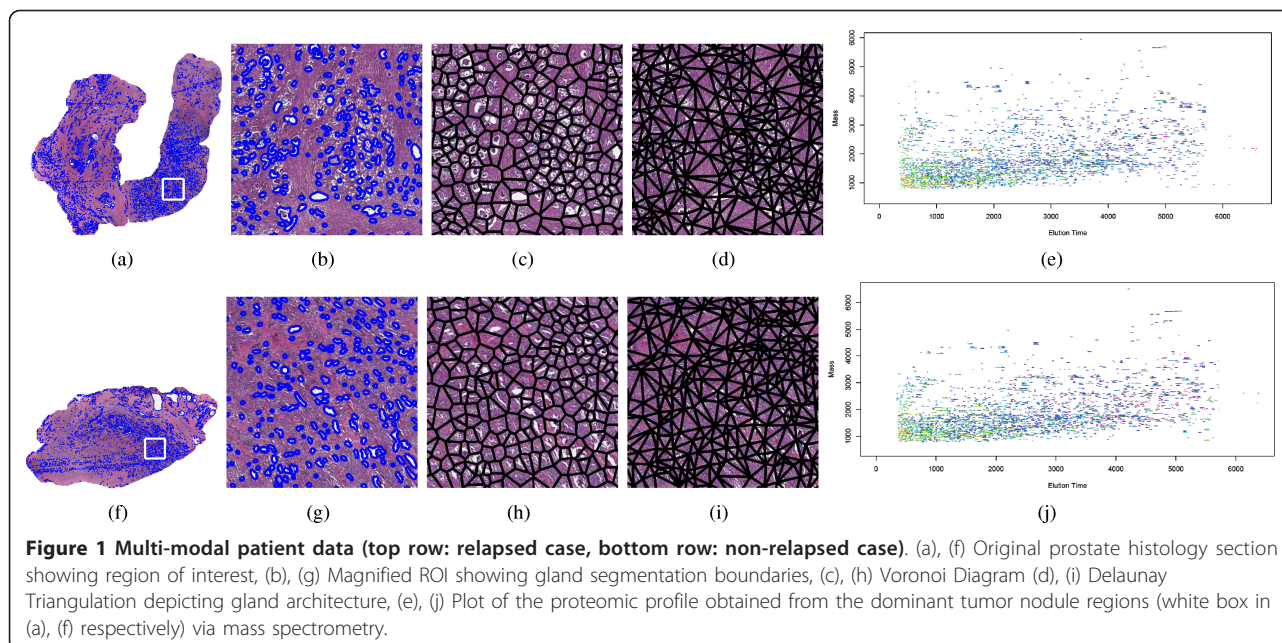
Figure 1 Multi-modal patient data (top row: relapsed case, bottom row: non-relapsed case). (a), (f) Original prostate histology section showing region of interest, (b), (g) Magnified ROI showing gland segmentation boundaries, (c), (h) Voronoi Diagram (d), (i) Delaunay Triangulation depicting gland architecture, (e), (j) Plot of the proteomic profile obtained from the dominant tumor nodule regions (white box in (a), (f) respectively) via mass spectrometry.

H}. φ^P was reduced to 25 features as ranked by the t-test, with a p-value cutoff of p = .05, using a leave-one-out validation strategy. For CCA, φ^P and φ^J were used as the two multidimensional variables, X and Y, as mentioned above in Section 2. For RCCA and SRCCA, φ^P and φ^J were used in a manner similar to CCA except they are tested with regularization parameters λ_x and λ_y evenly spaced from θ_1 = .001 to θ_2 = .2 with ν = 200.

The top d = 3 embedding components (which were experimentally found to meet the criteria of λ_0 = .99 for all SRCCA on all 3 multimodal combinations) were produced from CCA, RCCA, SRCCA$_{TT}$, SRCCA$_{WRST}$, and SRCCA$_{WLT}$. The classification accuracies were determined with the classifiers K-Nearest Neighbor, denoted

via φ^{KNN} [20], with K = 1, and Random Forest, denoted via φ^{RF} [21], with 50 Trees. Both these classifiers were used because of their high computational speed. Accuracies were determined using leave-one-out validation, which was implemented because of the small sample size. In this process, 18 samples were used for the initial feature pruning, determining the optimal regularization parameter and training the classifier while the remaining sample was used as the testing set for evaluating the classifier. This procedure was repeated till all the samples were used in the testing set.

Experiment 2 - Comparing SRCCA with PCA and PLSR

In addition to the steps performed in Experiment 1, metaspaces were also produced with PCA and PLSR. φ^P

Table 2 Description of 25 Proteomic Features, 100 Morphological, and 51 Architectural

Proteomic	#	Description
Proteins Identified	25	Some include: CSNK2A1 protein, Dihydroxyacetone kinase, Dynamin-2, Glycogenin-1, Mitochondrial PDHA1, Mu-crystallin homolog, Nit protein 2, Nucleolin, Synaptonemal complex protein 1 Putative uncharacterized protein RPL3
Morphological		**Description**
Gland Morphology	100	Area Ratio, distance Ratio, Standard Deviation of Distance, Variance of Distance, Distance Ratio, Perimeter, Ratio, Smoothness, Invariant Moment 1-7, Fractal Dimension, Fourier Descriptor 1-10 (Mean, Std. Dev, Median, Min/Max of each)
Architectural		**Description**
Voronoi Diagram	12	Polygon area, perimeter, chord length: mean, std. dev., min/max ratio, disorder
Delaunay Triangulation	8	Triangle side length, area: mean, std. dev., min/max ratio, disorder
Minimum Spanning Tree	4	Edge length: mean, std. dev., min/max ratio, disorder
Nearest Neighbors	27	Density of nuclei, distance to nearest nuclei

and φ^I were concatenated and PCA was then performed on this new data matrix. For PLSR, a regression of φ^I on φ^P was performed.

Similarly, using the top d = 3 embedding components produced from PCA, PLSR, $SRCCA_{TT}$, $SRCCA_{WRST}$, and $SRCCA_{WLT}$, the classification accuracies of φ^{KNN}, with K = 1, and φ^{RF}, with 50 Trees, were determined using leave-one-out validation.

Experiment 3 - Comparing classifier accuracy for PCA, PLSR and CCA variants using metaspace representations

Using the 10 different values for d ∈ {1, 2, ..10}, and the 3 fusion schemes considered (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H), 30 different embeddings were obtained for PCA, PLSR, CCA, RCCA, $SRCCA_{TT}$, $SRCCA_{WRST}$, and $SRCCA_{WLT}$. The maximum and median of these 30 different measurements for each classifier were calculated.

In addition, we denote as $\alpha_1(i)$, the classification accuracy obtained by the DF scheme i, where $i \in$ {PCA, PLSR, CCA, RCCA} and $\alpha_2(j)$ as the accuracy obtained by the DF scheme i, where $j \in$ {$SRCCA_{TT}$, $SRCCA_{WRST}$, $SRCCA_{WLT}$}. A two paired student t-test was employed to identify whether there were statistically significant improvements in the 3 SRCCA variants by comparing the classification accuracies with the null hypothesis:

$$H_o : \alpha_1(i) = \alpha_2(j) \tag{13}$$

for all $i \in$ {PCA, PLSR, CCA, RCCA} and for all $j \in$ {$SRCCA_{TT}$, $SRCCA_{WRST}$, $SRCCA_{WLT}$}.

Experiment 4 - Computational consideration for RCCA and SRCCA

We measured the 3 individual single run completion times for RCCA and SRCCA to fuse (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H), with the null hypothesis:

$$H_o : \text{completion time of RCCA} = \text{completion time of SRCCA} \tag{14}$$

These experiments were performed on a quadcore computer with a clock speed of 1.8 GHz, and the programs were written on MATLAB(R) platform.

Results and Discussion

Experiment 1

Across both classifiers for d = 3, the 3 SRCCA variants, $SRCCA_{TT}$, $SRCCA_{WRST}$, $SRCCA_{WLT}$, had a combined median classification accuracy of 80% compared to 60% for CCA and 42% for RCCA. SRCCA also performed better in all 36 of 36 direct comparisons with CCA and RCCA (see Tables 3 and 4). The higher classification accuracy results indicate that SRCCA produces a metaspace, where the samples are more stratified, compared to CCA and RCCA. This also seems to indicate that the supervised scheme of choosing regularization parameters, by the 3 SRCCA variants, is a more appropriate

Table 3 Experiment 1: Classification Accuracy with K-Nearest Neighbor

Dataset (φ^P, φ^I)	CCA	RCCA	$SRCCA_{TT}$	$SRCCA_{WRST}$	$SRCCA_{WLT}$
(φ^P, φ^M)	53%	37%	**80**%	79%	79%
(φ^P, φ^A)	58%	47%	**74**%	68%	**74**%
(φ^P, φ^H)	63%	47%	**74**%	**74**%	**74**%

Classification accuracies obtained for fusing (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H), with CCA, RCCA, $SRCCA_{TT}$, $SRCCA_{WRST}$, and $SRCCA_{WLT}$ using the top d = 3 components, using φ^{KNN} with K = 1 neighbor and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.

scheme for classification purposes compared to the ridge regression scheme used by RCCA.

These results, which seem to suggest that SRCCA outperforms the other two CCA based approaches for this dataset, CCA and RCCA, are observable in the embedding plots of Figure 2, which show the metaspace produced by CCA, RCCA, $SRCCA_{TT}$, $SRCCA_{WRST}$ and $SRCCA_{WLT}$ with d = 2 components. It may be seen that because CCA lacks regularization, the corresponding covariance matrices are singular and lack inverses. For this reason, in Figure 2 the embedding components are not orthogonal but are highly correlated to each other and yield the same information. RCCA overcomes this regularization problem but still does not produce the same level of discrimination between patient classes compared to the 3 variations of SRCCA. Note that $SRCCA_{TT}$, $SRCCA_{WRST}$ and $SRCCA_{WLT}$ chose similar regularization parameters, λ_x and λ_y, and have similar embedding plots.
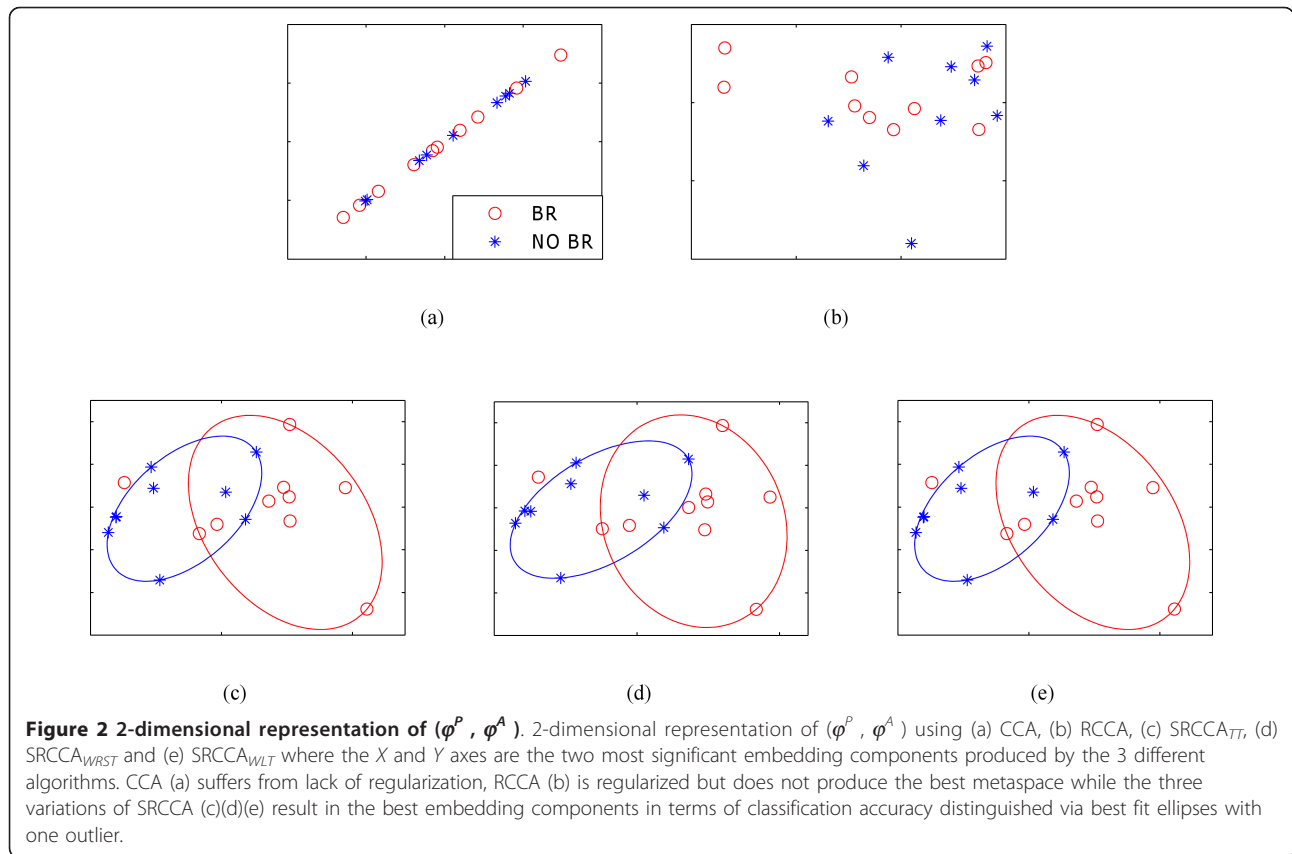
Experiment 2

We see that $SRCCA_{TT}$, $SRCCA_{WRST}$, $SRCCA_{WLT}$ are able to outperform PCA and PLSR in all 36 of 36 direct comparisons (see Tables 5 and 6). Even though, across both classifiers for d = 3, PCA and PLSR have median classification accuracies of 64% and 61%, which is higher than the accuracies for CCA and RCCA, it is still much lower than the 80% for $SRCCA_{TT}$, $SRCCA_{WRST}$, $SRCCA_{WLT}$. These results also seem to indicate that $SRCCA_{TT}$, $SRCCA_{WRST}$, $SRCCA_{WLT}$ could also create a more appropriate metaspace than, not only CCA and RCCA, but also PCA and PLSR.

Table 4 Experiment 1: Classification Accuracy with Random Forest

Dataset (φ^P, φ^I)	CCA	RCCA	$SRCCA_{TT}$	$SRCCA_{WRST}$	$SRCCA_{WLT}$
(φ^P, φ^M)	37%	42%	83%	81%	**84**%
(φ^P, φ^A)	74%	30%	81%	77%	**83**%
(φ^P, φ^H)	62%	42%	91%	89%	**93**%

Classification accuracies obtained for fusing (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H), with CCA, RCCA, $SRCCA_{TT}$, $SRCCA_{WRST}$, and $SRCCA_{WLT}$ using the top d = 3 components, using φ^{RF} with 50 trees and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.

Figure 2 2-dimensional representation of (φ^P, φ^A). 2-dimensional representation of (φ^P, φ^A) using (a) CCA, (b) RCCA, (c) SRCCA$_{TT}$, (d) SRCCA$_{WRST}$ and (e) SRCCA$_{WLT}$ where the X and Y axes are the two most significant embedding components produced by the 3 different algorithms. CCA (a) suffers from lack of regularization, RCCA (b) is regularized but does not produce the best metaspace while the three variations of SRCCA (c)(d)(e) result in the best embedding components in terms of classification accuracy distinguished via best fit ellipses with one outlier.

Experiment 3

In Tables 7 and 8 we see that the maximum and median φ^{KNN} and φ^{RF} of the 3 SRCCA variants for fusion of (φ^I, φ^J) were much higher than the corresponding values of PCA, PLSR, CCA or RCCA. We also see that SRCCA$_{WLT}$ attains a maximum classifier accuracy of 93.16% (see Table 7). In Tables 9 and 10, the 3 SRCCA variants are statistically significantly better than PCA, PLSR, CCA or RCCA even at the $p = .001$ level using either classifiers, φ^{KNN} or φ^{RF}. We further see that SRCCA$_{WLT}$ tends to marginally outperform SRCCA$_{TT}$ and SRCCA$_{WRST}$. However given the small sample size it is difficult to draw any definitive conclusions about which of SRCCA$_{TT}$, SRCCA$_{WRST}$, or SRCCA$_{WLT}$ might be the better SRCCA variant.

In Figures 3 and 4, we see the classification accuracies of the 7 DF methods, PCA, PLSR, CCA, RCCA, SRCCA$_{TT}$, SRCCA$_{WRST}$, or SRCCA$_{WLT}$ over a range of $d \in \{1, 2, ..10\}$ embedding components for the fusion (φ^P, φ^H). Importantly, we see that the SRCCA$_{TT}$, SRCCA$_{WRST}$, and SRCCA$_{WLT}$ all outperform PCA, PLSR, CCA and RCCA for a majority of the embedding dimensions, across both the φ^{KNN} and φ^{RF} classifiers.

Experiment 4

Figure 5 reveals that the completion time of SRCCA is significantly lower than the completion time of RCCA. Even though the differences in these times are visibly different, a p-value of 1.9×10^{-3} even with just 3

Table 5 Experiment 2: Classification Accuracy with K-Nearest Neighbor

Dataset (φ^P, φ^J)	PCA	PLSR	SRCCA$_{TT}$	SRCCA$_{WRST}$	SRCCA$_{WLT}$
(φ^P, φ^M)	68%	57%	**80**%	79%	79%
(φ^P, φ^A)	63%	47%	**74**%	68%	**74**%
(φ^P, φ^H)	53%	53%	**74**%	**74**%	**74**%

Classification accuracies obtained for fusing (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H), with CCA, RCCA, SRCCA$_{TT}$, SRCCA$_{WRST}$, and SRCCA$_{WLT}$ using the top $d = 3$ components, using φ^{KNN} with K = 1 neighbor and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.

Table 6 Experiment 2: Classification Accuracy with Random Forest

Dataset (φ^P, φ^J)	PCA	PLSR	SRCCA$_{TT}$	SRCCA$_{WRST}$	SRCCA$_{WLT}$
(φ^P, φ^M)	64%	75%	83%	81%	**84**%
(φ^P, φ^A)	50%	64%	81%	77%	**83**%
(φ^P, φ^H)	64%	67%	91%	89%	**93**%

Classification accuracies obtained for fusing (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H), with CCA, RCCA, SRCCA$_{TT}$, SRCCA$_{WRST}$, and SRCCA$_{WLT}$ using the top $d = 3$ components, using φ^{RF} with 50 trees and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.

Table 7 Experiment 3: Maximum φ^{KNN} and φ^{RF} of DF schemes across d ⌊ {1, 2, ..10}

Classifier	PCA	PLS	CCA	RCCA	SRCCA$_{TT}$	SRCCA$_{WRST}$	SRCCA$_{WLT}$
φ^{KNN}	**84.21%**	**84.21%**	73.68%	68.42%	**84.21%**	**84.21%**	**84.21**%
φ^{RF}	84.21%	84.21%	80.20%	68.42%	91.05%	88.95%	**93.16**%

Maximum classification accuracies obtained for fusing (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H), with PCA, PLSR, CCA, RCCA, SRCCA$_{TT}$, SRCCA$_{WRST}$, and SRCCA$_{WLT}$ across d ∈ {1, 2, ..10} components, using two classifiers, φ^{KNN}, with K = 1, and φ^{RF}, with 50 trees, and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.

Table 8 Experiment 3: Median φ^{KNN} and φ^{RF} of DF schemes across d ⌊ {1, 2, ..10}

Classifier	PCA	PLS	CCA	RCCA	SRCCA$_{TT}$	SRCCA$_{WRST}$	SRCCA$_{WLT}$
φ^{KNN}	52.63%	57.89%	57.89%	47.37%	**68.42**%	**68.42**%	**68.42**%
φ^{RF}	51.58%	62.37%	58.42%	37.37%	72.89%	69.47%	**74.21**%

Median classification accuracies obtained for fusing (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H), with PCA, PLSR, CCA, RCCA, SRCCA$_{TT}$, SRCCA$_{WRST}$, and SRCCA$_{WLT}$ across d ∈ {1, 2, ...10} components, using two classifiers, φ^{KNN}, with K = 1, and φ^{RF}, with 50 trees, and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.

Table 9 Experiment 3: Statistical Significance (p-value) of SRCCA for φ^{KNN}

Classifier	SRCCA$_{TT}$	SRCCA$_{WRST}$	SRCCA$_{WLT}$
PCA	5.9×10^{-10}	9.0×10^{-09}	4.7×10^{-8}
PLS	6.0×10^{-7}	9.2×10^{-5}	2.2×10^{-6}
CCA	3.0×10^{-8}	1.3×10^{-6}	4.0×10^{-9}
RCCA	4.0×10^{-10}	4.5×10^{-10}	7.1×10^{-11}

p-values for the twelve comparisons of every scheme in {PCA, PLSR, CCA, RCCA} to every scheme in {SRCCA$_{TT}$, SRCCA$_{WRST}$, SRCCA$_{WLT}$} for fusing (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H) across d ∈ {1, 2, ...10} components, using two classifiers, φ^{KNN}, with K = 1, and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.

Table 10 Experiment 3: Statistical Significance (p-value) of SRCCA for φ^{RF}

Classifier	SRCCA$_{TT}$	SRCCA$_{WRST}$	SRCCA$_{WLT}$
PCA	1.7×10^{-13}	4.7×10^{-12}	1.4×10^{-10}
PLS	1.3×10^{-5}	8.5×10^{-3}	1.6×10^{-4}
CCA	6.8×10^{-7}	5.4×10^{-6}	2.1×10^{-7}
RCCA	3.4×10^{-9}	1.8×10^{-9}	3.6×10^{-16}

p-values for the twelve comparisons of every scheme in {PCA, PLSR, CCA, RCCA} to every scheme in {SRCCA$_{TT}$, SRCCA$_{WRST}$, SRCCA$_{WLT}$} for fusing (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H) across d ∈ {1, 2, ...10} components, using two classifiers, φ^{RF}, with 50 trees, and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.

samples, indicates that SRCCA appears to be statistically significantly faster compared to RCCA.

Note that the canonical factorization stage is the most time consuming part of the of the algorithm. The Feature Selection stage computation, in comparison, is not as time consuming. SRCCA$_{TT}$, SRCCA$_{WRST}$, and SRCCA$_{WLT}$ (whose results are reported in Figure 5) all have similar execution times.

Conclusions

In this paper, we presented a novel data fusion (DF) algorithm called Supervised Regularized Canonical Correlation Analysis (SRCCA) that, unlike CCA and RCCA, is (1) able to fuse with a feature selection (FS) scheme, (2) regularized, and (3) computationally cheap. We demonstrate how SRCCA can be used for quantitative integration and representation of multi-scale, multi-modal imaging and non-imaging data. In this work we leveraged SRCCA for the purpose of constructing a fused quantitative histologic-proteomic classifier for predicting which prostate cancer patients are at risk for 5 year biochemical recurrence following surgery. We have demonstrated that SRCCA is able to (1) produce a metaspace, where the samples are more stratified than
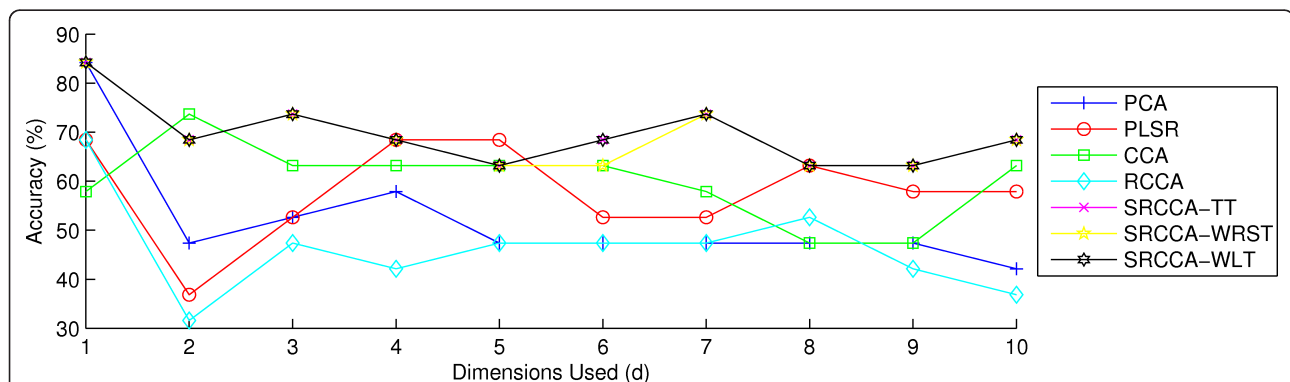

Figure 3 Classification accuracies of (φ^P, φ^H) across dimensions d ⌊ {1, 2, ..10} using the classifier φ^{KNN}. Accuracies were obtained for fusing (φ^P, φ^H), with PCA, PLSR, CCA, RCCA, SRCCA$_{TT}$, SRCCA$_{WRST}$, and SRCCA$_{WLT}$ across d ∈ {1, 2, ...10} components, using φ^{KNN}, with K = 1, and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.

Figure 4 Classification accuracies of (φ^P, φ^H) across dimensions d ∟ {1, 2, ..10} using the classifier φ^{RF}. Accuracies were obtained for fusing (φ^P, φ^H), with PCA, PLSR, CCA, RCCA, SRCCA$_{TT}$, SRCCA$_{WRST}$, and SRCCA$_{WLT}$ across d ∈ {1, 2, ...10} components, using φ^{KNN}, with K = 1, and leave-one-out validation to identify patients at the risk of biochemical recurrence from those who are not.
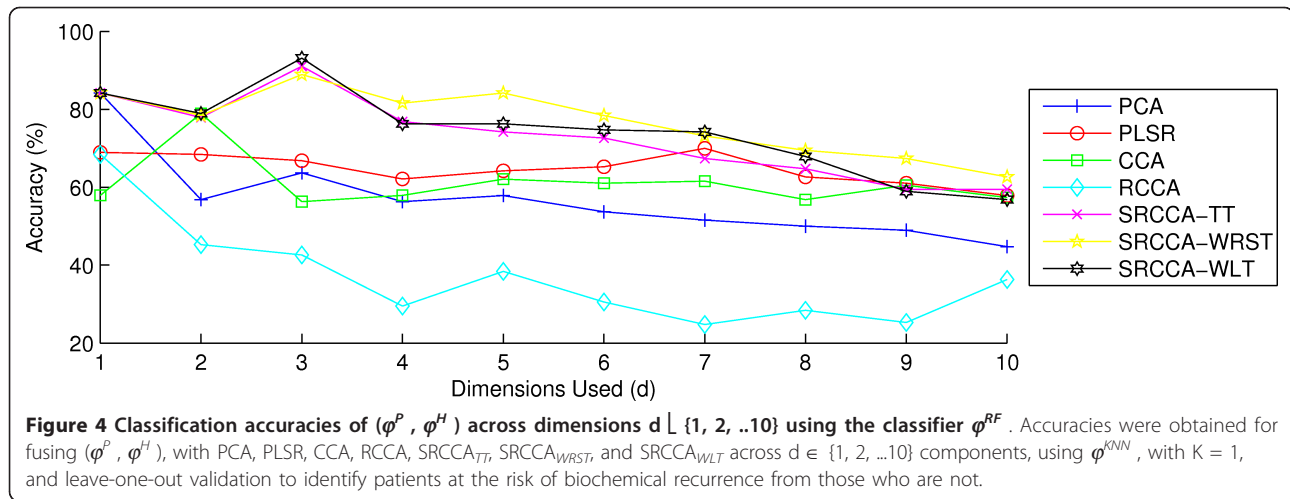
the metaspace produced by CCA or RCCA, (2) better identify patients at the risk of biochemical recurrence compared to Principal Component Analysis (PCA), Partial Least Squares Regression (PLSR), CCA or RCCA, (3) perform regularization, all the while being statistically significantly faster compared to RCCA.

While the fused prognostic classifier for predicting biochemical recurrence in this work appears to be promising, we also acknowledge the limitations of this work: (1) As previously mentioned, the cost of mass spectrometry limited this study to only 19 datasets. By using a minimum sample size derivation model [70,71], we were able to determine that our fused SRCCA classifier would yield an accuracy of 93%, more than 95% of the time if our dataset were expanded to 56 studies. We intend to evaluate our classifier on such a cohort in the future. (2) Ideally, a randomized cross validation strategy should have been employed for the training and evaluation of the classifier. Unfortunately, this was also limited

by the size of the cohort. While both parametric and non-parametric feature selection strategies were employed in this work, the availability of a larger dataset for classification in conjunction with SRCCA would allow for employment of parametric selection strategies, assuming that the underlying distribution can be estimated. For small sample datasets, a non-parametric feature selection strategy might be more approrpriate. In future work, we also plan to apply SRCCA in the context of data fusion for other imaging and non-imaging datasets in the context of other problem domains and applications.

Abbreviations

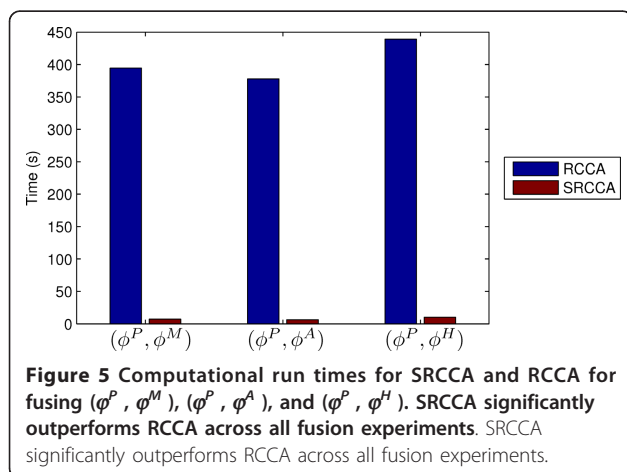### Acknowledgements

Author details

[1]Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey, USA. [2]Department of Biomedical Engineering, Rutgers University, Piscataway, New Jersey, USA. [3]Department of Pathology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [4]The Wistar Institute, Philadelphia, Pennsylvania, USA.

Authors' contributions

AM and AG devised the methodology and formulated the experiments. AG drafted the manuscript in collaboration with GL. AM edited the manuscript. SRM, MDF, JET, and JWS provided the data and the clinical expertise. All authors have read and approved the final manuscript.

Figure 5 Computational run times for SRCCA and RCCA for fusing (φ^P, φ^M), (φ^P, φ^A), and (φ^P, φ^H). SRCCA significantly outperforms RCCA across all fusion experiments. SRCCA significantly outperforms RCCA across all fusion experiments.

References

1. Madabhushi A, Agner S, Basavanhally A, Doyle S, Lee G: **Computer-aided prognosis: Predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data.** *CMIG* 2011.
2. Lanckriet GRG, Deng M, Cristianini N, Jordan MI, Noble WS: **Kernel-based data fusion and its application to protein function prediction in yeast.** *Proceedings of the Pacific Symposium on Biocomputing* 2004, 300-311.
3. Tiwari P, Viswanath S, Lee G, Madabhush A: **Multi-Modal Data Fusion Schemes for Integrated Classification of Imaging and Non-imaging Biomedical Data.** *ISBI* 2011, 165-168.
4. Duda RO, Hart PE: *Pattern Classification and Scene Analysis* John Wiley & Sons, New York; 1973.
5. Lee G, Monaco J, Doyle S, Masters S, Feldman M, Tomaszewski J, Madabhushi A: **A knowledge representation framework for integration, classification of multi-scale imaging and non-imaging data: Preliminary results in predicting prostate cancer recurrence by fusing mass spectrometry and histology.** *ISBI* 2009, 77-80.
6. Viswanath S, Rosen M, Madabhushi A: **A consensus embedding approach for segmentation of high resolution in vivo prostate magnetic resonance imagery.** *SPIE Med Imag* 2008, **6915**(1), 69150U.
7. Tiwari P, Kurhanewicz J, Rosen M, Madabhushi A: **Semi Supervised Multi Kernel (SeSMiK) Graph Embedding: Identifying Aggressive Prostate Cancer via Magnetic Resonance Imaging and Spectroscopy.** *MICCAI* 2010, **6363**:666-673.
8. Wu Y, Chang EY, Chang KCC, Smith JR: **Optimal Multimodal Fusion for Multimedia Data Analysis.** *ACM Conference on Multimedia* 2004, 572-579.
9. Freund Y, Schapire RE: **A decision-theoretic generalization of on-line learning and an application to boosting.** *Proceedings of the Second European Conference on Computational Learning Theory* London, UK: Springer-Verlag; 1995, 23-37.
10. Lewis DP, Jebara T, Noble WS: **Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure.** *Bioinformatics* 2006, **22**(22):2753-2760.
11. Hardoon DR, Szedmak S, Shawe-Taylor J: **Canonical correlation analysis: an overview with application to learning methods.** *Neural Comput* 2004, **16**(12):2639-2664.
12. Simonson DG, Stowe JD, Watson CJ: **A Canonical Correlation Analysis of Commercial Bank Asset/Liability Structures.** *Journal of Financial and Quantitative Analysis* 1983, **18**(01):125-140.
13. Chaudhuri K, Kakade SM, Livescu K, Sridharan K: **Multi-View Clustering via Canonical Correlation Analysis.** *Proceedings of the 26th Annual International Conference on Machine Learning* 2009, 129-136.
14. Bie TD, Moor BD: **On the Regularization of Canonical Correlation Analysis.** *ICA 2003* 2003.
15. Gou Z, Fyfe C: **A canonical correlation neural network for multicollinearity and functional data.** *Neural Networks* 2004, **17**(2):285-293.
16. Eaton ML, Perlman MD: **The Non-Singularity of Generalized Sample Covariance Matrices.** *The Annals of Statictics* 1973, **1**(4):710-717.
17. Hoerl AE, Kennard RW: **Ridge Regression: Biased Estimation for Nonorthogonal Problems.** *Technometrics* 1970, **12**:55-67.
18. Gonzalez I, Dejean S, Martin PGP, Baccini A: **CCA: An R Package to Extend Canonical Correlation Analysis.** *Journal of Stat Software* 2008, **23**(12):1-14.
19. Kakade SM, Foster DP: **Multi-View Regression via Canonical Correlation Analysis.** *In Proceedings of Conference on Learning Theory* 2007, 82-96.
20. Cover T, Hart P: **Nearest neighbor pattern classification.** *Information Theory, IEEE Transactions on* 1967, **13**:21-27.
21. Breiman L: **Random Forests.** *Machine Learning* 2001, **45**:5-32.
22. Hotelling H: **Analysis of a complex of statistical variables into principal components.** *Journal of Educational Psychology* 1933, **24**(7):498-520.
23. Wold S, Sjostrom M, Eriksson L: **PLS-regression: a basic tool of chemometrics.** *Chemometrics and Intelligent Laboratory Systems* 2001, **58**(2):109-130.
24. Tiwari P, Kurhanewicz J, Viswanath S, Sridhar A, Madabhushi A: **Multimodal Wavelet Embedding Representation for data Combination (MaWERiC): Integrating Magnetic Resonance Imaging and Spectroscopy for Prostate Cancer Detection.** *NMR in Biomedicine* 2011.
25. Janes KA, Kelly JR, Gaudet S, Albeck JG, Sorger PK, Lauffenburger DA: **Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data.** *Journal of computational biology a journal of computational molecular cell biology* 2004, **11**(4):544-561.
26. Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC: **Natural History of Progression After PSA Elevation Following Radical Prostatectomy.** *JAMA: The Journal of the American Medical Association* 1999, **281**(17):1591-1597.
27. Roberts SG, Blute ML, Bergstralh EJ, Slezak JM, Zincke H: **PSA doubling time as a predictor of clinical progression after biochemical failure following radical prostatectomy for prostate cancer.** *Mayo Clinic Proceedings* 2001, **76**(6):576-81.
28. Pisansky TM, Kozelsky TF, Myers RP, Hillman DW, I Blute M, Buskirk SJ, Cheville JC, Ferrigni RG, Schild SE: **Radiotherapy for Isolated Serum Prostate Specific Antigen Elevation After Prostatectomy For Prostate Cancer.** *The Journal of Urology* 2000, **163**(3):845-850.
29. Chrouser K, Lieber M: **Extended and saturation needle biopsy for the diagnosis of prostate cancer.** *Current Urology Reports* 2004, **5**:226-230.
30. Welch H, Fisher E, Gottlieb D, Barry M: **Detection of prostate cancer via biopsy in the medicare-seer population during the PSA era.** *Journal of the National Cancer Institute* 2007, **99**:1395-1400.
31. Veenstra TD: **Global and targeted quantitative proteomics for biomarker discovery.** *Journal of Chromatography B* 2007, **847**:3-11.
32. Chan DW, Sokoll LJ: **Prostate-specific antigen: update 1997.** *Journal of the International Federation of Clinical Chemistry* 1997, **9**:120-125.
33. Partin AW, Oesterling JE: **The clinical usefulness of percent free-PSA.** *Urology* 1996, **48**:1-3.
34. Gleason DF: **Classification of prostatic carcinomas.** *Cancer Chemother Rep* 1966, **50**:125-128.
35. Stephenson AJ, Kattan MW, Eastham JA, Bianco FJ, Yossepowitch O, Vickers AJ, Klein EA, Wood DP, Scardino PT: **Prostate cancer specific mortality after radical prostatectomy for patients treated in the prostate-specific antigen era.** *Journal of Clinical Oncology* 2009, **27**:4300-4305.
36. Montironi R, Mazzucchelli R, Scarpelli M, Lopez-Beltran A, Fellegara G, Algaba F: **Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: contemporary approach, current clinical significance and sources of pathology discrepancies.** *BJU International* 2005, **95**(8):1146-1152.
37. Allsbrook WC, Mangold KA, Johnson MH, Lane RB, Lane CG, Amin MB, Bostwick DG, Humphrey PA, Jones EC, Reuter VE, Sakr W, Sesterhenn IA, Troncoso P, Wheeler TM, Epstein JI: **Interobserver reproducibility of Gleason grading of prostatic carcinoma: Urologic pathologists.** *Human Pathology* 2001, **32**:74-80.
38. King CR: **Patterns of prostate cancer biopsy grading: Trends and clinical implications.** *International Journal of Cancer* 2000, **90**(6):305-311.
39. Doyle S, Hwang M, Shah K, Madabhushi A, Tomaszewski J, Feldman M: **Automated Grading of Prostate Cancer using Architectural and Textural Image Features.** *IEEE International Symposium on Biomedical Imaging (ISBI)* 2007, 1284-87.
40. Tabesh A, Teverovskiy M, Pang HY, Kumar V, Verbel D, Kotsianti A, Saidi O: **Multifeature Prostate Cancer Diagnosis and Gleason Grading of Histological Images.** *Medical Imaging, IEEE Transactions on* 2007, **26**(10):1366-1378.
41. Sved PD, Gomez P, Manoharan M, Kim SS, Soloway MS: **Limitations Of Biopsy Gleason Grade: Implications For Counseling Patients With Biopsy Gleason Score 6 Prostate Cancer.** *The Journal Of Urology* 2004, **172**:98-102.
42. Fredolini C, Liotta LA, Petricoin EF: **Application of proteomic technologies for prostate cancer detection, prognosis, and tailored therapy.** *Critical Reviews in Clinical Laboratory Sciences* 2010, **47**(3):125-138.
43. Ornstein DK, Tyson DR: **Proteomics for the identification of new prostate cancer biomarkers.** *Urologic Oncology: Seminars and Original Investigations* 2006, **24**(3):231-236.
44. Veenstra TD, Conrads TP, Hood BL, Avellino AM, Ellenbogen RG, Morrison RS: **Biomarkers: Mining the Biofluid Proteome.** *Molecular & Cellular Proteomics* 2005, **4**(4):409-418.
45. Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer PF, Yasui Y, Feng Z, Wright GL: **Serum Protein Fingerprinting Coupled with a Pattern-matching Algorithm Distinguishes Prostate Cancer from Benign Prostate Hyperplasia and Healthy Men.** *Cancer Research* 2002, **62**(13):3609-3614.
46. Al-Ruwaili JA, Larkin SE, Zeidan BA, Taylor MG, Adra CN, Aukim-Hastie CI, Townsend PA: **Discovery of Serum Protein Biomarkers for Prostate Cancer Progression by Proteomic Analysis.** *Cancer Genomics - Proteomics* 2010, **7**(2):93-103.

47. Tolonen TT, Tammela TL, Kujala PM, Tuominen VJ, Isola JJ, Visakorpi T: Histopathological variables and biomarkers enhancer of zeste homologue 2, Ki-67 and minichromosome maintenance protein 7 as prognosticators in primarily endocrine-treated prostate cancer. *BJU International* 2011.

48. Borga M, Landelius T, Knutsson H: A Unified Approach to PCA, PLS, MLR and CCA. 1997, Tech. rep., Report LiTH-ISY-R-1992, ISY, SE-581 83 Linkoping, Sweden.

49. Abdi H: Partial least squares (PLS) regression. *Encyclopedia of Social Sciences Research Methods* 2003, 1-7.

50. Rosipal R, Kramer N: Overview and Recent Advances in Partial Least Squares. *Subspace, Latent Structure and Feature Selection* 2006, **3940**:34-51.

51. Hotelling H: Relations between two sets of variants. *Biometrika* 1936, **28**:321-377.

52. Sun L, Ji S, Ye J: A least squares formulation for canonical correlation analysis. *ICML* 2008, **33**:1024-1031.

53. Vinod HD: Canonical ridge and econometrics of joint production. *Journal of Econometrics* 1976, **4(2)**:147-166.

54. Leurgans SE, Moyeed RA, Silverman BW: Canonical Correlation Analysis when the Data are Curves. *Journal of the Royal Statistical Society Series B (Methodological)* 1993, **55(3)**:725-740.

55. Guo Y, Hastie T, Tibshirani R: Regularized linear discriminant analysis and its application in microarrays. *Biostatistics* 2007, **8**:86-100.

56. Yates RD, Goodman D: *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers* John Wiley and Sons; 2005.

57. Jafari P, Azuaje F: An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors. *BMC Medical Informatics and Decision Making* 2006, **6**:27.

58. Thomas JG, Olson JM, Tapscott SJ, Zhao LP: An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res* 2001, **11**:1227-1236.

59. Hwang D, Schmitt WA, Stephanopoulos G, Stephanopoulos G: Determination of minimum sample size and discriminatory expression patterns in microarray data. *Bioinformatics* 2002, **18**:1184-1193.

60. Foster DP, Kakade SM, Zhang T: Multi-view dimensionality reduction via canonical correlation analysis. *Technical Report TR-2008-4, TTI-Chicago* 2008.

61. Borga M, Friman O, Lundberg P, Knutsson H: Blind Source Separation of Functional MRI Data. *SSBA* 2002.

62. Heaton K, Master S: Peptide Extraction from Formalin-Fixed Paraffin-Embedded Tissue. *Current Protocols in Protein Science, supplement 65, Unit 23.5* 2011.

63. Wisniewski JR, Zougman A, Nagaraj N, Mann M: Universal sample preparation method for proteome analysis. *Nature Methods* 2009, **6(5)**:359-362.

64. Rappsilber J, Mann M, Ishihama Y: Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* 2007, **2(8)**:1896-1906.

65. Doyle S, Feldman M, Tomaszewski J, Shih N, Madabhushi A: Cascaded Multi-Class Pairwise Classifier (CascaMPa) For Normal, Cancerous, And Cancer Confounder Classes In Prostate Histology. *IEEE International Symposium on Biomedical Imaging (ISBI)* 2011, 715-718.

66. Sparks R, Madabhushi A: Novel Morphometric based Classification via Diffeomorphic based Shape Representation using Manifold Learning. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Volume 6363* Springer Verlag, Beijing, China: Springer Verlag; 2010, 658-665.

67. Monaco J, Tomaszewski J, Feldman M, Moradi M, Mousavi P, Boag A, Davidson C, Abolmaesumi P, Madabhushi A: Detection of Prostate Cancer from Whole-Mount Histology Images Using Markov Random Fields. *Workshop on Microscopic Image Analysis with Applications in Biology (in conjunction with MICCAI)* New York, NY; 2008.

68. Monaco J, Tomaszewski J, Feldman M, Hagemann I, Moradi M, Mousavi P, Boag A, Davidson C, Abolmaesumi P, Madabhushi A: High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. *Medical Image Analysis* 2010, **14(4)**:617-629.

69. Basavanhally A, Ganesan S, Agner S, Monaco J, Feldman M, Tomaszewski J, Bhanot G, Madabhushi A: Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. *IEEE Transactions on Biomedical Engineering* 2010, **57**:642-653.

70. Mukherjee S, Tamayo P, Rogers S, Rifkin R, Engle A, Campbell C, Golub TR, Mesirov JP: Estimating Dataset Size Requirements for Classifying DNA Microarray Data. *Journal of Computational Biology* 2003, **10(2)**:119-142.

71. Basavanhally A, Doyle S, Madabhushi A: Predicting Classifier Performance With a Small Training Set: Applications to Computer-Aided Diagnosis and Prognosis. *IEEE International Symposium on Biomedical Imaging (ISBI)* IEEE, Rotterdam, NL: IEEE; 2010, 229-232.