Supervised Multi-View Canonical Correlation Analysis: Fused Multimodal Prediction of Disease Diagnosis and Prognosis

Asha Singanamalli[a], Haibo Wang [a], George Lee [a], Natalie Shih [b], Mark Rosen [b], Stephen Master [b], John Tomasewski [c], Michael Feldman [b], Anant Madabhushi[a],

[a]Case Western Reserve University, Cleveland, OH; [b] University of Pennsylvania, Philadelphia, PA; [c]University of Buffalo, Buffalo, NY.

ABSTRACT

While the plethora of information from multiple imaging and non-imaging data streams presents an opportunity for discovery of fused multimodal, multiscale biomarkers, they also introduce multiple independent sources of noise that hinder their collective utility. The goal of this work is to create fused predictors of disease diagnosis and prognosis by combining multiple data streams, which we hypothesize will provide improved performance as compared to predictors from individual data streams. To achieve this goal, we introduce supervised multi-view canonical correlation analysis (sMVCCA), a novel data fusion method that attempts to find a common representation for multiscale, multimodal data where class separation is maximized while noise is minimized. In doing so, sMVCCA assumes that the different sources of information are complementary and thereby act synergistically when combined. Although this method can be applied to any number of modalities and to any disease domain, we demonstrate its utility using three datasets. We fuse (i) 1.5 Tesla (T) magnetic resonance imaging (MRI) features with cerbrospinal fluid (CSF) proteomic measurements for early diagnosis of Alzheimer's disease (n = 30), (ii) 3T Dynamic Contrast Enhanced (DCE) MRI and T2w MRI for *in vivo* prediction of prostate cancer grade on a per slice basis (n = 33) and (iii) quantitative histomorphometric features of glands and proteomic measurements from mass spectrometry for prediction of 5 year biochemical recurrence post-radical prostatectomy (n = 40). Random Forest classifier applied to the sMVCCA fused subspace, as compared to that of MVCCA, PCA and LDA, yielded the highest classification AUC of 0.82 +/- 0.05, 0.76 +/- 0.01, 0.70 +/- 0.07, respectively for the aforementioned datasets. In addition, sMVCCA fused subspace provided 13.6%, 7.6% and 15.3% increase in AUC as compared with that of the best performing individual view in each of the three datasets, respectively. For the biochemical recurrence dataset, Kaplan-Meier curves generated from classifier prediction in the fused subspace reached the significance threshold (p = 0.05) for distinguishing between patients with and without 5 year biochemical recurrence, unlike those generated from classifier predictions of the individual modalities.

1. INTRODUCTION

Increasing accessibility to multiscale, multimodal biomedical data has begun to pave the way for personalized medicine. In particular, the advent of high-throughput molecular assays has yielded a plethora of molecular markers associated with diagnosis and prognosis in a number of different disease domains.[1] However, few of these markers have translated into clinical practice.[2] Alternatively, quantitative imaging features are now beginning to be considered as potential biomarkers, a term that has most commonly been associated with molecular signatures thus far. As a result, a number of promising quantitative imaging markers such as textural features on T2w magnetic resonance imaging (MRI)[3] and textural kinetic features[4] on dynamic contrast enhanced (DCE) MRI are beginning to emerge for disease characterization (e.g. prostate and breast cancer localization). Availability of multiple, complementary markers and data streams now presents an opportunity to fuse different sources of information in order to potentially improve prediction of disease diagnosis and prognosis as compared to any individual marker or data stream.

Although a number of data fusion strategies have been developed in the context of computer vision,[5] only a few generalized techniques are available for fusion of heterogeneous biomedical data types such as imaging and non-imaging, and structural and functional imaging which present unique challenges. Previous approaches to

Medical Imaging 2014: Biomedical Applications in Molecular, Structural, and Functional Imaging,
edited by Robert C. Molthen, John B. Weaver, Proc. of SPIE Vol. 9038, 903805
© 2014 SPIE · CCC code: 1605-7422/14/$18 · doi: 10.1117/12.2043762

data fusion can generally be categorized based on the level at which information is combined: (i) raw data level (low level fusion), (ii) feature level (intermediate level fusion) or (iii) decision level (high level fusion).[2] Integration at the raw data level is limited to homogeneous data sources and is thus inapplicable for multiscale, biomedical data. Alternatively, decision level strategies[6] bypass challenges associated with fusion of heterogeneous data types by combining independently derived decisions from each data source. In doing so, the information available at the intersection of different data channels may remain unexploited.[7,8] On the other hand, feature level integration involves converting raw data into quantitative feature representations which can then be combined using concatenation based,[9,10] kernel based[11] or dimensionality reduction based methods.[8] These methods transform quantitative features obtained from each data channel into an alternate, joint subspace termed metaspace where a meta-classifier is applied to distinguish between groups of patients with different diagnosis and/or prognosis. However, feature level fusion is complicated by differences in dimensionality as well as the 'curse of dimensionality'.[12] For instance, data sources residing at different scales often have significantly different dimensionalities which render simple concatenation of quantitative features sub-optimal as high dimensional modalities such as 'omics' are likely to dominate the joint-representation on account of the quantity, not necessarily the quality of data it provides.[13] Furthermore, biomedical datasets often comprise small sample size as a result of which concatenation based methods that increase data dimensionality are not suitable as they are prone to the curse of dimensionality.[12] Curse of dimensionality states that the sample size required to build a good predictor increases exponentially with the number of features. Kernel-based methods[11,14,15] alternatively transform raw data from the original space to a high dimensional embedding space where the different data types are more homogeneously represented thereby making them more amenable for fusion. However, such methods are prone to overfitting[16,17] particularly given the small sample size and the noise associated with each of the biomedical data sources which, if unaccounted for, may drown the increase in signal achievable by fusion. As such, we seek a data fusion method that is able to extract information pertinent to the task of interest while accounting for various sources of noise and reducing dimensionality.

Dimensionality reduction methods have emerged as effective means of fusing data.[8,17] Canonical correlation analysis (CCA)[18] is a linear dimensionality reduction method commonly used for data fusion as it accounts for relationships between multiple input variables. By capturing correlations between modalities, CCA seeks to identify the underlying structure common to the two views thereby creating a subspace that is robust to modality-specific noise. As a result, CCA has been popular in the computer vision community for applications in image retrieval from text query,[19] color demosaicing[20] as well as imaging and non-imaging data fusion.[17] Multi-view CCA (MVCCA) has emerged as an extension of traditional CCA for more than two views.[21] MVCCA generalizes CCA by finding the linear subspace where pairwise correlations between multiple (more than two) modalities can be maximized. However, both CCA and MVCCA are unsupervised and thus do not guarantee a subspace that is optimal for class separation. Previously, Golugula et al.[17] have attempted to incorporate supervision into the CCA framework as a regularization step, which although was shown to improve class separability in the reduced subspace, is computationally expensive. Alternatively, previous work has shown that embedding class labels as one of the two variable sets in CCA is equivalent to the supervised linear dimensionality reduction method, linear discriminant analysis (LDA).[22] LDA seeks to find a linear subspace that is optimal for classification by maximizing euclidean distance between classes and minimizing the distance within each class.[23] However, LDA, unlike CCA and MVCCA, is unable to account for relationships between multiple modalities, which may therefore result in overfitting.

In this work, we present a novel supervised multiview canonical correlation analysis (sMVCCA) scheme that combines properties of both MVCCA and LDA to provide a common, low dimensional subspace representation for fusing any number of heterogeneous forms of multidimensional, multimodal biomedical data. sMVCCA simultaneously maximizes correlations between multiple modalities and optimizes class separation by treating class labels as one of the views of MVCCA. In doing so, sMVCCA quantitatively transforms data into an alternate, reduced dimensional subspace that: (i) is able to ignore modality-specific noise thereby retaining information about the object of interest which is (ii) pertinent for the classification task under consideration. As all views capture information pertaining to the same object, information overlap is likely to increase with increasing number of views. While supervision enhances class discriminability in the joint-space, the correlation based representation ensures robustness to noise. We demonstrate the utility of sMVCCA in the context of learning fused predictors of (i) structural MRI and proteomics for early diagnosis of Alzheimer's disease, (ii)

DCE MRI and T2w MRI for *in vivo* determination of prostate cancer grade and (iii) histology and proteomics for prediction of 5-year biochemical recurrence associated with prostate cancer post surgery.

The rest of this paper is organized as follows. Section 2 reviews the theory and background of CCA and MVCCA which is then followed in Section 3 by detailed description of our methodology, sMVCCA. Section 4 describes the experimental design the results of which are presented and discussed in Section 5. We then conclude with a summary of the work and principal findings in Section 6.

2. THEORY AND REVIEW OF CCA AND MVCCA

We briefly introduce canonical correlation analysis (CCA) and its extension multiview canonical correlation analysis (MVCCA), which provide the theoretical framework for supervised MVCCA (sMVCCA). Table 1 lists all the notations used in subsequent formulations for reference.

Symbol	Description
n, N	samples, total number of samples; $n \in \{1, \ldots, N\}$
k, K	modalities, total number of modalities; $\mathbf{x}_k$, $k \in \{1, \ldots, K\}$
m, M_k	features, total number of features in each modality; $m \in \{1, \ldots, M_k\}$
M	total number of features over all modalities; $M = \sum_k M_k$
$\mathbf{x}_k^{n \times M_k}$	sample n described by modality k with M_k features
$\mathbf{x}_k$	data vector of all features $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M]$, $\mathbb{R}^{1 \times M}$
$\mathbf{X}$	concatenated data matrix containing all features from all modalities $[\mathbf{x}_1, \ldots, \mathbf{x}_K]$, $\mathbb{R}^{n \times (M_1 + \ldots + M_K)}$
$\mathbf{w}_k$	weight vector for modality k, $\mathbb{R}^{M_k \times 1}$
$\mathbf{W}_k$	weight matrix for modality k, $\mathbb{R}^{M_k \times n}$
$\mathbf{w}$	concatenated weight vector over all modalities $[\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_K^T]^T$, $\mathbb{R}^{M \times 1}$
$\mathbf{W}_x$	weight matrix for all modalities $[\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_M]$, $\mathbb{R}^{M \times n}$
$\mathbf{Y}$	label matrix $\mathbb{R}^{n \times g}$
g	number of classes
$\mathbf{W}_y$	notation used in sMVCCA to denote $\mathbf{W}$ for all labels $\mathbb{R}^{g \times n}$

Table 1: Summary of Notations

2.1 Canonical Correlation Analysis

Provided a dataset $x_k^{n \times m}$ with $n \in \{1, 2, ..., N\}$ samples and $k \in \{1, 2, ..., K\}$ modalities, each of which comprises $m \in \{1, 2..., M_k\}$ features. Canonical correlation analysis (CCA) considers two sets of variables ($K = 2$), $x_1^{n \times M_1}$ and $x_2^{n \times M_2}$, and projects them onto basis vectors, $\mathbf{w}_1$ and $\mathbf{w}_2$, such that correlation between projections of variables onto these basis vectors are mutually maximized. Formally, this can be expressed as

$$\arg\max_{\mathbf{w}_1, \mathbf{w}_2} \frac{\mathbf{w}_1^T \mathbf{C}_{12} \mathbf{w}_2}{\sqrt{\mathbf{w}_1^T \mathbf{C}_{11} \mathbf{w}_1 \mathbf{w}_2^T \mathbf{C}_{22} \mathbf{w}_2}}, \tag{1}$$

where $\mathbf{C}_{12} \in \mathbb{R}^{M_1 \times M_2}$, $\mathbf{C}_{11} \in \mathbb{R}^{M_1 \times M_1}$, $\mathbf{C}_{22} \in \mathbb{R}^{M_2 \times M_2}$ are covariance matrices of $\mathbf{x_1}$ and $\mathbf{x_2}$, $\mathbf{x_1}$ and $\mathbf{x_1}$, and $\mathbf{x_2}$ and $\mathbf{x_2}$, respectively.

2.2 Multi-View Canonical Correlation Analysis (MVCCA)

Multiview CCA (MVCCA) can be derived by extending the CCA formulation to account for more than two sets of variables ($K > 2$). Since the joint correlation of more than two variables does not formally exist, MVCCA maximizes the sum of correlations between each pair of modalities. Thus, MVCCA can be expressed as generic form of Equation 1.

$$\arg\max_{\mathbf{w}_1, \ldots \mathbf{w}_k \ldots, \mathbf{w}_K} \sum_{k \neq j} \sum \frac{\mathbf{w}_k^T \mathbf{C}_{kj} \mathbf{w}_j}{\sqrt{\mathbf{w}_k^T \mathbf{C}_{kk} \mathbf{w}_k \mathbf{w}_j^T \mathbf{C}_{jj} \mathbf{w}_j}}. \tag{2}$$

The scaling of $\mathbf{w}$ does not affect the arg max solution, allowing Equation 2 to be written as:

$$\begin{aligned} \underset{\mathbf{w}_1,\ldots,\mathbf{w}_K}{\arg\max} \quad & \sum\sum_{k\neq j} \mathbf{w}_k^T \mathbf{C}_{kj} \mathbf{w}_j \\ s.t. \quad & \mathbf{w}_1^T \mathbf{C}_{11} \mathbf{w}_1 = 1, \ldots, \mathbf{w}_K^T \mathbf{C}_{KK} \mathbf{w}_K = 1. \end{aligned} \tag{3}$$

Previously, Equation 2 has been solved by sequentially considering correlations of each pair of variables.[21] However, such an approach is sub-optimal as it requires iterative optimization, which is inefficient and can be susceptible to the order in which pairs of variable sets are chosen. Here, we present an alternative pairwise MVCCA approach by expressing correlations of all modalities in a combined correlation matrix which can be solved using eigenvalue decomposition method.

Letting $\mathbf{w} = [\mathbf{w}_1^T \ \mathbf{w}_2^T ... \mathbf{w}_K^T]^T$, $\mathbf{w} \in \mathbb{R}^{M\times 1}$ allows us to rewrite Equation 3 in a compact matrix form:

$$\begin{aligned} \underset{\mathbf{w}}{\arg\max} \quad & \mathbf{w}^T \bar{\mathbf{C}} \mathbf{w} \\ s.t \quad & \mathbf{w}^T \bar{\mathbf{C}}_d \mathbf{w} = 1 \\ & \mathbf{w}_1^T \mathbf{C}_{11} \mathbf{w}_1 = \ldots = \mathbf{w}_K^T \mathbf{C}_{KK} \mathbf{w}_K, \end{aligned} \tag{4}$$

where

$$\begin{aligned} \bar{\mathbf{C}} = & \begin{bmatrix} \mathbf{0} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1K} \\ \mathbf{C}_{21} & \mathbf{0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{C}_{(K-1)K} \\ \mathbf{C}_{K1} & \cdots & \mathbf{C}_{K(K-1)} & \mathbf{0} \end{bmatrix}, \\ \bar{\mathbf{C}}_d = & \begin{bmatrix} \mathbf{C}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{KK} \end{bmatrix}. \end{aligned} \tag{5}$$

In more general terms where $\mathbf{W} \in \mathbb{R}^{M\times n}$, Equation 4 reduces to

$$\begin{aligned} \underset{\mathbf{W}}{\arg\max} \quad & trace(\mathbf{W}_x^T \bar{\mathbf{C}} \mathbf{W}_x) \\ s.t \quad & \mathbf{W}_x^T \bar{\mathbf{C}}_d \mathbf{W}_x = \mathbf{I} \\ & \mathbf{w}_1^T \mathbf{C}_{11} \mathbf{w}_1 = \ldots = \mathbf{w}_K^T \mathbf{C}_{KK} \mathbf{w}_K, \end{aligned}$$

where $\mathbf{I}$ is an $n \times n$ identity matrix and the weight matrix is defined as $\mathbf{W}_x = [\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_K] \in \mathbb{R}^{M\times n}$

3. SUPERVISED MULTI-VIEW CANONICAL CORRELATION ANALYSIS (SMVCCA)

Although MVCCA subspace provides information about the underlying object, it does not guarantee a representation that is optimal for class separation. We hereby present supervised MVCCA (sMVCCA) that explicitly accounts for class labels and thus attempts to provide fused representation that selectively captures discriminative information of the underlying object. Previous work has shown that LDA is a special case of CCA where the correlation between data samples $\mathbf{X}$ with corresponding class labels $\mathbf{Y}$ are maximized.[22] sMVCCA leverages this idea with pairwise MVCCA to improve class seperability.

3.1 Formulation

We define our data in a compact matrix form as $\mathbf{X} \in \mathbb{R}^{n \times M}$. We extend the MVCCA formulation to incorporate an additional term that maximizes the correlation of $\mathbf{X}$ with class labels $\mathbf{Y}$, which can be expressed as follows:

$$\begin{aligned}
\underset{\mathbf{W}_x,\mathbf{W}_y}{\arg\max} \quad & trace(\mathbf{W}_x^T \bar{\mathbf{C}} \mathbf{W}_x) + 2 \times trace(\mathbf{W}_x^T \mathbf{X}^T \mathbf{Y} \mathbf{W}_y) \\
= \quad & trace(\left[\begin{array}{cc} \mathbf{W}_x^T & \mathbf{W}_y^T \end{array} \right] \left[\begin{array}{cc} \bar{\mathbf{C}} & \mathbf{X}^T\mathbf{Y} \\ \mathbf{Y}^T\mathbf{X} & \mathbf{0} \end{array} \right] \left[\begin{array}{c} \mathbf{W}_x \\ \mathbf{W}_y \end{array} \right]) \\
= \quad & trace(\hat{\mathbf{W}}^T \hat{\mathbf{C}} \hat{\mathbf{W}})
\end{aligned}$$

$s.t.$

$$\left[\begin{array}{cc} \mathbf{W}_x^T & \mathbf{W}_y^T \end{array} \right] \left[\begin{array}{cc} \bar{\mathbf{C}}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}^T\mathbf{Y} \end{array} \right] \left[\begin{array}{c} \mathbf{W}_x \\ \mathbf{W}_y \end{array} \right] = \mathbf{I}$$

$$\Leftrightarrow \hat{\mathbf{W}}^T \hat{\mathbf{C}}_d \hat{\mathbf{W}} = \mathbf{I} \tag{6}$$

$$\mathbf{W}_1^T \mathbf{C}_{11} \mathbf{W}_1 = \ldots = \mathbf{W}_K^T \mathbf{C}_{KK} \mathbf{W}_K = \mathbf{W}_y^T \mathbf{Y}^T \mathbf{Y} \mathbf{W}_y. \tag{7}$$

where $\mathbf{Y}$ is a matrix in which class labels are encoded using Soft-1-of-Class strategy.[22]

Solving Equation 6 consists of two steps: (i) Ignoring the constraint in (7) leaves us with a quadratic programming problem, whose $\mathbf{W}^*$ corresponds to eigenvectors of the n-largest eigenvalues of a generalized eigenvalue system: $\mathbf{C_{xy}W} = \lambda \mathbf{C_{d_{xy}}W}$; (ii) Imposing constraint (7) upon obtaining the optimal eigenvectors $\mathbf{W}^*$ by normalizing the corresponding section of each modality: $\mathbf{W_j^{**}} = \mathbf{W_j^*}(\mathbf{W_j^{*T}C_{jj}W_j^*})^{-\frac{1}{2}}, j = 1, ..., k$.

4. EXPERIMENTAL DESIGN

To evaluate the presented sMVCCA method, we chose three unique datasets that enabled us to address some of the most relevant clinical problems in two different disease domains. Fusion tasks for the three datasets considered in this work can be categorized as (1) Radiology-Proteomics fusion (2) Structural-Functional data fusion and (3) Histomorphometric-Proteomics fusion. In each case, the objective was to develop a fused predictor with a higher predictive performance as compared to that of individual data streams.

4.1 Dataset 1: Radiology-Proteomics Fusion for Early Diagnosis of Alzheimer's Disease

Structural T1w MRI and cerebrospinal fluid (CSF) proteomic measurements were acquired for 30 adults between the ages of 55 and 90, among whom 12 were diagnosed with Alzheimer's disease while the remaining 18 were healthy volunteers. This data was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database*. Provided that Alzheimer's is an irreversible disease, early detection of the disease may provide an opportunity to develop new treatments that may extend the patient's life of quality.

Structural T1w MRI scans were acquired from 1.5T scanners at multiple sites across United States and Canada. The imaging sequence was a 3-dimensional saggital magnetization prepared rapid gradient-echo (MPRAGE). Additional details on the acquisition and pre-processing of MRI scans can be found in.[25] The pre-processed MRI scans were subjected to FreeSurfer, a documented, freely available image analysis suite†, to extract features from 34 cortical ROIs in each hemisphere using the atlas detailed in Morra et al.[26]For each ROI, the cortical thickness average (TA), standard deviation of thickness (TS), surface area (SA) and cortical volume (CV) were calculated as features. SA was calculated as the area of the surface layer equidistant between the gray/white matter and gray matter/CSF surfaces. CV at each vertex over the whole cortex was computed by the product of the SA and thickness at each surface vertex. Left and right hemisphere SA and total intracranial volume (ICV) were also included. For each subcortical structure, the subcortical volume (SV) was extracted. A number of these features were previously found to be correlated with neurodegenerative processes associated with Alzheimer's disease.[27]

*http://www.loni.ucla.edu/ADNI

†http://surfer.nmr.mgh.harvard.edu/

Additionally, CSF proteomic biomarkers have previously shown promising results for diagnosis of Alzheimer's disease.[28] As such, we consider patients who have CSF proteomic measurements in addition to T1w MRI scans. For these patients, baseline CSF samples were obtained through lumbar puncture at all participating sites. Detailed protocols of CSF collection and transportation have previously been reported[28] and is available on the ADNI website[†]. A total of 80 CSF concentrations of different proteins (such as Adiponectin, Angiopoietin, and Cortisol) were collected. Examples of features extracted across T1w MRI and protein expression data are provided in Table 2.

4.2 Dataset 2: Structural-Functional Imaging Fusion for Prostate Cancer Grading

T2w MRI and DCE MRI were acquired prior to radical prostatectomy (RP) for 16 patients with biopsy confirmed prostate cancer. Provided that this dataset comprised intermediate Gleason score patients, the objective was to distinguish between primary Gleason grades 3 and 4 on MRI at a per slice basis. As such, we considered 2-3 MRI slices containing the most dominant tumor nodule in each patient, totaling 33 slices over 16 patients.

All MRI studies were performed on a 3T scanner (Verio, Siemens; Erlangen, GE) using a dedicated endorectal coil (Medrad, Pittsburgh, PA). Axial T1 and T2w imaging was performed with 3 mm slice thickness and 1.0 mm gap. DCE MRI was performed with T1w VIBE imaging at 3 mm slice thickness, with 24 cm FOV and matrix 256 by 192. Temporal resolution varied based on the number of prescribed slices. Twenty phases of imaging were performed, with IV gadolinium injection beginning 30 seconds after scan initialization. Surgical specimens were fixed in formalin and were subsequently sectioned into 3-4 mm slices, each of which was sectioned into 4 quadrants, stained with H&E and digitized at 20x magnification using Aperio scanner. An expert pathologist provided cancer annotations and determined the Gleason grades on each slice. 2-3 slices with the largest dominant tumor nodule was selected for analysis in each case. Ground truth cancer annotations were mapped from histologic sections onto MRI protocols via co-registration.[29] Closest corresponding sections between the histologic and T2w MRI slices were determined by radiologist and pathologist. Manually selected landmarks were used to co-register histologic slices with corresponding T2w MRI and DCE MRI slices using thin plate splines (TPS), which then allowed for mapping of the tumor on MRI.[29]

Textural features and kinetic features from T2w and DCE MRI, respectively were extracted from tumor voxels as detailed in Table 2. Previous work[3] has shown that textural features are able to distinguish between cancerous and benign voxels on T2w MRI. We extract the same features to distinguish between aggressive and indolent tumors as we anticipate that textural features, which generally capture heterogeneity in local neighborhoods, will reflect the heterogeneity of the tumor, a characteristic known to be associated with aggressive tumors. To complement textural features, we extract kinetic features from signal intensity vs. time curves, which were previously shown to be associated with Gleason grades[30] and a number of quantitative microvessel attributes.[29]

4.3 Dataset 3: Histomorphometry-Proteomics fusion for Early Prediction of 5-year Biochemical Recurrence in Prostate Cancer

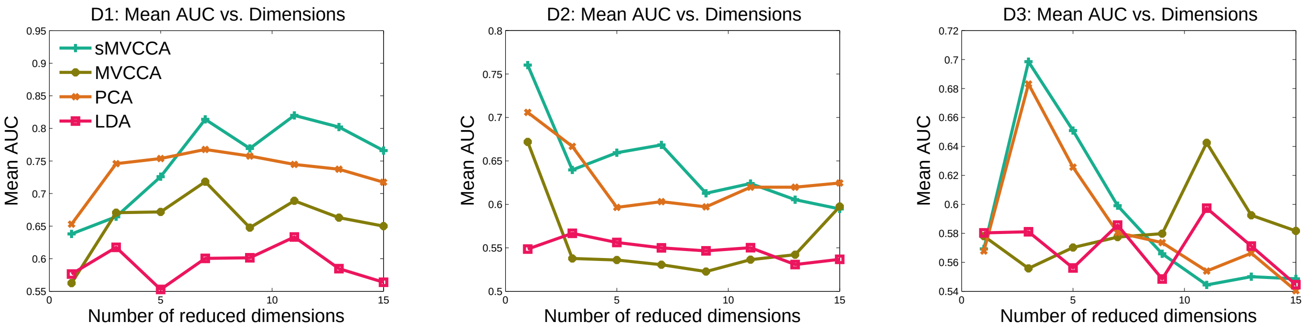
40 biopsy confirmed prostate cancer patients with intermediate Gleason scores underwent radical prostatectomy. Patients were followed up and monitored for 5 years. Among all the patients, 21 experienced biochemical recurrence within 5 years of surgery while the other 19 did not experience biochemical recurrence.

Surgical specimens were sectioned and a representative slice containing the most dominant tumor nodule in each specimen was digitized at 20x magnification. Representative tumor areas as determined and annotated by a pathologist on H&E sections were collected via needle dissection, and formalin cross-links were removed by heating at 99 degree Celsius. After peptide purification, samples were analyzed using C-18 reverse phase liquid chormatography/tandem mass spectrometry (nLC-MS/MS) on a LTQ Orbitrap mass spectrometer. Following data acquisition, a label free MaxQuant peptide identification package was used to extract ion chromatograms allowing for quantification of protein abundance. Proteins quantifiable in at least 50% of the studies were considered which thereby resulted in 650 proteomic expression values for each patient. Data imputation methods were used to replace missing values.

Proteomic expression values resulting from MaxQuant analysis of the raw mass spectrometry data was considered for analysis. On histology, previous work[31,32] has shown that quantitative histomorphometric features of glands may be able to predict the aggressiveness of tumor. As such, quantitative histomorphometric features

Dataset	Modality	Feature Type (num)	Examples/ Description
D1	T1w MRI	34 ROIs extracted (30)	cortical thickness average, standard deviation of surface area (SA), cortical volume, left and right hemisphere SA, and total intracranial volume
	Proteomics	Proteomics obtained from CSF (83)	Fatty Acid-Binding Protein, Resistin, Interleukin-3, Vascular Endothelial Growth Factor
D2	T2w MRI	Gradient & Gray-level statistics (25)	Features capturing summary statistics such as mean, standard deviation and derivative features of pixel values within a localized neighborhood computed using Sobel and Kirsch filters.
		Gabor wavelet transform (48)	Textural representation obtained via convolution of an image with Gabor filter bank, which comprises filters with different frequencies and orientations.
		Haralick statistics (39)	Statistics of gray-level co-occurrence matrices such as angular second moment, contrast and difference entropy.
	DCE MRI	Per-voxel kinetic curve statistics (24)	Statistics such as mean, median and variance from characteristics of the signal-intensity vs. time curves including maximum uptake and rate of washout computed over all tumor voxels
		ROI Modified Standard Logistic Fitted (MSLF) SI-Time Curve (14)	Signal intensity vs. time curves of all pixels within the tumor region were averaged and fitted to a modified standard logistic function. Features including the curve fitting parameters, maximum uptake, rate of washout and initial area under the curve were computed from this single summary kinetic curve.
D3	Histology	Gland Morphology (100)	Statistics of gland area, boundary length, distance, perimeter, smoothness, fractal dimensions and descriptors of invariant moments and Fourier transforms.
		Gland Architecture (51)	Statistics of graphical constructs such as Voronoi diagram, Delaunay Triangulation and Minimum Spanning Tree where gland centroids serve as nodes thereby capturing characteristics of global glandular distribution.
		Co-occurring Gland Tensors (39)	Gland orientation is quantified by measuring the angle of the principal axis of segmented gland boundaries following. Statistics of co-occurrence matrices that capture gland directionality in local neighborhoods then serve as features.
		Gland Subgraphs (26)	Statistics such as eccentricity and connected component coefficients of local subgraphs of gland distributions constructed using probabilistic decay function.
		Haralick Texture (26)	Second order statistics computed from a symmetric co-occurrence matrix of neighboring pixel intensities within a given window size around a pixel.
	Proteomics	Mass Spectrometry protein measurements (650)	Expression values of proteins that were expressed in more than 50% of the samples which included heat shock protein, 40S ribosomal protein and a number of Ras proteins.

Table 2: Summary of features extracted from the various modalities across datasets



(a) (b) (c)

Figure 1: Mean AUC as a function of the number of dimensions in reduced subspace for (a) dataset 1, (b) dataset 2, and (c) dataset 3

capturing gland morphology, orientation as well as local and global architecture were extracted.[17,31,32] Summary of extracted features is provided in Table 2.

4.4 Feature Selection

Wilcoxon rank sum test (WRST) was used to select features from all modalities within each dataset. Features were ranked using training samples within each cross validation fold based on their p-values, where the feature with the lowest p-value was ranked the highest. The number of top features to retain was empirically determined separately for each dataset.

4.5 Experimental Evaluation

Top ranked features from all modalities in each dataset were transformed into a reduced dimensional subspace via sMVCCA or other comparative strategies, which included MVCCA, LDA and PCA. Note that unlike sMVCCA and MVCCA, PCA and LDA are designed for single feature set. As such, selected features from all modalities were concatenated into a single input matrix prior to the application of PCA and LDA. In the reduced subspace, random forest(RF) classifier was used to evaluate the various fused and individual modality representations. RF is a widely used, well-established decision tree ensemble method that combines outputs of multiple decision trees. Three fold cross validation was performed for datasets 1 and 2, and ten fold stratified cross validation was performed for dataset 3 over 10 trials.

Experiments 1, 2 and 3 were conducted to (i) explore the effect of parameters associated with fused representations (ii) determine the value of considering the relationship between modalities (as in CCA and MVCCA) as well as relationship with class labels (as in LDA) as is done by our method, sMVCCA, and to (iii) test our hypothesis that combination of multiple sources of information yields better predictive power than any individual data source alone, respectively.

4.5.1 Experiment 1: Exploration of predictive performance vs. number of reduced dimensions

Dimensionality of reduced subspace is the only parameter that requires tuning to compute sMVCCA, MVCCA, PCA and LDA fused representations. Thus, classification performance was evaluated across a range of dimensions.

4.5.2 Experiment 2: Comparing sMVCCA vs. other linear dimensionality reduction methods for fusion

At the dimensionality providing the highest performance in Experiment 1, which we will denote as d^*, sMVCCA was compared with other supervised and unsupervised linear dimensionality reduction based fusion methods which included MVCCA, PCA and LDA. In comparing sMVCCA with MVCCA and LDA, we test our assumption that considering associations between modalities as well as with class labels improves predictive power over considering either one of the two criteria individually.

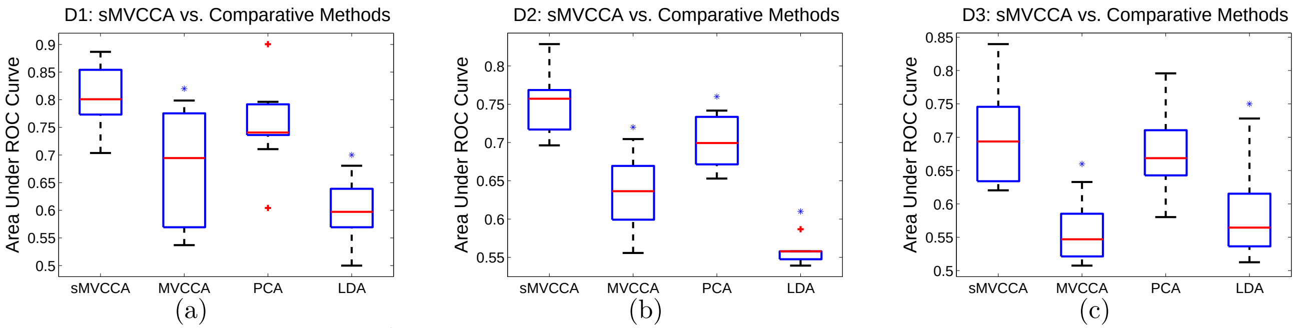


Figure 2: Box-and-whisker plots of AUC obtained over 10 runs of three, three and ten fold cross validations for datasets 1, 2 and 3, respectively using a random forest classifier on sMVCCA, MVCCA, PCA and LDA subspaces. The lower and upper bounds of each box indicate the 25th and 75th percentile of AUC whereas the red bar indicates the median AUC values. The dashed lines extend from the box to the maximum and minimum values. The red plus signs refer to outliers while the blue asterisk indicates statistically significant difference in AUC from that of sMVCCA, as determined by Tukey honest significance difference criterion.

4.5.3 Experiment 3: Comparing sMVCCA fused representation vs. individual modalities

At dimensionality d^*, RF classifier performance in the sMVCCA subspace was compared against that of individual modalities. For evaluation of individual modalities, raw features from each modality were first reduced to a PCA reduced subspace of d^* dimensions where the classifier was applied.

4.5.4 Performance Metrics

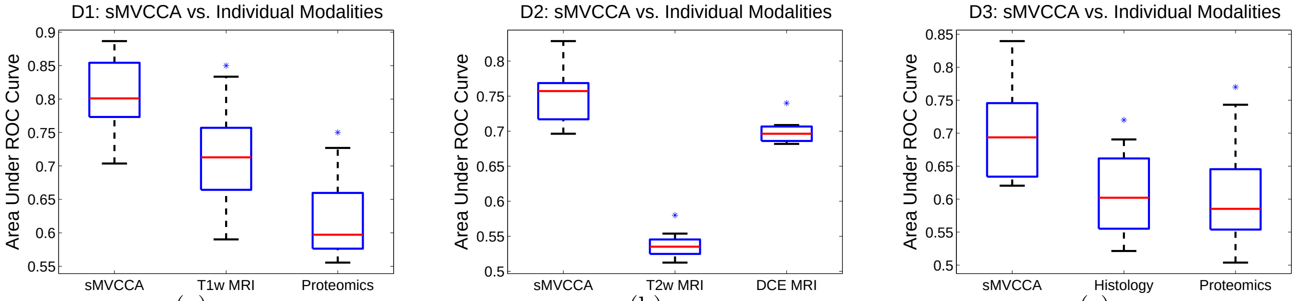
For all datasets, area under the curve (AUC) were computed over all folds. The mean and standard deviation of AUC were evaluated across 10 trials. AUCs across experimental conditions were compared via one-way analysis of variance (ANOVA), which tested the null hypothesis that the means of AUC across all experimental conditions were equal. An alpha of 0.05 was used to reject the null hypothesis. Following ANOVA, post-hoc test was performed using Tukey's honest significant difference (HSD) criterion to determine the means that were significantly different from that of sMVCCA.

For D3, time to recurrence was available for 30 out of 40 patients. For these patients, Kaplan-Meier (KM) analysis with logrank significance test was used to evaluate the predictability of biochemical recurrence free survival using the individual and sMVCCA combined modalities. KM curves provide an alternate, independent measure of performance that allowed us to better assess which patients were being misclassified by accounting for the time to recurrence. In general, we would expect that more errors would occur in predicting the early recurrence patients and thereby would have overlapping recurrence and non-recurrence KM curves at earlier time points. The goal however is to correctly predict both early and late recurrence patients which would result in non-overlapping KM curves with significantly different trajectories for the recurrence and non-recurrence groups.

5. RESULTS AND DISCUSSION

5.1 Experiment 1: Exploration of predictive performance vs. number of reduced dimensions

Figure 1 shows mean AUC as a function of the number of reduced dimensions. While D1 has a slow trajectory upward and reaches a plateau after the first few dimensions, AUC values in D2 peak at the first dimension after which they reach a plateau. D3 shows a different trajectory altogether where the AUC peaks within the first few dimensions after which it quickly drops significantly. PCA closely follows the path of sMVCCA particularly for D3 which indicates that direction of correlation across various modalities is the same as the direction of variance within the data, suggesting that the modalities in D3 may be highly redundant.



(a) (b) (c)

Figure 3: Box-and-whisker plots of AUC obtained over 10 runs of three, three and ten fold cross validations for datasets 1, 2 and 3, respectively using a random forest classifier on sMVCCA fused subspace and the individual subspaces. The lower and upper bounds of each box indicate the 25th and 75th percentile of AUC whereas the red bar indicates the median AUC values. The dashed lines extend from the box to the maximum and minimum values. The red plus signs refer to outliers while the blue asterisk indicates statistically significant difference in AUC from that of sMVCCA, as determined by Tukey honest significance difference criterion.

5.2 Experiment 2: Comparing sMVCCA vs. other linear dimensionality reduction methods for data fusion

Figure 2 summarizes the performance of sMVCCA and comparative fusion strategies at the dimensionality that provided the highest mean AUC in Experiment 1. sMVCCA achieves the highest classification AUC for all three datasets while PCA consistently emerges as the next best performing method. The improved performance of sMVCCA over that of PCA is evident in D2, as indicated by the blue asterisk which denotes statistically significant difference in AUC values as compared to that of sMVCCA. Although the difference in performance between sMVCCA and PCA is not significant in other datasets, we would like to note that, for D1, AUCs derived from PCA comprise outliers which suggest the unreliability of PCA provided fused embedding.

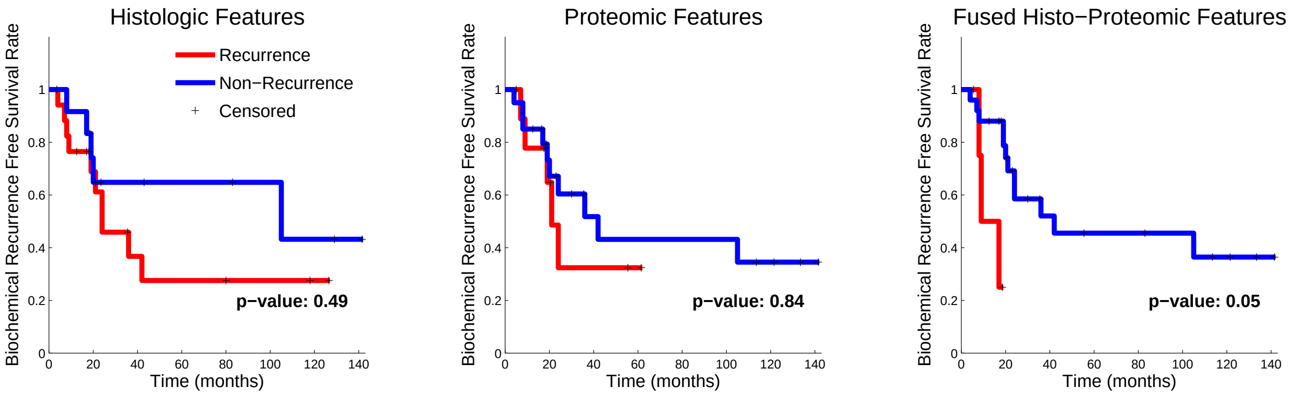
As compared to MVCCA, sMVCCA shows significantly higher performance across all datasets, suggesting that supervising the construction of correlated subspace is likely to improve class discriminability. At the same time, we note that the supervised comparative method, LDA, where features from all modalities are concatenated prior to computing the low dimensional embedding, has has significantly lower AUCs across all datasets. This in turn emphasizes the importance of intelligently combining heterogeneous data streams while exposing label information so as to avoid over-fitting.

5.3 Experiment 3: Comparing sMVCCA fused representation vs. individual modalities

Figure 3 shows the distribution of AUCs achieved by the sMVCCA fused subspace as well the individual views, reduced to the number of dimensions that achieved maximum AUC value in Figure 1. Classification in the sMVCCA fused subspace consistently results in significantly better predictive performance as compared to individual modalities across all datasets. The three datasets achieve 13.6%, 7.6% and 15.3% increase in mean AUC in the sMVCCA subspace as compared with that of the best performing individual view in D1, D2 and D3, respectively.

Unlike performances of individual views in D2, which appear to be highly different with respect to each other, the individual views in D1 and D3 show similar performances. Although ANOVA indicated that the performance of sMVCCA and the individual views were statistically significant, post-hoc pairwise comparison test using Tukey's honest significance difference indicated that the performance of the individual views in D1 and D3 were not significantly different from each other. Nonetheless, fusion appears to marginally but significantly improve the classification AUC in these datasets. Note that, no significant differences were found between PCA and sMVCCA in the same two datasets which suggests that although sMVCCA is driven by correlation or redundancies across views, it possibly converges to PCA when the views are highly redundant.

Furthermore, Kaplan-Meier analysis of 5 year biochemical recurrence free survival in D3 showed that the fused representation was better able to distinguish between the biochemical recurrence and non-recurrence groups as compared to the individual modalities. As shown in Figure 4, close to significant (p=0.05) differences were found



(a) (b) (c)

Figure 4: Kaplan Meier Analysis of biochemical recurrence free survival rate using histologic features, proteomic features and fused features in sMVCCA subspace.

between the Kaplan-Meier curves generated from the predicted biochemical recurrence and non-recurrence groups when both histology and proteomic data were fused using sMVCCA, whereas no significant differences were found when features from a single modality were used for classification.

6. CONCLUDING REMARKS

In this work, we introduced a novel supervised multi-view canonical correlation analysis (sMVCCA) method for multimodal data fusion in the context of combining (i) radiology and proteomics for early diagnosis of Alzheimer's disease, (ii) structural and functional MRI for prostate cancer grading, and (iii) histomorphometry and proteomics for early prediction of 5-year biochemical recurrence post radical prostatectomy. sMVCCA leverages associations between multiple modalities as well as with class labels to provide a fused low dimensional representation that captures the most discriminatory attributes of the underlying biological state, as reflected in the various data channels. In the experimental evaluation, sMVCCA was compared against other linear dimensionality reduction based fusion methods to determine the optimal joint-subspace for classification. In addition, we evaluated if sMVCCA fused subspace provides improved class discriminability as compared to the individual modalities. The following principal findings were discovered as a result of our experimental evaluation:

- Considering relationships (i) between modalities as well as (ii) with class label, as is done by sMVCCA, yields a more predictive subspace than considering either one of the two criteria alone
- Fused representation provides greater predictive power as compared to any individual modality

Although this work introduces a promising platform for quantitative fusion of heterogeneous data channels, the work is limited in a number of ways. All datasets used have small sample sizes and provide two modalities. One of the strengths of sMVCCA is that it is able to fuse any number of data channels, a property that remains experimentally unexplored on account of the datasets considered. In addition, sMVCCA representation is dependent on the input features from each modality, which were selected using a feature selection strategy. For datasets with small sample size, it is well known that feature selection strategies provide less than optimal features sets[33] which is likely to result in a sub-optimal fused subspace. Despite these limitations, current findings indicate that sMVCCA provides a promising framework for fusion of multiscale, multimodal data and that it may be important to incorporate the properties of sMVCCA in future biomedical data fusion strategies.

7. ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award numbers R01CA136535-01, R01CA140772-01, and R21CA167811-01; the National Institute of Diabetes and Digestive and Kidney Diseases under award number R01DK098503-02, the DOD Prostate

Cancer Synergistic Idea Development Award (PC120857); the QED award from the University City Science Center and Rutgers University, the Ohio Third Frontier Technology development Grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. In addition, data collection and sharing for the Alzheimer's Disease Dataset was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from a number of associations and companies [‡].

REFERENCES

[1] Lee, J. W., Figeys, D., and Vasilescu, J., "Biomarker assay translation from discovery to clinical studies in cancer drug development: quantification of emerging protein biomarkers," *Advances in cancer research* **96**, 269–298 (2006).

[2] Kern, S. E., "Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures," *Cancer research* **72**(23), 6097–6101 (2012).

[3] Viswanath, S., Bloch, N., Chappelow, J., Toth, R., Rofsky, N., Genega, E., Lenkinski, R., and Madabhushi, A., "Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 tesla endorectal, in vivo T2-weighted MR imagery," *Journal of Magnetic Resonance Imaging* (2012).

[4] Agner, S. C., Soman, S., Libfeld, E., McDonald, M., Thomas, K., Englander, S., Rosen, M. A., Chin, D., Nosher, J., and Madabhushi, A., "Textural kinetics: a novel dynamic contrast-enhanced (dce)-mri feature for breast lesion classification," *Journal of Digital Imaging* **24**(3), 446–463 (2011).

[5] Khaleghi, B., Khamis, A., Karray, F. O., and Razavi, S. N., "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion* (2011).

[6] Rohlfing, T. and Maurer, C. R., "Multi-classifier framework for atlas-based image segmentation," *Pattern Recognition Letters* **26**(13), 2070–2079 (2005).

[7] Tiwari, P., Viswanath, S., Lee, G., and Madabhushi, A., "Multi-modal data fusion schemes for integrated classification of imaging and non-imaging biomedical data," in [*Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*], 165–168, IEEE (2011).

[8] Lee, G., Doyle, S., Monaco, J., Madabhushi, A., Feldman, M. D., Master, S. R., and Tomaszewski, J. E., "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," in [*Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*], 77–80, IEEE (2009).

[9] Verma, R. Zacharaki, E. O. Y. a. a., "Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of mr images," *Academic Radiology* **15(8)**, 966–977 (2008).

[10] Chan, I., Wells III, W., Mulkern, R. V., Haker, S., Zhang, J., Zou, K. H., Maier, S. E., and Tempany, C. M., "Detection of prostate cancer by integration of line-scan diffusion, t2-mapping and t2-weighted magnetic resonance imaging; a multichannel statistical classifier," *Medical physics* **30**, 2390 (2003).

[11] Tiwari, P., Kurhanewicz, J., Rosen, M., and Madabhushi, A., "Semi supervised multi kernel (SeSMiK) graph embedding: identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy," *MICCAI* **13**(Pt 3), 666–73 (2010).

[12] Bellman, R. E., [*Adaptive control processes: a guided tour*], vol. 4, Princeton university press Princeton (1961).

[13] Madabhushi, A., Agner, S., Basavanhally, A., Doyle, S., and Lee, G., "Computer-aided prognosis: predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data," *Comput. Med. Imaging and Graph.* **35**(7-8), 506–14 (2011).

[14] Lanckriet, G. R., Deng, M., Cristianini, N., Jordan, M. I., Noble, W. S., et al., "Kernel-based data fusion and its application to protein function prediction in yeast.," in [*Pacific symposium on biocomputing*], **9**, 300–311 (2004).

[‡] http://adni.loni.usc.edu/about/funding/

[15] McFee, B., Galleguillos, C., and Lanckriet, G., "Contextual object localization with multiple kernel nearest neighbor," *Image Processing, IEEE Transactions on* **20**(2), 570–585 (2011).

[16] Lewis, D. P., Jebara, T., and Noble, W. S., "Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure," *Bioinformatics* **22**(22), 2753–2760 (2006).

[17] Golugula, A., Lee, G., Master, S. R., Feldman, M. D., Tomaszewski, J. E., Speicher, D. W., and Madabhushi, A., "Supervised Regularized Canonical Correlation Analysis: integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery," *BMC Bioinformatics* **12**, 483 (2011).

[18] Hotelling, H., "Relations between two sets of variates," *Biometrika* **28**(3/4), 321–377 (1936).

[19] Hardoon, D. R., Szedmak, S., and Shawe-Taylor, J., "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation* **16**(12), 2639–2664 (2004).

[20] Hel-Or, Y., "The canonical correlations of color images and their use for demosaicing," *HP Laboratories Israel, Tech. Rep. HPL-2003-164R1* (2004).

[21] Kettenring, J. R., "Canonical analysis of several sets of variables," *Biometrika* **58**(3), 433–451 (1971).

[22] Sun, T. and Chen, S., "Class label versus sample label-based cca," *Applied Mathematics and computation* **185**(1), 272–283 (2007).

[23] Fisher, R. A., "The use of multiple measurements in taxonomic problems," *Annals of eugenics* **7**(2), 179–188 (1936).

[24] Bartlett, M. S., "Further aspects of the theory of multiple regression," in [*Proceedings of the Cambridge Philosophical Society*], **34**, 33–40, Cambridge Univ Press (1938).

[25] Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L Whitwell, J., Ward, C., et al., "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging* **27**(4), 685–691 (2008).

[26] J., M., Z., T., L., A., A., G., C., A., S., M., N., P., X., H., A., T., C., J., M., W., and P., T., "Validation of a fully automated 3d hippocampal segmentation method using subjects with alzheimer's disease mild cognitive impairment, and elderly controls.," *Neuroimage* **43**, 59–68 (2008).

[27] Dickerson, B. C., Feczko, E., Augustinack, J. C., Pacheco, J., Morris, J. C., Fischl, B., and Buckner, R. L., "Differential effects of aging and alzheimer's disease on medial temporal lobe cortical thickness and surface area," *Neurobiology of aging* **30**(3), 432–440 (2009).

[28] Shaw, L. M., Vanderstichele, H., Knapik-Czajka, M., Clark, C. M., Aisen, P. S., Petersen, R. C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., et al., "Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects," *Annals of neurology* **65**(4), 403–413 (2009).

[29] Singanamalli, A., Sparks, R., Rusu, M., Shih, N., Ziober, A., Tomaszewski, J., Rosen, M., Feldman, M., and Madabhushi, A., "Identifying in vivo dce mri parameters correlated with ex vivo quantitative microvessel architecture: A radiohistomorphometric approach," in [*SPIE Medical Imaging*], 867604–867604, International Society for Optics and Photonics (2013).

[30] Vos, E. K., Litjens, G., Kobus, T., Hambrock, T., Kaa, C. A., Barentsz, J. O., Huisman, H., and Scheenen, T. W., "Assessment of prostate cancer aggressiveness using dynamic contrast-enhanced magnetic resonance imaging at 3 t," *European urology* (2013).

[31] Lee, G., Sparks, R., Ali, S., Madabhushi, A., Feldman, M. D., Master, S., Shih, N., and Tomaszewski, J., "Co-occurring gland tensors in localized cluster graphs: Quantitative histomorphometry for predicting biochemical recurrence for intermediate grade prostate cancer," in [*Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*], 113–116, IEEE (2013).

[32] Lee, G., Ali, S., Veltri, R., Epstein, J. I., Christudass, C., and Madabhushi, A., "Cell orientation entropy (core): Predicting biochemical recurrence from prostate cancer tissue microarrays," in [*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*], 396–403, Springer (2013).

[33] Sima, C. and Dougherty, E. R., "What should be expected from feature selection in small-sample settings," *Bioinformatics* **22**(19), 2430–2436 (2006).