Consensus-Locally Linear Embedding (C-LLE): Application to Prostate Cancer Detection on Magnetic Resonance Spectroscopy

Pallavi Tiwari[1], Mark Rosen[2], and Anant Madabhushi[1]

[1] Department of Biomedical Engineering, Rutgers University, USA, [*]
`pallavit@eden.rutgers.edu, anantm@rci.rutgers.edu`,
[2] Department of Radiology, University of Pennsylvania, USA.

Abstract. Locally Linear Embedding (LLE) is a widely used non-linear dimensionality reduction (NLDR) method that projects multi-dimensional data into a low-dimensional embedding space while attempting to preserve object adjacencies from the original high-dimensional feature space. A limitation of LLE, however, is the presence of free parameters, changing the values of which may dramatically change the low dimensional representations of the data. In this paper, we present a novel Consensus-LLE (C-LLE) scheme which constructs a stable consensus embedding from across multiple low dimensional unstable LLE data representations obtained by varying the parameter (κ) controlling locally linearity. The approach is analogous to Breiman's Bagging algorithm for generating ensemble classifiers by combining multiple weak predictors into a single predictor. In this paper we demonstrate the utility of C-LLE in creating a low dimensional stable representation of Magnetic Resonance Spectroscopy (MRS) data for identifying prostate cancer. Results of quantitative evaluation demonstrate that our C-LLE scheme has higher cancer detection sensitivity (86.90%) and specificity (85.14%) compared to LLE and other state of the art schemes currently employed for analysis of MRS data.

1 Introduction

Due to inherent non-linearities in biomedical data, non-linear dimensionality reduction (NLDR) schemes such as Locally Linear Embedding (LLE) have begun to be employed for data analysis and visualization. LLE [1] attempts to preserve geodesic distances between objects, while projecting the data from the high to the low dimensional feature spaces unlike linear dimensionality reduction (DR) schemes such as Principal Component Analysis (PCA) which preserve the Euclidean distances between objects. The low dimensional data representations obtained via LLE are a function of κ, a parameter controlling the size of the neighborhood within which local linearity is assumed and used to approximate geodesic distances. Since LLE is typically used in an unsupervised context for visualizing and identifying object

clusters, *a priori*, the optimal value of κ is not-obvious. In [1], Roweis and Saul suggest that varying κ over a wide range of values, still yields stable, consistent low dimensional embeddings for dense synthetic datasets. Our own experiments on real biomedical data, suggest otherwise [2]. Further, for sparsely populated datasets, the most common failure of LLE is to map faraway points to adjacent locations in the embedding space depending on the choice of κ [1].

Automatically estimating κ is largely an open problem, though some researchers have attempted to adaptively determine a globally optimal κ value [3], [4]. However, these experiments were limited to dense synthetic datasets. We argue that in general no single global optimal value of κ can be applied to learning the low dimensional manifold over the entire data space. We further contend that different values of κ are required in different regions of the data space to optimally reconstruct locally linear neighborhood.

In this paper, we present a novel Consensus-LLE (C-LLE) algorithm for creating a stable low dimensional representation of the data, in a manner analogous to building classifier ensembles such as Breiman's Bagging scheme [5]. Instead of attempting to estimate a single globally optimal κ value as in [3], [4] to be applied to the entire dataset, our scheme aims to estimate the true pairwise object adjacency $\mathcal{D}(c,d)$ in the low dimensional embedding between two objects $c, d \in C$. We formulate the problem of estimating object distances $\mathcal{D}(c,d)$ as a Maximum Likelihood Estimation problem (MLE) from multiple approximations $\mathcal{D}_{\kappa}(c,d)$ obtained by varying κ, which we assume are unstable and uncorrelated. Our scheme thus differs from related work in two fundamental ways: (a) C-LLE attempts to reconstruct the true low dimensional data manifold by learning pairwise object distance across the entire data space and avoids the κ estimation problem, and (b) C-LLE learns the low dimensional manifold in a locally adaptive fashion, compared to [3], [4] that attempt to learn an optimal κ value which is then uniformly applied to learning the manifold across the entire data space.

Prostate MRS is a non-invasive imaging technique used to detect changes in the concentration of the specific metabolites (choline, creatine and citrate) which have been shown to be representative of prostate cancer (CaP). Peak detection algorithms to automatically determine metabolite peaks have been found to be sensitive to presence of noise and other biomedical signal artifacts. In [6] we presented a novel hierarchical clustering algorithm employing NLDR to automatically distinguish between normal and diseased spectra within the prostate. In this paper, we apply the newly developed C-LLE algorithm to distinguishing between benign and cancerous MR spectra on a total of 18 studies and compare C-LLE's performance with LLE and 3 other state of the art MRS analysis methods. A secondary contribution of this work is the use of Independent Component Analysis (ICA) to (a) automatically determine search ranges within which to perform peak detection, and (b) validating the clusters obtained via C-LLE.

2   Consensus-Locally Linear Embedding (C-LLE)

2.1   Issues with LLE

The objective behind LLE is to non-linearly map objects $c, d \in C$ that are adjacent in the M dimensional ambient space ($\boldsymbol{F}(c)$, $\boldsymbol{F}(d)$) to adjacent locations in

the low dimensional embedding $(S(c), S(d))$, where $(S(c), S(d))$ represent the m-dimensional dominant eigen vectors corresponding to c, d ($m << M$). If d is in the κ neighborhood of $c \in C$, then $c, d \in C$ are assumed to be linearly related. LLE attempts to non-linearly project each $\boldsymbol{F}(c)$ to $S(c)$ so that the κ neighborhood of $c \in C$ is preserved. LLE is sensitive to the choice of κ since different values of κ will result in different low dimensional data representations.

2.2 Relationship between C-LLE and Bagging Classifiers

The aim behind constructing ensemble classifiers such as Bagging [5] is to reduce the variance and bias across weak classifiers. In Bagging [5], for an input object $c \in C$, a sequence of weak predictors $\phi(c, \mathcal{S}_k)$ are generated from K bootstrapped training sets $\mathcal{S}_k$ where $1 \leq k \leq K$. A strong Bagged classifier $\phi^{Bag}(c)$ is obtained by averaging or voting over the multiple weak classifiers $\phi(c, \mathcal{S}_k)$, $k \in \{1, ...K\}$. An analogous idea is used for C-LLE whereby we combine several weak embeddings, $S_\kappa(c)$ across different values of $\kappa \in \{1, ...K\}$ to obtain a comprehensive stable low dimensional data embedding, with lower variance and bias compared to individual weak embeddings. Our hypothesis is that for any $c, d \in C$, the pairwise object distance in the low dimensional space is faithfully represented in the stable consensus embedding $\hat{S}(c)$, for $c \in C$.

2.3 Maximum Likelihood Estimation (MLE) of Object Adjacency

The spirit behind C-LLE is the direct determination of pairwise object adjacencies in the low dimensional embedding space as opposed to κ estimation. For each c, d the aim is to find the true distance $\hat{\mathcal{D}}^\psi(c, d)$ between $c, d \in C$ in some lower dimensional embedding space, where ψ is an appropriately defined distance metric. Given multiple lower dimensional embeddings, the distance between c, d can be expressed as a distribution $\mathcal{D}_\kappa(c, d)$ where for brevity the metric notation has been dropped. The problem of determining $\hat{\mathcal{D}}(c, d)$ can be posed as a MLE problem. Thus we can rewrite this problem as,

$$p(\varphi_\mathcal{D}|\hat{\mathcal{D}}) = \prod_{\kappa=1}^{K} p(\mathcal{D}_\kappa|\hat{\mathcal{D}}); \forall c, d \in C, \tag{1}$$

where $\varphi_\mathcal{D}$ is a set of low dimensional distance estimates between $c, d \in C$, and based on the assumption that the lower dimensional embeddings obtained for $\kappa \in \{1, ...K\}$ are independent. We endeavor to find the MLE of $\mathcal{D}$, $\tilde{\mathcal{D}}$ that maximizes $\ln p(\varphi_\mathcal{D}|\mathcal{D})$ for $c, d \in C$. Intuitively this corresponds to computing the peak (mode) of the distribution $p(\varphi_\mathcal{D}|\hat{\mathcal{D}})$.

2.4 Algorithm for C-LLE

Step 1. Multiple lower dimensional embeddings are generated by varying $\kappa \in \{1, ..K\}$ using LLE. Each embedding $S_\kappa(c)$ will hence represent adjacencies between objects $c_i, c_j \in C, i, j \in \{1, ...|C|\}$, where $|C|$ is the cardinality of C. Thus $||S_\kappa(c_i) - S_\kappa(c_j)||_\psi$ will vary as a function of κ.

Step 2. Obtain MLE of pairwise object adjacency: A confusion matrix $W_\kappa \in \Re^{|C| \times |C|}$ representing the adjacency between any two objects $c_i, c_j \in C, i, j \in \{1, ..., |C|\}$ in the lower dimensional embedding representation $S_\kappa(c)$ is calculated as:

$$W_\kappa(i, j) = \mathcal{D}_\kappa(c_i, c_j) = \|S_\kappa(c_i) - S_\kappa(c_j)\|_\psi, \quad\quad (2)$$

where $c_i, c_j \in C$, for $i, j \in \{1, \ldots, |C|\}, \kappa \in \{1, \ldots, K\}$, and ψ in our case is the L2 norm. MLE of $\mathcal{D}_\kappa(c_i, c_j)$ is estimated as the mode of all adjacency values in $W_\kappa(i, j)$ over all κ. This $\hat{\mathcal{D}}$ for all $c \in C$ is then used to obtain the new confusion matrix $\hat{W}$.

Step 3. Multidimensional scaling (MDS): MDS [7] is applied to $\hat{W}$ to achieve the final combined embedding $\tilde{S}(c)$ for $c \in C$. MDS is implemented as a linear method that preserves the Euclidean geometry between each pair of objects $c_i, c_j \in C, i, j \in \{1, ..., |C|\}$. This is done by finding optimal positions for the data points c_i, c_j in lower-dimensional space through minimization of the least squares error in the input pairwise distances in $\hat{W}$.

3 Prostate cancer detection on MRS via C-LLE

3.1 Notation

We define a spectral scene $\mathcal{C} = (C, f)$ where C is a 3D grid of spatial locations. For each spatial location $c_i \in C, i \in \{1, \ldots, |C|\}$, there is an associated 256-dimensional valued spectral vector $\boldsymbol{F}(c) = [f_u(c) \mid u \in \{1, ..., 256\}]$, where $f_u(c)$ represents the concentration of different biochemicals (such as creatine, citrate, and choline) at every spatial location c.

3.2 Data Description

A total of 18 1.5 T *in vivo* endorectal T2-weighted MRI and MRS ACRIN studies[3] were obtained prior to prostatectomy. Partial ground truth for the CaP extent on MR studies is available in the form of approximate sextant locations and sizes for each study. The maximum diameter of the tumor is also recorded in each of the 6 prostate sextants (left base, left midgland, left apex, right base, right midgland, right apex). The tumor size and sextant locations were used to identify a *potential cancer space* used for performing a semi-quantitative evaluation of our CAD scheme. Additional details on identifying this cancer space are provided in [6].

3.3 C-LLE and consensus clustering for CaP detection on MRS

Figure 1 shows the flowchart demonstrating the different steps comprising our prostate MRS detection scheme. The adjacency matrix (W) is constructed across K embeddings of the MR spectral space so that for any $c_i, c_j \in C$, where $i, j \in \{1, ...|C|\}$, $W_\kappa(i, j)$ represents the distribution $\mathcal{D}_\kappa(c_i, c_j)$ of the low dimensional distance between MR spectra $\boldsymbol{F}(c_i), \boldsymbol{F}(c_j)$, for $k \in \{1, ...K\}$. As described in Section

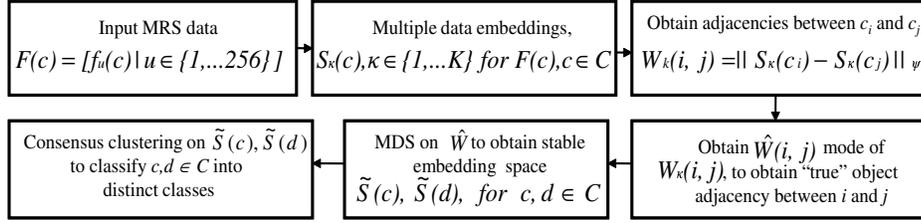| Input MRS data $F(c) = [f_u(c) \mid u \in \{1,...256\}]$ | Multiple data embeddings, $S_\kappa(c), \kappa \in \{1,...K\}$ for $F(c), c \in C$ | Obtain adjacencies between c_i and c_j $W_k(i,\ j) = \|\ S_\kappa(c_i) - S_\kappa(c_j)\ \|_\psi$ |
| Consensus clustering on $\tilde{S}(c), \tilde{S}(d)$ to classify $c, d \in C$ into distinct classes | MDS on $\hat{W}$ to obtain stable embedding space $\tilde{S}(c),\ \tilde{S}(d),\ for\ c, d \in C$ | Obtain $\hat{W}(i,\ j)$ mode of $W_\kappa(i,\ j)$, to obtain "true" object adjacency between i and j |

Fig. 1. Flowchart showing the C-LLE algorithm and consensus clustering scheme for CaP detection on prostate MRS.

2.4, the stable spectral distance matrix $\hat{W}$ is then obtained and MDS applied to obtain the stable embedding representation of the spectra, $\tilde{S}(c)$ for each $c \in C$.

To overcome the instability associated with centroid based clustering algorithms, we generate multiple weak clusterings $V_t^1, V_t^2, V_t^3, t \in \{0, \ldots, T\}$ by repeated application of k-means clustering on the combined low dimensional manifold $\tilde{S}(c)$, for all $c \in C$. We assume that each prostate spectra could be classified as one of the three classes: cancer, benign and other tissue classes (e.g. benign hyperplasia (BPH)). Each cluster, V_t is a set of objects which has been assigned the same class label by the k-means clustering algorithm. As the number of elements in each cluster tends to change for each such iteration of k-means, we calculate a co-association matrix H with the underlying assumption that objects belonging to a *natural* cluster are very likely to be co-located in the same cluster for each iteration. Co-occurrences of pairs of objects $c_i, c_j \in C$ in the same cluster V_t are hence taken as votes for their *association*. $H(i, j)$ thus represents the number of times $c_i, c_j \in C$, for $i, j \in \{1, ...|C|\}$, were found in the same cluster over T iterations. We apply MDS [7] to H followed by a final unsupervised classification using k-means, to obtain the final stable clusters $\hat{V}^1, \hat{V}^2, \hat{V}^3$.

In the following two sections we describe 3 other commonly used MRS analysis schemes.

3.4 Model based peak integration scheme via Independent Component Analysis (ICA)

Peak detection on prostate MRS is a difficult problem due to noise and spectral contributions from extra-prostatic regions. A secondary contribution of this paper is a model based approach to localize choline, creatine and citrate peaks based on Independent Component Analysis (ICA). ICA is a multivariate decomposition technique which linearly transforms the observed data into statistically maximally independent components (ICs). For a set of voxels identified offline as cancer, $\chi_{CaP} \subset C$, we obtain A independent components $\mathcal{F}_\alpha^{IC}, \alpha \in \{1, ...A\}$ which represent spectral contributions of choline, creatine and citrate for prostate cancer. The parts per million (ppm) ranges (ν_{cc}, ν_{cr}) on the X-axis are then learnt for choline+creatine and citrate from $\mathcal{F}_\alpha^{IC}, \alpha \in \{1, ...A\}$. Peak detection is then performed on $\mathcal{C}$ to

---

[3] http : //www.acrin.org/6659 protocol.html

identify choline, creatine and citrate peaks within the ranges ν_{cc} and ν_{cr}. Area under the choline+creatine peak (V_{cc}) and under citrate peak (V_{cr}) is obtained via integration for all voxels $c \in \mathcal{C}$. Zakian index ($\gamma(c)$) [8] is then calculated as the ratio of $V_{cc}(c)/V_{cr}(c)$. A pre-defined threshold determined by radiologists [8] is used to classify the spectra as cancer/benign based on $\gamma(c)$ for $c \in C$.

3.5 z-score and Principal Component Analysis (PCA)

z-score is a statistical measure defined as the ratio of the difference between the population mean and individual score to the population standard deviation. For a set of voxels, Φ^{tr} of c, $\Phi^{tr} \subset C$, the mean spectral vector $\boldsymbol{F}^{\mu} = [f_u^{\mu}|u \in \{1,..256\}]$ is obtained and the corresponding standard deviation vector $\boldsymbol{F}^{\sigma} = [f_u^{\sigma}|u \in \{1,...256\}]$, where $f_u^{\mu} = \frac{1}{|\Phi^{tr}|} \sum_{c \in \Phi^{tr}} f_u(c)$ and $f_u^{\sigma} = \sqrt{\frac{1}{|\Phi^{tr}|} |\sum_{c \in \Phi^{tr}} [f_u(c) - f^{\mu}(c)]^2}$. The z-score at each $c \in C$ is given as $z(c) = \frac{||\boldsymbol{F}(c) - \boldsymbol{F}^{\mu}||_2}{||\boldsymbol{F}^{\sigma}||_2}$, where $|\Phi^{tr}|$ is the cardinality of Φ^{tr}. A predefined threshold θ_z is then used to identify each $c \in C$ as cancerous or not based on whether $z(c) \geq \theta_z$. PCA attempts to find the orthogonal axes that contain the greatest amount of variance in the data using eigenvalue decomposition. Similar to our C-LLE scheme, each $c \in C$ is described by 5 principal components, $S^{PCA}(c)$ which contain 98% of the data variance. Consensus clustering is then applied on $S^{PCA}(c)$, to cluster each $c \in C$ into one of 3 classes.

4 Results and Discussion

4.1 Qualitative Results

Figure 2 shows CaP detection results on a prostate MRS study via C-LLE, LLE, PCA, z-score, and ICA based peak detection. Figures 2(a)-(c) show the comparison of C-LLE (c) with LLE for (a) $\kappa = 5$, and (b) $\kappa = 7$ on a single 2D T2 weighted MRI slice. Figures 2(d), (e), (f) show the clustering results obtained via ICA peak detection, z-score and PCA respectively. At each spatial location on the MRI slice a spectral signature was analyzed and that corresponding location was assigned one of 3 colors (for C-LLE, PCA) and one of the two colors (for z-score, ICA based peak detection) based on clustering/classifier results. The white box superposed on 2(a)-(f) show the potential cancer space for corresponding slices. In each of Figures 2(a)-(f) the red cluster represents the one identified as cancer by each of the different methods. Note in Figure 2(c) that the C-LLE result shows excellent sensitivity and specificity and also appears to reduce the variance and instability in the individual weak embeddings shown in 2(a), (b). Figures 2(d)-(f) corresponding to results obtained via ICA-based peak detection, z-score method and PCA respectively, show low sensitivity and specificity compared to our C-LLE scheme (Figure 2(c)). To assess the validity of C-LLE and PCA, we employed ICA to isolate independent components (IC) from clusters identified as CaP by the two schemes. Figure 2(h) shows an IC obtained from the cluster identified as CaP by C-LLE (shown as red in 2(c)); Figure 2(i) shows the corresponding result obtained via PCA. Note the strong correlation between the ICs obtained via C-LLE (2(g)) and a known CaP
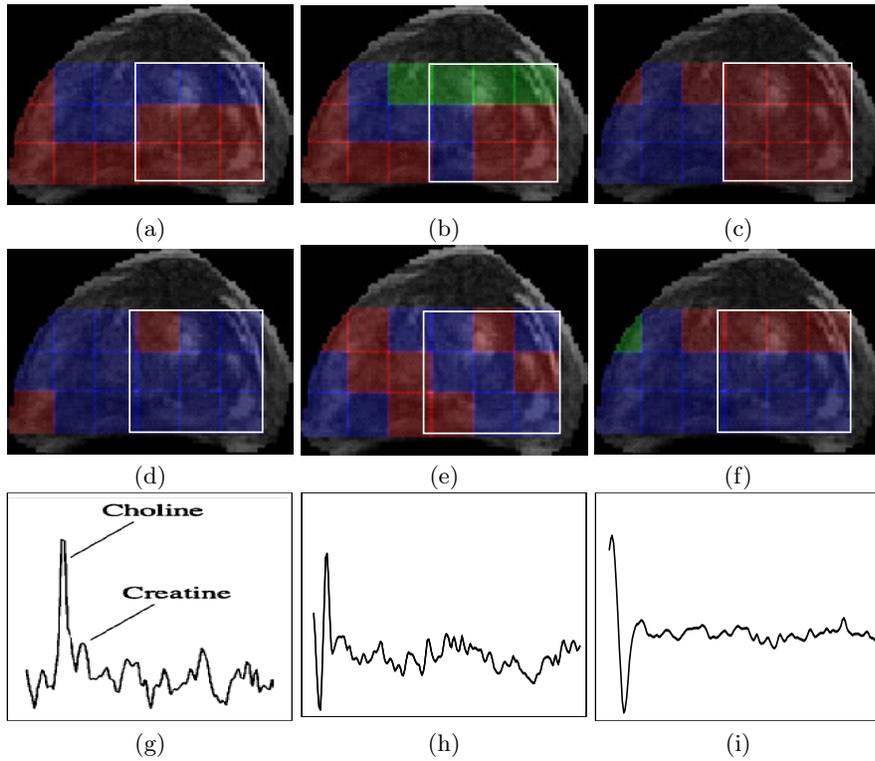
Fig. 2. Clustering results for (a) LLE ($\kappa = 5$), (b) LLE ($\kappa = 7$), (c) C-LLE, (d) ICA based peak detection, (e) z-score, and (f) PCA. The white box superposed on the T2 MR image corresponds to the ground truth region. In each image the red cluster corresponds to the locations identified as cancer by each of the methods. Figure 3(g) Typical CaP MR spectra, (h) IC obtained from clusters identified as CaP in C-LLE, and (i) corresponding IC obtained via PCA.

MRS signature 2(h) according to the 5-point scale defined in [9]. Note also the dissimilarity between the spectra obtained by PCA and shown in 2(i) compared to those in 2(g),(h).

4.2 Quantitative Results

Table 1 shows average CaP detection sensitivity and specificity over 18 studies obtained from C-LLE ($m = 4$), ICA based peak detection, PCA and z-score. Table 1 (b) shows the sensitivity and specificity results averaged across 18 datasets for C-LLE ($m = 3, 4, 5$) compared to LLE by varying the number of dimensions. Note that the C-LLE scheme has a higher sensitivity and specificity across all dimensions which suggests the efficacy of the scheme. The effectiveness of our scheme for detection of prostate cancer is evident from the quantitative results (Table 1) with both sensitivity and specificity of close to 87% and 85% respectively compared to current state of the art methods peak detection, PCA and z-score. Table 1(b) reveals

Method	Sensitivity	Specificity
C-LLE	86.90	85.14
Peak detection	45.29	76.62
PCA	66.95	77.89
z-score	74.74	49.75

(a)

m	Sensitivity		Specificity	
	C-LLE	LLE	C-LLE	LLE
3	83.20	82.07	84.81	81.89
4	86.90	83.38	85.14	81.77
5	84.88	82.10	85.60	81.70

(b)

Table 1. (a) Average CaP detection Sensitivity and Specificity of C-LLE ($m = 4$), compared to ICA based peak detection, z-score, and PCA, averaged over a total of 18 MRS studies using the top 7 eigen values. Table 1(b). Average CaP detection Sensitivity and Specificity results of C-LLE compared to LLE for dimensions 3, 4 and 5.

that C-LLE consistently outperforms traditional LLE across multiple dimensions ($m = 3, 4, 5$).

5   Concluding Remarks

In this paper we presented a novel C-LLE algorithm which aims to overcome the limitations of traditional LLE by obtaining a stable low dimensional representation of high dimensional data by integrating object adjacencies from across multiple low dimensional unstable data projections. The spirit behind C-LLE, is rooted in the classifier ensembles literature where multiple weak classifiers with high variance and bias are combined to create a strong classifier with low bias and variance. Experimental results demonstrate excellent performance of our algorithm for CaP detection on prostate MRS compared to ICA based peak detection scheme, PCA, z-score, and traditional LLE. One issue we have not addressed in this work is the intrinsic dimensionality at which to combine the multiple weak embeddings in C-LLE. While, C-LLE is computationally more expensive compared to traditional LLE, it is not computationally prohibitive except for very large or dense datasets. In future work we will address some of these issues and also explore the applicability of C-LLE to other classification problems.

References

1. Roweis, S., Saul, L.: Fit Locally Think Globally. JMLR **4** (2003) 119–155.
2. Lee, G., Rodriguez, C., Madabhushi, A.: Investigating the Efficacy of Nonlinear Dimensionality Reduction Schemes in Classifying Gene- and Protein-Expression Studies. IEEE/ACM Transactions on Computational Biology and Bioinformatics (2008)
3. Lin, J., Zha, J.: Riemann Manifold Learning. IEEE Transactions on pattern Analysis **30** (2008) 796–809.
4. Wang, J., Zhang, Z., Zha, H.: Adaptive Manifold Learning. NIPS (2004)
5. Breiman, L.: Bagging Predictors. Machine Learning **123-140** (1996) 123–140.
6. Tiwari, P., Rosen, M., Madabhushi, A.: A Hierachical Unsupervised Spectral Clustering Scheme for Detection of Prostate Cancer from Magnetic Resonance Spectroscopy. MICCAI, Part II, LNCS 4792 (2007) 278–286.
7. Venna, J., et al.: Local Multidimensional Scaling. Neural Networks **19(6).** (2006)
8. Zakian, K., et al.: Correlation of Proton MRSI with gleason score based on sept-section Pathologic Analysis after Radical Prostatectomy. Radiology **234** (2005) 804:814.
9. Kurhanewicz, J., et al.: Prostate Depiction at Endorectal MR Spectroscopic Imaging: Investigation of a Standardized Evaluation System. Radiology **233** (2004) 701–708.